Physics-based compact model of nanoscale MOSFETs – Part I: Transition from drift-diffusion to ballistic transport

Giorgio Mugnaini
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa
Giuseppe Iannaccone
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni, Università di Pisa

G. Mugnaini, G. Iannaccone, *Physics-based compact model of nanoscale MOSFETs – Part I: Transition from drift-diffusion to ballistic transport*, IEEE Transactions on Electron Devices, **52**, pp.1795-1801 (2005).

Physics-Based Compact Model of Nanoscale MOSFETs—Part I: Transition From Drift-Diffusion to Ballistic Transport

Giorgio Mugnaini and Giuseppe Iannaccone, *Member, IEEE*

***Abstract*—In this paper, we present a physics-based analytical model for nanoscale MOSFETs that allows us to seamlessly cover the whole range of regimes from drift-diffusion (DD) to ballistic (B) transport, taking into account quantum confinement in the channel. In Part I we focus on MOSFETs with ultrathin bodies, in which quantum confinement is structural rather than field-induced, and investigate in detail an analytical description of the transition from drift-diffusion to B transport based on the Büttiker approach to dissipative transport. We first start from the derivation of a closed form analytical expression of the Natori model for B MOSFETs, and show that a MOSFET with finite scattering length can be described as a suitable chain of B MOSFETs. Then, we are able to compact the behavior of the B chain in a simple analytical model. In the derivation, we also find a similarity between the B limit in the chain and the saturation velocity effect, that leads us to propose an alternative implementation of the saturation velocity effect in compact models.**

***Index Terms*—Ballistic transport, compact models, Lambert $\mathcal{W}$-function, nanoscale MOSFETs, quantum confinement.**

I. Introduction

MOSFET models included in circuit simulators can be classified in analytical models and table look-up models. According to [1], most MOSFET analytical models are based either on the regional approach, or on surface potential formulations. Models based on the regional approach such as the BSIM model [2] use different sets of equations to describe the weak and strong inversion regions, generally bridged by using some nonphysical curve fitting. Such models suffer from several problems because of the large number of parameters and the unphysical nature of vertical electrostatics modeling that leads to the wrong behavior of intrinsic capacitances and transconductances, as discussed in [3]. Models based on a surface potential formulation such as MOSFET Model 11 (MM11) [4], Hisim [5], and SP [1] are inherently continuous; however, they need the solution of implicit equations for the surface potential. In recent years, much work has been done to describe MOSFETs with simplified surface potential models, using electrostatics linearization schemes together with an approximate solution for vertical electrostatics, as in the EKV and "EKV-like" models [6]–[11] for bulk and fully depleted silicon-on-insulator (FDSOI) MOSFETs and in the model SPBM [1] for bulk MOSFETs.

We call "EKV-like" those models defined as "q_i-based models" in [1] and built on the basis of the inversion charge relation obtained by the linearization of mobile charge in terms of the surface potential. Such descriptions typically represent a tradeoff between a physically accurate and an analytically treatable model. EKV-like models must be mentioned because they are relatively simple, intrinsically continuous and symmetrical on all operating regions. Remarkably, they have a small number of physically based parameters. The SP model presented in [1] is linked with EKV-like models, because charge linearization is at the basis of the model, but it includes a more refined vertical electrostatics and therefore it represents a remarkable contribution to the evolution of compact models, but we believe it is not fully adequate to describe two—dimensional quantum confinement and far-from-equilibrium transport, because quasi-ballistic (B) transport is treated as an "effective" interpolation between B and drift-diffusion (DD) transport.

Recently proposed compact models describe transport through an approximate energy transport model, namely the models by Baccarani and Reggiani [12], University of Florida Double Gate (UFDG) [13] and University of Florida Partially Depleted/Bulk (UFPDB) models. We believe that such approaches can not be totally realistic descriptions, because they do not describe B transport, but a DD regime subject to overshoot due to electron heating, that does not hold for every short channel.

In this paper, we present a different treatment of the transition from DD to B transport, inspired by the Büttiker model of virtual probes [14], [15]. We will show that the effect of the B limit acts similarly to a saturation velocity effect, suggesting a fundamental relationship between such aspects. On the basis of the Büttiker probe approach, we will build a physically based model valid in any operating regime and for any ratio of the channel length L to the mean-free path λ, thus describing the whole intermediate regime from DD to B transport in MOSFETs.

