Euro. Jnl of Applied Mathematics (2006), *vol.* 17, *pp.* 465–489. © 2006 Cambridge University Press
doi:10.1017/S0956792506006656 First published online 1 November 2006 Printed in the United Kingdom

Nano-scale MOSFET device modelling with quantum mechanical effects

ELLIS CUMBERBATCH[1], SHIGEYASU UNO[1] and HENOK ABEBE[2]

[1]*School of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711, USA*
[2]*USC Information Sciences Institute, MOSIS Service, 4676 Admiralty Way, Marina del Rey, CA 90292, USA*
email: abebeh@mosis.org

(*Received 19 July 2005; revised 14 June 2006*)

The continuing down-scaling trend of CMOS technology has brought serious deterioration in the accuracy of the SPICE (Simulation Program with Integrated Circuit Emphasis) device models used in the design of chip functions. This is due to in part to hot electron and quantum effects that occur in modern nano-scale MOSFET devices [13, 25, 28, 33, 34]. The focus of this paper is on modeling quantum confinement effects based on the Density-Gradient (DG) model [6, 9, 14], for application in SPICE. Analytic 1-D quantum mechanical (QM) effects correction formulae for the MOSFET inversion charge and electrostatic potential are derived from the DG model using matched asymptotic expansion techniques. Comparison of these new models with numerical data shows good results.

1 Introduction

The semi-conductor industry manufactures devices of amazing complexity and miniaturization. The minimum feature size (usually channel length: L in Figure 1) of the transistor in an integrated circuit (IC) design is currently around 100 nm, and the 2003 forecast of the International Technology Roadmap for Semiconductors shows further shrinkage, to 50 nm in 2009. Each of these devices is a variant on a simple structure, the *p–n* junction. This is usually a silicon crystal which has been "doped": atoms of different atomic structure are implanted in the silicon to alter its electrical properties. Typically, for a *p–n* junction, phosphorus atoms are implanted on one side of an interface and boron atoms are implanted on the other. In addition to silicon, other common semiconductors used in the industry are germanium and gallium arsenide. A silicon atom has four electrons in the valence band; phosphorus has five such electrons, and boron three. At moderate temperatures (including room temperature) the doped material is ionized: the phosphorus atoms lose an electron to accommodate their silicon neighbors, creating a "donated" population of free electrons (of density n, for negative) and leaving a (static) positive charge density, N_D, at the phosphorus sites. At the boron sites there is "acceptance" of an electron, resulting in a negative charge density N_A. The transport of *these* electrons is modelled in terms of a number density, p, denoting the absence of electrons from the valence band at other sites. This is referred to as the hole density. The overall space charge

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

density is therefore

$$\rho = q(p + N_D - n - N_A), \tag{1.1}$$

where q is the charge on an electron. Charge neutrality is at $\rho = 0$. As stated above, it is usually assumed that all implanted sites are ionized so that N_D and N_A correspond to the number of phosphorus and boron atoms implanted in the silicon. The physics, and mathematical modeling, of the transport and generation/recombination of electrons and holes is complicated. It is also subject to revision as devices are made smaller, and effects that are of minor importance in larger devices become non-negligible in the current technology. One such effect is impact ionization, which occurs when electrons are subject to increased force fields (the same voltage applied over a much smaller distance) and the consequent increase in acceleration causes them to dislodge new electrons on their impact with atomic sites. Another effect, the subject of this paper, is the revision of the basic classical model for electron and hole transport (the drift-diffusion model) when quantum effects intrude. This occurs, for instance, when the width of electrical confinement in the vicinity of the silicon/insulator interface is reduced to the order of the de Broglie wavelength. Electrons are then repelled away from this interface, compared with the classical case, which leads to a reduction of current at given gate voltage and to resultant deterioration in performance of the device. The loss of current needs to be quantified for the electronic engineer designing circuits. Before introducing the specific problem addressed in this paper, we continue with a brief history of device technology, as it serves to explain the context for the solution approach.

When a voltage is applied to contacts affixed to the ends of a *p–n* junction, current flows. The device acts as a diode: at one polarity little current is conducted, but a change in polarity generates a much bigger current. A technology upgrade is afforded by a sandwich structure: *p–n–p* or *n–p–n*, with a voltage applied to the filling of the device providing the third electrode controller. The transistors are referred as *pnp* or *npn* transistors and are called bi-polar junction transistors (BJTs).

Prior to the introduction of semiconductors, electronic circuits used vacuum tubes to obtain amplification. In 1947 the first transistor was demonstrated at Bell Laboratories. Amplification was obtained in a semiconductor three-terminal device many times smaller in size than the vacuum tube, and eventually much lower in cost. The theory of operation of the transistor relies on the physics of the *p–n* junction. The development of the transistor started a major revolution in the electronics industry. However, the BJT has one very undesirable characteristic: low input impedance associated with its base-emitter junction. For years scientists searched for a solution that would combine the high input impedance of the vacuum tube with the many other advantages of the transistor.

The technological advance that solved this problem may be viewed in topological terms: the sandwich is turned into a pizza. The bipolar devices are opened up from the middle, the filling of the sandwich made to form a base or substrate, and the sandwich ends are now located on the top as surface variations. Such devices are called MOSFETs (metal-oxide-silicon field-effect transistors). A schematic is shown in Fig. 1. In contrast to the bipolar transistor, which uses bias current between base and emitter to control conductivity, the FET uses voltage to control an electrostatic field within the transistor. These devices are fabricated by doping applied at the top surface. A substrate may

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

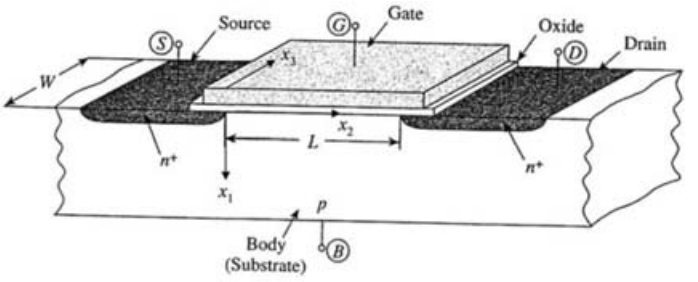


FIGURE 1. MOSFET device.

be doped first with boron, then dominating phosphorus doping is introduced over the source and drain regions, giving an n–p–n structure (NMOS device). The reverse, or complement of this, gives a p–n–p (PMOS device). The CMOS technology (C standing for complementary), which is widely used, has the NMOS and PMOS devices alternating on the chip. The feature of alternating device structure helps electronically isolate the devices from each other.

To describe semi-conductor modeling one must delineate the context. Consider the following four problem areas:

- P1: process modelling, which describes the physics of fabrication, the extent to which a region (source or drain, say) is laid down to a prescribed doping concentration and to a prescribed depth.
- P2: device modelling, which attempts to ascertain the fluxes of holes or electrons, and hence the current, when a set of voltages is applied at the terminals of a single transistor.
- P3: computation of the current flows between interacting devices when a set of these devices is fabricated to form a chip. The electrical behavior of the IC as a whole is calculated by coupling the functioning of individual transistors, using the standard Kirchhoff laws.
- P4: the associated chip design problem of the optimal placement of devices and their interconnections in order to facilitate prescribed results for the overall chip function.

In this paper we are concerned with P2.

In the simplest model the fluxes of holes and electrons in a transistor are governed by three nonlinear partial differential equations (in n, p and ψ, the electrostatic potential) with boundary conditions given on the boundary of a complicated region. Clearly, a complete, accurate solution is achievable only by a high-level numerical solution. This approach is totally unsuited for P3, i.e. for the simulation of electronic circuits, since the latter now contain 10^5 to 10^6 individual devices. As a consequence, solutions to P2 separate into two broad categories related to the context of the application. The first encompasses accurate solutions incorporating the full physics and device geometry (i.e. numerical solutions) which are used to investigate single device properties of new geometries, doping levels, materials, etc. However for application to P3, solutions must be given in terms of easily computed functions, and so approaches to P2 are restricted to simple geometries, reduced physics, etc. This is referred to as "compact modelling" in the semi-conductor industry.

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

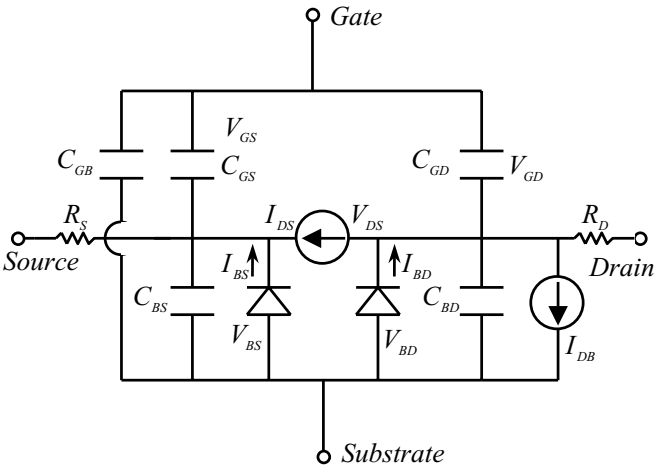


FIGURE 2. A MOSFET equivalent circuit.

