# Current-Voltage Characteristics from an Asymptotic Analysis of the MOSFET Equations

Ellis Cumberbatch, [a] Henok Abebe, [b] and Hedley Morris [c]

[a] *Department of Mathematics, Claremont Graduate University, 121 E. Tenth Street, Claremont, CA 91711*

[b] *Information Sciences Institute,4676 Admiralty Way, Marina del Rey, California 90292-6695 USA*

[c] *Department of Mathematics, San Jose State University, One Washington Square, San Jose, CA 95192*

**Abstract**

Phenomenological formulae for the current-voltage characteristics of a MOSFET are derived by extending the asymptotic method of Ward [12]. Practical methods for combining the perturbation approximations and numerically implementing the Ward equations are developed. A detailed comparison with real MOSFET data is presented and the model is shown to be effective over a range of device geometries.

## 1 Introduction

A semi-conductor chip is made by fabricating up to $10^6$ transistors and their inter-connections on the surface of a silicon wafer. The transistors are usually MOSFET's (metal-oxide-silicon-field-effect-transistors), drawn from a small set of designs ($n$-type, $p$-type) and of various sizes (lengths and widths). Each chip has circuit functions, and in order to design the circuits using the transistor building blocks, the VLSI industry uses a program called SPICE. SPICE is the most common program for the simulation of electronic circuits. The original SPICE was developed in 1971 at U.C. Berkeley, and updated versions exist in many forms both commercial and Public Domain. See [5] for more information. For SPICE to produce a design the software must access the current-voltage relation for *each* transistor a number of times. In order that this be effected in a reasonable time the current-voltage relationships for the transistors must have simple form.

The classical approximation governing the flow of electrons and holes in a semi-conductor relevant to the low field-strength operation of a transistor has been in use for over forty years [10],[11]. It results in a set of non-linear partial differential equations for the electron density, $n$, the hole density, $p$, and the electric field $\mathbf{E}$. Even in the time-independent case, which will be considered here and which is sufficient for many applications, the non-linear character and the non-trivial geometry of the MOSFET preclude approaches that would generate exact solutions, least of all any that have simple current-voltage relationships. Hence there has resulted a variety of approximate solutions. A typical MOSFET geometry is shown in Figure 1..

The first approximation, which allows a reduction from a PDE system to an ODE system has the designation "long channel" or "quasi-one-dimensional" (called 1-D in the sequel). As indicated, this approximation has validity when the channel length, L, is relatively large. (See Figure 1 for the coordinate system, for identification of various regions on the device, etc. Also refer to Appendices 1, 2 for variable definitions). When voltages are applied at the gate and or drain, electrons and holes are attracted or repelled from these locations. The depletion region, bereft of either electrons or holes, plays an important role; a measure of its size is $L_d(\frac{\ln\lambda}{\lambda})^{\frac{1}{2}}$ where $L_D$ is the Debye length and $\lambda$ is a measure of the doping level relative to the intrinsic level. ($\lambda$ is large, typically $10^6 - 10^7$.) The ratio of these lengths, which plays the role of the aspect ratio of the device, is

$$\epsilon = \frac{L_d}{L}\sqrt{\frac{\ln\lambda}{\lambda}} \tag{1}$$

When $\epsilon$ is small most of the region under the gate is uniform in the $x_2$ -direction and the $1-D$ approximation is exact for the fields there; the approximation breaks down in the source-gate and drain-gate corners where the fields lose their $1-D$ character. When $\epsilon$ is large the corner regions influence the solution across the whole gate region and a two-dimensional solution is required. Ward, [12], made the $1-D$ approximation into an explicit asymptotic expansion by scaling $x_2$, the variable along the channel with $L$ and scaling $x_1$, the perpendicular variable, with $L_D$. Series expansions in powers of $\epsilon^2$ give the $1-D$ approximation as the leading term of these series. In a subsequent paper, [13], Ward obtained some solutions for the corner regions and matched them to the $1-D$ approximation in the central region under the gate. This enhanced solution would be expected to be more accurate down to smaller channel lengths, but the solution obtained is restricted to small drain voltages. For a typical device $L_d = 33$ microns, $\lambda = 10^6$, and for $L = 10$ microns, $\epsilon = 0.012$. For $\epsilon$ this small we expect the $1-D$ approximation to be accurate; this is borne out by the data. However we show that by adjustments of some of the parameters we can make the $1-D$ approximation match the data at much reduced chan-

nel lengths. At the time of writimg, the standard industry channel length is 0.25 micron, and the SPICE software, based on the 1-D approximation (but improved impirically), is still being used.

The $1-D$ approximation has the descriptor "quasi" because the ODE in the $x_1$ direction contains $x_2$ as a parameter via the quasi-Fermi potential, $\varphi$. The electric potential, $\psi$, satisfies the second-order non-linear ODE in $x_1$, which contains $\varphi$, and therefore its solution depends on $x_2$ parametrically. The form of $\varphi(x_2)$ is determined subsequently when the current flowing across the device is calculated. The ODE allows one integration in the case of constant doping. The current can then be expressed as a double integral called the Pao-Sah integral, see [6]. An approximation allows reduction to a single integral [7]. Further approximations valid over different regimes of the applied voltages $V_{ds}, V_{gs}$ then yield explicit formulae for $I_{ds}(V_{ds}, V_{gs})$. These latter approximations are motivated by neglect of various physical effects in one or other regimes of operation. A comprehensive survey of these approximations and their inter-relations is found in [11] and [4]. These separate formulae, valid over distinct ranges of voltages, were found to be accurate for long channel use. As the need for denser packing of transistors on chips has intensified, and the technology has been able to satisfy those needs by reducing lengths and widths of transistors, these formulae have been adapted in various ways for use at the reduced sizes. Often these adaptations are little more than "fudge factors," which have parameters valid only over limited ranges of size and voltage regimes. Hence these parameters are changed regime to regime ("binning"). Current technology uses of the order of 150-200 parameters for $I_{ds}(V_{ds}, V_{gs})$ formulae for use in SPICE. Most of these parameters are not universal constants, and they must be "identified" for each fabrication batch. This means considerable testing and data collection, followed by numerical optimization for parameter identification. All of this is expensive, both in time and money. There are pressures to improve the models so that they are less cumbersome and based more on physics. The work reported here, done by the Claremont Graduate University Mathematics Clinic for the Information Sciences Institute MOSIS Program, has these more efficient models as its goal. Related goals were (a) comparison of the asymptotic results with data to determine the domain of validity of the asymptotic approximations, and (b) improvement of solution "blending" by which solutions valid over different voltage regimes are patched together.

The work presented in here generates explicit $I_{ds}(V_{ds}, V_{gs})$ formulae based on the $1-D$ approximation. It takes further the asymptotic results for large $\lambda$ derived by Ward in (Ward 1990, 1992). These results in turn were based on the asymptotic scheme originated by Please, [8], for the $p-n$ junction. This scheme, based on the matched asymptotic expension concept, [1], obtains solutions valid in different regions and matches them across overlapping regions to provide a composite expansion valid over the whole region (to the order com-

puted). In the $1 - D$ MOSFET approximation there are three main regions: the inversion layer, a thin boundary layer containing a profusion of holes or electrons under some voltage regimes, the depletion layer, and the substrate which is determined by conditions far from the channel. The separate solutions in these regions are, by themselves, not difficult to obtain. However, matching them, and satisfying the boundary condition at the gate, are non-trivial exercises. Ward successfully accomplished these, in [12] for constant doping and for variable doping of constant signature, and in [13] for variable doping of opposite signature.

The asymptotic solution achieved by Ward requires numerical work (to solve non-linear transcendental equations) and the $I_{ds}(V_{ds}, V_{gs})$ formula remains an integral to be computed numerically. Hence its use in the semi-conductor industry (via SPICE) is precluded. We briefly outline Ward's asymptotic results, and in Section 3 we show how the Ward analysis may be taken further (in the constant doping case) to achieve an *explicit* $I_{DS}(V_{ds}, V_{gs})$ formula. Part of our analysis is justifiable as a continuation of the asymptotics and it agrees with the numerical solution of the exact ODE/BC system. A second approximation is not justifiable asymptotically. It approximates by a constant a function that is varying inside an integral; with this, though, the integral can be evaluated analytically, and the two approximations combined to produces a formula for the current.

