Analytical Thermal Model for Self-Heating in Advanced FinFET Devices With Implications for Design and Reliability

Chuan Xu, Member, IEEE, Seshadri K. Kolluri, Member, IEEE, Kazuhiko Endo, Member, IEEE, and Kaustav Banerjee, Fellow, IEEE

Abstract—A rigorous analytical thermal model has been formulated for the analysis of self-heating effects in FinFETs under both steady-state and transient stress conditions. 3-D self-consistent electrothermal simulations, tuned with experimentally measured electrical characteristics, were used to understand the nature of self-heating in FinFETs and calibrate the proposed model. The accuracy of the model has been demonstrated for a wide range of multithin devices by comparing it against finite element simulations. The model has been applied to carry out a detailed sensitivity analysis of self-heating with respect to various FinFET parameters and structures, which are critical for improving circuit performance and electrical overstress/electrostatic discharge (ESD) reliability. The transient model has been used to estimate the thermal time constants of these devices and predict the sensitivity of power-to-failure to various device parameters, for both long and short pulse ESD situations. Suitable modifications to the model are also proposed for evaluating the thermal characteristics of production level FinFET (or Tri-gate FET) structures involving metal-gates, body-tied bulk FinFETs, and trench contacts.

Index Terms—Analytical model, device design, electrothermal, electrostatic discharge (ESD), FinFETs, reliability, self-heating, thermal modeling, Tri-gate FET.

I. INTRODUCTION

FinFET devices have shown tremendous potential for overcoming CMOS scaling limitations beyond the 22 nm node due to their superior electrostatics. However, the self-heating effects in these ultrascaled structures are becoming an increasing concern due to the low thermal conductivity of the buried oxide (BOX) and the interlayer dielectric (ILD) materials, as well as the physical confinement of the Si fin in the 3-D FinFET device geometry [1]. In addition, the reduced thermal conductivity of the thin Si film [2] leads to increased self-heating resulting in performance and reliability degradation [3]. As FinFET process technology and integration progresses rapidly [4]–[7], FinFET parameters are being optimized to maximize their performance [8], [9]. While experimental extraction techniques for self-heating effect are being developed [10], there is a need to concurrently formulate accurate and fast electrothermal modeling and simulation capability. Such capability is crucial for simultaneously evaluating the impact of these parameters on the thermal characteristics, which in turn impact the reliability and the electrical performance of FinFETs. Self-heating is already known to degrade the negative bias temperature instability (NBTI) and oxide reliability of FinFET devices [11], [12]. It is also known that self-heating degrades the drain current in silicon-on-insulator (SOI) and strained-Si devices by around 15% due to the presence of a BOX layer in SOI type devices or a SiGe graded layer in strained-Si devices that increases the thermal resistance of the device due to low thermal conductivity of these materials [13], [14]. In SOI FinFETs, the problem increases manifold due to the small and confined dimensions of the fin that reduces its thermal conductivity (due to reduced phonon mean free path), and thereby increases the thermal resistance, which makes heat transport out of the device more difficult. Although not as severe as in SOI FinFETs, the thermal issue in the bulk FinFET [15] is more critical than that in the bulk CMOS, due to the large thermal resistance of the narrow fin. Moreover, it has been recently reported that the electrostatic discharge (ESD) reliability of FinFET-based NMOS protection devices is strongly modulated by self-heating in these structures [16], [17]. Hence, it is important to thoroughly investigate and study electrothermal issues in these devices to understand their implications for the reliability as well as design optimization in order to realize their maximum potential.

3-D self-consistent electrothermal technology computer aided design (TCAD) simulations have been studied in the literature [18]–[20]. While they are the most accurate way to capture the self-heating effects in FinFETs and predict the temperature profiles in the device, these 3-D electrothermal
simulations are computationally very expensive and are not practical for design optimization purposes. To reduce the complexity of this problem, a coupled numerical simulation methodology of 2-D electrical transport and 3-D thermal transport in FinFETs has been proposed in [21]. However, this work does not provide any verification of said methodology with self-consistent electrothermal simulations and severely under-predicts the temperature variation within each fin (<2 K). In addition, the proposed methodology is still not practical for the use of design engineers as it requires the use of device simulators for solving both electrical and thermal transport problems. Pop et al. [22] proposed a compact thermal model for ultrathin body SOI devices, which is later extended by Swahn et al. [23], for multi-fin FinFET devices, to model the heat flow in the FinFET device through a thermal resistive network. However, these models have assumed constant temperature boundary conditions at the source, drain and gate contacts, whereas most of the heat flow is through the BOX layers to the heat sink at the bottom, even for SOI device structures, as observed in [24], [25]. Recent work by Bansal et al. [26], provides compact models for gate-level thermal analysis of FinFETs, ignoring the temperature variations from source to drain within the fin. This work also assumes that the majority of the heat flow is through the interconnects, as opposed to the observations in [24]. A similar argument can be found in [20], but the treatment of the input, output, ground (Gnd), and VDD terminals of a minimized inverter as heat sink, is invalid (the heat sink should be the silicon substrate). It should be noted, that for the worst-case situations where all the neighboring devices are in the ON-state, or for the large FinFETs with multiple fins, interconnects would be at a higher temperature than the device itself and offer negligible heat sinking, causing most of the heat to flow through the BOX layer (for SOI FinFETs) or through the narrow body-tie connecting the silicon fin and the substrate (for bulk FinFETs). Electrothermal simulations providing insights into the ESD failure mechanisms in MultiGFETs have been recently reported in [17].

In this work, we have formulated an analytical method to understand and quantify self-heating effects in FinFETs under both steady-state and transient conditions. Using our method, we can better understand the physics of thermal transport and the effect of various device parameters on modern FinFET architectures. Our method is used for studying the ESD performance of FinFETs, which is a strong function of temperature distribution in the device. In addition, our method is extended to handle production level FinFET (or Tri-gate FET) structures including metal-gate FinFET, bulk FinFET, and trench contact FinFET. We have also included a section on how we can extend our model to incorporate the thermal boundary resistance (TBR). Our method has the potential to be extended to more compact thermal models. In addition, as our model is derived from thermodynamics, it can be applied to scaled devices in more advanced FinFET technologies.