The transistor-based IC circuit simulations are obtained by replacing each transistor by an equivalent circuit. A typical representation for SPICE (Simulation Program with Integrated Circuit Emphasis) for transient analysis is shown in Fig. 2; other circuits are also used for AC and for noise analysis. The subscript notation for the parameters in Fig. 2 is: S, D, G, B, refer to source, drain, gate, base (substrate), respectively, so, for instance, C_{GB} indicates the capacitance between the gate and base contacts. The symbol I_{DS} refers to the current flowing drain to source. This current flows when a positive voltage V_{DS} is applied, and when V_{GS} is sufficiently positive that electrons have been attracted to collect below the gate insulator. (This is called the inversion charge.) Some of the component "lumped" parameters of Fig. 2 are non-constant, being dependent on the applied voltages, and these variations are larger in the smaller devices. Hence the description that accompanies a representation such as Fig. 2 for the circuit components has extensive features, as each of the components requires detailed specification of its dependence on the applied voltages.

SPICE is the most common program for the simulation of electronic circuits. The original SPICE was developed at U.C. Berkeley in 1971, and updated versions exist (some for specific applications) both in the public and the commercial domain. For SPICE to produce a circuit simulation, the software must access the description of an individual transistor many times, and for IC circuits containing many transistors this description must have a simple form in order for the simulation to be completed in a reasonable time. These device descriptions have a long evolution, concurrent with developments in IC fabrication technology. Early MOSFET devices in the CMOS technology were modelled using a quasi 1-D approximation so that, for instance, I_{DS} was calculated from the solutions of ordinary differential equations. Even so, the result is in the form of an integral and approximations were made to reduce it to simple algebraic expressions valid in different voltage regimes. (This may cause discontinuities as differentiation of this function is required.) These approximations were accurate when transistor channel lengths were several micrometers,

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

and the more complex physical mechanisms had negligibly small effects. As feature sizes reduced, these approximations lost accuracy, and since our ability to produce simple, analytic solutions to non-linear, multi-dimensional PDEs is somewhat lacking (certainly at the pace the technology was changing), the industry was forced to rely on empirical approximations. The current situation for the implementation of circuit simulation via the industry standard SPICE BSIM3v3 model is that over 250 device parameters must be provided for the various functions used in the circuit element descriptions such as Fig. 2. Most of these parameters do not have direct physical identification and must be obtained from current-voltage (I–V) and capacitance-voltage (C–V) data fitting (see Liu [16]). The data collection, and ensuing accurate parameter extraction, is a time-consuming and expensive operation. BSIM3v3 is suitable for both digital and analogue circuit design applications, in part because of improved modelling of the output conductance, when compared to previous MOS device models. But the physical foundation of BSIM3v3 is compromised by the very large number of non-physical correction (binning) parameters that are necessary for the model to work over a large range of channel dimensions in a deep sub-micron process. Most of the simulation problems encountered with BSIM3v3 are created by these correction parameters.

Efforts to re-align the situation just described, to provide SPICE-compatible solutions for device descriptions based on the physics, have been attempted, and there has been success for some SPICE models, in particular for analogue designs, e.g. the EKV device model [12]. However, these models are limited to specific circuit design applications.

Our approach in this direction has mainly involved the use of asymptotic analysis to solve the boundary value problems. Of direct relevance here are Cumberbatch *et al.* [11] and Abebe & Cumberbatch [1, 2]. The former achieves an explicit formula for the drain to source current I_{DS} characteristics based on asymptotic approximation of solutions of the drift-diffusion equations, and the latter provide amendments to this formula necessary to take into account quantum effects. Both formulae are shown to be accurate in comparison with experimental data and with simulations based on more exact physics. We have also developed accurate models for gate capacitance [22], impact ionization [21], and an I–V model for the SOI (silicon-on-insulator) device [20]. Other relevant publications in these areas are Quan *et al.* [26] and Ramey & Ferry [27].

Please [24] was the first to provide a non-dimensional version of the drift-diffusion equations so that the importance of the various physical entities in the equations are made transparent, and the technique of matched asymptotic expansions (MAE) may then be employed. The application was the $p-n$ junction and his PhD thesis [23] contains a number of results on current flow not contained in Please [24]. Ward [31, 32] continued the approach, providing: (i) the asymptotic justification for the quasi 1-D approximation; (ii) a formula for I_{DS} for the MOSFET based on MAE; and (iii) an attempt to extend the 1-D analysis to include 2-D effects by formulating (simplified) PDEs in the corners of the device near the source and drain regions. Markowich *et al.* [18, 19] present various problems approached by asymptotic techniques (as well as a broad introduction to the modelling of semi-conductor physics), including the Ward analysis. Our work [11] made further approximations in the Ward solution for I_{DS} to obtain the explicit formula mentioned above. To our knowledge there has been no progress made in extending the quasi 1-D solutions on which these results are based (using the approach in (iii) above, or

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

otherwise), even though current devices are well outside the parameter regime for which the 1-D approach can be justified.

A full treatment using quantum mechanics requires QM solutions in the gate polysilicon material and in the oxide insulator, as well as in the silicon base. The oxide, having a large barrier height, would generate only a small electron density. In this case the electrons are said to have leaked or tunneled into the oxide. In the treatment here, the electron density in the insulator is taken to be zero. The quantum treatment requires that the electron density be continuous so that the boundary condition at the oxide/silicon interface is zero electron density, whereas the classical theory based on the drift-diffusion model using the quasi 1-D approximation gives maximum charge density at this interface. This is the main discrepancy between the classical and quantum models. Several models have received attention. Full solutions of the Schrödinger and Gauss equations, and various other approximations, all require high level numerical simulations. This multi-dimensional microscopic solution is inappropriate for practical circuit application [6, 9].

The classical description, which is conventionally the drift-diffusion (DD) current density model, may be obtained by assuming that the internal energy densities of the electron and hole gases have a logarithmic dependence on the charge densities. A more general series expansion of the density using the kinetic theory of gases gives that the energy depends not only on the density but also on the spatial gradient of gas density [6, 7]. This is the central assumption of the DG theory and it is derived from electron and hole kinetics by applying hydrodynamic theory. The association with quantum mechanical effects is made by interpreting the extra DG term as a quantum term, and suitably changing the boundary conditions. However, the DG model may also be generated directly from the quantum mechanics for electrons by expressing the Schrödinger wave function in amplitude/phase form, and this is briefly explained in § 2.2. The DG model has been solved numerically for a variety of devices and it has shown good accuracy compared with more complete models, in particular with quantum microscopic solutions for tunneling and confinement effects [6, p. 7964; 4, p. 1228; 5, p. 9537].

The quantum term in the DG model has a higher order derivative than the terms in the DD model and it is multiplied by a numerically small factor. This indicates a boundary layer behavior which is confirmed by the numerical results: the inversion charge density is reduced significantly in a small layer close to the silicon/silicon-oxide interface, but the charge behaviour outside this layer is similar to the non-quantum, classical solution; see Fig. 3. The maximum of the classical charge distribution is shifted away from the oxide interface under the influence of quantum effects, and its value may be reduced by up to two orders of magnitude. This loss of electrons results in diminished source-to-drain current, and it affects the gate capacitance circuit element (C_{GB} in Fig. 2). (The shift in the maximum density may be viewed as an increased oxide thickness.) Also, in the case when the electron mobility is position dependent, an effective value is obtained by averaging with respect to the electron density, and this value is amended due to the classical to quantum profile change shown in Fig. 3. Using a boundary layer approach we are able to generate explicit, analytic expressions for the potential and electron density profiles that include the quantum effects and that are SPICE-compatible. Ancona [3] noted the boundary layer structure of the DG equations, but did not pursue this approach in detail. Since the governing equations for electron and hole current flow are similar, our work

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

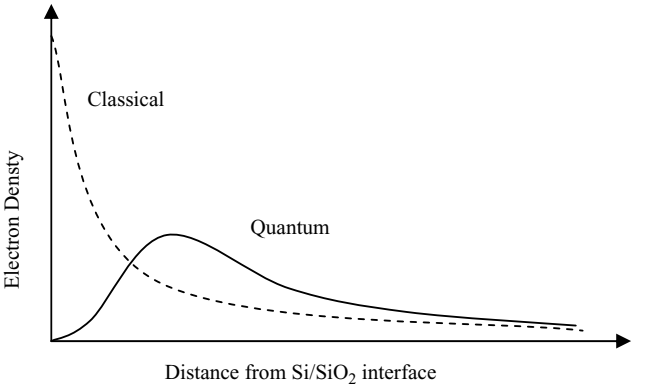


FIGURE 3. Comparison of classical and quantum solutions for the inversion charge.

here is concentrated only on the silicon based NMOS device, in which electrons are the dominant carriers.

