A Physically Based Compact Model of Partially Depleted SOI MOSFETs for Analog Circuit Simulation

Mike S. L. Lee, Bernard M. Tenbroek, Member, IEEE, William Redman-White, Member, IEEE, James Benson, and Michael J. Uren

Abstract—In this paper, the Southampton Thermal Analogy (STAG) compact model for partially depleted (PD) silicon-on-insulator (SOI) MOSFETs is presented. The model uses a single expression to model the channel current, thereby ensuring continuous transition between all operating regions. Furthermore, care has been taken to ensure that this expression is also infinitely differentiable, resulting in smooth and continuous conductances and capacitances as well as higher order derivatives. Floating-body effects, which are particular to PD SOI and which are of concern to analog circuit designers in this technology, are well modeled. Small geometry effects such as channel length modulation (CLM), drain-induced barrier lowering (DIBL), charge sharing, and high field mobility effects have also been included.

Self-heating (SH) effects are much more apparent in SOI devices than in equivalent bulk devices. These have been modeled in a consistent manner, and the implementation in SPICE3f5 gives the user an additional thermal node which allows internal device temperature rises to be monitored and also accommodates the modeling of coupled heating between separate devices. The model has been successfully used to simulate a variety of circuits which commonly cause problems with convergence. Due to its inherent robustness, the model can normally achieve convergence without recourse to the setting of initial nodal voltage estimates.

Index Terms—Analog integrated circuits, circuit simulation, MOSFETs, semiconductor device modeling, silicon-on-insulator technology.

I. INTRODUCTION

In addition to its established niche markets of radiation-hard and high-temperature applications, silicon-on-insulator (SOI) CMOS is increasingly viewed as a serious prospect for mainstream VLSI. Lower parasitic capacitances lead to higher speeds at lower power levels, and packing densities can also be improved. Starting materials are becoming more readily available and reliable and many large scale designs have been reported [1]. Technologies generally divide into those where the silicon below the transistors is fully depleted, and those where there is a neutral body region between the channel and the buried oxide (partially depleted). The latter type is at present more reliably manufacturable, due to practical limitations in the control of the very thin film thicknesses needed for fully depleted devices [2]; it is to this type that the modeling work of this paper is addressed.

The electrical behavior of partially depleted (PD) SOI devices where the body node of the device is left floating has been extensively studied over many years [3]. Static and transient variations in body potential due to capacitive coupling and ionization can lead to data history sensitivity in DRAMS [4] and clock frequency dependent propagation delay in inverter circuits [5], [6]. In logic circuits, these problems can often be accommodated within design tolerances. The effects of the floating body are far more serious for analog design; the well-known kink effect [3] leads to large sensitivity to drain bias, and more seriously, the drain conductance exhibits bias and frequency dependent variations [7], often of more than an order of magnitude within normal operating regimes. These problems can be reduced by the use of a body tie, at the cost of some additional area and parasitic capacitance. However, unavoidable series resistance in the contact often make this approach only a partial solution, and significant deviations from ideal behavior can be observed even when a body tie is present [8].

A further deviation from the bulk case is the need to consider the self-heating of individual devices. This arises from the poor thermal conductivity of the underlying buried oxide, and device channel temperatures can rise dynamically many tens of degrees above ambient during normal operation [9]. Consequently, device characteristics become significantly modified, with negative output conductance often observable at higher drain currents [10]. Such heating is not instantaneous, but has a transient response to bias changes, with typical time constants of the order of $10^{–5}$ s [11]. While there do exist some issues relating to accurate parameter extraction (insofar as the behavior without self-heating is not directly observable, and hence the electrical and thermal contributions are not easily separated), thermal behavior is not generally significant for digital circuits, due to the low mean power dissipations, and clock frequencies normally being well above the thermal time constants [12]. Analog circuits, however, can be significantly affected. Output conductances can be low or even negative at low frequencies and then rise with frequency [13], leading to unforeseen gain and phase
variations [14]. Matching can also be disturbed by heat flow from adjacent devices [14].

Digital designers have a route for circuit simulation, either by fitting bulk models, or by using a SOI model suitable for digital applications, such as UFSOI [5], [15]. To make mixed signal design in SOI CMOS a practical proposition, the analog designer requires models which can not only handle the special physics of this technology, but which also satisfy more stringent criteria such as a high order of equation continuity and accurate conductance modeling. One candidate for analog design is BSIMSOI [16]. This takes the more traditional approach of using an explicit threshold rather than the increasingly popular surface potential approach [5], [17]–[20], and uses parameter binning for width and length dependencies, rather than the ideal of a physically based model and parameter set.

In this paper, we present the Southampton Thermal AnaLogue (STAG) SOI CMOS model. In developing this model, we have paid particular attention to the handling of body node and thermal behavior in the context of a highly continuous and robust MOS formulation, using some conventional techniques where appropriate. The model has been kept physically based wherever practical, to permit straightforward parameter extraction without excessive fitting procedures. The recommended procedures for extracting the parameters used in the STAG model are discussed in some detail in [21].