We want to remark that we have dedicated significant effort not to introduce unphysical smoothing functions in the proposed model because we believe that smoothing functions hide the real nature of device physics and lead to bad analytical behavior. In Part I, we focus on nondegenerate carrier statistics and on rectangular quantum confinement, such as that obtainable in ultrathin body devices. In Part II, we will extend the model to de-

Manuscript received April 7, 2005. This work was supported in part by the MIUR through the PRIN program, in part by the Fondazione Cassa di Risparmio di Pisa, and in part by the EU through the Network of Excellence Silicon Nano Devices. The review of this paper was arranged by Editor S. Datta.

The authors are with the Dipartimento di Ingegneria dell'Informazione, Università di Pisa, Pisa I-56122, Italy (e-mail: giorgio.mugnaini@iet.unipi.it.e-mail; g.iannaccone@iet.unipi.it).

Digital Object Identifier 10.1109/TED.2005.851827

generate statistics in the limit of only one occupied subband in the channel, for both the cases of rectangular and triangular quantum confinement.

II. Ballistic Transport

Let us consider the case of double-gate (DG) MOSFETs with rectangular confinement in the silicon film, neglecting the effects of Fermi–Dirac statistics and assuming that the bottom of the conduction band is approximately flat. We expect that this type of confinement is predominant for MOSFETs with a silicon thickness $t_{\rm si} < 3$ nm, though it can be useful for MOSFETs with larger $t_{\rm si}$ in the subthreshold regime, in order to describe the quantum shift of threshold voltage. Extensions to include a more realistic field-dependent quantum confinement is possible, but is beyond the scope of this paper. For thick film and thick oxide SOI MOSFETs, when the quantum levels form a continuum, the model recovers the conventional EKV electrostatics.

In the case of a 100-oriented silicon film, quantum confinement and the anisotropy of effective mass give three different sets of subbands, two of which have the same subband profile. In the case of a rectangular quantum well, with flat conduction band edge in the vertical direction, the eigenvalues of the Schrödinger equation are

$$\epsilon_{n,k} \equiv E_{n,k} - E_{\rm co} = \frac{n^2\pi^2\hbar^2}{2m_k t_{\rm si}^2} \tag{1}$$

where $E_{\rm co}$ is the bottom of the conduction band at center of the film, $E_{n,k}$ denotes the energy of the nth k-type subband, and the subscript $k = l, t$ indicates that the subband has longitudinal or transversal effective mass m_k in the confinement direction, respectively. The effective density of states in the nth l-type and t-type subbands are

$$N_{n,l} = 2\frac{k_B T m_t}{\pi\hbar^2} \tag{2}$$

$$N_{n,t} = 4\frac{k_B T\sqrt{m_t m_l}}{\pi\hbar^2}. \tag{3}$$

In order to obtain simple equations, it is useful to consider the mobile charge density Q_m as localized at a fixed distance z_I from the oxide interface, known as inversion layer centroid as defined in [16], that can be considered approximately independent of V_g. Following the considerations in [16], the simplified electrostatics for the DGMOSFET in Fig. 1, if the silicon film is uniformly doped and fully depleted, is described by

$$Q_m = 2C_g\left[V_g - \phi_m + \chi - \phi_c - \frac{qN_A t_{\rm si}}{2}\left(\frac{1}{C_{\rm ox}} + \frac{t_{\rm si}}{4\epsilon_{\rm si}}\right)\right] \tag{4}$$

where: $\phi_c \equiv -E_{\rm co}/q$ is the central electrostatic potential, $V_g = -E_{Fg}/q$ is the gate voltage, ϕ_m is the gate workfunction, χ is the silicon electron affinity, N_A is the acceptor dopant density, and

$$C_g = \left(\frac{1}{C_{\rm ox}} + \frac{1}{C_d}\right)^{-1} \tag{5}$$

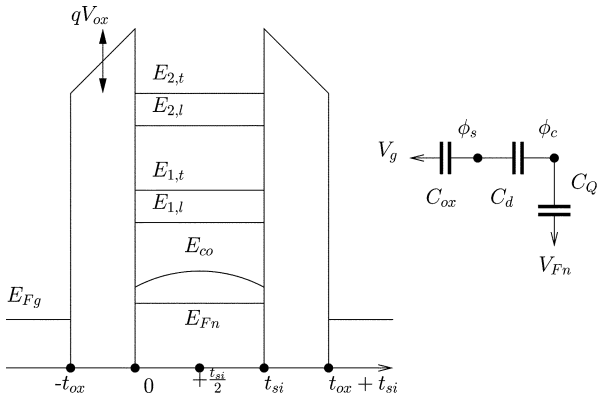

Fig. 1. Band diagram of a DG-MOSFET. The equivalent circuit includes the geometric capacitances $C_{\rm ox}, C_d$ and the quantum capacitance C_Q, which includes degeneracy capacitance and inversion layer capacitance. Maxwell–Boltzmann statistics allows us to aggregate the subband quantum capacitances.