The work presented in Abebe & Cumberbatch [1, 2] used an approximation to the electrostatic potential that allowed the equation for the electron density in the quantum layer to be integrated analytically. This solution was then successfully matched with an exterior classical solution to provide a profile for the DG quantum electron density similar to the one shown in Fig. 3. However this model was incomplete as it required the quantum change in the surface potential as an external input. In Abebe & Cumberbatch [1] we used the surface potential as a fitting parameter in comparing the analytical I-V model with the numerical data. In Abebe & Cumberbatch [2], a more complete model was provided by incorporating the quantum correction of the surface potential based on a formula for the change in the semi-conductor band gap due to the high fields at the oxide interface. In the current paper the DG equations are solved by the matched asymptotic expansion technique, the expansion parameter being related to Planck's constant. A version of the asymptotic solution gives good results compared with numerical data. The quantum layer is treated as an inner layer of the inversion region, and asymptotic expansions for both the electron density and the electrostatic potential are obtained there and matched to expansions valid in the inversion layer. Satisfying the boundary conditions at the silicon/insulator interface, and in the bulk (via a small change in the location of the depletion layer) then provides a comprehensive solution. This solution is self-contained – it depends only on the physics involved in the DG quantum model, and it avoids data fitting or external empirical formulae.

2 Model equations and parameter scaling

2.1 The drift-diffusion equations

The motion of electrons and holes is assumed to be governed both by the electric field and by diffusion using the standard Fick's law. This gives for the electron and hole fluxes

$$\left.\begin{aligned}\mathbf{J}_n &= q(n\mu_n\mathbf{E} + D_n\nabla n)\\ \mathbf{J}_p &= q(p\mu_p\mathbf{E} - D_p\nabla p)\end{aligned}\right\} \tag{2.1}$$

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

where $\mathbf{J}_n, \mathbf{J}_p$ are the electron and hole current densities, q is the magnitude of the electronic charge, n, p are the electron and hole concentrations and D_n, D_p are the electron and hole diffusivities. The electron hole mobilities μ_n and μ_p depend in general on doping levels and the electrostatic field $\mathbf{E}$, but usually are taken as constant. With the Einstein relation:

$$D = \frac{kT\mu}{q}, \tag{2.2}$$

and with

$$E = -\nabla\psi, \tag{2.3}$$

we can write

$$\left.\begin{aligned} \mathbf{J}_n &= n\mu_n\nabla(-\psi q + kT\ln n) = -qn\mu_n\nabla\Phi_n \\ \mathbf{J}_p &= -p\mu_p\nabla(\psi q + kT\ln p) = -qp\mu_p\nabla\Phi_p \end{aligned}\right\} \tag{2.4}$$

where k is Boltzmann's constant, T is the lattice temperature (assumed constant), ψ is the electrostatic potential, Φ_n and Φ_p are the "electro-chemical quasi-Fermi potentials" for electrons and holes. These definitions give

$$n = n_i \exp\left[(\psi - \Phi_n)/V_{th}\right] \quad \text{and} \quad p = n_i \exp\left[(-\psi + \Phi_p)/V_{th}\right], \tag{2.5}$$

where $V_{th} = kT/q$ is called the thermal voltage. Here n_i is a constant called the intrinsic carrier density, and it represents the density of both holes and electrons in undoped silicon.

Gauss' equation for charge conservation and the conservation equations for holes and electrons can now be written

$$\left.\begin{aligned} &\varepsilon\nabla\cdot\mathbf{E} = -\varepsilon\nabla^2\psi = \rho = q(p - n + N) \\ &\tfrac{\partial n}{\partial t} - \tfrac{1}{q}\nabla\cdot\mathbf{J}_n = G_n - U_n \\ &\tfrac{\partial p}{\partial t} + \tfrac{1}{q}\nabla\cdot\mathbf{J}_p = G_p - U_p \end{aligned}\right\} \tag{2.6}$$

where ε is the material dielectric constant, and $N = N_D - N_A$ is the static doping, comprising donors and acceptors, G_n, G_p are the electron and hole generation rates, and U_n, U_p are the electron and hole recombination rates. See Sze [29] for expressions specifying the generation and recombination terms. Equations (2.6), together with (2.4, 2.5), constitute three non-linear partial differential equations for the potentials ψ, Φ_n and Φ_p, in the most general context. This is the standard drift-diffusion model which has been the basis for most of the technical development of the semi-conductor industry [29, 19].

In the application to current flow in an n-channel MOSFET, the holes are considered to be in thermal equilibrium, implying that Φ_p is constant, taken to be zero. Also the generation/recombination terms are omitted and only the case of time-independent currents is considered. For SPICE application this steady-state approximation is good for circuits operating at frequencies less than 100 MHz. For high frequency circuit applications SPICE steady-state solutions are corrected by engineering approximations to include high frequency effects.

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

2.2 The density gradient model for the quantum effect

The model employed in this paper is referred to in the semi-conductor literature as the DG model. This appellation derives from continuum mechanics [6, 5]: extensions from kinetic theory to the basic laws for a continuum which assume that the internal energy is a function of density give that the energy depends also on the spatial gradient of density. In the semiconductor context a similar extension to the DD model to include a quantum term may be obtained directly from the Schrödinger equation and this is briefly explained here. In some references this is given the name of the Madelung description [17]. The Schrödinger equation is written as

$$i\hbar\frac{\partial \Psi}{\partial t} + \frac{\hbar^2}{2m}\nabla^2\Psi = V\Psi \tag{2.7}$$

where $\hbar$ is Planck's constant, divided by 2π, V is the potential energy field encountered by an electron, and m is the effective electron mass. The Schrödinger wave function Ψ is written in terms of amplitude and phase as

$$\Psi = A\exp(iS/\hbar) \tag{2.8}$$

which, when substituted into (2.7), gives the two equations

$$\left.\begin{aligned} 2m\left(\tfrac{\partial S}{\partial t} + V\right) + |\nabla S|^2 - \hbar^2(\nabla^2 A)/A = 0 \\ 2m\tfrac{\partial A}{\partial t} + 2\nabla A\cdot\nabla S + A\nabla^2 S = 0. \end{aligned}\right\} \tag{2.9}$$

The second of these equations can be written as

$$\frac{\partial}{\partial t}(A^2) + \nabla\cdot(A^2\mathbf{v}) = 0 \quad \text{where} \quad \mathbf{v} = \frac{1}{m}\nabla S, \tag{2.10}$$

and this allows an identification with the conservation of mass in a continuum, the density being $|A|^2$ and the velocity being $\mathbf{v}$. The first equation in (2.9) may now be written as

$$m\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla V + \frac{\hbar^2}{2m}\nabla\left(\frac{1}{A}\nabla^2 A\right) \tag{2.11}$$

and this has the equation form: rate of change of momentum (material time derivative) equals applied forces. The right-hand side consists of the classical term

$$-\nabla V = -q\nabla\psi = q\mathbf{E} \tag{2.12}$$

representing the force experienced by the electron due to the electrostatic field $\mathbf{E}$, plus the quantum term, which is also a gradient of a potential called the Bohm potential, see Bohm [10]. The DG quantum model may now be viewed as an extension of the classical DD model for which the force potential is Φ_n given by (2.4a). The extension is obtained by including a quantum force on an electron represented by the Bohm potential. In (2.11) the amplitude A in the Bohm potential has the interpretation that $|A|^2$ is the probability density function for an electron. In adapting the Bohm potential to the continuum model, A in (2.11) is replaced by $\sqrt{n}$, n being the continuum density of electrons. The classical

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

quasi-Fermi potential in (2.5a), whose gradient is proportional to the force field extant in the continuum is, in the DG quantum extension, replaced by

$$\Phi_n = \psi - V_{th} \ln(n/n_i) + \Phi_{qn} \tag{2.13}$$

where

$$\Phi_{qn} = 2b_n(\nabla^2\sqrt{n})/\sqrt{n}, \quad b_n = \hbar^2/12m^*q$$

The extra factors introduced in the Bohm potential in (2.13) that differ from those in (2.11) are necessary to account for aspects of quantum mechanics, such as energy quantization, not present in this development. In addition the factor m^* represents the electron effective mass normal to the interface (the electrons are not quantized transverse to the interface). This factor is usually used as a fitting parameter [7]. A similar expression to (2.13) holds for the DG quantum model for holes, p replacing n in the appropriate places. An alternative interpretation of the addition of Φ_{qn} to the quasi-Fermi potential in (2.13) is that Boltzmann statistics have been corrected by quantum mechanical effects.

2.3 The boundary value problem

Equations (2.6), amended by the assumptions discussed at the end of Section 2.1, together with equations (2.4a), (2.13), relating Φ_n with n, comprise the field equations in the silicon $0 < x_1 < \infty$, $0 < x_2 < L$, where L is the channel length (see Fig. 1). No variation with x_3 is considered. The silicon region is adjoined by the oxide, $-t_{ox} < x_1 < 0$, in which no charges are assumed to reside, giving $\nabla\cdot\,\mathbf{E} = 0$ there.

The following boundary conditions on ψ and Φ_n are standard [32]:

$$\psi|_{x_1=-t_{ox}} = V_{GS} - V_{FB}, \psi|_{x_1\to\infty} = -V_{th}\ln(N_A/n_i)$$

$$(\psi, \Phi_n)|_{x_2=0} = (V_{th}\ln(N_A/n_i), 0)$$

$$(\psi, \Phi_n)|_{x_2=L} = (V_{DS} + V_{th}\ln(N_A/n_i), V_{DS}) \tag{2.14}$$

In the above V_{GS}, V_{FB} and V_{DS} are gate, flat-band and drain voltages, N_D and N_A are the source/drain and silicon doping levels, respectively, and t_{0x} is the oxide thickness. No current leaves the device across $x_1 = 0$ or at $x_1 = \infty$. Since the PDE system described above contains two more derivatives than the classical system, two extra boundary conditions must be prescribed. These are taken to be

$$n = 0 \text{ at } x_1 = 0 \text{ and } n \to n_i^2/N_A \text{ as } x_1 \to \infty \tag{2.15}$$

Condition (2.15*a*) is consistent with quantum continuity of electron density and the assumption of zero density in the oxide. Condition (2.15*b*) is adopted on the basis that, though quantum effects alter the solution in the inversion and depletion regions, the solution in the silicon bulk returns to its classical value. It is noted that in simulations using the full Schrödinger wave function Ψ the boundary conditions are taken to be $\Psi = 0$ at $x_1 = 0$ and at $x_1 = \infty$. The first condition is equivalent to (2.15a) and the second is not dis-similar to (2.15*b*) since n_i^2/N_A is many orders of magnitude smaller than the electron densities in the inversion region.