## 2 Equations, Boundary Conditions, and the Quasi-1-D Asymptotic Approximation

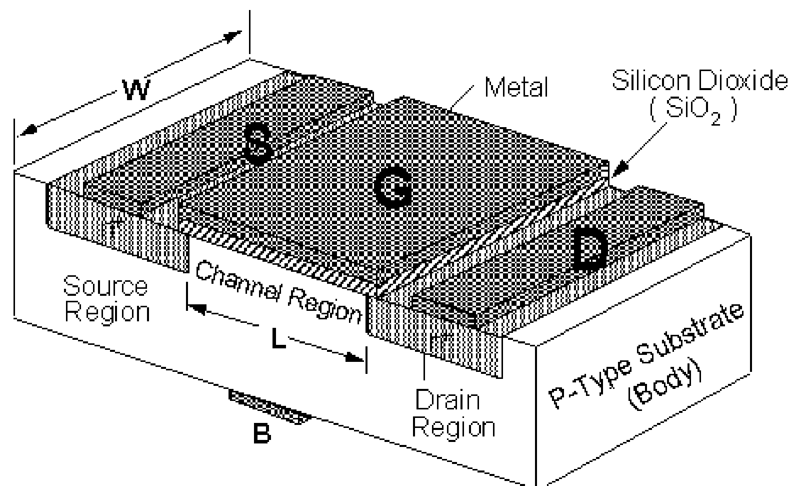


Fig 1. A schematic diagram of a MOSFET.

Voltages may be applied at the source, drain, gate and substrate contacts (S,D,G,B in Figure 1); we shall assume that the substrate and source are at the same voltage. The voltages drain-to-source and gate-to-source are denoted by $V_{ds}, V_{gs}$, respectively. We shall treat an n-channel device (in inversion the channel is profuse in electrons). Here the silicon has been doped with acceptor atoms (typically the number ionized is $N_A^- \sim 10^{16}$–$10^{17}$). Source, drain regions have been overlaid with donor atoms ($N_D^+ \sim 10^{19}$). An insulated layer (silicon dioxide) separates the gate contact from the silicon. As the gate voltage is raised, holes are repelled from the region under the gate, creating a depleted region. Further increase in $V_{gs}$ attracts electrons to the insulator-substrate interface, creating the "channel". Voltages at the drain then cause a current to flow from source to drain. The aim of analysis of the device is to obtain the current created by this flow of electrons, $I_{ds}$, as a function of the applied voltages $V_{ds}$ and $V_{gs}$. In addition, $I_{ds}$ also depends on the parameters of the device: its size, the doping levels, the mobility of electrons, etc.

The standard model for the MOSFET comprises Gauss' equation for charge

$$\epsilon_s \boldsymbol{\nabla} \cdot \mathbf{E} = -\rho = q(p - n + N) \tag{2}$$

together with the drift-diffusion model for the motion of electrons

$$\mathbf{J}_n = q\mu_n \left(\frac{kT}{q} \boldsymbol{\nabla} n + n\mathbf{E}\right) \tag{3}$$

and the static conservation equation with no recombination/generation for the flux of electrons

$$\boldsymbol{\nabla} \cdot \mathbf{J}_n = 0. \tag{4}$$

The assumption that the holes remain in thermal equilibrium allows the hole density to be obtained as

$$p = n_i e^{-q\frac{\psi}{kT}} \tag{5}$$

It is convenient to write

$$n = n_i e^{q\frac{\psi - \varphi_n}{kT}} \tag{6}$$

with $\varphi_n$ replacing $n$ as a primary variable; it is called the quasi-Fermi potential. For other notation and definitions see Appendix A.

The boundary conditions comprise

(i) specification of the electric potential at the gate contact:

$$\psi = V_{gs} - V_{fb} \text{ at } \mathrm{x}_1 = -\mathrm{t} \tag{7}$$

where $V_{fb}$ is the "flat-band voltage" which is a voltage that results from terminal contacts with the substrate [10].

(ii) continuity of electrical potential and electric displacement at the oxide-substrate interface:

$$\psi^- = \psi^+ \text{ and } \epsilon_\mathrm{i}\frac{\partial\psi^-}{\partial \mathrm{x}_1} = \epsilon_\mathrm{s}\frac{\partial\psi^+}{\partial \mathrm{x}_1}\text{at } \mathrm{x}_1 = 0 \tag{8}$$

(iii) specification of the electric and quasi-Fermi potentials at the source-substrate and drain-substrate boundaries. These are given by Ward, [12], and will not be written out here. They are required for the specification of $\psi, \varphi_n$ at these boundaries, and they will be introduced when needed subsequently.

(iv) current flow occurs only across the source-substrate and drain-substrate boundaries. A consequence of this is that that the total current flowing across a plane $x_2 = \text{const}$ does not vary with $x_2$ for $0 \leq x_2 \leq L$.

The independent and dependent variables are scaled, see Appendix B. The differential scaling of the independent variables introduces the parameter $\epsilon$, equation (1), into the field equations (2) and (4). The quasi-one-dimensional approximation is obtained for the solution in the mid-channel region (away from the source and drain corners) by the asymptotic expansion for the scaled electric and quasi-Fermi potentials

$$w(x,y) = w^0(x,y) + \varepsilon^2 w^1(x,y) + ... \tag{9}$$

$$\varphi(x,y) = \varphi^0(x,y) + \varepsilon^2 \varphi^1(x,y) + ... \tag{10}$$

The electron mobility in (3) is not necessarily constant: it may vary spatially and with field strength. The Ward assumption is

$$\mu_n = \mu_n(x, \frac{\partial\varphi^0}{\partial y}) > 0 \tag{11}$$

To leading order

$$\frac{\partial^2 w^0}{\partial x^2} = \frac{1}{\lambda}(\exp((w^0 - \varphi^0)\ln(\lambda)) - \exp(-w^0\ln(\lambda)) + d(x) \tag{12}$$

$$\frac{\partial^2 \varphi^0}{\partial x^2} = -(\frac{1}{\mu_n}\frac{\partial\mu_n}{\partial x} + \ln(\lambda)(\frac{\partial w^0}{\partial x} - \frac{\partial\varphi^0}{\partial x}))\frac{\partial\varphi^0}{\partial x} \tag{13}$$

The equation for $\varphi^0$ can be integrated, and with the no-flux boundary condition it follows that

$$\frac{\partial \varphi^0}{\partial x} = 0 \text{ and } \varphi^0 = \varphi(\mathrm{y}) \tag{14}$$

where we have dropped the zero superscript for convenience.

The order $\epsilon^2$ terms give for $\varphi^1$

$$\frac{\partial^2 \varphi^1}{\partial x^2} + (\frac{1}{\mu_n}\frac{\partial \mu_n}{\partial x} + \ln(\lambda)\frac{\partial w^0}{\partial x})\frac{\partial \varphi^1}{\partial x} = -(\varphi'' + (\frac{1}{\mu_n}\frac{\partial \mu_n}{\partial y} + \ln(\lambda)(\frac{\partial w^0}{\partial y} - \varphi^{'}))\varphi') \tag{15}$$

where primes represent derivatives with respect to y. By means of an integrating factor

$$\frac{\partial}{\partial x}(\mu_n \exp(\ln(\lambda) w^0)\frac{\partial \varphi^1}{\partial x}) = -\mu_n \exp(\ln(\lambda) w^0)(\varphi'' + (\frac{1}{\mu_n}\frac{\partial \mu_n}{\partial y} + \ln(\lambda)(\frac{\partial w^0}{\partial y} - \varphi'))\varphi') \tag{16}$$