is the effective gate capacitance, $C_{\rm ox} = \epsilon_{\rm ox}/t_{\rm ox}$ is the oxide capacitance per unit area, $t_{\rm ox}$ is the oxide thickness, $C_d = \epsilon_{\rm si})/(z_I$, as in [16]. It is easy to see that if the electron density is uniform in the vertical direction, (4) reduces to the vertical electrostatics of [12].

In B transport, carriers move without inelastic scattering along the channel and therefore at the subband peak the carriers that propagate toward the drain ("forward states") come from the source only and the carriers propagating toward the source ("reverse" states) come from the drain, as discussed in [17]. As a consequence of the separated two equilibrium populations on the subband peak, we have the superposition of two hemi-maxwellian distributions in the assumption of nondegenerate statistics:

$$Q_m = q\sum_{n=1}\sum_{k=t,l}\frac{N_{n,k}}{2}\left[e^{\frac{-\epsilon_{n,k}+\phi_c-V_s}{\phi_t}} + e^{\frac{-\epsilon_{n,k}+\phi_c-V_d}{\phi_t}}\right] \tag{6}$$

where V_s and V_d are the source and drain Fermi potentials, and ϕ_t is the thermal voltage. Substituting (6) into (4), we have

$$2C_g\left(V_g - \phi_m + \chi + \frac{qN_A t_{\rm si}}{2}\left(\frac{1}{C_{\rm ox}} + \frac{t_{\rm si}}{4\epsilon_{\rm si}}\right) - \phi_c\right) = \frac{qN_c}{2}\left[e^{-\frac{V_s}{\phi_t}} + e^{-\frac{V_d}{\phi_t}}\right]e^{\frac{\phi_c}{\phi_t}}. \tag{7}$$

It is mathematically convenient to define a "threshold voltage" V_T, as in [12]

$$V_T \equiv \phi_m - \chi - \frac{qN_A t_{\rm si}}{2}\left(\frac{1}{C_{\rm ox}} + \frac{t_{\rm si}}{4\epsilon_{\rm si}}\right) + \frac{k_B T}{q}\log\left(\frac{Q_n}{qN_c}\right) \tag{8}$$

where N_c is the effective density of states in all the conduction subbands

$$N_c \equiv \sum_{n=1}\sum_{k=t,l} N_{n,k}e^{-\frac{\epsilon_{n,k}}{\phi_t}} \tag{9}$$

and Q_n is the normalization charge density: $Q_n = 2C_g\phi_t$. With these definitions and by the means of Lambert $\mathcal{W}$-function [18], we can write

$$Q_m = Q_n\mathcal{W}\left[\frac{1}{2}e^{\frac{V_g-V_s-V_T}{\phi_t}} + \frac{1}{2}e^{\frac{V_g-V_d-V_T}{\phi_t}}\right] \tag{10}$$

and correspondingly ϕ_c. Concerning transport, if we suppose a superposition of two hemi-Maxwellian distributions on the peak, we have for forward I^+ and reverse I^- currents, respectively [17]

$$I^+ = \frac{qN_c v_{\rm th}}{2}e^{\frac{\phi_c-V_s}{\phi_t}} \tag{11}$$

$$I^- = \frac{qN_c v_{\rm th}}{2}e^{\frac{\phi_c-V_d}{\phi_t}}. \tag{12}$$

Note that here and in the following, the currents are normalized to the device width.

In (11) and (12), $v_{\rm th} \equiv \sqrt{2kT/\pi m_t}$ is the unidirectional thermal velocity, because for simplicity we assume that $v_{\rm th}$ is the same in any subband in order to obtain a simpler closed form. This simplifying approximation can be easily removed, but it is useful to remark the fact that similar assumptions are present in [19]–[21]. Therefore, the nondegenerate B transistor is described by

$$\begin{aligned} I_{\rm ds} &= I^+ - I^- = \frac{qN_c}{2}v_{\rm th}e^{\frac{\phi_c}{\phi_t}}\left(e^{-\frac{V_s}{\phi_t}} - e^{-\frac{V_d}{\phi_t}}\right) \\ &= Q_n\mathcal{W}\left[\frac{1}{2}e^{\frac{V_g-V_s-V_T}{\phi_t}} + \frac{1}{2}e^{\frac{V_g-V_d-V_T}{\phi_t}}\right]v_{\rm th}\tanh\left(\frac{V_{\rm ds}}{2\phi_t}\right). \end{aligned} \tag{13}$$

Equation (13) is similar to the equation previously proposed in [22]. However, the result in [22] is valid for a classical B transistor only in inversion, while it is worth noticing that (13) is valid both below and above threshold.