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

2.4 Scaled variables

In order to facilitate the use of the MOSFET I–V results obtained in Cumberbatch *et al.* [11] for the non-quantum case, we employ the scaling used there; it was introduced by Ward [32]. The scaling is

$$\begin{gathered}(x_1, t_{ox}) = (x, t)L_D\sqrt{2(\ln\lambda)/\lambda},\ x_2 = yL, (\psi, \Phi_n) = (w, \varphi)V_{th}\ln\lambda \\ V_{DS} = V_{ds}V_{th},\ V_{GS} - V_{FB} = V_{gs}V_{th}\ln\lambda\end{gathered} \tag{2.16}$$

where $L_D = \left(\frac{kT\varepsilon_{si}}{2n_i q^2}\right)^{1/2}$ is called the intrinsic Debye length, and $\lambda = N_A/n_i$. This definition of L_D, in accord with current practice in the semi-conductor literature, has a $1/\sqrt{2}$ factor not present in previous publications [32, 31, 11, 1, 2]. In the above, N_A is the substrate doping, typically of order $10^{15} - 10^{17}\,\mathrm{cm}^{-3}$, giving $\lambda \sim 10^5 - 10^7$. The scaling in (2.16) yields $O(1)$ changes in the scaled potential over the $O(1)$ changes in x that represent the depletion depth.

The reduction to ordinary differential equations is afforded by the quasi-one-dimensional approximation, which is valid for large aspect-ratio devices ($L \gg$ depletion depth) [32, 11]. In this limit the approximation predicts that the quasi-Fermi potential, φ, in (2.16), is a function only of y, and w depends on this coordinate only through its dependence on φ.

With this scaling, the quasi-1-D drift-diffusion equations reduce to

$$\left.\begin{aligned} w_{xx} &= 0 && \text{in} \quad -t < x < 0 \\ w_{xx} &= (n-p)/N_A + 1 && \text{in} \quad x > 0 \end{aligned}\right\} \tag{2.17}$$

with p given by

$$p = n_i \exp(-w\ln\lambda). \tag{2.18}$$

The quantum revision to n, equation (2.13), becomes

$$w - \varphi = \frac{1}{\ln\lambda}\ln\left(\frac{n}{n_i}\right) - \frac{\lambda\beta^2}{(\ln\lambda)^2}\frac{1}{\sqrt{n}}\frac{d^2\sqrt{n}}{dx^2} \tag{2.19}$$

where $\beta^2 = 2b_n/V_{th}L_D^2$. Since φ depends only on y it may be taken as constant in the solution of the differential equations above.

Equation (2.17a) gives a linear potential in the oxide so that the continuity of electrostatic potential and displacement at the silicon/silicon dioxide interface yield boundary conditions there for equations (2.17b)–(2.19) as

$$n = 0 \text{ and } \frac{\partial w}{\partial x} = c(w - V_{gs}) \text{ at } \quad x = 0 \tag{2.20}$$

where $c = \sqrt{2\ln\lambda/\lambda}\epsilon_{ox}L_D/\epsilon_{si}t_{ox}$ and $\varepsilon_{si}, \varepsilon_{ox}$ are the dielectric constants of silicon and silicon dioxide. Boundary conditions in the silicon bulk are

$$n \to 0, \text{ and } w \to -1 \text{ as } x \to \infty, \tag{2.21}$$

In (2.19) the factor $\beta^2\lambda/(\ln\lambda)^2$ is $O(10^{-4})$ for $\lambda = 10^7$. This indicates that the quantum correction term is significant in a layer much smaller than the depletion length and a

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

fraction of the length scale of the inversion layer. Outside this narrow quantum layer this term has the effect of shifting the classical solution away from the interface. Numerical solutions confirm this behaviour showing that the QM effect on the electron density is substantial in a narrow layer close to the oxide interface, reducing it from high values to zero at the interface. (See Figure 3). The electron density is zero at the oxide interface with the inversion charge peak at $5\mathring{A}$ to $15\mathring{A}$. The solution of (2.19) in this narrow layer is called the inner solution.

3 Solution by matched asymptotic expansions

3.1 Introduction

In the MAE technique [8, 15], approximate solutions to the differential equations, valid in adjacent but overlapping regions, are blended together by suitably matching integration constants, resulting in a composite solution valid over the whole region of interest. Explicit solutions for the classical solution to equations (2.17)–(2.21) were developed in Cumberbatch *et al.* [11], based on the MAE formulation presented earlier [31, 32]. We shall repeat a brief version of that material next, as it is necessary to establish the size of the inversion layer as it pertains to the size of the quantum layer. In this we shall see a surprising result (in MAE terms) of the inversion layer thickness.

The Classical Solution. The classical solution, which provides the first term in the outer expansion, is found by omitting the quantum term in (2.19), giving

$$n_{out} = n_i e^{(w-\varphi)\ln\lambda} \tag{3.1}$$

Equations (2.17), (3.1) give for the potential

$$w_{xx} = e^{(w-\varphi-1)\ln\lambda} - e^{-(w+1)\ln\lambda} + 1, \quad \text{in} \quad x > 0 \tag{3.2}$$

Since $\lambda \gg 1$, each of the terms on the RHS of (3.2) can dominate according to values of w. We shall be treating the strong inversion mode for which $w(0) = w_s > 1 + \varphi$, and the first exponential on the RHS is dominant in a neighborhood of $x = 0$. This occurs when V_{gs} is sufficiently large in the B.C. (2.20). This voltage attracts a dense layer of electrons to $x = 0$. Outside this layer the value of w reduces and both exponential terms are negligible. This is the depletion region, so called because it is depleted of both holes and electrons. Further out, $w \sim -1$, with holes present, in balance with the doping level.

The solution regimes for (3.2) are therefore:

$$\text{BULK:} \ w \sim -1; \tag{3.3}$$

$$\text{DEPLETION:} \ -1 < w < 1 + \varphi, \ \ w = x^2/2 + ax + b; \tag{3.4}$$

$$\text{INVERSION:} \ w > 1 + \varphi, \ w_{xx} = e^{(w-\varphi-1)\ln\lambda}. \tag{3.5}$$

Ward treated the inversion region as a thin layer adjacent to $x = 0$. However the scalings used in Ward *et al.* [31, 32] are different. We examine these expansions here as they apparently indicate different inversion layer thicknesses, and we wish to provide a

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

comparison with the quantum layer thickness. In Ward *et al.* [32] the expansion is

$$w - \varphi = w_0(\tilde{x}) + w_1(\tilde{x})/(\ln\lambda)^2 + \cdots \text{where } \tilde{x} = x\ln\lambda. \tag{3.6}$$

The first term satisfies

$$w_0'' = e^{(w_0-1)\ln\lambda}/(\ln\lambda)^2 \text{ with } w_0(0) = w_s. \tag{3.7}$$

The expansion in Ward *et al.* [31] is

$$w - \varphi = w_s + v_0(\tilde{x})/\ln\lambda + \cdots \text{where } \tilde{x} = x\lambda^{(w_s-1)/2}\sqrt{\ln\lambda}. \tag{3.8}$$

The first term satisfies

$$v_0'' = e^{v_0} \text{ with } v_0(0) = 0 \tag{3.9}$$

The solutions to (3.7), (3.9) give identical first-order solutions.

It is conventional wisdom in the MAE approach that scaling giving rise to $O(1)$ relationships represents the correct order-of-magnitude of the dependent and independent variables for the phenomena under investigation. With λ not present in (3.9) (as opposed to (3.7)), one would therefore infer from (3.8) that the inversion layer thickness is $O(\lambda^{(1-ws)/2}/\sqrt{\ln\lambda})$, and not $O(1/\ln\lambda)$ as (3.6) implies. However this is not supported by numerical evidence, nor by the analysis that follows. The reference length, $L_D\sqrt{2(\ln\lambda)/\lambda}$, has the value 123 nm at $\lambda = 10^6$. A fraction $1/\ln\lambda$ of this gives 9 nm, which looks reasonable for the regions in Figs. 5 and 6, in which inversion takes place. At $w_s = 1.5$ the fraction $\lambda^{(1-w_s)/2}/\sqrt{\ln\lambda}$ of the reference length gives 1 nm which is clearly too small. Additionally the inversion regime does not show appreciable thinning as w_s increases, which is implied by the scaling in (3.8).