If we integrate each side of (16) from $x = x^*$ ($x^*$ is a point in the substrate where there is no current flux across $x = x^*$) we obtain for the left-hand side

$$[(\mu_n \exp(\ln(\lambda) w^0)\frac{\partial \varphi^1}{\partial x})]_0^{x^*} = 0 \tag{17}$$

as $\frac{\partial \varphi^1}{\partial x} = 0$ when $x = 0$ and $x = x^*$. It follows that we must have

$$\varphi'' + \frac{\partial}{\partial y}\ln(\int_0^{x^*} \mu_n \exp(\ln(\lambda)(w^0 - \varphi))dx)\varphi' = 0 \tag{18}$$

The notation

$$\sigma_m = \int_0^{x^*} \mu_n \exp(\ln(\lambda)(w^0 - \varphi))dx \tag{19}$$

allows the equation determining the y dependence of $\varphi$ to be integrated once to give

$$\varphi' \sigma_m \ln(\lambda) = C \tag{20}$$

## 3 Ward's $\lambda \gg 1$ Asymptotic Expansion

We copy the 1-D scaled charge conservation equation (12), written in the form

$$w_{xx} = \exp((w - \varphi - 1)\ln\lambda) \; - \exp(-(w+1)\ln\lambda) \; +1 \quad \text{for x} > 0 \tag{21}$$

$$\text{(1)} \qquad\qquad \text{(2)} \qquad\qquad\qquad\qquad \text{(3)} \qquad\qquad \text{(4)}$$

The second line provides notation for the identification of terms: terms (2),(3) and (4) on the RHS represent charge densities due to electrons, holes and the static impurity doping, respectively. The impurity doping is taken to be constant: for an n-channel device and with the scaling adopted in Appendix B, it has the value unity. The electric potential, $w$, asymptotes to the value $w = -1$ in the substrate (terms (3) and (4) balance) and this provides the boundary condition at large $x$. There are no charges in the oxide layer, $-t < x_1 < 0$, giving a potential linear in $x$. This together with the boundary condition that the potential is $V_{gs} - V_{fb}$ at $x_1 = -t$ and the continuity of electric displacement at $x = 0$ means that the latter can be written

$$\frac{\partial w}{\partial x}|_s = c_{ox}\sqrt{\frac{\ln\lambda}{\lambda}}(w_s - V_{gs}^*) \tag{22}$$

where $s$ refers to values taken at $x = 0^+$. Figure 2. shows $w(x)$. (See also Figure 2 in Ward 1990.)

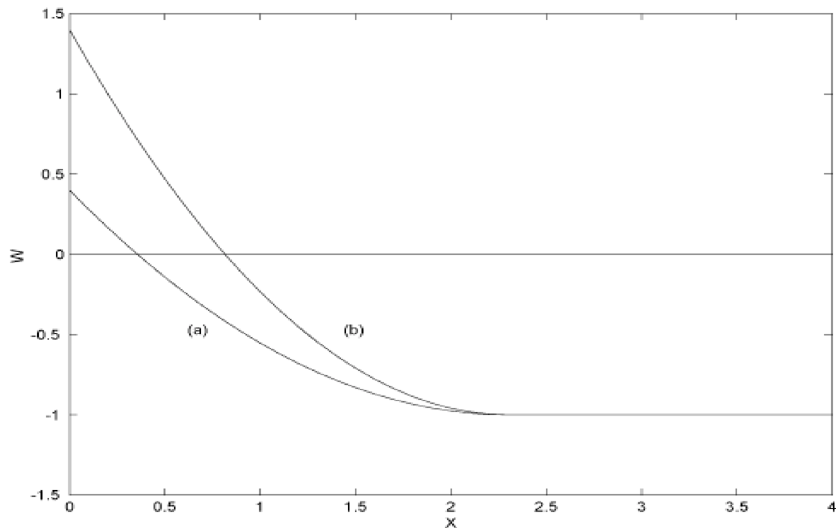


Fig 2. The function $w(x)$ in (a) inversion and (b) depletion.

In equation (21) $\varphi$ is a function only of $y$ which is determined subsequently. Equation (21) allows one integration which together with the boundary condition in the substrate yields

$$w_x^2/2 = \{\exp[w - \varphi - 1) \ln\lambda] + \exp[-(w+1)\ln\lambda]\} / \ln\lambda + w + 1 + 0(\lambda^{-2}) \quad (23)$$

Evaluating this at $x = 0^+$ relates $w_s, w_{xs}$ and $\varphi$ while (22) relates $w_s, w_{xs}$ and $V_{gs}^*$. Elimination of $w_{xs}$ yields a relation between $w_s, \varphi$ and $V_{gs}^*$. The formula for the current requires $w_s(\varphi)$ in an integration. Hence this may be accomplished numerically.

An alternative to the procedure just outlined is to solve (21) by the method of matched asymptotic expansions. In this, the spatial domain is separated into $\lambda$-dependent regions in which various terms in (21) may be neglected. More amenable solutions are achieved, then suitably "matched." There are three regimes of operation depending on whether term (2) is totally absent, present, or present only over part of the $y$ domains. Since $\varphi(0) = 0, \varphi(1) = V_{ds}/V_{th}\ln\lambda$, and $\varphi(y)$ is monotonic, the first two regimes categorize easily:

**Weak inversion - depletion:**

$$-1 < w_s < 1 + \frac{V_{ds}}{V_{th}\ln\lambda}, \quad \text{for all } y \quad (24)$$

**Strong inversion:**

$$w_s > 1 + \frac{V_{ds}}{V_{th}\ln\lambda}, \quad \text{for all } y \quad (25)$$

The last regime has part of the device (adjacent to the source) in strong inversion and part (adjacent to the drain) in weak inversion-depletion and the change-over depends on $\varphi(y)$. This is called pinch-off. Designating the pinch-off location by the subscript *po* the third regime is:

**Pinch off: weak inversion - depletion**

$$-1 < w_s < 1 + \varphi_{po}, \quad \text{for } y_{po} < y < 1 \quad (26)$$

In the weak inversion-depletion regime there is an $x$-domain adjacent to $x = 0$ in which the dominant terms in (21) are (1) and (4). This solution must be matched to the solution in the substrate via a transition layer (in which (1) = (3) + (4)). Note that $\varphi$ plays no role in this solution. In the strong inversion regime there is an $x$-domain adjacent to $x = 0$ in which the dominant terms in (21) are (1) and (2). (The latter term represents proliferation of electrons

in a narrow layer near $x = 0$ – the channel – and it is flow of these electrons along the channel that gives strong current flow). This inverted region must be matched to a depletion solution and at larger $x$ to the substrate solution. All these solutions and matchings are obtained in Ward; only the results will be quoted here.

**Strong Inversion Regime:** $w_s > 1 + \varphi$

From the solution given by Ward 1990, eq. (6.1),

$$\frac{\partial w}{\partial x}|_s = -\sqrt{2}\alpha_0 \coth \gamma \tag{27}$$

where

$$\alpha = \sinh \gamma, \;\; \alpha = \alpha_0 (\ln \lambda)^{1/2} \lambda^{(1-w_s+\varphi)/2} \tag{28}$$

$$\alpha_0^2 = 2 + \varphi + K - 1/\ln \lambda \tag{29}$$

$$K = \frac{\ln(\ln \lambda)}{\ln \lambda} + \frac{2}{\ln \lambda}\left[\ln(2\alpha_0) - \gamma\right]. \tag{30}$$

**Pinch-off: Asymptotic Solution**

In the pinch-off regime a portion of the device is in weak inversion-depletion. The asymptotic solution (Ward 1990, Section 4.1) is substituted into the boundary condition (22) giving

$$(w_s)_{po} = V_{gs}^* + \frac{1}{c^2 \ln \lambda} - \frac{1}{c^2 \ln \lambda}\left\{2c^2 \ln \lambda \left(V_{gs}^* + 1 - \frac{1}{\ln \lambda}\right) + 1\right\}^{1/2}. \tag{31}$$

Since the device becomes inverted for $w_s > 1 + \varphi$ (term (2) on the R.H.S. of (21) becomes dominant) the pinch-off boundary is at $\varphi_{po}$ where

$$\varphi_{po} = (w_s)_{po} - 1. \tag{32}$$

The asymptotic solution (27) is a replacement for the exact relations between $w_s$,$w_{xs}$ and $\varphi$ obtained from (23). The elimination of $w_{xs}$ from (22) and (27) to generate $w_s(\varphi)$ remains a numerical task; Sarvas and Spanier obtained results by Newton iteration. In the following section we take the asymptotics a little further so that $w_s(\varphi)$ is obtained in analytic form.