II. FINFET DEVICE FABRICATION AND CHARACTERIZATION

SOI FinFET devices were fabricated according to the process flow outlined in Fig. 1, at the National Institute of Advanced Industrial Science and Technology (AIST), in Tsukubu, Japan. The starting material for the FinFET was (100) oriented silicon-on-insulator (SOI) wafer. A conventional reactive ion etching was used to etch the Si-Fin. The 3-nm gate oxide was formed at 850 °C and then covered with the n+ poly-Si. A shallow As implantation into the extension of the source/drain (S/D) was performed, followed by a sidewall formation and a S/D implantation. Finally, the S/D was activated at 950 °C and the devices were annealed at 450 °C in 3% H₂ ambient after the metallization. Fully processed wafers were then used to measure the electrical characteristics of these devices. In the entire paper, we treat this research level fabricated SOI FinFET as the structure to be simulated and modeled, except for studying the variations in geometrical and material parameters in Section IV-E and the divergences toward production level FinFETs in Section VII.

Fig. 2 shows measured current-voltage characteristics for one sample, and the calibrated curve from electrothermal device simulation using SDevice [27] (details of the simulation are illustrated in Section III). Fig. 3 shows the TEM image of the fabricated multifinger SOI FinFET device. Important materials in the FinFET cross-section and different planes describing the symmetry in the multifinger structure are identified in the figure. For the worst-case simulations, where all the fingers are assumed to be in the ON-state, the simulation domain can be reduced on grounds of electrical and thermal symmetry, as indicated in the figure and as explained in Section III.
Fig. 3. Fabricated multifinger device and its reduction to a simpler structure for device simulation. The structure simplification is based on electrical and thermal symmetry across indicated axes.

Fig. 4. 3-D schematic of a multifinger SOI FinFET structure.

III. DEVICE SIMULATION

Fig. 4 shows the 3-D schematic view of the multifinger SOI FinFET structure used for our study. Different regions in the device and geometric parameters related to them are identified in the figure. Fig. 4 also shows the coordinate system used throughout this work and reduced simulation domain obtained for the worst-case scenario when all the fingers are in the ON-state, as explained below.

The simulation structure for the schematic shown in Fig. 4 is as drawn in Fig. 5(a). It can be observed that the reduction of Fig. 3 along the axes of symmetry for multifinger devices (marked in figure) simplifies the simulation domain to the one shown in Fig. 5(a). Further it can be noted that even this single-finger structure is symmetric, both electrically and thermally, across a plane perpendicular to the $Y$ direction and passing through the line marked $XX'$ in Fig. 5(a). The reduction of simulation domain across this plane results in the final simulation structure, as shown in Fig. 5(b). Dielectric passivation layer (SiO$_2$), BOX, and Si substrate are also included in the simulation, though they are not shown in the figure for clarity.

The dimensions of the simulated structure are same as the dimensions of the fabricated SOI FinFETs. The symmetry arguments discussed above would lead to adiabatic boundary conditions at $y = 0$ and $y = 0$. Length of the source and drain regions ($L_{SD}$) is chosen to be sufficiently large ($\sim 550$ nm), so that the adiabatic boundary conditions at their ends do not affect the temperature observed in the device. Appropriate thermal conductivity values for thin-film silicon and BOX layer, obtained from existing literature, are used for our simulations and are summarized along with other device parameters in Table I.

Please note that in all sections in the entire paper, the temperature difference between the heat sink and the very top of the silicon substrate has not been considered, as we are focused on the localized temperature variations in FinFETs. In real circuits, the very top of the silicon substrate can be several tens of degrees higher than that of the heat sink. Readers can refer to [29] and [30] for more details on how to simulate the temperature variation across the whole die. The actual temperatures of FinFETs are the superposition of the localized

| Device Parameters Used for the Simulated Structure Shown in Fig. 5(b) |
|------------------|------------------|------------------|------------------|
| $L_g$ (nm)       | $L_{ext}$ (nm)   | $L_{SD}$ (nm)    | $W_{SD}$ (nm)    |
| 50               | 125              | 550              | 500              |
| $J_{sc}$ (mA)    | $W_{ch}$ (mm)    | $W_{dx}$ (mm)    | $H_{silicon}$ (mm) |
| 3                | 20               | 100              | 100              |
| $K_{poly}$ (W/m-K) | $K_{BOX}$ (W/m-K) | $K_{Si}$ (W/m-K) in source/drain (S/D) |
| 45               | 0.8 (Ref. [28])  | 6.2 (Ref. [22])  |
| $K_{LID}$ (W/m-K) | $K_{Si}$ (W/m-K) in channel and S/D extension |
| 0.21             | 25 (Ref. [27])   |

These are the equivalent thermal conductivities, which include the TBR (for $K_{BOX}$) and the impact of phonon-boundary scattering (for $K_{Si}$).
temperature variations and the temperature variations across the whole die.

Fig. 6 shows the simulation structure and the 3-D temperature profile in the channel, source/drain (S/D) regions of the device through the self-consistent electrothermal device simulation.

The first step in our approach involved the calibration of the 3-D electrothermal device simulator (SDevice [27]), against measured I-V characteristics of the SOI FinFET structures. We assume constant heat generation across the entire cross-section of Si-fin in both ANSYS and analytical model. But in SDevice simulations, it is strongly dependent on the y-coordinate (or the distance to the Si-SiO2 interface), as indicated in Fig. 8. Only the integration of the heat generation across a cross-section normal to the XX′ line from SDevice is comparable to ANSYS and analytical model.