We now take up this theme analytically. We examine only the inversion equation, and replace $w - \varphi$ by w. The scaling adopted is

$$\tilde{x} = x\delta\ln\lambda \tag{3.10}$$

so that (3.6) is replicated by $\delta = 1$ and (3.8) is replicated by $\delta = (\lambda^{w_s}/\lambda\ln\lambda)^{1/2}$. The inversion equation is

$$w'' = e^{w\ln\lambda}/\delta^2\lambda(\ln\lambda)^2 \tag{3.11}$$

which has the solution

$$w = 1 + \frac{1}{\ln\lambda}\ln(\delta^2\alpha_0^2\ln\lambda) - \frac{2}{\ln\lambda}\ln[\sinh(\frac{\alpha_0\tilde{x}}{\sqrt{2}} + \gamma)] \tag{3.12}$$

$$\text{with } \sinh\gamma = \lambda^{(1-w_s)/2}\delta\alpha_0\sqrt{\ln\lambda}.$$

This reproduces the solutions presented in Ward *et al.* [31, 32] using the appropriate choice for δ. This solution is now matched with the adjacent solution in the depletion layer, (3.4), which in turn is matched to the bulk solution (3.3). This is done in Ward *et al.* [32], and we present that analysis here for the solution (3.12). The relationships are

$$a = -\sqrt{2}\alpha_0\delta, \quad b = 1 + \{\ln(\ln\lambda) + 2\ln(2\delta\alpha_0) - 2\gamma\}/\ln\lambda \tag{3.13}$$

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

with

$$\alpha_0\delta = \{2 + (\ln\ln\lambda + 2\ln(2\delta\alpha_0) - 2\gamma - 1)/\ln\lambda\}^{\frac{1}{2}} \tag{3.14}$$

The boundary condition (2.20) then gives

$$\left\{2\left(\lambda^{w_s-1}\Big/\sqrt{\ln\lambda} + \delta^2\alpha_0^2\right)\right\}^{\frac{1}{2}} = c\{V_{gs} - w_s\} \tag{3.15}$$

Equations (3.14), (3.15) provide two relations from which to determine w_s and α_0 in terms of the given parameters.

These equations reveal that α_0 and δ appear only as the product $\delta\alpha_0$. Hence $\delta\alpha_0$ has the same dependence on w_s as α_0 does when the value $\delta = 1$ is chosen, as in Ward *et al.* [32]. It was noted there that α_0 is "highly insensitive to the surface potential w_s.". Our numerical evidence supports this statement, showing that $\alpha_0 = 1.5$ over the inversion region. Note that the dependence on x in the solution (3.12) has the factor $\delta\alpha_0\ln\lambda/\sqrt{2}$. This, together with the statements just made concerning the dependence of $\delta\alpha_0$ on w_s, establishes that the inversion layer thickness is $O(1/\ln\lambda)$ with only a weak dependence on w_s. As stated previously this behaviour is evident in the numerical results, as shown in Figs. 5 and 6. The scaling introduced in Ward *et al.* [31] (see (3.9), (3.10)), though appealing on mathematical terms, is seen to be misleading.

3.2 Inner and outer expansions

We now take up the solutions to equations (2.17)–(2.19) based on formal expansions in $\varepsilon \ll 1$, where ε is defined by

$$\varepsilon^2 = \lambda\beta^2/2\ln\lambda \tag{3.16}$$

In this section only the quantum effect on the inversion layer is investigated: the hole and static doping terms (p and 1) in (2.17) are neglected, since $w > 1$ in the quantum/inversion region. Subsequently, perturbations of the depletion layer solution, in which these terms play a role, are included. The system (2.17)–(2.19) is in conventional MAE format, with the small parameter multiplying the highest derivative term, indicating that the quantum layer thickness is much smaller than the reference length. For $\lambda = 10^5, 10^6, 10^7$ the reference length is 354 nm, 123 nm, 42 nm, respectively, and the inversion layer ($O(1/\ln\lambda)$ of the reference length) is then 30 nm, 9 nm, 2.6 nm for these cases. We also assume that the quantum layer thickness (1–3 nm in numerical solutions) is smaller than the inversion layer depth. This assumption looks tenuous at the higher doping levels for which the inversion layer thins. This is borne out by error comparisons which we take up in § 4.

Quantum Layer : Inner Expansion. For this expansion we prefer to use w and

$$T = \sqrt{n/n_i} \quad \text{or} \quad Y = \ln\left(n/n_i\right)/\ln\lambda \tag{3.17}$$

as dependent variables. The latter are equivalent and are used interchangeably, depending on which gives advantages in ease of solution. For the quantum inner layer the length scale is $O(\varepsilon)$ and the inner variable X given by

$$X = x/\varepsilon \tag{3.18}$$

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

is approriate. With (3.16)–(3.18), equations (2.17) and (2.19) in terms of X become

$$\frac{d^2w}{dX^2} = \varepsilon^2 \exp\left[(Y-1)\ln\lambda\right], \tag{3.19}$$

$$\text{and} \quad T^{-1}\frac{d^2T}{dX^2} = \ln T - \ln\lambda(w-\varphi)/2 \tag{3.20}$$

$$\text{or} \quad \frac{d^2Y}{dX^2} + \frac{1}{2}\ln\lambda\left(\frac{dY}{dX}\right)^2 + w - \varphi - Y = 0. \tag{3.21}$$

The expansions for w, T, Y are written as

$$(w, T, Y) = (W_0, T_0, Y_0) + \varepsilon(W_1, T_1, Y_1) + \varepsilon^2(W_2, T_2, Y_2) + \cdots \tag{3.22}$$

Substituting these into (3.19), (3.20) gives

$$W_0'' = 0 \rightarrow W_0 = W_{0s} + s_0 X, \tag{3.23}$$

$$W_1'' = 0 \rightarrow W_1 = W_{1s} + s_1 X, \tag{3.24}$$

$$W_2'' = T_0^2/\lambda + 1, \tag{3.25}$$

$$T_0'' - T_0\ln T_0 + \ln\lambda\,(W_0 - \varphi)\,T_0/2 = 0, \tag{3.26}$$

$$Y_1'' + \ln\lambda\, Y_0' Y_1' + W_1 - Y_1 = 0. \tag{3.27}$$

together with the relationships

$$T_0 = \exp\left[\frac{1}{2}Y_0\ln\lambda\right] \quad \text{and} \quad T_1 = \frac{1}{2}T_0 Y_1 \ln\lambda. \tag{3.28}$$

The boundary condition (2.20), and (3.23), (3.24), yield

$$\begin{aligned} \tfrac{\partial W_0}{\partial X}\big|_{X=0} &= s_0 = 0, \\ \tfrac{\partial W_1}{\partial X}\big|_{X=0} &= s_1 = c\left(W_{0s} - V_{gs}\right), \\ \tfrac{\partial W_2}{\partial X}\big|_{X=0} &= cW_{1s}, \end{aligned} \tag{3.29}$$

Outer Expansion. Here the quantum term is considered to be a small perturbation of the system (2.17), and the outer expansion reads

$$\begin{aligned} w &= w_0 + \varepsilon w_1 + \varepsilon^2 w_2 + \cdots \\ \sqrt{n/n_i} &= \tau_0 + \varepsilon\tau_1 + \varepsilon^2\tau_2 + \cdots \end{aligned} \tag{3.30}$$

The first terms satisfy

$$\frac{d^2w_0}{dx^2} = \frac{n_i}{N_A}\tau_0^2, \qquad (w_0 - \varphi)\ln\lambda = 2\ln\tau_0, \tag{3.31}$$

which, when combined, are equivalent to (3.7). The second terms satisfy

$$\frac{d^2w_1}{dx^2} = 2\frac{n_i}{N_A}\tau_0\tau_1, \qquad w_1\ln\lambda = 2\tau_1/\tau_0, \tag{3.32}$$

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

giving

$$\frac{d^2w_1}{dx^2} = \frac{\ln\lambda}{\lambda}\tau_0^2 w_1 \tag{3.33}$$

The outer solutions for w and Y are identical to $O(\varepsilon)$, see (2.19).

3.3 Solutions and matching

The solution to (3.31) is

$$w_0 = \varphi + 1 + \frac{1}{\ln\lambda}\left\{\ln\left(\alpha_0^2\ln\lambda\right) - 2\ln\left[\sinh\left(x\alpha_0\ln\lambda\Big/\sqrt{2}+\gamma\right)\right]\right\} \tag{3.34}$$

with $\sinh\gamma = \alpha_0\left\{\lambda^{1+\varphi-w_s}\ln\lambda\right\}^{1/2}$. The subscript "s" has been used in the inner solution to denote values at $x = 0$; it is used in (3.34) to denote values as $x \to 0$, since that limit will be used in matching.