## 4 Further Asymptotics

If we combine the expressions for $\alpha_0$ and $K$ in (28 - 30) we obtain

$$\alpha_0^2 = 2 + \varphi + \frac{\ln(\ln\lambda)}{\ln(\lambda)} + 2\frac{\ln(2\alpha_0) - \gamma - \frac{1}{2}}{\ln(\lambda)} \tag{33}$$

Using the definition of $\gamma$ we can write this as

$$\alpha_0^2 = 2 + \varphi + \frac{\ln(\ln\lambda)}{\ln(\lambda)} + \frac{2}{\ln(\lambda)}\ln(2\frac{z\sqrt{\ln(\lambda)}}{1+\sqrt{\frac{z^2}{\alpha_0^2}\ln(\lambda)+1}}) - \frac{1}{\ln(\lambda)} \tag{34}$$

Also we write

$$c_{ox} = c\sqrt{\lambda} \tag{35}$$

In Ward $c_{ox} = 450$, and in the data supplied by ISI $c_{ox} = 1244$; for $\lambda \backsim 10^6$ this gives $c$ to be an $O(1)$ quantity.

Elimination of $w_{xs}$ between (22) and (27), and use of (35) yields

$$\left\{2\left(1+\alpha_0^2\ln\lambda\ \ \lambda^{1-w_s+\varphi}\right)\right\}^{1/2}\lambda^{(w_s-1-\varphi)/2} = c\ln\lambda(V_{gs}^* - w_s) \tag{36}$$

This complicated relationship between $w_s$ and $\varphi$ can be solved by iteration. It is convenient to introduce the notation:

$$r = w_s - 1 - \varphi\ \ , s = V_{gs}^* - 1 - \varphi,\ \ z = \lambda^{r/2}/\ln\lambda \tag{37}$$

**Note:** $z = \alpha_0/\alpha\sqrt{\ln\lambda}$ , but it helps to keep the notation just introduced.

With the substitutions (37), equation (36 ) reads

$$\sqrt{2(z^2+\alpha_0^2/\ln\lambda)} = c\left(s - 2\frac{\ln\ln\lambda}{\ln\lambda} - \frac{2\ln z}{\ln\lambda}\right) \tag{38}$$

Since $\ln\lambda \gg 1$, an iterative solution to (34) and (38 ) can be constructed in the following way. Define the sequences $\{\beta_j\}_{j=0}^{\infty}$ and $\{z_j\}_{j=0}^{\infty}$ by the recurrence formulae

$$\beta_{j+1}^2 = 2 + \varphi + \frac{\ln(\ln\lambda)}{\ln(\lambda)} + \frac{2}{\ln(\lambda)}\ln(2\frac{z_j\sqrt{\ln(\lambda)}}{1+\sqrt{\frac{z_j^2}{\beta_j^2}\ln(\lambda)+1}}) - \frac{1}{\ln(\lambda)} \tag{39}$$

$$z_{j+1} = \frac{c}{\sqrt{2}}(s - 2\frac{\ln(\ln\lambda)}{\ln(\lambda)} - 2\frac{\ln(\ln z_j)}{\ln(\lambda)})(1 + \frac{\beta_j^2}{z_j^2\ln(\lambda)})^{-\frac{1}{2}} \tag{40}$$

for $j \geq 0$ and start the sequences with the values

$$\beta_0^2 = 2 + \varphi + \frac{\ln(\ln\lambda)}{\ln(\lambda)} \tag{41}$$

$$z_0 = \frac{c}{\sqrt{2}}(s - 2\frac{\ln(\ln\lambda)}{\ln(\lambda)}) \tag{42}$$

Provided these sequences are convergent we have

$$z_j \to z \text{ and } \beta_j \to \alpha_0 \text{ as } j \to \infty \tag{43}$$

These values of $z$ give $r$ and hence $w_s$, in terms of $s$, and hence $\varphi$, with $V_{gs}^*$ as a parameter.

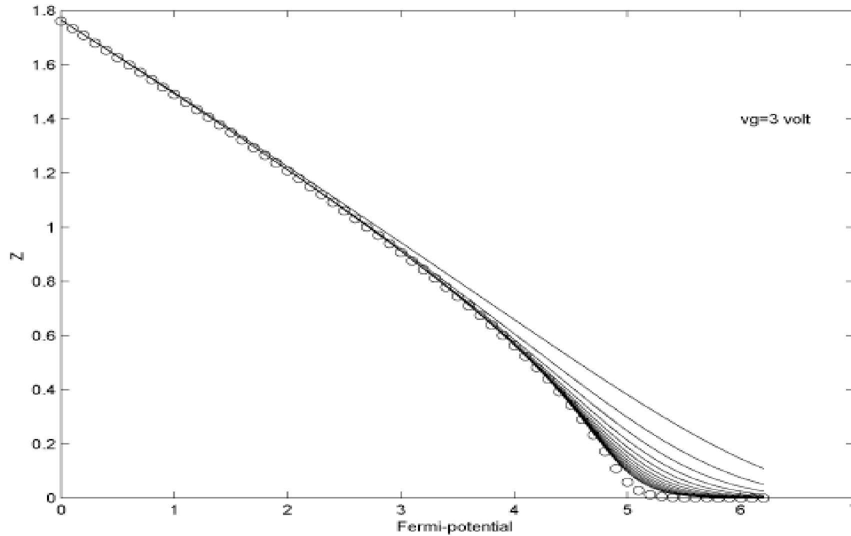


Fig 3. The first eleven iterations ($z_0$ to $z_{10}$) .

Figure 3. shows the first eleven iterations ($z_0$ to $z_{10}$) given by (40) for the solution of (38) in the case $\overline{V}_{gs} = 70$, and shows excellent convergence to the exact solution. Similar good results were obtained for other values of $\overline{V}_{gs}$. Since there is virtually no cost in running the iteration most subsequent calculations used the fourteenth iteration which differs very little from the exact solution. The latter is obtained from Sarvas and Spanier, [9], who solved (22), (23) numerically (and provided us with the MATLAB programs to reproduce their results). What is remarkable is the degree of similarity between the iterated solution we have obtained and the exact solution, since the former neglects

some terms and can only be considered as the first term in the asymptotic expansion for large $\lambda$.

Using the $z$-iterate the relation $w_s(\varphi)$ may be obtained from the relations (37) in the inverted region, and $(w_s)_{po}$ from (31) extends it for $\varphi > \varphi_{po}$. These results are shown in Figure 4. for $\overline{V}_{gs} = 70, 80, 90$, respectively.

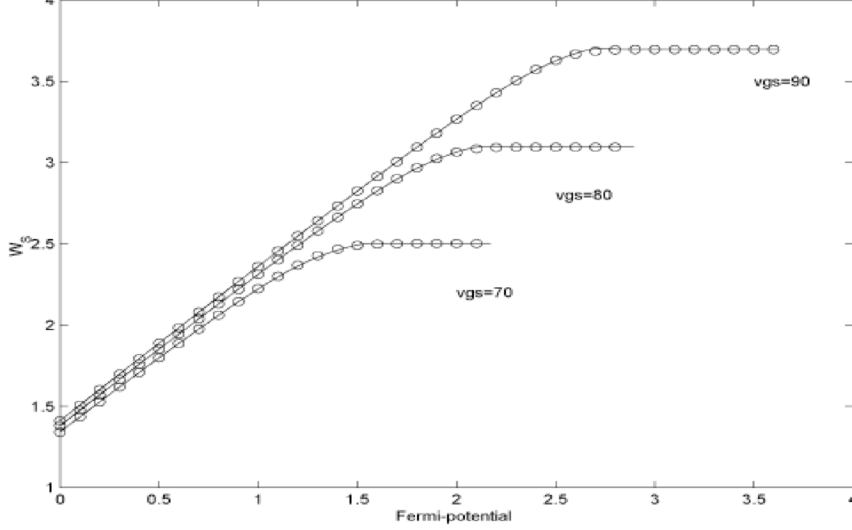


Fig 4. $w_s(\varphi)$ for $\overline{V}_{gs} = 70, 80, 90$.