An important point is that the saturation of current occurs when $V_{\rm ds} \gtrsim 2\phi_t$, because for these values of $V_{\rm ds}$ the term $\tanh(V_{\rm ds}/2\phi_t)$, which is responsible for the B limit, saturates.

III. DD Transport in Terms of a Chain of B Devices

It is interesting to investigate the behavior of a finite chain of B MOSFETs. While a series of N DD MOSFET channels of length L/N can be thought as a single DD MOSFET of length L, in the case of a series of B transistors (B chain), this simple rule does not apply. Indeed, in a B chain of N transistors, $N+1$ different Fermi levels can be defined at the $N+1$ contacts. It is clear that this behavior is incompatible with the transport regime in a single B MOSFET, where no local quasi-Fermi level can be defined, except for source and drain contacts.

This behavior is well understandable, if we consider the widely diffused interpretation of inelastic scattering represented by the Büttiker approach of virtual thermalizing probes [14], which can be used to describe transport in any regime. Within this approach, carriers are removed from the device and injected into a "virtual" reservoir where they are thermalized and re-injected into the device so that the number of electrons is conserved. Therefore, the internal contacts of a B chain behave as Büttiker probes, and then we expect that a B chain can realistically describe a MOSFET in any transport regime. We define V_k as the Fermi potential at the kth contact, and we suppose that the kth contact is placed at $x_k = k\lambda$ with $k = 1,\ldots,N$. That is equivalent to place N B MOSFETs of channel length λ in series, as reported in Fig. 2. Since the current in all MOSFETs is $I_{\rm ds}$, we have the following $N+1$ equations

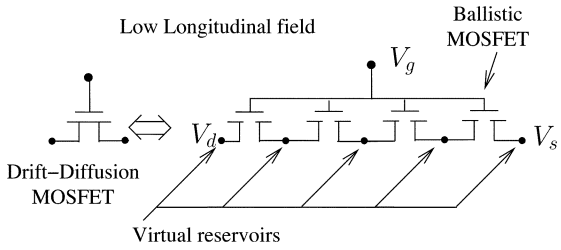

Fig. 2. Long chain of B MOSFETs. The contacts act as thermalizing reservoir. We also show the symbol we use for a B MOSFET in the following figures.

$$\begin{aligned} I_{\rm ds} = Q_n v_{\rm th}\mathcal{W} &\times \left[\frac{e^{(V_g-V_k-V_T)/\phi_t} + e^{(V_g-V_{k+1}-V_T)/\phi_t}}{2}\right] \\ &\times \tanh\left(\frac{V_{k+1}-V_k}{2\phi_t}\right) \end{aligned} \tag{14}$$

where $k = 0,\ldots,N$ and $V_0 = V_s$ and $V_N = V_d$. This set of equations is constituted by $N+1$ transcendental equations. At this point, if N is large enough, we are tempted to define a continuous quasi-Fermi level V_{Fn} subject to these conditions

$$\begin{aligned} V_{Fn}(x_{k+1/2}) &\equiv (V_{k+1}+V_k)/2 \\ \frac{dV_{Fn}}{dx}(x_{k+1/2}) &\equiv (V_{k+1}-V_k)/\lambda \end{aligned} \tag{15}$$

where we have defined $x_{k+1/2} \equiv (x_k + x_{k+1})/2$. Therefore, (14) becomes

$$\begin{aligned} I_{\rm ds} = v_{\rm th}Q_n\mathcal{W}&\left[e^{\left(\frac{V_g-V_{Fn}-V_T}{\phi_t}\right)}\right. \\ &\left.\times\cosh\left(\frac{\lambda}{2\phi_t}\frac{dV_{Fn}}{dx}\right)\right]\tanh\left(\frac{\lambda}{2\phi_t}\frac{dV_{Fn}}{dx}\right) \end{aligned} \tag{16}$$

with the obvious conditions

$$V_{Fn}(0) = V_s, \quad V_{Fn}(L) = V_d. \tag{17}$$

If we make the hypothesis that a low longitudinal bias is present on every B MOSFET, i.e., that

$$V_{k+1} - V_k \simeq \lambda\frac{dV_{Fn}}{dx} \ll 2\phi_t \tag{18}$$

we easily obtain that every B MOSFET is in the linear region and therefore in (14), the term $\cosh(\cdot)$ can be approximated