The target technology was a 0.7-μm process; some deep-submicron effects have not yet been included. The paper is structured as follows: in Section II we present the core model based on a surface potential derivation of channel current. We then present the auxiliary model which handles geometry dependence and body effects. In Section IV we describe how self-heating is incorporated into the model itself (rather than as an external macro model). Section V describes the evaluation of the model against measurements obtained from individual devices and simple analog circuits, and conclusions are presented in Section VI.

II. CORE MODEL

The model is presented for an n-channel device, though a similar analysis can be performed for a pMOS device if the polarity of the voltages and the direction of currents are reversed.

For a truly partially depleted SOI MOSFET, there is no back gate coupling effect [22]. Furthermore, in the case of analog circuits, the back gate (or substrate) is invariably tied to ground. With a threshold voltage of typically of the order of tens of volts, the back device is operating deep within the subthreshold regime and will pass negligible current. This also implies there is negligible charge at the back interface, as well as a much smaller gate capacitance due to the thicker oxide, leading to negligible impact on transient behavior.

With the justifications above, the formulation of the STAG model makes the assumption that the back device has no significant effect on the total device behavior. Thus the initial analysis for the STAG model proceeds as for bulk MOSFETs.

A. Channel Current

Fig. 1 shows a schematic of an SOI MOSFET and the direction of the displacement variables used in this analysis. Using standard assumptions (unipolar device, infinitesimally thin inversion layer, single direction of current flow [23]–[25]), the charge sheet model expression for a MOSFET may be obtained

\[ I_{CH}(y) = -W \mu_s \left( q_c(y) \frac{d\psi_s(y)}{dy} - \phi_y \frac{d\psi_y(y)}{y} \right) \] (1)

where \( I_{CH} \) is the channel current, \( W \) is the channel width, \( \mu_s \) is the surface mobility of the carriers, \( q_c \) is the channel charge per unit area, and \( \psi_y \) is the surface potential. By assuming that the current is constant along the channel, and that the mobility can be represented by a constant value \( \mu_s \) at some average gate and drain field, (1) may be integrated to give an equation of the general form [25]

\[ I_{CH} = \frac{W}{L} \mu_s C_{ef} \left( f(\psi_{SL}) - f(\psi_{SO}) \right) \] (2)

where \( L \) is the effective channel length, \( C_{ef} \) is the front gate capacitance per unit area, and \( \psi_{SL} \) and \( \psi_{SO} \) are the surface potentials at the drain and source end of the channel, respectively. The function \( f \) will depend on the exact expression used to model \( q_c \) in terms of the surface potential \( \psi_y \).

B. Channel and Body Charge

The channel charge consisting of mobile carriers and the immobile depletion charge together constitute the total charge in the body of the PD SOI MOSFET. This relationship may be expressed as

\[ q_{tot} = q_c + q_b \] (3)

where \( q_{tot} \) denotes the total charge density within the body of the device and \( q_b \) is the density of the depletion charge. By applying Gauss’ Law and the potential balance equation across the front gate oxide, we get

\[ V_{GFB} = V_{FB} + \eta_b \psi_y(y) = \frac{q_{tot}(y)}{C_{ef}} \] (4)

where \( V_{GFB} \) is the front gate-body voltage, \( V_{FB} \) is the front flat band voltage, and \( \eta_b \) accounts for the influence of the fast surface states at the silicon–oxide interface [23]. By using (4) in (3), the channel charge density can be expressed in terms of the terminal voltages and the surface potential if \( q_b \) is known.
1) Body Charge: This is analogous to the bulk charge in a standard MOSFET and (by using the depletion approximation) is generally taken to be [23], [24]

\[ q_b(y) = -\gamma C_{ox} \sqrt{\psi_s(y)} \tag{5} \]

However, the presence of the square root term causes problems when deriving expressions for the saturation voltage and intrinsic capacitances [23]. In the STAG model, \( q_b \) has been linearized about an initial estimate for the surface potential at the source end \( \psi_{sto} \)

\[ q_b(y) = -\gamma C_{ox} \left\{ \sqrt{\psi_{sto}} + \delta(\psi_s(y) - \psi_{sto}) \right\} \tag{6} \]

where

\[ \delta = \frac{1}{2\sqrt{1 + \psi_{sto}}} \tag{7} \]

\( \delta \) has been chosen to be well defined even for \( \psi_{sto} = 0 \). The exact value of \( \psi_{sto} \) will be given later.

2) Channel Charge: By combining (3), (4), and (6), we obtain an expression for the channel charge as

\[ q_c(y) = -C_{ox} \left( V_{GB} - V_{FB} - \eta_s \psi_s(y) \right) - q_b(y) \]

\[ = -C_{ox} \left[ V_{GT} - \alpha \psi_s(y) \right] \tag{8} \]

where

\[ V_{GT} = V_{GB} - V_{FB} - \gamma \left( \sqrt{\psi_{sto}} - \delta \psi_{sto} \right) \tag{9} \]

\[ \alpha = \eta_s + \gamma \delta \tag{10} \]

Using (8), the function \( f \) in (2) can now be written as

\[ f(\psi_s) = \left\{ V_{GBT} - \frac{\alpha}{2} \psi_s \right\} \psi_s \tag{11} \]

where

\[ V_{GBT} = V_{GT} + \psi_s \tag{12} \]

All that now remains is for the estimate \( \psi_{sto} \) and the surface potentials at the source and drain ends (\( \psi_{sto} \) and \( \psi_{sL} \)) to be found.