The approximations pre-and post-pinch-off have been blended using the blending $Y$, see Appendix D. Also shown are the exact results obtained by Sarvas and Spanier, which indicate that a very high level agreement between the asymptotic and exact numerical approaches has been achieved.

## 5 Current

The definition of the quasi-Fermi potential $\varphi_n$, (6), allows the current flux $\mathbf{J}_n$, (3), to be expressed as

$$\mathbf{J}_n = q\mu_n \frac{kT}{q} n \boldsymbol{\nabla} \varphi_n \tag{44}$$

The current flowing from source to drain, $I_{ds}$, is obtained:

(a) by first integrating the $x_2$ - component of $\mathbf{J}_n$ in the $x_1$ - direction to obtain the current flowing across a plane $x_2 =$ a constant. This integral may be evaluated using the asymptotic representation of the solution, [12].
(b) by performing a second integration in the $x_2$ - direction from source to drain.

Since, as has been noted already, the current across a plane $x_2$ = a constant does not vary with $x_2$, this integration of $I_{ds}$ yields on the left-hand side $LI_{ds}$, whilst on the right-hand side the conjunction $\frac{d\varphi_n}{dx_2}dx_2 = d\varphi_n$ yields an integration with respect to $\varphi_n$. Since the solution is known in terms of $\varphi_n$, this artifice avoids determining $\varphi_n$ as a function of $x_2$.

We obtain expressions for $I_{ds}$ in the regimes when the whole device is inverted, and when it is in pinch-off (partially inverted, partially depleted)

**Linear Regime:**

When the whole device is inverted the current rises approximately linearly with $V_{ds}$ : hence the title. For constant mobility the current is found from Ward 1990 (equations (3.6,7), (6.5,6)) as

$$\frac{I_{ds}}{\mu I_c} = \sqrt{2}\ln\lambda \int_0^{\overline{V}_{ds}/\ln\lambda} \exp\left[\frac{1}{2}(w_s - \varphi)\ln\lambda - \gamma\right] d\varphi \tag{45}$$

The substitutions (37) allow us write the current integral in the form

$$\frac{I}{\mu I_c} = \sqrt{2\lambda}(\ln(\lambda))^2 \int_{z_d}^{z_s} H(z, \alpha_0^2(z))dz \tag{46}$$

where

$$H(z, \alpha_0^2(z)) = z\frac{ds}{dz}e^{-\gamma} \tag{47}$$

Equation (38) yields

$$\frac{ds}{dz} = \left\{\frac{\sqrt{2}}{c}\frac{z}{\sqrt{z^2 + \alpha_0^2/\ln\lambda}} + \frac{2}{z\ln\lambda}\right\}\left\{1 + \frac{1}{c\ln\lambda(2z^2 + 2\alpha_0^2/\ln\lambda)^{1/2}}\right\}^{-1} \tag{48}$$

Using the relations (30) we can express

$$e^{-\gamma} = (1 + \alpha_0^2/z^2\ln\lambda)^{1/2} - \alpha_0/z\sqrt{\ln\lambda} \tag{49}$$

The functions $\frac{ds}{dz}$ and $e^{-\gamma}$ can be seen to be decreasing functions of $\alpha_0^2$ for fixed $z$ by direct observation of their functional forms in (48) and (49). Consequently, the current integrand $H(z, \alpha_0^2(z))$ is monotonically decreasing as a function of

$\alpha_0^2$ for fixed $z$. The function $\alpha_0^2(z)$ is similarly monotonically decreasing and bounded on $[z_d, z_s]$

$$m \leq \alpha_0^2(z) \leq M \quad \text{for } \mathrm{z} \in [\mathrm{z_d}, \mathrm{z_s}] \tag{50}$$

Define the function

$$W(t) = \int_{z_d}^{z_s} H(z,t)dz \quad \text{for } \mathrm{t} \in [\mathrm{m}, \mathrm{M}] \tag{51}$$

This is monotonic in $t$ and lies in the interval $[W(M), W(m)]$. As $H(z, \alpha_0^2(z))$ is monotonically decreasing as a function of $\alpha_0^2$ for fixed $z$ we can write

$$W(M) \leq \int_{z_d}^{z_s} H(z, \alpha_0^2(z))dz \leq W(m) \tag{52}$$

By the intermediate value theorem there is a value of $t$ for which $W(t)$ takes any value in $[W(M), W(m)]$. In particular there is a value $t = t^* \in [m, M]$ for which

$$\int_{z_d}^{z_s} H(z, \alpha_0^2(z))dz = \int_{z_d}^{z_s} H(z, t^*)dz \tag{53}$$

As $\alpha_0^2(z)$ is monotonic decreasing on $[z_d, z_s]$ there exists a $z^* \in [z_d, z_s]$ for which

$$\alpha_0^2(z^*) = t^* = \overline{\alpha}_0^2 \tag{54}$$

From (34) the equation for $\overline{\alpha}_0^2$ is

$$\overline{\alpha}_0^2 = 1 + V_{gs}^* - \frac{\sqrt{2}}{c}\sqrt{(z^*)^2 + \frac{\overline{\alpha}_0^2}{\ln(\lambda)}} - \frac{\ln(\ln\lambda)}{\ln(\lambda)} + \frac{2}{\ln(\lambda)}\ln\left(\frac{2}{\sqrt{e}}\frac{\sqrt{\ln(\lambda)}}{1 + \sqrt{\frac{(z^*)^2}{\overline{\alpha}_0^2}\ln(\lambda) + 1}}\right) \tag{55}$$

which can be written in the form

$$\overline{\alpha}_0^2 = P - \frac{\sqrt{2}}{c}\sqrt{(z^*)^2 + \frac{\overline{\alpha}_0^2}{\ln(\lambda)}} \tag{56}$$

involving

$$P = 1 + V_{gs}^* - \frac{\ln(\ln\lambda)}{\ln(\lambda)} + \frac{\kappa}{\ln(\lambda)} \tag{57}$$

and

$$\kappa = 2\ln(\frac{2}{\sqrt{e}}\frac{\sqrt{\ln(\lambda)}}{1 + \sqrt{\frac{(z^*)^2}{\overline{\alpha}_0^2}\ln(\lambda) + 1}}) \tag{58}$$

Assuming that $z^*$ is small we can write the solution for $\overline{\alpha}_0$ in the form

$$\overline{\alpha}_0^2 \backsim P - \frac{\sqrt{2}}{c}\sqrt{\frac{P}{\ln(\lambda)}} \tag{59}$$

with $P$ given by (57). The unknown value $z^*$ has a dependence upon $V_{gs}^*$ but this is expected to be weak. In the fitting process (see section 6) $\kappa$ is treated as a free parameter and determined by the least-squares optimization process. As $\alpha_0$ can be treated as a constant $\overline{\alpha_0}$ in (46) the integral can be evaluated explicitly. Having done this, the final result, for the current in the linear regime, is

$$\frac{I_{ds}}{\mu I_c} = \left\{2\lambda(\ln\lambda)^3\right\}^{1/2}\overline{\alpha_0}\left\{I\left(\frac{z_s\sqrt{\ln\lambda}}{\overline{\alpha_0}}\right) - I\left(\frac{z_d\sqrt{\ln\lambda}}{\overline{\alpha_0}}\right)\right\} \tag{60}$$

where $I(x)$ is defined by

$$\begin{aligned} I(x) \quad &= \tfrac{\overline{\alpha_0}}{c}\sqrt{\tfrac{2}{\ln\lambda}}\left(\tfrac{1}{2}y^2 - \beta^2 y + \beta^4\ln(y+\beta^2)\right) \\ &+ \tfrac{2}{\ln\lambda}\left(y + \tfrac{\ln(y-1)}{2(1+\beta^2)} - \tfrac{\ln(y+1)}{2(1-\beta^2)} + \tfrac{\beta^6}{1-\beta^4}\ln(y+\beta^2)\right) \\ &- \tfrac{\overline{\alpha_0}}{c}\sqrt{\tfrac{2}{\ln\lambda}}\left(y - \beta^2\ln(y+\beta^2)\right) \\ &- \tfrac{2}{\ln\lambda}\left(\tfrac{\ln(y-1)}{2(1+\beta^2)} + \tfrac{\ln(y+1)}{2(1-\beta^2}) + \tfrac{\beta^4}{\beta^4-1}\ln(y+\beta^2)\right) \end{aligned} \tag{61}$$

where $y = \sqrt{1+x^2}$ and $\beta^2 = 1/c\overline{\alpha_0}\sqrt{2\ln\lambda}$.