Matching the inner and outer expansions follows conventional lines: the expansions are expressed in terms of the intermediate variable $x_\eta = x/\eta$ and the limit $\eta \to 0$, $\eta/\varepsilon \to \infty$ is taken. The first terms satisfy

$$w_{0s} = W_{0s},\ w_{1s} = W_{1s}, \tag{3.35}$$

$$T_0(\infty) = \tau_{0s} = \exp\left[\frac{1}{2}(w_{0s}-\varphi)\ln\lambda\right], \tag{3.36}$$

$$\text{and } s_1 = -\sqrt{2}\alpha_0\coth\gamma, \tag{3.37}$$

the latter obtained from the solution for w_0, (3.34). The result for s_1, together with (3.34) identifies w_0 as the classical, non-quantum potential. The equation for the first term in the inner solution, (3.26), when combined with (3.23), (3.29), (3.35) and (3.36) can be written as

$$S_0'' = S_0\ln S_0 \qquad \text{where} \qquad S_0 = T_0/\tau_{0s} \tag{3.38}$$

This has the solution

$$X = \int_0^{S_0}\left\{s^2\ln s + \left(1-s^2\right)/2\right\}^{-1/2} ds \tag{3.39}$$

in which integration constants have been chosen so that $S_0(\infty) = 1$, to satisfy (3.36) and $S_0(0) = 0$, to give zero electron density at the silicon/oxide interface, see (2.20).

For matching purposes, the asymptotic behaviour of S_0 is required. This is found to be

$$\begin{gathered} S_0 \sim 1 - k_1 e^{-X}/2 \quad \text{as} \quad X \to \infty \\ \text{where} \quad \ln(k_1/2) = \int_0^1 \left\{\left[s^2\ln s + \left(1-s^2\right)/2\right]^{-1/2} - (1-s)^{-1}\right\} ds = 0.244 \end{gathered} \tag{3.40}$$

The equation for the second term in the inner solution (T_1 or Y_1, see (3.28)) is (3.27), which, with (3.24) and (3.38), becomes

$$Y_1'' + 2\left(S_0'/S_0\right)Y_1' - Y_1 = -W_{1s} - s_1 X \tag{3.41}$$

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

With $S_0(X)$ available only in inverse form, (3.39), we adopt an approximation for the variable coefficient. It is straightforward to show that

$$S_0'/S_0 \sim 1/X \text{ for } X \ll 1, \text{ and } S_0'/S_0 \sim e^{-X} \text{ for } X \gg 1, \tag{3.42}$$

and that S_0 has the shape of the tanh function. We use the approximation

$$S_0 = \tanh(X/2) \tag{3.43}$$

which has the required behaviour at $X \ll 1$, and asymptotic decay for $X \gg 1$. In addition it has the property of yielding an explicit solution to (3.41), which is welcome for SPICE application. This is

$$\begin{aligned} Y_1 = \tfrac{1}{3}e^{-X}\left(1-e^X\right)^{-1}\Big\{&E_1\left(1+3e^X\right)+E_2\left(1+e^X\right)^3-W_{1s}\left[2+3e^X+3e^{2X}\right]\\ &+s_1\left[2e^X+2e^{2X}+3Xe^X+3Xe^{2X}+2Xe^{3X}\right.\\ &\left.-2\left(1+3e^X+3e^{2X}+e^{3X}\right)\ln\left(1+e^X\right)\right]\Big\} \end{aligned} \tag{3.44}$$

Since the E_2 term is exponentially large for $X \gg 1$, we have $E_2 = 0$. In order to retain zero electron density at $X = 0$, it is sufficient to require that Y_1 be finite there, see (3.28). This is achieved in (3.44) by requiring that

$$E_1 = 2W_{1s} + s_1(1 - 4\ln 2) \tag{3.45}$$

For matching purposes, the behaviour of Y_1 as $X \to \infty$ is noted:

$$Y_1 = W_{1s} + s_1 X + O\left(Xe^{-X}\right) \quad \text{as } X \to \infty. \tag{3.46}$$

The solution for W_2, (3.25), may be written as

$$W_2 = \frac{\tau_{0s}^2}{\lambda}\int_0^X\int_0^s S_0^2(t)\,dtds + \frac{X^2}{2} + W_{2s}'X + W_{2s}. \tag{3.47}$$

To the order that the expansions are calculated here, the explicit integration in (3.47) is not required; however its asymptotic behaviour as is required for matching purposes. The asymptotic behaviour of S_0 is given in (3.40) and with this it can be shown that

$$\int_0^X\int_0^s S_0^2(t)\,dtds = X^2/2 + (k_2 - k_1)X + k_1 + k_3 + o(1) \text{ as } X \to \infty, \tag{3.48}$$

where
$k_2 = \int_0^\infty \left\{S_0^2(t) - 1 + k_1 e^{-t}\right\} dt = 0.877$, $k_3 = \int_0^\infty\int_s^\infty \left\{S_0^2(t) - 1 + k_1 e^{-t}\right\} dtds = 0.476$. We now return to the outer expansion to find the second term, governed by (3.33). Using (3.31) and (3.34) the equation for w_1 becomes

$$\frac{d^2w_1}{dr^2} = 2w_1\text{cosech}^2(r) \quad \text{where } r = x\alpha_0\ln\lambda\Big/\sqrt{2} + \gamma \tag{3.49}$$

which has the solution

$$w_1 = D_1\coth r + D_2(r\coth r - 1). \tag{3.50}$$

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

The undetermined constants in the solution above are determined by matching. The outer expansion (3.30) for which we have found and ((3.34), (3.50)) is matched on the one side with the quantum inner solution, and on the other side with the depletion solution. The latter is given by (3.4). The quantum effect on this solution is modelled by taking expansions of the coefficients a and b:

$$(a, b) = (a_0, b_0) + \varepsilon\,(a_1, b_1) + \varepsilon^2\,(a_2, b_2) + \cdots \tag{3.51}$$

The depletion/inversion layer matching follows that performed in Ward *et al.* [32, p. 1113] and yields

$$\begin{aligned} &a_0 = -\sqrt{2}\alpha_0,\ b_0 = w_{0s} + 2\,(\ln(2\alpha_0) - \gamma)/\ln\lambda \\ &a_1 = \alpha_0 D_2 \ln\lambda \Big/ \sqrt{2}\ , \quad b_1 = D_1 + (\gamma - 1)\,D_2. \end{aligned} \tag{3.52}$$

Further matching of the depletion solution with the bulk solution ($w = -1$) is accomplished by expressing the quantum effect on the depletion layer thickness as

$$x_d = x_{d0} + \varepsilon x_{d1} + \cdots \tag{3.53}$$

and this gives

$$\begin{aligned} &x_{d0} = -a_0 - k_4 \Big/ \sqrt{2\ln\lambda}, \qquad x_{d1} = -a_1 = -b_1/a_0 \\ &\text{where } a_0 = -\sqrt{2}\left(b_0 + 1 - 1/\ln\lambda\right)^{1/2}, \\ &k_4 = \int_1^\infty \left[(y-1)^{-1/2} - \left(e^{-y} + y - 1\right)^{-1/2}\right] dy = 0.818. \end{aligned} \tag{3.54}$$

These relationships yield

$$D_1 = k_5 D_2 \quad \text{where} \quad k_5 = \left(1 - \gamma - \alpha_0^2 \ln\lambda\right). \tag{3.55}$$

The preliminary matching across the quantum/inversion layers, accomplished in (3.35)–(3.37) can now be continued. Matching the η^2 and $\varepsilon\eta$ terms requires the quadratic and linear terms from (3.47) and similar terms from the expansion of w_0 from (3.34). The η^2 terms match identically, and matching the $\varepsilon\eta$ terms requires that

$$w_{1s}' = (k_2 - k_1)\,\tau_{0s}^2/\lambda + W_{2s}'. \tag{3.56}$$

Combining this with the boundary condition (3.29c) and the matching condition (3.35b) gives

$$cw_{1s} - w_{1s}' = (k_1 - k_2)\,\tau_{0s}^2/\lambda, \tag{3.57}$$

Note that this is a Robin boundary condition similar to the one, (3.29b), governing the first term in the outer solution. Using the solution for w_1, (3.50), in (3.57) then allows D_2 to be found from

$$D_2(ck_6 - k_7) = (k_1 - k_2)\,\tau_{0s}^2/\lambda \tag{3.58}$$

where $k_6 = (k_5 + \gamma)\coth\gamma - 1$, $\quad k_7 = \alpha_0 \ln\lambda\left[\coth\gamma - (\gamma + k_5)\,\mathrm{cosech}^2\gamma\right]\Big/\sqrt{2}\ D_1$ is then available from (3.55).

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

The above analysis completes the goals we have:

- The first term outer solution is given by (3.34), the constants there being found as per the prescription given in Cumberbatch *et al.* [11].
- The first term inner solution, given by (3.39).
- First-order corrections to these solutions, being given by (3.44) and (3.50).

3.4 Composite solution

The quantum layer and outer solutions have been calculated separately. A composite solution is achieved by adding them together and subtracting the common terms, i.e. the terms that match in the intermediate layer. This is given by

$$Y = \ln(n/n_i)/\ln\lambda = w_0 + \varepsilon w_1 + Y_0 + \varepsilon Y_1 - \varepsilon(w_{1s} + s_1 X) + O(\varepsilon^2) \tag{3.59}$$

where w_0, w_1, Y_0 and Y_1 are given by (3.34), (3.50) and (3.28), respectively.

4 Results

In the following, results from the asymptotic analysis (labeled MAE in the Figures) are compared with numerical results (labeled SP). The latter are obtained from a program called SCHRED, available on the website www.nanohub.org . This program solves the Schrödinger equation for the wave function Ψ (see (2.7)), from which the electron density is calculated. This, in turn, is substituted into the Gauss equation, (2.6), to obtain the electrostatic potential, ψ. The latter is then available to be used in (2.7), since $V = -q\psi$. These solutions must be made to conform – the so-called "self-consistent" solution – and this is usually done iteratively.