C. Ideal Surface Potential

In most modern circuits, the accumulation region \( (V_{GB} < V_{FB}) \) is seldom used. By neglecting this region, the solution to the one-dimensional Poisson’s equation (applied across the interface between gate oxide and the body) can be simplified to [24]

\[ \psi_s(y) = 2\phi_F + V_{ch}(y) \]

\[ + \phi_F \ln \left\{ \frac{1}{\phi_F} \left[ \frac{1}{2\sqrt{\left( V_{GB} - V_{FB} - \eta_s \psi_s(y) \right)^2 - \psi_s(y)} \right] \right\} \tag{13} \]

where \( \phi_F \) is the Fermi-potential of the device, \( V_{ch} \) is the channel potential with \( V_{ch}(0) = V_{GB} \) and \( V_{ch}(L) = V_{DB} \) in strong inversion, \( \phi_F = kT/q \) is the thermal voltage, and \( \gamma = \sqrt{2e^2qN_A/C_{ox}} \) is the body factor. This is an implicit equation and no closed-form solution for the surface potential can be derived.

Various methods for obtaining the solution to (13) have been proposed. Numerical techniques can be employed [17], but they involve iterative algorithms which can be computationally expensive. Storing precalculated solutions in a two-dimensional (2-D) array and using interpolation for points in between stored values has also been suggested [26]. Both of these approaches require the partial derivatives of \( \psi_s \) (which are needed for the Newton–Raphson technique employed by most circuit simulators) to be derived numerically, thus consuming additional CPU time. A third approach is the approximation of the solution using cubic spline functions [27]. This method is more efficient computationally as the derivatives can be obtained analytically, but it only guarantees continuity of derivatives up to second order. The approach used in the STAG model to tackle this challenging problem is presented in the next section.

D. Surface Potential Determination

The use of a function which approximates the true solution to (13) has been proposed for standard MOSFETs [18], [28]. The technique involves finding very good approximate solutions in the subthreshold and strong inversion regions and then joining them in a smooth manner. Unfortunately, the proposed function in [28] possesses a discontinuous first derivative, while the function presented in [18] possesses continuity up to the first derivative only. It is now well understood that a single function, valid in all regions of operation, and with a continuous first-order derivative, is very desirable for use in circuit simulators if good convergence behavior is to be obtained [17], [29], [30]. Furthermore, continuity in higher order derivatives is also desirable if certain analyses are to be performed (e.g., third-order intermodulation studies [31]). The function developed for use within the STAG model will be shown to be both continuous and also infinitely differentiable.

1) Continuity of Model Equations: Problems often occur within models due to the fact that standard simplifications in device physics during the initial model development usually lead to device equations containing singularities or other numerical problems. Examples are division by quantities which could become zero for certain bias conditions, and arguments of square roots or logarithm functions which could become negative. In the STAG model, singularities and other problem points are identified in all expressions and transform functions are then used to prevent numerical problems. The basic transform function used throughout is

\[ x' = \epsilon \ln \left\{ 1 + \exp \left( \frac{x}{\epsilon} \right) \right\} \tag{14} \]

2) Subthreshold: In this region, the channel charge is negligible compared with the body charge, and \( \psi_s \) is almost constant. Denoting \( \psi_s(y) \) by the constant \( \psi_{sto} \), (13) can then be rearranged and simplified to

\[ \psi_{sto} = \left[ -\frac{\gamma}{2\eta_s} + \sqrt{\frac{\gamma^2}{4\eta_s^2} + \frac{V}{\eta_s}} \right]^2 \tag{15} \]
where \( V_g = V_{GFB} - V_{FB} \). Note that \( \psi_{ss} \) is the saturation potential for the charge sheet model without the linearization of the body charge. This is because \( \psi_{ss} \) is the solution for \( q_c = 0 \) with \( q_b = -\gamma C_{ox} \sqrt{\psi} \).

Problems can occur if the term inside the square root becomes negative, but note that if \( V_g < 0 \) then (13) is not valid (the accumulation condition was specifically excluded in order to simplify the one-dimensional solution of Poisson’s equation). However, during Newton–Raphson iterations, the gate voltage may venture into this region. Therefore, (14) is used to ensure that \( V_g \) remains positive.

3) Strong Inversion: In strong inversion, \( \psi_s \) on the right-hand side of (13) can be replaced with \( 2\phi_F + V_{cb} \) as a good estimate.

\[
\psi_{st} = 2\phi_F + V_{cb} + \phi_T \ln \left\{ \frac{1}{\phi_T} \left[ \frac{1}{\gamma} \left[ V_g - \eta_s (2\phi_F + V_{cb}) \right]^2 - (2\phi_F + V_{cb}) \right] \right\}.
\]

This expression causes numerical problems when the argument of the logarithm becomes negative. One way of avoiding this is to take the modulus of this argument [28], but this leads to a discontinuity in the first derivative which may cause poor convergence behavior. Note though that this problem only occurs in subthreshold, a regime for which the expression is no longer a good approximation to the true solution. Therefore, in the STAG model, the gate voltage within the logarithm term is transformed to avoid the problem region.