This is a critical step where the simulator was calibrated so that it would match the experimentally measured saturation current for the same values of applied bias-voltages in each case. This would ensure that the temperature profiles obtained correspond to the final operating point (maximum drain and gate voltage) in each of the measured curves. This calibrated simulator is then used as a reference for verifying the accuracy of our analytical thermal model. Self-consistent hydrodynamic-based 3-D electrothermal simulations (in this paper, we use SDevice [27]) are required for this calibration as they offer the sole means to accurately capture the nature of self-heating in the FinFETs and predict the temperature profiles in the device. They also provide the most accurate understanding of the nature of heat-generation in the FinFETs, which is required as an input to the analytical thermal model.

Fig. 7 shows the temperature and heat generation profiles along the fin (X-direction), obtained for one particular value of Y and Z coordinates. Please note that our hydrodynamic transport-based simulations have taken into account the hot-electron injection effect. In other words, the electron temperature is different from the lattice temperature, and the electrons relax most of their energy (with heat generation) in the drain extension (DE) region, adjacent to the channel, instead of at the drain-side channel. Though the variation in temperature was observed to be within 2%–4% along the fin-thickness, dimension (Y-direction), there is significant variation of heat generation along the fin-thickness as shown in Fig. 8. This figure shows only the simulated half of the fin, and the actual channel would also consist of another half reflected along the XX′ axis shown in the figure.

3-D electrothermal device simulation (SDevice simulation) is time consuming and does not easily converge. To further understand the nature of heat-flow in the FinFETs, 3-D thermal simulations have been performed for the SOI FinFET structure, using a commercially available finite element solver, ANSYS [31]. Fig. 7 also shows a comparison between the temperature distribution from SDevice and from ANSYS. Good agreement is achieved between the two simulators, which indicates the validity of using ANSYS.

ANSYS thermal simulations give us more flexibility than 3-D electrothermal device simulations, in terms of controlling the heat generation in the device and studying the nature of heat fluxes in different regions of the device, and so on; hence it is very useful for the verification of our thermal model.

IV. STEADY-STATE THERMAL MODEL

The analytical thermal model divides the SOI FinFET into different regions, as shown in Fig. 9, and solves the heat...
transfer equations in each of those regions, coupling the solutions from the neighboring regions with temperature and heat flux continuity boundary conditions. The steady-state thermal model, developed based on these boundary conditions, is later extended to transient conditions for studying electrical overstress/electrostatic discharge (EOS/ESD) applications.

Fig. 9 also shows various regions of the SOI FinFET structure along with different directions of heat spreading from the Si-fin and polysilicon gate regions to the substrate, which are accounted for in our model. As shown in Fig. 9, the rectangular Si-fin region of width \( W_{\text{fin}} \) is surrounded by ILD in regions 2 and 4 and thin-oxide-layer followed by polysilicon gate in region 3. Regions 1 and 5 comprise only silicon as the material. BOX and substrate layers are at the bottom of each of these regions. It was observed from the simulations in Section III that the temperature variation in the bottom of the BOX layer is negligible within the whole routing area of the SOI FinFET. Hence, in our analytical thermal model and the simulations in the following sections, the bottom of the BOX layer is assumed to be at a constant temperature \( T_0 \) (in this paper, \( T_0 = 300 \text{K} \)). In addition, due to the much higher thermal conductivity in the Si-fin than in the BOX and ILD, the variation in temperature along \( Z \)-direction in the Si-film at any \((x,y)\) is negligible compared to the average temperature rise in the Si-film at \((x,y)\), in all five regions in the device. Hence, we assume that the temperature of Si-fin is uniform along the \( Z \)-direction in all the five regions. These assumptions allow us to reduce the 3-D heat transfer problem to either 2-D or 1-D, as explained in the following subsections.

We first introduce part of the nomenclature used in the following sub-sections. Other terms are defined when they are introduced in different equations. \( K_i \) denotes the thermal conductivity of the material in the \( i^{th} \) region, in which the heat transfer equations are being solved. Thermal conductivity of Si-fin varies from region to region based on the thickness of the film in that region, as given in Table I. Temperature in the \( i^{th} \) region is denoted by either \( T_i(x) \) or \( T_i(x,y) \), depending on whether it is 1-D or 2-D respectively and \( T_0 \) denotes the reference temperature (300 K), as mentioned before. Heat generation rate per unit volume is considered to be position dependent and for the \( i^{th} \) region, it is denoted by \( q_{g,i} \).

### A. Solution in the Source Extension and Drain Extension Regions

In the source extension (SE) and drain extension (DE) regions (regions 2 and 4, respectively), due to the small thickness of the fin and high conductivity of the Si-fin region compared to the other regions, temperature can be assumed to be uniform in the \( X \)-direction. Hence, in these regions, at any particular value of \( X \), entire cross-section of the Si-fin normal to the \( XX' \) direction can be assumed to be isothermal. Fig. 10 shows the \( X \)-component of heat flux in different regions of the device, obtained from ANSYS simulations. It can be observed that, in the SE and DE regions, heat-flow in the \( X \)-direction is mainly due to thin Si-fin region and thermal flux in the \( X \)-direction through the surrounding ILD material can be neglected because of low thermal conductivity of that region. For the same reason, heat flux in the \( X \)-direction through the BOX layer can also be ignored. However, there is a lateral heat spreading in the \( Y-Z \) plane through the ILD and BOX layers, at any particular value of \( X \). This 2-D conduction can be reduced to 1-D by the use of an increased effective thermal conductivity of BOX layer \( (K_{\text{BILT}}) \) to account for fringing heat spreading toward the substrate. \( K_{\text{BILT}} \) values used in this work were extracted from 2-D finite element calculations to accurately account for fringing heat spreading through the BOX and passivation layers (Fig. 11).
The parameters in Table I and the simulation of Fig. 7. The figure shows the ANSYS simulation. The simulation is this figure uses the same structure and heat-loss terms (for formulated as a 1-D heat equation with heat-generation and height), though approximate formulations for the same can be reduced to the following standard form:

\[ \frac{\partial^2 T_i(x,y)}{\partial x^2} + \frac{\partial^2 T_i(x,y)}{\partial y^2} = m_i(T_i(x,y) - T_0) = 0. \]  

(6)

The boundary conditions for the source and drain regions for the equation represented by (6) are listed below. The parameters such as \( X_{D}, X_{SD}, \) and \( W_{SD}, \) used in these boundary conditions are defined in Fig. 9(b) and for simplicity of analysis, this heat generation \( q_{g,i}(x) \) is assumed to be a constant value plus a Gaussian distribution along X-direction with the peak in DE region. The Gaussian distribution should be normalized so that the total heat in DE region is equal to \( I_D V_{DS} \), while the peak value is equal to the maximum heat generation.