Results derived from the MAE solution, (3.59), and the SCHRED quantum solution are shown in Figures 4–9 for typical strong inversion cases. The inner first term and the combined first and second outer terms of the asymptotic solution show good comparison with the numerical results in their respective domains of validity. However, there are two problem areas with (3.59) (see Fig. 4): (1) The composite solution is not accurate in the domain bridging the inner and outer regions, and (2) the second term in the inner solution has too large an influence outside the inner region, spoiling the composite. Our goal is to provide an accurate approximation valid across the full range of voltages, and we shall develop an amended MAE that has this property in what follows.

MAE examples in the standard texts [8, 15] show various discrepancies from the exact solution especially in the bridging region, this error diminishing with reducing values of the expansion parameter. Hence it is plausible to assert that the values of ε that we use in the examples considered are just not small enough for the asymptotic technique to provide close accuracy. Our problem is exceedingly stiff even in its classical mode though, so this conjecture may not be the whole story and the fault may be that we have adopted only simple expansions in ε and assumed that the role of the large parameter λ has been organized by the classical solution. The direct cause of problem (ii) is the size of s_1 in strong inversion (s_1 is proportional to the classical surface electric field), so that the second term in the inner solution in (3.59) involves the direct interaction of ε with λ.

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

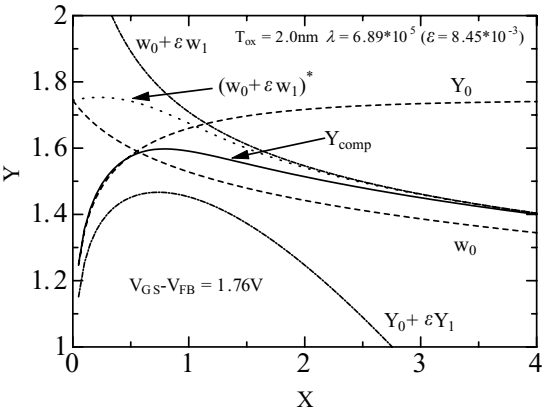


FIGURE 4. Various component terms of the composite solution for Y, (3.59). The amended outer solution is $(w_0 + \varepsilon w_1)^*$, defined in (4.1). The function Y_{comp} is defined in (4.2).

Being mindful that the goal of this analysis is to find a formula acceptable for engineering use, the fact that the inner first term and the outer two terms are each good approximations except over the bridging region is exploited. Forming a composite solution between these two is not possible directly since $w_0 + \varepsilon w_1$ has a limit as $x \to 0$ above the limit of Y_0 as $X \to \infty$ (see Fig. 4). Hence we amend $w_0 + \varepsilon w_1$ in the quantum layer to have the correct limit for matching with Y_0. After some experimentation we chose

$$(w_0 + \varepsilon w_1)^* = \left\{w_0 + \varepsilon w_1(1 - e^{-1\cdot 3X})\right\}\left\{1 + e^{-3(X-0.5)}\right\}^{-1} + w_{0s}\left\{1 + e^{3(X-0.5)}\right\}^{-1} \tag{4.1}$$

and then use, instead of (3.59), the composite solution

$$Y = (w_0 + \varepsilon w_1)^* + Y_0 - w_{0s} \tag{4.2}$$

Figures 5–9 show the results using (4.2) in comparison with SCHRED results, with Fig. 5 indicating the strong effect the quantum term has in modifying the classical electron density profile in the quantum layer, and Fig. 6 showing the quantum density profiles at various gate voltages. The ordinate in Fig. 7 is $\int n(x)dx$ over the inversion region and is therefore a measure of the electrons available for conduction from source to drain. It shows the overall reduction in electron charge due to quantum confinement.

Figures 8 and 9 show error results. Figure 8 shows the average relative error in the electron density profile as a function of w_s. The average relative error is calculated using the following expression

$$\text{Relative Error} = \frac{1}{L}\int_0^L \frac{n_{\text{SP}}(x) - n_{\text{MAE}}(x)}{n_{\text{SP}}(x)}dx$$

where n_{SP} refers to SCHRED results and n_{MAE} refers to results using equation (4.2). The length L was taken large enough not to affect the result. The quantity w_s is the classical surface potential, and values larger than approximately 1.0 indicate the inversion regime. Note that the curves plotted for various device specifications follow a general trend, and the error increases both as w_s drops to the inversion threshold and at large values. In the

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

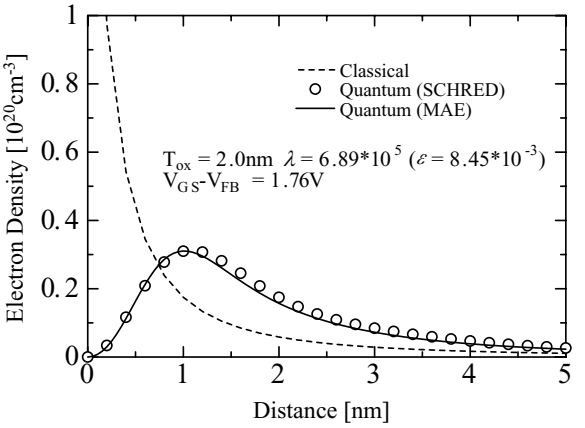


FIGURE 5. Electron density distribution with the MAE quantum solution based on (4.2).

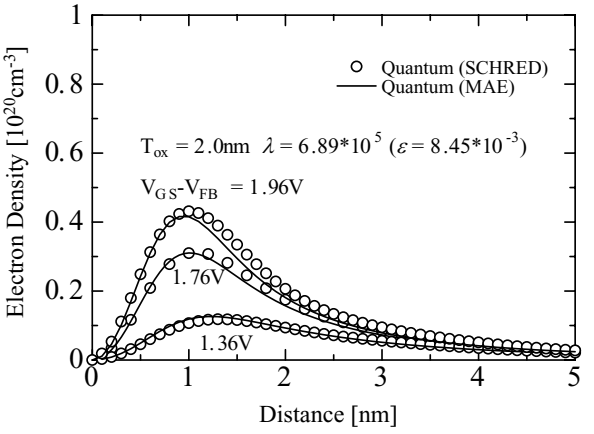


FIGURE 6. Same as Figure 5 at different gate voltages.

former the adoption of an outer inversion profile is losing validity, and in the latter we suspect that the inversion layer has thinned to such an extent at large gate voltage that the assumption that the quantum layer is a small fraction of the quantum layer is no longer satisfied.

Figure 9 explores one of the cases of Fig. 8 by separating the relative error into component parts from inner, transition and outer regions of the solution. These parts were roughly assessed to be in the domains (0.0, 0.75 nm), (0.75, 1.5 nm) and (1.5, 5.0 nm), respectively. This data shows that the inner solution has efficiently captured the quantum effect, but that the outer solution (to the 2 terms computed) is less well approximated. The shift of the quantum profile away from the oxide interface (see Figure 5) looks to be well approximated in the outer regions, but the small differences contribute significantly to the error calculated. The problem may be that the classical solution which comprises the first term in the outer solution computed in Section 3 is not a good enough approximation to

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

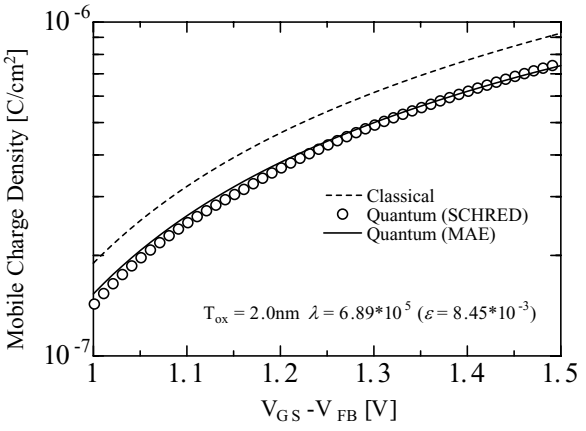


FIGURE 7. Mobile charge density (integral of $w(x)$ over the inversion region) as a function of gate voltage, showing the overall loss of electrons due to the quantum effect.

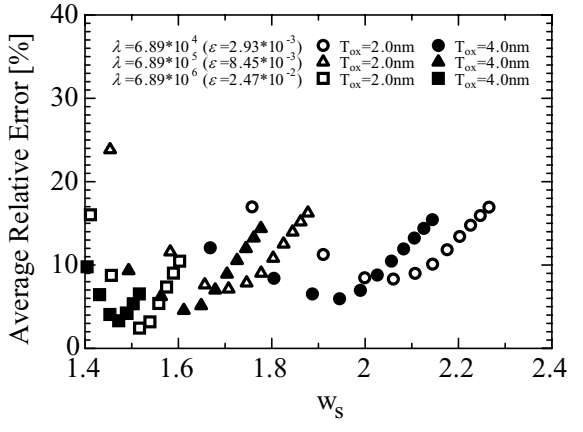


FIGURE 8. Averaged relative error as a function of w_s plotted for different oxide thicknesses and doping densities.

the exact solution, and addition of one more term in the asymptotic series is not sufficient to bring it close enough.