4) Single-Piece Model: Having obtained good approximations in weak and strong inversion, the following expression can be used to ensure that each is reduced smoothly to insignificance outside of their respective regions of validity, while avoiding any abrupt truncation [28].

\[
\psi_s = \psi_{st} - \phi_T \ln \left( 1 + \exp \left( \frac{\psi_{ss} - \psi_{st}}{\phi_T} \right) \right).
\]

This expression relies on the fact that in strong inversion, \( \psi_{ss} \gg \psi_{st} \), and in subthreshold, \( \psi_{ss} \ll \psi_{st} \). Note also that all the transforms used are continuous and infinitely differentiable, as are \( \psi_{st} \) and \( \psi_{ss} \), and so \( \psi_s \) is itself continuous and infinitely differentiable.

The maximum possible value of \( \psi_s \) for a given \( V_{GFB} \) is \( \psi_{ss} \), i.e., saturation occurs when the channel potential has reached the subthreshold condition. This definition will be used again to define the saturation potential when the effects of carrier velocity saturation are included.

Often the quantity \( \sqrt{\psi_s} \) will be used to model the body/bulk charge and so it is imperative that \( \psi_s > 0 \). Indeed (13), from which we have derived (15) and (16), is only valid for \( \psi_s > 0 \). Consider now the source end of the channel where \( V_{cb} = V_{SB} \). In bulk, \( V_{SB} = 0 \) and so no problems are usually encountered. However, in an SOI MOSFET with a floating body, often \( V_{SB} < 0 \) and this can cause \( \psi_s \) in (17) to become negative, especially during Newton–Raphson iterations. For use in a PD SOI model, the surface potential expression has to be made more robust.

This following transform has the same function as (17) but also ensures that \( \psi_s > 0 \).

\[
\psi_{st} = \phi_T \ln \left[ 1 + \exp \left( \frac{\psi_{ss} - \psi_{st}}{\phi_T} \right) \right].
\]

Fig. 2 shows this single-piece function in comparison with the ideal surface potential. Equation (18) is valid for long devices where the effects of velocity saturation can be neglected. For the source end, however, this equation provides a very good estimate of the true surface potential, and so it is used for the estimate \( \psi_{st} \), i.e.

\[
\psi_{st0} = \psi_{st} (V_{GFB}, V_{SB}).
\]

E. High Field Mobility Effects

When deriving (2), it was assumed that the mobility \( \mu_s \) was constant along the channel. This was done to make the integration of (1) more tractable. In reality, the relative contributions of the transverse and longitudinal electric fields will differ along the length of the channel, and so the mobility will be a function of channel position. To account for this, we use an established effective mobility model [17].

\[
\mu_{eff} = \mu_0 \left[ 1 + \theta \left( \frac{V_g - \psi_{sat} + \psi_{st0}}{2} \right) + \frac{\mu_0}{\psi_{sat}} \left( \frac{\psi_{sat} - \psi_{st0}}{L} \right)^{-1} \right].
\]

where \( \mu_0 \) is the low field mobility, \( \theta \) is the vertical field mobility degradation parameter, and \( \psi_{sat} \) is the carrier saturation velocity. The first term in (20) accounts for the normal field influence and the second term accounts for the velocity saturation of carriers. This expression is included in the core model because it alters the saturation potential, which will be discussed in the next section.

F. Saturation Model

The basic charge sheet model is not suitable for use with small geometry devices due to the fact that inclusion of carrier velocity
saturation greatly reduces the saturation voltage and hence the current drive of the device in saturation. The basic concept of defining a saturation potential [17] for the drain end is used in the STAG model. This is, in effect, the maximum value that $\psi_{sL}$ can attain in the model. Assuming that $V_{DS} \geq 0$, the surface potential at the drain end cannot be less than that at the source end, being $\psi_{s0} \approx \psi_{st0}$. Also, if no velocity saturation effect were to be included, then the drain saturation potential would be attained when the channel charge density $q_s$ vanishes. Using these criteria and the linearized channel charge expression in (8), the upper bound of the saturation potential is found to be $V_{GT}/\alpha$. Therefore, the saturation potential including high field mobility effects may be defined as [17]

$$\psi_{sL_{sat}} = \psi_{st0} + \frac{V_{GT}}{\alpha} - \frac{2}{\alpha S}$$

(21)

where $S \geq 1$ is a quantity dependent on mobility parameters. For no velocity saturation, $S = 1$, while for carriers with a low saturation velocity, $S \rightarrow \infty$.