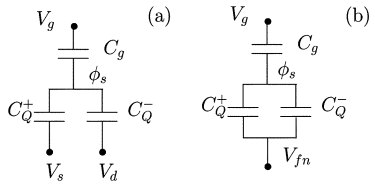

Fig. 3. Effect of transport transition on the vertical electrostatics. In (a), vertical electrostatics valid in B transport is reported. The source and drain Fermi potentials and the respective contributions C_Q^+, C_Q^- to the charge on the peak of the channel, are put in evidence. In (b), vertical electrostatics valid in DD transport is reported. In this case, local equilibrium occurs and the V_s and V_d potentials are equal to the local quasi-Fermi potential V_{Fn}.

with unity and the term $\tanh(\cdot)$ can be substituted with its argument. Consequently, the local $I_{\rm ds}$ and Q_m become, respectively

$$I_{\rm ds} = \frac{\lambda}{2\phi_t} v_{\rm th} Q_n \mathcal{W}\left(e^{\frac{V_g - V_{Fn} - V_T}{\phi_t}}\right) \frac{dV_{Fn}}{dx} \tag{19}$$

and

$$Q_m = Q_n \mathcal{W}\left[e^{\frac{V_g - V_{Fn} - V_T}{\phi_t}}\right] \tag{20}$$

consistent with Fig. 3. Remarkably, such a description of the electrostatics is equivalent to that used in EKV and EKV-like compact models. Indeed, we can easily rewrite the vertical electrostatics (20) as

$$V_g - V_T - V_{Fn} = \frac{Q_m}{2C_g} + \phi_t \log \frac{Q_m}{qN_c} \tag{21}$$

which has the same form of similar equations in [6]–[10]. It is interesting to observe that (19) resembles a DD equation if we define a low field mobility, as in [23]

$$\mu_{\rm no} \equiv \frac{\lambda v_{\rm th}}{2\phi_t} \tag{22}$$

that forces us to identify λ with the mean-free path. In order to obtain a closed-form of current from (19) and (17), we can integrate (19) from $x = 0$ to $x = L$ exploiting current continuity along the channel

$$I_{\rm ds} L = \mu_{\rm no} \int_0^L Q_m \frac{dV_{Fn}}{dx} dx = \mu_n \int_{V_s}^{V_d} Q_m dV_{Fn}. \tag{23}$$

Using the following property of the Lambert $\mathcal{W}$-function [18]:

$$\int \mathcal{W}(e^x) dx = \frac{\mathcal{W}^2(e^x)}{2} + \mathcal{W}(e^x) \tag{24}$$

we have

$$I_{\rm ds} = \frac{\mu_{\rm no} Q_n \phi_t}{L} \left[\frac{Q_{\rm ms}^2 - Q_{\rm md}^2}{2Q_n^2} + \frac{Q_{\rm ms} - Q_{\rm md}}{Q_n}\right] \tag{25}$$

where $Q_{\rm ms}$ and $Q_{\rm md}$ are obtained from (21) taking $V_{Fn} = V_s$ and $V_{Fn} = V_d$, respectively. We note that (21) and (25) constitute the basic structure of compact models like EKV, ACM, USIM, and this fact suggests us that such compact models are more fundamental than others, because they can be obtained from the behavior of a long enough B chain, subject to the aboveseen simplified electrostatics. We have also shown in a rigorous way that a DD transistor can be described as a long chain of B MOSFETs of length λ, working in their linear region.

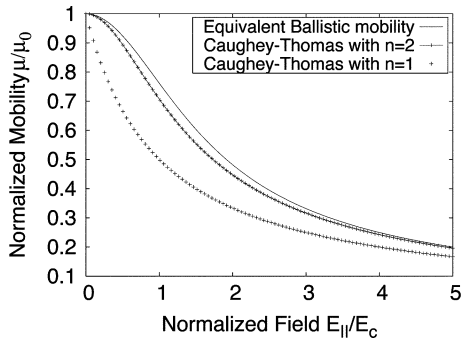

Fig. 4. Comparison of mobility (26) with the Caughey–Thomas models for electrons and holes. The "field" and the mobility are normalized to the critical field, and the low-field mobility respectively.