### Saturated Regime

When $V_{ds}$ is large enough a region near the drain changes from inverted to depleted. In this region the elevated values of $\varphi$ eliminate from importance term (2) in equation (21) and the results used for inverted $w_s(\varphi)$ are no longer

valid. The integral for current is separated into two parts: (1) that valid in the inverted region and available from the results of the previous section, merely substituting $z_{po}$ for $z_d$ in the upper limit in (46), and (2) that valid in the depleted region for which we can use Ward's equations [12] (4.6), (4.7). This gives for the current in the saturated regime the formula

$$\frac{I_{ds}}{\mu I_c} = \left\{2\lambda(\ln\lambda)^3\right\}^{1/2} \overline{\alpha}_0 \left\{ I\left(\frac{z_s\sqrt{\ln\lambda}}{\overline{\alpha}_0}\right) - I\left(\frac{z_{po}\sqrt{\ln\lambda}}{\overline{\alpha}_0}\right)\right\} + Q_c\left(\exp\left(-\overline{V}_{ds}\right) - \exp\left(-\varphi_{po}\ln\lambda\right)\right) \tag{62}$$

where $I(x)$ is given by (61), and $\varphi_{po}$ is given by (32), and $Q_c$ is given by (see Ward 1990, (4.7)):

$$Q_c = -e^{w_s \ln\lambda}(2\lambda\ln\lambda)^{-\frac{1}{2}}(w_s + 1 - 1/\ln\lambda)^{-\frac{1}{2}}\left(1 + \frac{1}{2(w_s + 1 - 1/\ln\lambda)\ln\lambda}\right) \tag{63}$$

with $w_s$ taking the value $w_{spo}$.

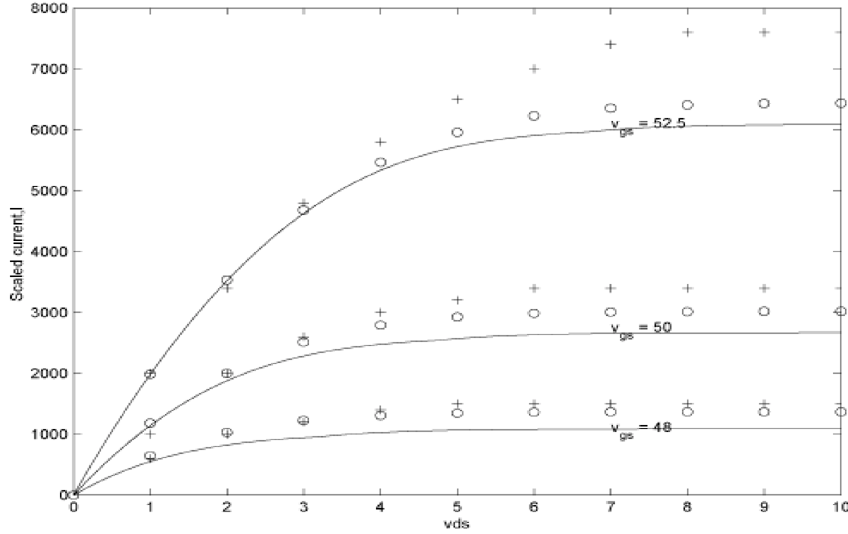


Fig 5. The scaled current for $\overline{V}_{gs} = 48, 50, 52.5$ (For $\lambda = 10^6$).

Figure 5. shows the scaled drain current $I_{ds}$ as a function of $\overline{V}_{ds}$ for values of $\overline{V}_{gs} = 48, 50, 52.5$. The + are Ward results, 0 are Spanier and Sarvas numerical results [9] and $-$ are the results of the calculations reported on here.These curves were obtained by blending the current in the linear region, (60), with that in the saturated region (63). The three approaches have the same shape as experimental data, but provide slightly different amplitudes. This is attributed to the following:

**Comment 1:** Since the agreement between our results and Sarvas and Spanier has been excellent so far, the difference in the results for current is attributed to the approximations made in generating the explicit formula (60) from the integral (45).

**Comment 2:** Divergence between the Sarvas and Spanier and Ward current results was also shown in [9]. The results generated in this paper show the possible source of the differences. Ward states "$\gamma$ is transcendentally small" and "$\alpha$ is transcendentally small in $\ln\lambda$" in the linear regime, and bases some approximations on these assertions. However the $w_s(\varphi)$ relation (see Figures 4 , 5) give $w_s - 1 - \varphi \cong 0.4$ and this gives

$$\alpha = \alpha_0 \sqrt{\ln\lambda}\ \ \lambda^{-0.2} \tag{64}$$

There is a negative power of $\lambda$ in $\alpha$, but it is small. Since $\alpha = \alpha_0/z\sqrt{\ln\lambda}$ and $z$ is $0(1)$ over a substantial portion of the linear regime (see Figure 40), $\alpha = 0(1/\sqrt{\ln\lambda})$ there, and may not be neglected without some penalty.

## 6 Optimized Fitting

A number of assumptions have been made in constructing the model of the MOSFET that has been used. It has been supposed that the mobility $\mu$ is a constant in the sense that it is independent of $z$. It does, however, have a dependence upon $V_{gs}^*$ that must be taken into account if we are to compare our formulae to real data. A common form used in SPICE is given by

$$\mu = \frac{1}{1 + \theta(V_{gs} - V_{th})} \tag{65}$$

and we adopt this also. The mobility scaling factor $\mu_0$ is also a parameter that has to be provided in SPICE and our model. These parameters can be extracted from experimental data but it is best to regard such extracted values as estimates. We shall take both these physiscal parameters, $\theta$ and $\mu_0$, as fitting parameters to be determined by the optimization process. In section five we were able to make a partial evaluation of the integrals by invoking the intermediate value theorem. This required the introduction of an addition non-physical parameter $z^*$. Although this cannot be extracted from the data we do have a sense of its value as it must lie in $[z_d, z_s]$ which we can estimate. Thus we have three variable and approximate value from which to start an optimization proces to obtain the best fit to the data. The method of least squares implemented in the form of the Marquardt algorithm [3] is used to

determine the values of these three parameters. We should remark that a similar parameter optimization technique is used in SPICE. None of the physical parameters is known well enough for typical fabricated devices to warrant any alternative approach.

Using .35 micron technology data provid by ISI we have applied our method to a number of devices of different channel length and width. Each device was separately fitted. We present the results as a series of plots. Figure 6. is a long and wide device for which we would expect the best agreement.