An expression for $S$ can now be obtained if the saturation condition is defined as that at which the drain conductance in the intrinsic model vanishes, i.e., $dI_{CH}/dV_{DS} = 0$. Because we must include high mobility effects when deriving an expression for the saturation condition, we do not use the unmodified channel current expression defined in (2). Instead, we replace the constant mobility $\mu_s$ in (2) with the effective mobility $\mu_{eff}$ defined in (20). Differentiating this high field current expression leads to the following result for $S$:

$$S = \frac{1}{2} \left( 1 + \sqrt{1 + \frac{2M_{mothrop}}{1 + \theta V_{GET}}} \right)$$

(22)

where

$$M_{mothrop} = \frac{\mu_0}{\psi_{sat} L} - \frac{\theta}{2}$$

$$\psi = \frac{V_{GT}}{\alpha} - \psi_{st0}$$

$$V_{GET} = V_{g} - \psi_{st0}$$

The saturation potential accounting for high field mobility effects is plotted in Fig. 3 in comparison with the ideal surface potential and $\psi_{ss}$.

G. Surface Potential Including High Field Mobility Effects

Analogous to (18), but using $\psi_{sL_{sat}}$ as the upper bound in order to account for high field mobility effects, the surface potential in the STAG model is given as

$$\psi_{sL_{STAG}} = \phi_t \ln \left[ 1 + \frac{\exp \left( \frac{\psi_{sat}}{\phi_t} \right)}{1 + \exp \left( \frac{\psi_{sat} - \psi_{sL_{sat}}}{\phi_t} \right)} \right].$$

(23)

The surface potential at the source and drain end are defined respectively as

$$\psi_{s0} = \psi_{sL_{STAG}}(V_{GB}, V_{DB})$$

$$V_{GB} = V_{gs}$$

$$V_{DB} = V_{ds}$$

Fig. 3. Saturation potential $\psi_{sL_{sat}}$ compared to $\psi_{ss}$ for $v_{sat} = 10^7$ and $\theta = 0.08$ (typical values).

III. AUXILIARY MODEL

The core model accounts for the intrinsic device including high field mobility effects, particularly in terms of the altered saturation voltage due to velocity saturation of carriers. In this section, the auxiliary model is presented, which handles the modeling of small geometry and floating-body effects. The dynamic charge equations also form part of the auxiliary model, but due to space limitations, discussion of the dynamic model is outside the scope of this paper. The approach taken closely follows that used in the SUSOS model [18]. Charge nonconservation is avoided by modeling the transient currents as the time derivatives of the nodal charges, as shown in Fig. 5.
A. Small Geometry Effects

The well-known short-channel and narrow-width effects are modeled via a modified body factor \[ \gamma_{\text{eff}} = \gamma \cdot \left( 1 - \frac{\Delta L,_{e}}{L_{c}} \right) \cdot \left( 1 + \frac{\Delta W}{W_{c}} \right) \] (25)

where \( \Delta L_{e} \) and \( \Delta W \) are left as model parameters. The modified \( \gamma_{\text{eff}} \) is then used throughout in the evaluation of the surface potentials.

The drain-induced barrier lowering (DIBL) is frequently modeled via an empirical threshold shift relation \[ V_{\text{FB}} = \frac{\sigma}{L} \cdot V_{DS} \] (26)

The effect of channel length modulation (CLM) is modeled using a variant of the Klaasen model \[ I_{CLM} = \frac{1}{L} \cdot \frac{1}{\eta_{d}} \cdot \frac{1}{\eta_{M1}} \cdot \frac{1}{\eta_{M2}} \] suitably modified to avoid numerical difficulties. An alternative linear “\( \lambda \)" model for CLM can also be used within STAG to facilitate fast parameter extraction at the expense of some accuracy.

B. Extrinsic Parasitics and Floating-Body Effects

Accurate modeling of floating-body behavior in dc, transient and small-signal simulations is essential for a PD SOI; in all three modes, the observable behavior will differ profoundly from familiar bulk characteristics [3], [7], [34]. Without a sound physical representation of the underlying phenomena, such accuracy cannot be achieved. Furthermore, since logic designs in particular are operated with floating-body devices [35], robustness and convergence performance under these conditions must be considered from the outset.

The equivalent circuit for the static part of the extrinsic parasitic components is shown in Fig. 6 alongside the channel current model. The impact ionization current is modeled by \( I_{\text{IMP}} \) for normal operation and by \( I_{\text{IMP}}^{\prime} \) for the case \( V_{DS} < 0 \). This arrangement ensures the symmetry of the model and at all times only one of \( I_{\text{IMP}} \) and \( I_{\text{IMP}}^{\prime} \) will be nonzero. Assuming normal operation, the expression for \( I_{\text{IMP}} \) is

\[ I_{\text{IMP}} = I_{CH} \left[ \frac{\alpha_{0}}{\beta_{0}} V_{m} \cdot \exp \left( \frac{I_{m} \cdot \beta_{0}}{V_{m}} \right) \right] \] (27)

where \( \alpha_{0} \) and \( \beta_{0} \) are the impact ionization constants, \( I_{m} \) is the length of the impact ionization region, and \( \eta \) is an empirical model parameter used to compensate for errors in the approximation of the lateral field. \( V_{m} \) and \( I_{m} \) are limited to ensure that impact ionization disappears smoothly in the linear region.