5 Conclusion

An approximate solution, based on a matched asymptotic expansion approach to the density gradient equations, has been found in order to quantify the reduction due to quantum effects in the electron density distribution in the standard MOSFET. The formulae obtained, though complicated, are given in terms of simple functions. In the expression for Y, (4.2), which is used for the results shown, the term $Y_0(X)$ is given by (3.28), (3.38), (3.39) as an inverse relation. However, the circuit simulators do not require the details of the electron density distribution. They use overall charge densities, which are integrals involving Y, and these formulae can be written down explicitly using (4.2).

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

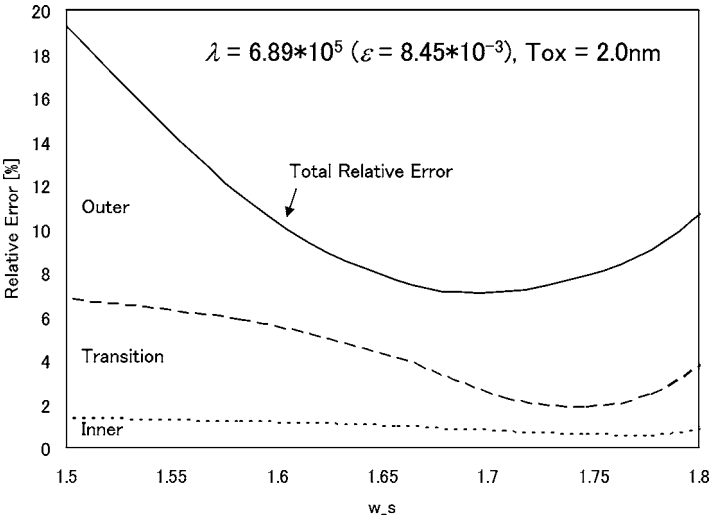


FIGURE 9. Relative error comprising inner, transition and outer region contributions.

Acknowledgements

Our device modeling efforts have been supported by the MOSIS service of ISI (see www.MOSIS.org) for a number of years, including post-doctoral support for Dr. Uno at Claremont Graduate University during the period when this work was completed. We would like to thank Cesar Pina and Vance Tyree from MOSIS for this support, and for numerous technical discussions.

References

[1] ABEBE, H. & CUMBERBATCH, E. (2003) Quantum mechanical effects corrections models for inversion charge and current-voltage (I-V) characteristics of the MOSFET device. *Proceedings 2003 Nanotechnology Conference*, San Francisco, USA, vol. 2, pp. 218–221.

[2] ABEBE, H. & AND CUMBERBATCH, E. (2004) Modeling quantum effects on MOSFET channel surface potential. *Proceedings International Conference on Computing, Communications and Control Technologies: CCCT'04*, Austin, TX, vol. 7, pp. 198–201.

[3] ANCONA, M. G. (1990) Asymptotic structure of the density-gradient theory of quantum transport. *Proc. Comp. Elec. Workshop*, Champaign-Urbana, IL.

[4] ANCONA, M. G. (1990) Macroscopic description of quantum-mechanical tunneling. *Phy. Rev. B*, **42**(2).

[5] ANCONA, M. G. & IAFRATE, G. J. (1989) Quantum correction to the equation of state of an electron gas in a semiconductor. *Phy. Rev. B*, **39**(13), May.

[6] ANCONA, M. G. & AND TIERSTEN, H. F. (1987) Macroscopic physics of the silicon inversion layer. *Phy. Rev. B*, **35**(15).

[7] ANCONA, M. G., YU, Z., LEE, W. C., DUTTON, R. W. & VOORDE, P. V. (1997) Density-Gradient simulations of quantum effects in ultra-thin-oxide MOS structures. *IEEE Proceedings. Simulation of Semiconductor Processes and Devices, International Conference, SISPAD'97*, pp. 97–100.

[8] BENDER, C. M. & ORSZAG, S. A. (1999) *Advanced Mathematical Methods for Scientists and Engineers*. Springer-Verlag.

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

[9] Biegel, B. A., Rafferty, C. S., Yu, Z., Dutton, R. W. & Ancona, M. G. (1998) Simulation of ultra-small MOSFETs using a 2-D quantum-corrected drift-diffusion model. *35th Annual Technical meeting of Society of Engineering Science*, Pullman, WA, pp. 53–64.
[10] Bohm, D. (1952) A suggested interpretation of the quantum theory in terms of "Hidden" variables. *Phys. Rev.* **85**(2), 166–179.
[11] Cumberbatch, E., Abebe, H. & Morris, H. (2001) Current-voltage characteristics from an asymptotic analysis of the MOSFET equations. *J. Eng. Math.* **39**, 25–46.
[12] Enz, C. C., Krummenacher, F. & Vittoz, E. A. (1995) Analytical MOS transistor model valid in all regions of operation and dedicated to low voltage and low current application. *Analog Integrated Circuits & Process*, **8**, 83–114.
[13] Ferry, D. K., Akis, R. & Vasileska, D. (2000) Quantum effects in MOSFETs: Use of an effective potential in 3D Monte Carlo simulation of ultra-short channel devices. *IEDM Tech. Dig.*, pp. 287–290.
[14] Iafrate, G. J., Grubin, H. L. & Ferry, D. K. (1981) Utilization of quantum distribution function for ultra-submicron device transport. *J de Phys*, **C7 10**(42), 307–312.
[15] Kevorkian, J. & Cole, J. D. (1996) Multiple scale and singular perturbation methods. *Appl. Math. Sci.* **114**, May.
[16] Liu, W. (2001) *MOSFET Models for SPICE Simulation.* John Wiley & Sons.
[17] Madelung, E. (1927) Quantentheorie in hydrodynamischer form. *Z. Phys.* **40**, 322.
[18] Markowich, P. A., Ringhofer, C. A. & Schmeiser, C. (1986) An asymptotic analysis of one-dimensional semiconductor device models. *IMA J. Appl. Math.* **37**, 1–24.
[19] Markowich, P. A., Ringhofer, C. A. & Schmeiser, C. (1998) *Semiconductor Equations.* Springer-Verlag.
[20] Morris, H. C., Cumberbatch, E., Phillips, T. & Hinderberger, B. (2000) Analytical results for the current-voltage characteristics of an SOI-MOSFET. *Proceedings Third International Conference on Modeling and Simulation of Microsystems*, San Diego, CA, March.
[21] Morris, H. C. & Abebe, H. (2004) MOSFET analytical substrate current model for circuit simulation. *Proceedings International Conference on Computing, Communications and Control Technologies: CCCT'04*, Austin, TX, vol. 1, pp. 162–165.
[22] Nadim, A., Cumberbatch, E., Attiyah, A., Dang, V., Mutlugun, C., SabolC. & Wong, C. (2002) *Gate capacitance modeling*. Claremont Graduate University and USC/ISI Mathematics Clinic report.
[23] Please, C. (1978) Some mathematical problems of semiconductor devices. DPhil Thesis, Oxford University.
[24] Please, C. (1982) An analysis of semiconductor P-N junctions. *IMA J. Appl. Math.* **28**, 301–318.
[25] Pregaldiny, F., Lallement, C. & Mathiot, D. (2003) Quantum surface potential model suitable for advanced MOSFETs simulation. *IEEE SISPAD Conference*, Boston, MA.
[26] Quan, W., Kim, D. M. & Lee, H.-D. (2002) Quantum C-V modeling in depletion and inversion: Accurate Extraction of electrical thickness of gate oxide in deep submicron MOSFETS. *IEEE Trans. Electron Dev.* **49**(5), May.
[27] Ramey, S. M. & Ferry, D. K. (2002) Modeling of quantum effects in ultrasmall FD-SOI MOSFETs with effective potentials and theree-dimensional Monte Carlo. *Physica B*, **314**, 350–353.
[28] Rios, R., Arora, N. D., Huang, C.-L., Khalil, N., Faricelli, J. & Gruber, L. (1995) A physical compact MOSFET model, including quantum mechanical effects, for statistical circuit design application. *IEEE IEDM Tech Dig.* pp. 937–940.
[29] Sze, S. M. (1981) *Physics of Semiconductor Devices.* John Wiley & Sons.
[30] Thompson, S., Packan, P. & Bohr, M. (1998) MOS scaling: transistor challenges for 21st century. *Intel Tech. J.* **1**(3).
[31] Ward, M . J. (1992) Singular perturbations and a free boundary problem in the modeling of field- effect transistors. *SIAM J. Appl. Math.* **52**, 112–139.
[32] Ward, M. J., Odeh, F. M. & Cohen, D. S. (1990) Asymptotic methods for metal oxide semiconductor field effect transistor modeling. *SIAM J Appl. Math*, **50**(4), 1099–1125.

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press

[33] Watling, J. R., Brown, A. R. & Asenov, A. (2002) Can the density gradient approach describe the source-drain tunneling in decanano double-gate MOSFETS? *J. Computational Electr.*, in press.

[34] Watling, J. R., Brown, A. R., Asenov, A. & Ferry, D. K. (2001) Quantum corrections in 3-D drift diffusion simulations of decanano MOSFETs using an effective potential. In: D. Tsoukalas and C. Tsamis (eds.), *Simulation of Semiconductor Processes and Devices*, pp. 82–85. Springer-Verlag

https://doi.org/10.1017/S0956792506006656 Published online by Cambridge University Press