B. Solution in the Source and Drain Regions

For the source and drain regions (regions 1 and 5, respectively), there is a significant variation of temperature both along the X- and Y- directions. Hence, we need to consider 2-D heat spreading in silicon, in addition to the transfer of heat to the substrate. As mentioned earlier, from Fig. 10, it can be observed that heat flux in the X-direction through the ILD regions is negligible when compared to that through the Si-fin. Hence, most of the heat flowing into the source and drain regions is through the thin Si-fin and the contribution of the ILD layers to the heat flowing into the source and drain regions is negligible. In the source and drain regions, under the assumption that temperature in the Si-film is uniform along the XD- and YD-directions, there is a significant variation of temperature both along the XD- and YD-directions, and for simplicity of analysis, this heat generation term \( q_{g,i}(x) \) is assumed to be a constant value plus a Gaussian distribution along X-direction with the peak in DE region. The Gaussian distribution should be normalized so that the total heat in DE region is equal to \( I_D V_{DS} \), while the peak value is equal to the maximum heat generation.

![Fig. 10. KBef, versus KILD for SE and DE regions in FinFET from 2-D (per-unit-length) ANSYS simulation. All geometrical and material parameters are assumed the same as those in Table I and Fig. 7, except KILD.](image)

![Fig. 11. KBef, versus KILD for SE and DE regions in FinFET from 2-D (per-unit-length) ANSYS simulation. All geometrical and material parameters are assumed the same as those in Table I and Fig. 7, except KILD.](image)
Under the above boundary conditions, (6) has a general solution of the form

\[
T_1(x, y) = T_0 + \sum_{n=0}^{\infty} A_{n,d} \left[ e^{jx \lambda_1} + e^{-jx \lambda_1} \right] \cos \left( \frac{2 \pi n y}{W_{SD}} \right)
\]

(13)

\[
T_3(x, y) = T_0 + \sum_{n=0}^{\infty} A_{n,d} \left[ e^{jx \lambda_1} + e^{-jx \lambda_1} \right] \cos \left( \frac{2 \pi n y}{W_{SD}} \right)
\]

(14)

where

\[
A_{n,d} = \begin{cases} \frac{2 \cosh \lambda_1}{W_{SD}} \frac{W_{fin}}{A_{g,n}} & \text{for } n = 0, \\ \frac{\sinh \lambda_1}{W_{SD}} \frac{W_{fin}}{A_{g,n}} & \text{for } n \neq 0. \end{cases}
\]

(15)

\[
\lambda_{g,n} = \sqrt{\cosh^2 \left( \frac{2 \pi n y}{W_{SD}} \right) - 1}.
\]

(16)

(17)

C. Solution in the Gate Region

Unlike the SE and DE regions, thermal conduction in X-direction cannot be neglected in the polysilicon gate material. There is a lateral spreading in the X-Y plane, in addition to the conduction toward the substrate in the Y-Z plane. Hence, like the source and drain regions, a 2-D heat equation with heat-loss terms (ignoring the heat-generation in the channel) should be solved in region 3. To account for the fringing heat-spreading in the X-Z plane through the ILD and BOX layers, an increased effective thermal conductivity of BOX layer \((K_{BOX})\) is extracted by ANSYS 2-D simulation (shown in Fig. 12). The gate region is actually a composite region consisting of three materials—Si channel, SOI gate-oxide, and the polysilicon gate. Since it is difficult to solve the 3-D heat transfer equations analytically in such a composite region, reasonable approximations need to be introduced. As an initial approximation, since the thickness of the gate oxide layer is very small (3 nm), the thermal resistance caused by the thin gate oxide layer has been ignored in our calculations. As an additional simplification, since \(W_{fin} \ll W_{SD}\) and the thermal conductivities of Si and polysilicon are of the same order, the thermal parameters of the Si in the channel region are assumed to be the same as those of polysilicon. These assumptions greatly simplify the calculation of temperature profile in the gate region, without introducing significant errors as evident from the results presented later in this section. The heat-flux into the channel from the DE region is modeled as being mainly through the Si-fin region, as the heat-flux in the X-direction through the ILD layers is negligible, as mentioned before. Under these assumptions, the governing heat equation in region 3 takes the following form:

\[
\frac{\partial T_3(x, y)}{\partial x} + \frac{\partial T_3(x, y)}{\partial y} - m_3^2 (T_3(x, y) - T_0) = 0
\]

(18)

where

\[
m_3 = \frac{K_{BESF}}{K_B H_{box} H_{box}}
\]

(19)

\(K_1\) is the thermal conductivity of poly-silicon gate, and \(H_{box}\) is the gate-height. In (18), we neglect the heat generation in the CH region. On the other hand, we move such heat generation into the Gaussian distribution shaped heat generation in the DE region, so that the total heat in SE and DE region in our model is equal to \(I_d V_d\). As from the SDevice simulations in Figs. 7 and 8, such assumption is valid since the total heat in CH is much smaller than that in DE, and the majority of heat generation in CH is near the CH/DE boundary. The thermal boundary conditions in the gate regions, for the equation represented by (18) are listed below: The parameters like \(X_{H}, W_{SD}, \ldots\) used in these boundary conditions are defined in Fig. 9(b)

\[
\left. \frac{\partial T_3(x, y)}{\partial y} \right|_{y=W_{SD}} = 0
\]