The body–source and body–drain junction diodes are modeled by \( I_{BS} \) and \( I_{BD} \) respectively. However, they also form part of the Ebers–Moll model used to account for the lateral parasitic bipolar transistor effect. The expression used for the body–source junction is

\[ I_{BS} = I_{S} \left[ \exp \left( \frac{V_{BS}}{\eta_{d} \cdot \phi} \right) - 1 \right] + I_{S1} \left[ \exp \left( \frac{V_{BS}}{\eta_{M1} \cdot \phi} \right) - 1 \right] \] (29)

where \( \eta_{d} \) and \( \eta_{M1} \) are the ideality factors with \( \eta_{d} \approx 1 \) and \( \eta_{M1} \approx 2 \). A similar expression is used for the body–drain diode.

For analog design, the model only needs to indicate if the device is being operated in a regime where bipolar action is dominant. Thus the parasitic bipolar model in STAG acts merely as a warning mechanism and is not a breakdown model involving positive feedback [36]. A modified Ebers–Moll model is used based on the body–source and body–drain diodes, indicated within the dashed lines in Fig. 6. The expressions for the current sources are

\[ I_{BD}^{\text{BJT}} = \left( \frac{\beta_{B} \cdot I_{CLM}}{I_{CLM} + 1} \right) I_{B1} \] (30)

where \( \beta_{B} \) is the current gain of the bipolar junction transistor (BJT). It has been found that \( \beta_{B} \) exhibits an inverse square dependence on the channel length [37]; therefore STAG uses the following expression:

\[ \beta_{B} = \frac{\beta_{B} \cdot I_{CLM}}{I_{CLM} + 1} \] (31)

where \( \beta_{B} \) is a model parameter.
Fig. 7. Device heating model embedded within the STAG electrical model.

Note that although a body tie connection from the body region to a convenient reference node (usually the source) will, in principle, eliminate all anomalous behavior related to the body node, in reality there is always some series resistance. This resistance is technology, layout and indeed bias dependent, and major departures from expected characteristics can arise from its presence [8]. The present model version does not include this series resistance, since at the time of writing, no convenient description suitable for compact modeling has been proposed. If body tie layouts are used, an external resistance should be added based on specific characterization data.

IV. MODELING OF SELF-HEATING

To ensure the consistency and robustness of the model in both static and dynamic analysis modes, the thermal effects have been included directly in the formulation, rather than as a macro model or as a post-processing function.

A. Modeling of Device Temperature Rise

Although the temperature rise in a device is strictly a distributed effect, satisfactory results can be obtained by assuming a uniform average temperature rise. A simple electrical analog is used in the STAG model involving the basic Joule heating model

$$\Delta T = \left[ I_{CH} \cdot V_{DS} + I_{CH}^2 (R_S + R_D) \right] \cdot R_T \tag{32}$$

where $R_T$ is the device thermal resistance. As up to three thermal time constants have been observed [38], the heating model is implemented as shown in Fig. 7 with

$$R_T = R_{T0} + R_{T1} + R_{T2} \tag{33}$$

The temperature rise node is made an external node of the device, allowing both monitoring of each individual device temperature rise and also thermal coupling of the device to neighboring devices [39].

B. Thermal Effect on Physical Parameters

The STAG model implements dependence of threshold voltage, mobility, and carrier saturation velocity on the local device temperature. Being a physically based model, these three quantities should be sufficient to model the thermal dependence of the intrinsic device [13], [23].

1) Threshold Voltage: The thermal dependence of the measured device threshold is accounted for by the use of an effective flat-band voltage.

$$V_{TH}^{eff} = V_{FB} - \frac{\sigma}{L} \cdot V_{DS} + \chi_{FB} \cdot \Delta T \tag{34}$$

where the shift due to DIBL is included for completeness.

2) Mobility: For a given temperature rise, this physical quantity has by far the most effect on the output characteristics of a device. It has been found that the surface mobility in a MOSFET varies as $T^{-k}$ [40], with $k$ reported between 1.4 to 1.8. In STAG, the mobility thermal dependence is written as

$$\mu_{oh} \equiv \mu_0(T) = \mu_0(T_{amb}) \cdot \left( 1 + \frac{\Delta T}{T_{amb}} \right)^{-k} \tag{35}$$

$k$ has been left as a model parameter in STAG, but it defaults to 1.5. There is some dependence of the vertical field mobility degradation on temperature, but this is weak and has been ignored.

3) Carrier Saturation Velocity: In STAG, it is assumed that a saturation velocity has been found at $T_{amb}$, being denoted by $v_{sat0}$, and this value is related to the value used in the model by

$$v_{sat} = v_{sat0} \cdot \frac{1 + 0.8 \exp \left( \frac{T_{amb}}{600} \right)}{1 + 0.8 \exp \left( \frac{T_{amb} + \Delta T}{600} \right)} \tag{36}$$

This expression reflects the functional temperature dependency [40]. Note that in the modeling of velocity saturation effects, the mobility is also used [see last term in (20)] and therefore its thermal dependence must also be included as per Section IV-B-2.

C. Thermal Effect on Parasitic Components

The local temperature rise $\Delta T$ will affect the parameters of the parasitic components of the device as well as the intrinsic device. Investigations into the impact ionization current [18], [30], [41] indicate that the only parameter which exhibits a temperature dependence is $\phi_T$. This dependence is linear and so the STAG model uses the following equation to incorporate this effect:

$$\beta_M = \beta_0 + \chi_{\beta} \cdot \Delta T \tag{37}$$

where $\chi_{\beta}$ is a model parameter. $\beta_M$ is then substituted for $\beta_0$ in (27).