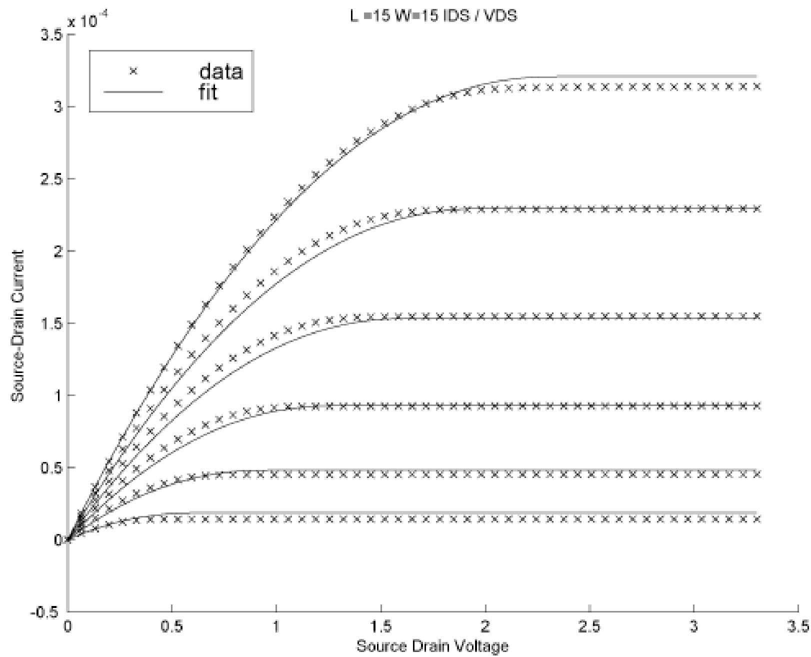


Fig 6. A long L = 15 and width W = 15 device ($\lambda = 10^7$).

the fit is excellent. This continues to be true as the channel length and width are reduced. Figure 7. shows an 5.4 micron length device with width 5.4 microns.

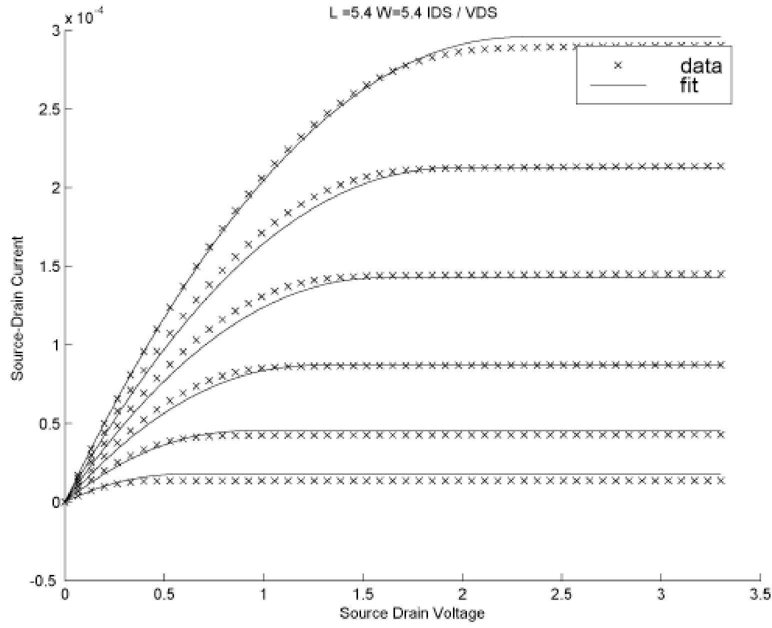


Fig 7. A L = 5.4 and width W = 5.4 device ($\lambda = 10^7$).

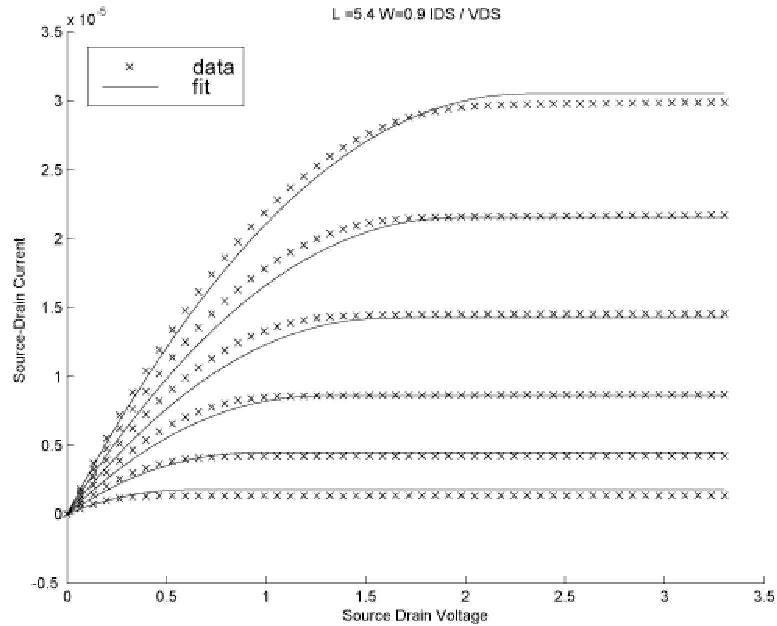


Fig 8. A L = 5.4 and width W = 0.9 device ($\lambda = 10^7$).

As figure 8. shows we can even reduce the width to 0.9 microns still retain an acceptable fit. The fits remain good down to about 1.5 microns as shown by figure 9.

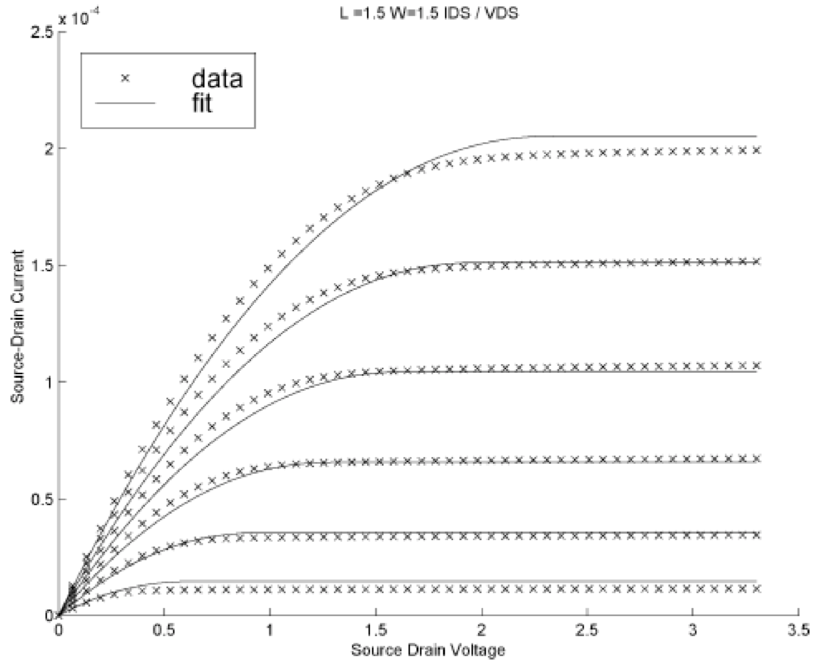


Fig 9. A L = 1.5 and width W = 1.5 device ($\lambda = 10^7$).

## A Primary Notation and Symbols

| | |
|---|---|
| $x_1, x_2$ | co-ordinates along and perpendicular to channel (See Figure 1) |
| $L, W$ | are channel length and width |
| $t$ | is oxide thickness |
| $n, p$ | are the electron and hole concentrations |
| $N$ | is the concentration of selected impurities or dopants |
| $\mathbf{J}_n$ | is the electron current density |
| $\mathbf{E}, \psi$ | are the electric field and electrostatic potential |
| $\phi_n$ | is the quasi-Fermi potential for electrons |
| $\mu_n, \mu_0$ | are the electron mobility, typical mobility magnitude |
| $\varepsilon_s$ | is the dielectric constant of the semiconducting material |
| $\varepsilon_i$ | is the dielectric constant of the oxide insulator |
| $\rho$ | is the space charge density |
| $n_i$ | is the intrinsic carrier concentration |
| $q$ | is the charge of a proton |
| $k$ | is the Boltzmann's constant |
| $T$ | is the lattice temperature (assumed constant) |
| $V_{th} = kT/q$ | is called the thermal voltage |
| $L_D = \left(\frac{kT\varepsilon_s}{n_i q^2}\right)$ | is called the Debye length |
| $V_{ds}$ | is drain-to-source voltage |
| $V_{gs}$ | is gate-to-source voltage |
| $V_{fb}$ | is flat-band voltage |
| $I_{ds}$ | is the source-to-drain current |