(20)

\[
\left. \frac{\partial T_3(x, y)}{\partial y} \right|_{y=W_{SD}/2} = \begin{cases} \beta_{g,n} & \text{for } 0 \leq y \leq W_{SD}/2 \\ 0 & \text{for } W_{SD}/2 \leq y \leq W_{SD}/2 \\ \end{cases}
\]

(21)

\[
\left. \frac{\partial T_3(x, y)}{\partial x} \right|_{x=X_{H}} = 0
\]

(22)

Under the two adiabatic boundary conditions given by (20), (18) has a solution of the form

\[
T_3(x, y) = T_0 + \sum_{n=0}^{\infty} \left( A_{g,n} e^{jx \lambda_1} + B_{g,n} e^{-jx \lambda_1} \right) \cos \left( \frac{2 \pi n y}{W_{SD}} \right)
\]

(23)

where

\[
\lambda_{g,n} = \sqrt{m_3^2 + \frac{2 \pi n}{W_{SD}}}.
\]

(24)

Solution of (23) along with heat-flux boundary conditions given by (21) and (22) for \(A_{g,n}\) and \(B_{g,n}\) yields

\[
A_{g,n} = \frac{e^{jx \lambda_1} \left( e^{jx \lambda_1} \beta_{g,n} - \beta_{g,n} \right)}{e^{jx \lambda_1} + 1}
\]

(25)
where

\[ \theta_n' = \frac{\beta_n' W_{\text{in}} / (W_{\text{fin}} k_{x,y})}{n\pi} \text{ for } n \neq 0, \]

and

\[ \theta_n'' = \frac{\beta_n'' W_{\text{in}} / (W_{\text{fin}} k_{x,y})}{n\pi} \text{ for } n \neq 0. \] (28)

The parameters \( \beta_n' \) and \( \beta_n'' \) used in the above equations are constants proportional to the heat-flux from the channel into the source and drain extension regions, respectively, and are determined from temperature and heat-flux continuity boundary conditions as described in Section IV-D.

D. Coupling Solutions from Different Regions

By combining the temperature profiles in different regions obtained using (4), (13), (14), and (23), we can obtain the steady-state temperature profile \( T(x) \) from source to drain in the FinFET device, as described below

\[
T(x) = \begin{cases} 
T_1(x, 0) & \text{for } -X_D \leq x \leq -X_{\text{DE}} \\
T_2(x) & \text{for } -X_D \leq x \leq X_D \\
T_3(x, 0) & \text{for } -X_{\text{SE}} \leq x \leq X_D \\
T_4(x, 0) & \text{for } X_{\text{SE}} \leq x \leq X_D 
\end{cases}
\] (29)

The eight unknown coefficients used in (4), (13), (14) and (23), which are required for complete specification of steady-state temperature profile as defined in (29), can be obtained by solving the following temperature and heat-flux continuity boundary conditions for different regions of the device

\[
T(-X_{\text{DE}}, 0) = T(-X_{\text{DE}}) \] (30)

\[
T(-X_{\text{D}}) = T(-X_{\text{D}}) \] (31)

\[
T(X_{\text{G}}, 0) = T_4(X_{\text{G}}) \] (32)

\[
T(X_{\text{DE}}) = T_4(X_{\text{DE}}) \] (33)

\[
K_1 \frac{\partial T_1(x, y)}{\partial x} \bigg|_{(x, y) = (0, 0)} = K_2 \frac{\partial T_2(x)}{\partial x} \bigg|_{x = -X_{\text{DE}}} \] (34)

\[
K_2 \frac{\partial T_2(x)}{\partial x} \bigg|_{x = X_{\text{G}}} = K_3 \frac{\partial T_3(x, y)}{\partial x} \bigg|_{(x, y) = (0, 0)} \] (35)

\[
K_3 \frac{\partial T_3(x, y)}{\partial x} \bigg|_{(x, y) = (0, 0)} = K_4 \frac{\partial T_4(x)}{\partial x} \bigg|_{x = X_{\text{DE}}} \] (36)

\[
K_4 \frac{\partial T_4(x)}{\partial x} \bigg|_{x = X_{\text{DE}}} = K_4 \frac{\partial T_4(x)}{\partial x} \bigg|_{x = X_{\text{G}}} \] (37)

where \( K_i \) denotes the thermal conductivity of the material in the \( i \)th region, as mentioned before.

E. Comparison With 3-D Thermal Simulations

Fig. 13(a) shows a comparison between the steady-state temperature profiles obtained using our analytical model and the ANSYS simulation. It can be observed that the analytical model agrees with the simulator in general, but there is an overestimation of temperature in CH and DE regions and an underestimation in source and drain regions. The observed error between the analytical model and ANSYS simulation is possibly because of the approximations made in the previous subsections: 1) the heat generation in CH region is moved to the Gaussian distribution near CH/DE boundary in DE region, which makes the peak temperature higher, and 2) the 2-D approximation is used in obtaining the effective thermal conductivity of BOX layer for SE, DE and gate regions, which underestimates the heat spreading from SE, DE or gate regions, and thereby overestimates the temperature in CH and DE regions and underestimates in source and drain regions.

Accuracy of the analytical method is further verified by comparing the peak temperature with ANSYS simulations under the variations in device parameters. All the device parameters and Joule heating profile, other than the one that is being varied, are kept constant at the values in Table I and
V. TRANSIENT THERMAL MODEL

The steady state heat transfer equations, formulated in the previous section for various regions of the SOI FinFET device, can be extended to the transient domain with the addition of heat capacity terms for the Si-fin region and the BOX layer. For example, the transient heat transfer equation for the SE and DE regions (regions 2 and 4) can be written from its steady-state counterpart (1) as follows:

\[
-T_i H_{fin} W_{fin} \frac{\partial^2 T_i(x, t)}{\partial x^2} - \frac{K_{BOX} W_{BOX}}{4} \frac{\partial T_i(x, t)}{\partial t} - \rho_{BOX} c_{p, BOX} A_{BEF,i} \frac{\partial T_i(x, t)}{\partial t} = -q_{g,i}(x, t) K_i
\]

where \( t \) denotes the time; \( T_i(x, t) \) and \( q_{g,i}(x, t) \) are transient equivalents of previously defined variables; \( \rho_i \) and \( c_{p,i} \) are the density and specific heat capacity of the material in the \( i \)-th region (silicon, in this case); \( \rho_{BOX} \) and \( c_{p, BOX} \) denote the density and the specific heat of the BOX layer; \( \rho_{BOX} c_{p, BOX} A_{BEF,i} \) denotes the effective heat capacity per unit length of BOX and ILD.