There are two quantities in the ideal diode equation which exhibit a temperature dependence. The obvious one is $\phi_T = kT/q$, and the other is the reverse saturation current $I_s$ and $I_{sat}$. It has been found [40] that

$$I_s \sim T^{(3+\gamma)/2} \cdot \exp \left( \frac{-E_g}{kT} \right) \tag{38}$$

where $\gamma$ is a constant as specified in [40]. At normal operating temperatures however, the exponential term dominates, and hence the STAG model uses a simple exponential temperature dependence.
V. MODEL EVALUATION

Several commercial processes have been used in the course of evaluating the performance of the model and its applicability. Unless otherwise stated, results presented here are from a commercial 0.7-μm process. A single parameter set, optimized over a range of channel lengths, was used to produce all the simulation results for this process. Owing to space considerations, results for just a single channel length of 0.7-μm are presented here. In some cases, the matching of simulation data to measurements could have been further improved by using a parameter set optimized specifically for each channel length. Furthermore, because our model is intended for use in analog applications, we have focused on obtaining good data matching for typical analog operating conditions. Accurate simulation at higher gate biases is less of a priority, and this is reflected in the less precise fit observed in some of the results at high gate biases. Due to time constraints, the simple CLM model has been used to account for the effect of channel length modulation.

A. Device Simulations

Simulations of drain current versus drain and gate bias are compared with measured data; Fig. 8 shows body-tied results, while Fig. 9 shows the corresponding data with the body left floating. The results show some of the shortcomings in the auxiliary model; the bipolar model implemented does not predict the onset of breakdown, and the simple CLM model lacks accuracy around saturation for higher gate biases in the floating-body case. However, the onset of kink is quite well predicted, as is the magnitude of the current step and its attenuation at high biases due to self-heating. Similar comparisons are made for the drain conductance obtained under identical conditions. The drain conductance is smooth and continuous, as can be seen from Figs. 10 and 11. Negative conductance at high drain currents, the result of self-heating, is correctly simulated.

Subthreshold characteristics at both low and high drain biases show excellent agreement when the body is tied to source, as shown in Fig. 12. The anomalous subthreshold slope which occurs when the body is left floating is reproduced in qualitative
terms, as can be seen from Fig. 13. However, sensitivity to the extracted impact ionization parameters is relatively strong and these parameters would require further optimization in order to produce a more accurate fit in this region.

Because the body node is externally accessible, body current may be directly observed in simulation. Comparisons between simulated and measured data are shown in Fig. 14. Good agreement is observed, with emphasis placed on results at low gate biases for analog circuit design. Another important modeling issue is the body effect, which causes a threshold shift with changing body–source voltage. This is of particular relevance to PD SOI devices, since the body–source junction can become forward biased under certain conditions [36]. From Fig. 15, it can be seen that the model can account physically for the body effect, and provides a good fit to the measured data.

The frequency dependence of floating-body effects [7] should also be reproducible in simulation. Fig. 16 shows how the frequency-dependent roll-off of the small-signal drain conductance in the kink region can be accurately simulated. Due to the unavailability of measured ac data for the standard 0.7-μm process, comparison is made with a commercial 0.35-μm process instead.

Fig. 12. Subthreshold current for $V_{GS} = 0.25$ mV and 2 V of a 0.7-μm body-tied device. Measured values are shown as circles, simulated values as solid lines.

Fig. 13. Subthreshold current for $V_{GS} = 0.25$ mV and 2 V of a 0.7-μm floating-body device. Measured values are shown as circles, simulated values as solid lines.

Fig. 14. Body current for $V_{GS} = 0.3, 0.4$ and 5 V of a 0.7-μm body-contacted device. Measured values are shown as circles, simulated values as solid lines.

Fig. 15. Drain current versus gate voltage for $V_{DS} = 0.0, 0.2, 0.4$ and 0.6 V, for a 0.5-μm body-contacted device. Measured values are shown as circles, simulated values as solid lines.

Fig. 16. Small-signal drain conductance versus drain voltage for $V_{GS} = 0.9$ V, over a frequency range of 10 Hz to 1 MHz. A 0.5-μm floating-body device was used. Measured values are shown as circles, simulated values as solid lines.
B. Circuit Simulations

The model has been used extensively, both for investigations into physical behavior [14], and for full chip designs, including continuous-time filters and switched-capacitor ΣΔ A/D converters [42]. Some of the important capabilities of the model are highlighted here, using simulations and measurements of small circuit cells. Robust convergence, floating-body behavior and transient self-heating capabilities are demonstrated using a five-stage current-controlled ring oscillator fabricated in a 0.7-μm technology. Fig. 17 shows the simulated output from start-up, observable via a buffer stage. Note that although all transistors have floating-body nodes, and all have self-heating calculation enabled, simulation of the ring is possible in the standard public-domain SPICE3 simulator, without need for node-setting or specifying initial conditions. Fig. 18 shows the corresponding body potentials of the n- and p-channel transistors in one of the inverters in the ring. Excursions beyond the power rails due to charge pumping are clearly evident. Fig. 19 shows the evolution of the channel temperatures of the same n-channel device versus time. The results for a simulated 51-stage oscillator are also provided for comparison. Devices can be seen to heat quickly during transitions and cool more gradually when no current flows, although the absolute temperature rise within the circuit is quite modest.