## B Scaled Variables

$\lambda = \max |N(x_1)/n_i|$
$x_1 = L_d \left(\frac{\ln\lambda}{\lambda}\right)^{1/2} x \qquad x_2 = Ly$
$d(x) = N(L_D(\frac{\ln\lambda}{\lambda})^{1/2} x)/n_i$

$(\psi, \phi_n) = V_{th} \ln\lambda \;\; (w, \phi)$
$c_{ox} = \varepsilon_i L_D / \varepsilon_s t$
$c = c_{ox}/\sqrt{\lambda}$
$V_{gs} - V_{fb} = V_{th}(\overline{V}_{gs} - \ln\lambda) = V_{th} \ln\lambda \;\; V^*_{gs}$
$V_{ds} = V_{th}\overline{V}_{ds}$

## C Numerical Values

$L_D = 33$ microns
$V_{th} = 0.0259$ volt
$V_{fb} = -0.6$ volt
$\mu_0 = 0.0407$ m$^2$ volt$^{-1}$ sec$^{-1}$

The table below shows the corresponding magnitudes of $\overline{V}_{gs}$ and $V_{gs}$ for $\lambda = 10^6$

| $\overline{V}_{gs}$ | 48 | 50 | 52.5 | 70.0 | 80.0 | 90.0 |
|---|---|---|---|---|---|---|
| $V_{gs}$ volt | 0.2854 | 0.3372 | 0.4019 | 0.8552 | 1.1142 | 1.3732 |

If the value of $\lambda$ is increased the values of $\overline{V}_{gs}$ are larger. For example if $\lambda = 10^7$ then $\overline{V}_{gs} = 155$ corresponds to $V_{gs} = 2.997$ volt.

## D Blending

In the theory of matched asymptotic expansions the construction of two approximate solutions each valid in one of two domains that are contiguous and overlapping is a typical situation. The construction of a composite expansion valid over the union of the domains is then indicated, and there are techniques available prescribing how to accomplish this. The composite expansion (to some order) will be accurate (to that order) uniformly over the whole domain and will have no reduction in smoothness from the component solutions. The desirability of this procedure is evident, and it transcends the domain of asymptotic analysis: one may have separate solutions valid in separate domains that one wishes to join or blend smoothly. These solutions may in some sense be asymptotic – and so complicated that one shrinks at the chore of finding the composite – or not. This section explores some ideas on how to blend such solutions in an empirical, ad-hoc way.

This topic was taken up by Wollkind, [14], and the analysis concerned prob-

lems in which (in the language of matched asymptotic expansions) the two component solutions consist of an inner solution valid near a boundary of the domain (a boundary layer) and an outer solution valid elsewhere. We have used the ideas presented by Wollkind were put to form a blended solution from the asymptotic formulae for separate "linear" and "saturated" components.

Let us start the discussion with a simple example. Suppose we have approximations (in $x > 0$)

$$y = \begin{cases} \alpha x \text{ close to } \ x = 0 \\ 1 \quad \text{away from } \ x = 0 \end{cases} \tag{D.1}$$

Following Wollkind, a blended solution is

$$y = (\alpha x - 1)e^{-x/\epsilon} + 1 \tag{D.2}$$

since it has the behavior of (D1), for $x \gg \epsilon$ and for $x \ll \epsilon$. (The Wollkind technique would find $\epsilon$ by least squares fitting (D2) to (D1) ($\alpha x$ joined to 1 at $x = 1/\alpha$)). Clearly this procedure gets more accurate the more closely this example resembles a boundary layer problem ($\alpha \gg 1$).

Let us now extend the discussion beyond the boundary layer context. Suppose we have approximations

$$y = L(x), R(x) \tag{D.3}$$

on the left and right, respectively, of $x = a$. The task is to blend $L, R$ in a smooth formula to provide a blended approximation, $y = Y(x)$.

A first attempt at a blend is given by

$$\begin{gathered} Y(x) = b_L(x)L(x) + b_R(x)R(x) \\ \text{where } b_R, b_L = \tfrac{1}{2}\left\{1 \pm \tanh \tfrac{x-a}{\epsilon}\right\} \end{gathered} \tag{D.4}$$

The blending functions $b_R, b_L$ are, for small $\epsilon$, essentially unity in $x \gtrless a$ and zero in the complement, and use of them assures that $Y_1$ gives (D3) outside a transition region around $x = a$. Since $Y_1$ is a convex linear combination of $L, R$ at each $x$ it lies between $L$ and $R$ and provides a blend.

## Comments

**Choice of $\epsilon$:** In problems in which asymptotic series have been developed for $L(x), R(x)$ there will be a small parameter which may be identified with $\epsilon$, in which case $L, R$ have been blended through a suitably chosen transition layer. This is the case in the applications in this paper. In other circumstances, $\epsilon$ may have to be determined numerically, say by a least squares fitting of the blended function to data.

**Choice of $a$:** In the theoretical development above, the value of $a$ was determined as the intersection point for $L, R$ equation (D3). Where $L, R$ do not intersect, as in the scenario of Figure (b), or where $a$ may not be found accurately, experimentation or extra numerical procedures may be required in order to determine $a$ in order to arrive at a good blending function.

We wish to acknowledge Cesar Piña and Vance Tyree of the University of Southern California Information Sciences Institute for technical discussions and provision of the MOSFET data that was used in this study.

## References

[1] Kervorkian, J., Cole, J. D., *Multiple Scale and Singular Perturbation Methods,* Applied Mathematical Sciences Vol 114,Springer-Verlag, May 1996.

[2] Maneesha, Subhasis Haldar, Manoj K. Khanna and R. S. Gupta, "Analytical Theory of Two-Dimensional Charge Sheet Model for Short Channel MOSFETs Under Non Linear Charge Control," *Solid-State Electronics*, Vol. 38, No. 1, 197-202, 1995.

[3] Marquardt, D.,"An Algorithm for Least-squares Estimation of Nonlinear Parameters," *SIAM J. Appl. Math.* Vol 11 pp. 431-441, 1963.

[4] Martelli, Mario, Ellis Cumberbatch, Alex Klafter, Stuart Pilorz, Rosa Loguercio Polosa, Guy Ruddock, Andrew Smith, for Jet Propulsion Laboratory, "Modelling Short Channel MOSFETS for Use in VLSI," Claremont Graduate School Final Clinic Report, May 1986.

[5] Massobrio, A., *Semiconductor Device Modeling with SPICE,* McGraw-Hill, 1993

[6] Pao, H., Sah, C.,"Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors," *Solid State Electronics* 10 pp. 927-937, 1966.

[7] Pierret, R., Shields, J.," Simplified long channel MOSFET theory," *Solid State Electronics* 26 pp. 143-147, 1983.

[8] Please, C.,"An Analysis of Semiconductor P-N Junctions," *IMA J. Appl. Math.* 28 pp. 301-318, 1982.

[9] Sarvas, J., and J. Spanier, "A Direct Approach to Solving the Drift-Diffusion Model Equations for Use in Certain MOSFET Devices," *Mathl. Comput. Modelling*, Vol. 22, No. 8, pp. 17-31, 1995.

[10] Sze, S., *Physics of semiconductor devices,* John Wiley, New York, 1969.

[11] Tsividis, Y., *Operation and Modeling of the MOS Transistor,* McGraw-Hill Book Company, 1987.

[12] Ward, M. J., F. M. Odeh, and D. S. Cohen, "Asymptotic Methods for Metal Oxide Semiconductor Field Effect Transistor Modeling," *SIAM J. Appl. Math*, Vol. 50, No. 4, pp. 1099-1125, August 1990.

[13] Ward, M. J., "Singular Perturbations and a Free Boundary Problem in the Modeling of Field-Effect Transistors," *SIAM J. Appl. Math.*, Vol. 52, No. 1, pp. 112-139, February 1992.

[14] Wollkind, David J., "Singular Perturbation Techniques: A Comparison of the Method of Matched Asymptotic Expansions with that of Multiple Scales," *SIAM Review*, Vol. 19, No. 3, pp. 502-516, July 1977.