Equation (38) indicates a lumped RC thermal circuit in YZ plane. It can be simplified to the following form:

\[
\frac{\partial^2 T_i(x, t)}{\partial x^2} - m_i^2 \left( T_i(x, t) - T_0 + \tau_i \frac{\partial T_i(x, t)}{\partial t} \right) = -q_{g,i}(x, t) K_i
\]

where \( \tau_i \) is the time constant in region \( i \), which can be obtained from the 10%-90% rise time in 2-D transient ANSYS simulation (\( \tau_i = \tau_{90\% - 90\%} / 2.2 \)). Equation (39) can be transformed into the frequency domain to yield the following equation:

\[
\frac{\partial^2 T_i(x, \omega)}{\partial x^2} - m_i^2(1 + j \omega \tau_i)(T_i(x, \omega) - T_0) = -q_{g,i}(x, \omega) K_i
\]

where \( \omega \) is the radial frequency. If we assume a step function of \( t \) for \( q_{g,i}(x, t) \), i.e., \( q_{g,i}(x, t) = \begin{cases} 0, & t < 0 \\ q_{g,i}, & t \geq 0 \end{cases} \right. \), then \( q_{g,i}(x, \omega) = 2q_{g,i} \omega \).

It can be observed that the above frequency domain model is very similar in its form to its steady-state counterpart (2), except for the addition of the \( \omega \tau_i \) term, which is independent of the spatial coordinates \((X, Y \) and \( Z)\) in which the heat equation is being solved. Hence, this equation can be solved in the frequency domain using the approach used for solving the steady-state equation and the solution can be converted back through inverse Fourier transform to give the transient temperature profile.

A similar approach can be followed for solving the transient heat equation in all the other regions of the device to yield the

---

1In general, converting a Maple script to compiled C or Fortran should speed it up significantly.

2The parameter sets we have used for the simulation in Fig. 13 (see Table I) indicate a very low thermal conductivity ILD, which are consistent with our previously published EEDM paper (Ref. [38]). In contrast to the low thermal conductivity low-K dielectrics used as back-end level ILD, front-end level ILDs (dielectrics surrounding the FinFETs) used in industrial processes employ SiO2 or even Si3N4, which has much higher thermal conductivity.
VI. APPLICATION OF TRANSIENT THERMAL MODEL FOR EOS/ESD SITUATIONS

Fig. 15 shows the transient thermal response of an SOI FinFET, with the dimensions given in Table I, predicted by the transient analytical model as compared to ANSYS simulations. The transient analytical model is less accurate than the steady-state analytical model. This incremental discrepancy is due to the lumped RC thermal circuit approximation (in YZ plane for regions 2 and 4, and in XZ plane for region 3) made in Section V: the heat dissipation and storage is actually a distributed RC thermal circuit. The transient thermal model has also been applied to derive the power-to-failure versus time-to-failure profiles and compared with ANSYS simulations as shown in Fig. 16 (failure is under the assumption that the maximum temperature reaches the melting point of Si). It was observed (as shown in Fig. 15) that the thermal time constants for SOI FinFETs of typical dimensions are of the order of 10–30 ns. Hence, in the ESD situations with pulse-widths more than 100 ns, failure prediction should be done assuming steady-state conditions for conservative estimates.

VII. APPLICATION FOR PRODUCTION STRUCTURES

The FinFET structures of production level [37] have several divergences from the structure of research level given in Figs. 3 and 4, including the introduction of metal gate [38], body-tied bulk FinFET [15], trench contacts [39], and so on. In this section, we analyze the steady-state thermal profile for each divergence separately and we illustrate the impact on the temperature profile or the peak temperature.

The analysis of the impact from metal gate is simple. The only thing to be changed in either simulation or analytical model is the thermal conductivity of the gate. In Fig. 17, we change the gate thermal conductivity to 100 W/m-K, as compared to 45 W/m-K in the assumption in Fig. 13. Comparing Fig. 17 with Fig. 13, the effect of employing metal gate instead of polysilicon gate is found to be marginal from thermal point of view, because the heat dissipation in gate region is primarily determined by the (lower thermal conductivity) surrounding dielectrics, not the (higher thermal conductivity) gate.

In the case of bulk FinFET, the BOX under thin-Si regions (source, SE, CH, DE and drain regions) is changed to high resistive silicon (HR-Si), while the BOX not under the thin silicon regions remains unchanged. The majority of the electrical current is still confined within the thin-Si regions, and not the HR-Si regions. Therefore, the heat generation within HR-Si regions is negligible. The analytical model for bulk FinFET is slightly different from that for SOI FinFET. Similar expressions are still valid for regions 2, 3 and 4 (Fig. 9). For regions 2 and 4, $K_i(j = 2$ or 4) in (1), (35), and (36) should
be replaced with $K_{eff,i}$, where

$$K_{eff,i} = \frac{K_i H_{fin} + H_{BOX}}{H_{fin}}.$$  \hspace{1cm} (41)