To demonstrate the small-signal capabilities, results are presented for a single-stage amplifier (fabricated in the same technology) having a low load resistance ac coupled to the output. Prior to high-frequency roll-off, the gain is seen to rise slightly with frequency, due to the reduction in the influence of self-heating on the transconductance (see Fig. 20). STAG's
embedded thermal model is able to correctly predict this unusual trend [43].

VI. CONCLUSION

A partially depleted SOI CMOS compact model suitable for the simulation of analog circuits has been presented. Floating-body and self-heating effects have been included from the outset in a consistent and robust fashion, thus providing insight into the impact of these phenomena in all analysis modes. The body node may be used connected directly to source, or via a series resistance, or left floating, without prejudicing convergence. Under such conditions, predictions of the body potential during circuit operation may be observed. Similarly, the external thermal node may be used as an observation point for internal channel temperature during circuit operation, or may be connected via a thermal impedance to heat sources in other adjacent devices, thereby facilitating the modeling of thermal coupling. Particular attention has been paid to model continuity and robustness; this is borne out in the evaluation of the model comparing simulations against fabricated circuits, since node-setting and other such aids to initial convergence are seldom required.

REFERENCES

Mike S. L. Lee received the degree in mathematics from the University of Cambridge, Cambridge, U.K., in 1989. After briefly working in insurance risk analysis, he received the M.Sc. degree in microelectronic systems design from the University of Southampton, Southampton, U.K., in 1991. He received the Ph.D. degree from the University of Southampton in 1997. From 1991 to 1998, he was a Researcher at the Microelectronics Centre, University of Southampton. From 1998 to 1999, he was with National Semiconductors, Santa Clara, CA, where he was involved in deep-submicron device characterization. He is now with AMD Semiconductors, Sunnyvale, CA, working on digital IC design.

Bernard M. Tenbroek (M’94) was born in The Netherlands in 1969. He received the M.Sc. degree in electronic engineering from the University of Twente, The Netherlands, in 1993. He received the Ph.D. degree from the University of Southampton, U.K., in 1997 for his work on the characterization of SOI MOSFETs. While working toward the M.Sc. degree, he worked for four months at the Telkom Development Institute, Pretoria, South Africa. From 1993 to 1998, he was a Researcher at the Microelectronics Centre, University of Southampton, U.K., on compact modeling and characterization of SOI MOSFETs and analog circuit design in SOI MOSFETs technologies, with a particular interest in self-heating and floating-body effects. Since 1998, he has been a Member of Technical Staff with AKM DesignTek, San Diego, CA, working on mixed-signal IC design.

Michael J. Uren received the B.A. degree in natural sciences (experimental physics) in 1978, and the Ph.D. for his research on electrical and magnetic studies of silicon inversion layers, both from the University of Cambridge, Cambridge, U.K. He carried out postdoctoral work at IBM, Yorktown Heights, NY, on infrared spectroscopy of MOSFETs before joining the Silicon Division at DERA, Malvern, U.K. (formerly RSRE) in 1982, where he is now a Division Fellow. His research career has spanned work on CHEMFETs, porous silicon formation layers mechanisms and devices, SOI devices and reliability, and in particular, telegraph noise, 1/f noise, and MOS oxide physics. His current research interests lie mainly in the area of wide bandgap materials and devices. He is the Manager for the DERA Silicon Carbide Programme.

Dr. Uren is a Fellow of the Institute of Physics.

James Benson received the B.Sc. degree in physics from Leicester University, Leicester, U.K., in 1994, and the M.Sc. degree in semiconductor science and technology from Imperial College, London, U.K., in 1995. Since 1997, he has been a Researcher at the Microelectronics Centre, Southampton University, Southampton, where his research interests include the modeling and characterization of silicon-on-insulator MOSFETs. From 1995 to 1997, he was with Philips Semiconductors, Southampton, U.K.

Mr. Benson is a member of the Institute of Physics.

William Redman-White (M’83) received the B.Sc. degree from Exeter University, Exeter, U.K., in 1974, and the M.Sc. and Ph.D. degrees from Southampton University, Southampton, U.K., in 1980 and 1983, respectively.

While maintaining a continuing academic activity, he has worked extensively in the semiconductor industry. From 1984 to 1990, he was a part-time Design Consultant for GEC Research, London, U.K., working on projects including silicon-on-sapphire design. During 1989, he was with Motorola, Switzerland, and since 1990, he has been with Philips Semiconductors, Southampton, where he acts as Analog Design Consultant for advanced consumer IC projects. His research interests are centered on analog and mixed-signal design issues, spanning data conversion, filtering, circuit design for space environments, and characterization and modeling of SOI CMOS.

Dr. Redman-White is a member of the IEE.