The reason behind this change is that the lateral heat dissipation in the HR-Si regions can no longer be ignored, since the HR-Si regions contribute a similar level of heat conduction as that of thin-Si regions. The temperature in regions 1 and 5 can be assumed to be the bulk temperature $T_b$, since they are directly connected through high thermal conductivity silicon. Therefore, (30) and (33) should be changed as

$$T_1(x, y) = T_b(x, y) = T_b$$  \hspace{1cm} (42)

and

$$T_7(x, y) = T_b(x, y) = T_b$$  \hspace{1cm} (43)

respectively. The coefficients $C_1, D_2, C_3, D_4, \beta_1^i$ and $\beta_2^i$ can be obtained from (31), (32), (42), (43) and modified versions of (35) and (36) with (41). The temperature profile for bulk FinFET is obtained using both ANSYS simulation and analytical model, which is also shown in Fig. 17. Good agreement has been achieved. Bulk FinFET exhibits significant advantage in thermal performance compared to SOI FinFET, as compared to the result from Fig. 13. It should be noted that our model predicts the temperature for bulk FinFETs better than that for SOI FinFETs. This can be explained by the validity of the 2-D approximation of fringing heat dissipation through ILD and BOX in regions 2 and 4 in our analytical model. While this approximation is clear in terms of the physics, it is not perfectly accurate. In bulk FinFETs, the heat dissipation is mostly constrained in the high-resistive silicon (HR-Si) region and the fringing heat dissipation through ILD and BOX is much less important. In SOI FinFETs, however, the fringing heat dissipation through ILD and BOX is more dominant.

The production architecture of trench contacts for FinFETs, which is proposed by Intel for reducing parasitic capacitance, is shown in Fig. 19. The various regions of the FinFET structure with trench contacts is shown in Fig. 19. The analytical model given in Section IV is generally valid for the FinFETs with trench contact. However, an effective BOX thermal conductivity ($K_{BOX,i}$) should be used for regions 1 and 5, since the length of the contact can be small and the fringing heat dissipation can be nonnegligible. $K_{BOX,i}$ can be extracted from ANSYS 2-D simulation, similar to the way in Sections IV-A and IV-C. In such a way, the fringing heat dissipation in regions 6 and 7 is included in the formulation for regions 1 and 5, respectively. In addition, $K_{eff,i}$ ($i = 1$ or 5) in Section IV-B and equations (34) and (37) should be replaced with the trench contact effective thermal conductivity $K_{eff,i}$

$$K_{eff,i} = \frac{K_{i,holes}/H_{fin}}{H_{fin}}.$$  \hspace{1cm} (44)

where $H_{holes}$ is the height (thickness of metal) of the trench contact. With these modifications, the analytical model in Section IV can be used.

![Fig. 18. Intel’s trench contact FinFET architecture.](image)

The comparison of predicted peak temperature by the modified analytical model and ANSYS simulation is shown in Fig. 20. Good agreement is achieved. It can be observed that reducing the widths of the source and the drain from 550 nm to 50 nm, increases the peak temperature slightly (about 30 K) under the same power consumption. At widths above 300 nm, the peak temperature tends to be saturated.

**VIII. Extension of Self-Heating Model to Account for Thermal Boundary Resistance**

According to [41], when there is heat flow across the boundary between two materials, there must be a temperature discontinuity between the two sides of the interface. The ratio of the temperature discontinuity across the interface to the heat flow per unit area across the interface is defined as the TBR, which is the inverse of the thermal boundary conductance. A
typical thermal boundary conductance of 1000W/(m² K) [42]–[44] is equivalent to a thermal conductive film with 10nm thickness and 1W/m·K thermal conductivity.

It is straightforward to account for TBR in regions 1, 2, 4 and 5 in our FinFET analytical thermal model. Specifically, the effective thermal conductivity of BOX layer (KB_eff,i) should take into account the contribution from TBR by extracting KB_eff,i from a proper cross-sectional 2-D simulation. In fact, the BOX thermal conductivity of 0.8W/m·K, which is used in the test-case setups in this paper, can be treated as an equivalent thermal conductivity from both the bulk thermal conductivity (1.3W/m·K) and the TBR of upper and lower boundaries of BOX (2.4 × 10⁻³ m²K/W). Nevertheless, to correctly extract KB_eff,i, proper cross-sectional 2-D simulation with TBR setup at the boundaries should be performed rather than assuming a bulk effective BOX thermal conductivity of 0.8W/m·K, due to the fringing heat dissipation. However, in this section, we skipped changing values of KB_eff,i, since such changes would not affect our modeling framework.

On the other hand, it requires a modified analytical model to take into account the TBR associated with region 3. Specifically, the ultrathin gate oxide can induce a large TBR and significantly impede the heat dissipation. In the rest of this section, we re-examine our research level SOI FinFET and analytical model for the SOI FinFET with consideration of TBR between channel-fin and gate.

Neglecting the temperature variation along x in the gate region, the heat transfer equation in the gate region can be expressed as

\[ T_{3g}(y) = T_{3g0} + \frac{K_{3g}W_{3g}H_{gate}K_{3g}}{H_{BOX}} \left( T_{3g0} + \frac{W_{finH_{fin}}}{R_{3c}} \right) \]

where \( T_{3g0} \) is the temperature of Si substrate, \( K_{3g} \) is the thermal conductivity from the gate through the BOX to the Si substrate, \( H_{gate} \) is the gate-height, and \( L_g \) has been defined in Fig. 5. In (45), we neglect the heat generation in the CH region. On the other hand, we move such heat generation into the Gaussian distribution shaped heat generation in the DE region, similar to what was done in Section IV-C. In addition, we treat \( T_{3g0} \) independent of \( x \), as a first-order approximation.

The solution to (45) can be expressed as

\[ T_{3g}(y) = T_{3g0} + C_3 e^{-\frac{y}{W_{SD}}} + D_3 e^{-\frac{y}{W_{SD}}} + \frac{(C_3W_{SD})^2 - (y - y_0)^2}{(C_3W_{SD})^2 - (y - y_0)^2} / R_{3c} \]

where \( C_3 \) and \( D_3 \) are unknown coefficients that are determined from temperature and heat-flux continuity boundary conditions (just like \( C_1 \) and \( D_1 \) defined in Section IV-C), and

\[ m_{3g} = \sqrt{\frac{2H_{3g}W_{3g}H_{gate}K_{3g}}{K_{3g}W_{fin}H_{fin}}} \]

Neglecting the temperature variation along \( x \) in the gate region, the heat transfer equation in the gate region can be expressed as

\[ K_{3g}H_{gate}L_g \frac{d^2T_{3g}(y)}{dy^2} = K_{3g}H_{gate}L_g (T_{3g}(y) - T_{3g0}) \]

where \( T_{3g}(y) \) is the temperature distribution along the gate, \( K_{3g} \) is the thermal conductivity of Si-fin in region 3c, \( K_{3g} \) is the effective thermal conductivity from the Si-fin channel through the BOX to the Si substrate, \( T_{3g0} \) is the temperature near the gate oxide of the gate region, and \( H_{gate}, W_{3g} \) and \( R_{3c} \) have been defined in Fig. 5. In (48), we neglect the heat generation in the CH region. On the other hand, we move such heat generation into the Gaussian distribution shaped heat generation in the DE region, similar to what was done in Section IV-C. In addition, we treat \( T_{3g0} \) independent of \( x \), as a first-order approximation.

The solution to (48) can be expressed as

\[ T_{3g}(y) = T_{3g0} + \frac{C_3 e^{-\frac{y}{W_{SD}}} + D_3 e^{-\frac{y}{W_{SD}}} + \frac{(C_3W_{SD})^2 - (y - y_0)^2}{(C_3W_{SD})^2 - (y - y_0)^2}} \]

where \( C_3 \) and \( D_3 \) are unknown coefficients that are determined from temperature and heat-flux continuity boundary conditions (just like \( C_1 \) and \( D_1 \) defined in Section IV-C), and

From the nine ([30]–[31]) boundary condition equations, the unknown coefficients can be obtained.
We re-examined the test case of Fig. 13 with the modification that we consider a TBR between the channel-fin and gate. The results from our analytical model are compared with the ANSYS simulation in Fig. 21, which show good agreement. The peak temperature rise in FinFET with consideration of TBR is proportional to the square of the feature size.

IX. Conclusion

In conclusion, an accurate analytical thermal model was developed, for the first time, for estimating self-heating in FinFETs (or Tri-gate FETs) under both steady-state and transient conditions. The model was verified against finite element simulations. Our analysis indicated that TBR should be considered for accurate self-heating estimation in advanced FinFET structures. The analytical modeling framework for FinFET self-heating developed in this work also established a platform for building more compact models for circuit-level simulation and analysis.

REFERENCES


[34] E. Pop, K. Banerjee, P. Sverdrup, B. Dutton, and K. Goodson, “Localized heat sources,” in IBM’s Electromagnetic Field Solver Suite of Tools called the EIP tools. Solving Award Based on the Use of the EIP Tool Suite,” which recognizes exceptional conference papers. He was the recipient of the “2008 IBM Problem Solving Award” based on the Use of the EIP Tool Suite, which recognizes outstanding contributions by students in solving the most interesting problems using IBM’s Electromagnetic Field Solver Suite of Tools called the EIP tools.


[37] M. Bohr and K. Misty. Intel’s Revolutionary 22 nm Transistor


Chuan Xu (S’08–M’12) received the B.S. degree in microelectronics in 2004 the M.S. degree in microelectronics and solid-state electronics in 2007, both from Peking University, Beijing, China, and the Ph.D. degree in electrical and computer engineering from the University of California at Santa Barbara, Santa Barbara, in 2012. In the summer of 2008, he was an Intern with the Electrical Intersocket and Packaging Group, IBM T. J. Watson Research Center, Yorktown Heights, NY, USA. During the summers of 2009–2011, he was with the Calibre Division, Mentor Graphics Corporation, St. Louis, France. In June 2012, he joined the Technology Development and Innovation Group in Marlin Integrated Products, Beaverton, OR. His current research interests include modeling of active devices, interconnects, and packages in very large scale integrated (VLSI) circuits.


Kamakshi Endo (M’91) received the Ph.D. degree in electrical engineering from Waseda University, Tokyo, Japan, in 1999. From 1993 to 2003, he was with Silicon Systems Research Laboratories, NEC Corporation, where he worked on the research and development of multilevel interconnects and high-k gate-stack technologies for ULSI. From August 1999 to August 2000, he was a Visiting Scholar at the Center for Integrated Systems, Stanford University, Stanford, CA, USA. He is currently a Senior Researcher with the Silicon Nanoscale Devices Group, National Institute of Advanced Industrial Science and Technology, Tsukuba, Japan. His current research interests include nanomaterials manufacturing for aggressively scaled multilevel devices in advanced VLSI technologies.

Kunstav Banerjee (S’02–M’09–SM’13–F’12) received the Ph.D. degree in electrical engineering and computer sciences from the University of California, Berkeley, in 1999. In 2002, he joined as a Faculty Member with the Department of Electrical and Computer Engineering, University of California, Santa Barbara (UCSB), where he has been a Full Professor since 2007. Prior to joining UCSB, he was a Research Associate with the Center for Integrated Systems, Stanford University, Stanford, CA from 1999 to 2001. His current research interests include nanomaterials and emerging nanoelectronics. He is intricately involved in exploring the physics, technology, and applications of low-dimensional nanomaterials for next-generation green electronics, photonics, and bioelectronics.

Seshadri Kolluri (S’07–M’12) received the B.Tech. degree in electrical engineering from the Indian Institute of Technology, Kanpur, India, in 2004, and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of California, Santa Barbara, in 2007 and 2011, respectively. He is currently with Efficient Power Conversion Corporation, El Segundo, CA. He has previously worked at Micron Technology Inc., San Jose, CA, USA, and Neomorphic Semiconductor Pvt. Ltd., Noida, India. His current research interests include nanoscale semiconductor device fabrication, characterization, modeling, and gallium nitride electronic devices.

His current research interests include nanoscale semiconductor device fabrication, characterization, modeling, and gallium nitride electronic devices.