IV. Intermediate Regime Between Drift-Diffusion and Ballistic Transport

We have discussed the behavior of the B chain if N is very large, so that any elementary B transistor is in the linear region. Clearly if the elementary longitudinal bias is comparable or larger than ϕ_t, approximation (18) does not hold and reconsidering (16), we can see that the effects of large longitudinal bias are striking: Q_m depends not only on V_{Fn}, but also on (dV_{Fn}/dx) through the term $\cosh(\cdot)$, and in addition transport is affected by the nonlinear behavior of the term $\tanh\left((\lambda/2\phi_t)(dV_{Fn}/dx)\right)$. Remarkably, such a nonlinear term can be interpreted as the product of a "field dependent" mobility and the gradient of the quasi-Fermi potential. Similarly to the arguments that provide the mobility expression (22) in Section III, now we can define a mobility

$$\mu_n = \mu_{\rm no} \frac{\tanh\left(\frac{\lambda}{2\phi_t}\frac{dV_{Fn}}{dx}\right)}{\left(\frac{\lambda}{2\phi_t}\frac{dV_{Fn}}{dx}\right)}. \tag{26}$$

From this point of view, the term $(2\phi_t/\lambda)$ can be interpreted as a "critical field", analogous to the critical field in the saturation velocity effect.

It is interesting to compare (26) with the Caughey–Thomas model [24] for saturation velocity effect in Fig. 4. The similarity of this behavior with the real saturation is striking. We want to underline that a somewhat similar mobility expression was independently proposed by Arora in [25] and [26]. However, his work was based on the Zukotynski and Howlett approximate distribution function [27], while our result is not based on any particular hypothesis, except B transport along a mean free path λ in a chain of MOSFETs. Moreover, Arora mobility differs from our "B chain mobility" because it is dependent on the electric field $\partial\phi_c/\partial x$, instead of the gradient of the quasi-Fermi potential. It is interesting to note the fact that in [12], a mobility is used that depends on the gradient of the quasi-Fermi potential in order to simplify the calculations. Previous considerations suggest that the saturation velocity effect is linked to the B limit in a fundamental way. Conversely, we believe that a model of

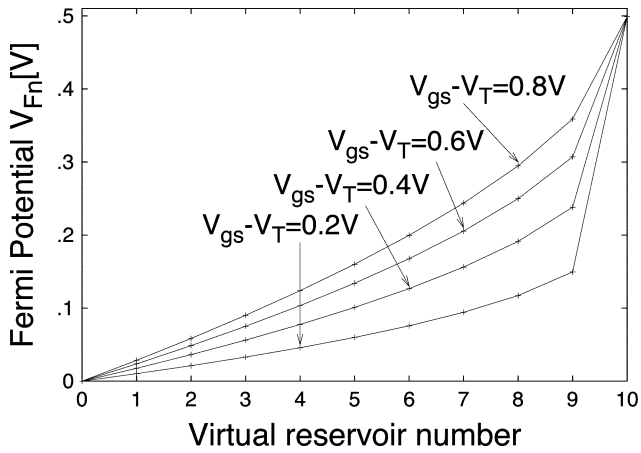

Fig. 5. Discrete quasi-Fermi potential obtained solving the equation set (14), (17) with $N = 10$ at fixed $V_{\rm ds} = 0.5$ V Fermi-Levels are defined only at points $x = k \times \lambda$ with k integer. $t_{\rm ox} = 2.5$ nm, $V_g - V_T = 0.2, 0.4, 0.6,$ and0.8 V.

mobility which includes the saturation velocity can reasonably describe this behavior in a compact model.

Following this suggestion, we can build an alternative compact model for MOSFETs where a degradation of the mobility caused by high longitudinal fields is present. It must be observed that it is not sufficient to introduce such a decreasing high field mobility in the compact model. Indeed, the velocity saturation effect and the current continuity impose a minimum mobile charge density at a given current and this fact must cause a modification in electrostatics given by (21) consistent with current continuity, but to our knowledge in every compact model [1], [2], [4], [6] the most common approach is continuing to use a "saturation-free" electrostatics and then introducing an "effective" $V_{\rm ds}$ artificially limited at $V_{\rm ds,sat}$ by some opportunely smoothed clamping function. Some other authors previously tempted to build more accurate physical models without artificial clamping, like [28] or [29]. In (16) this effect is contained in the expression of Q_m which can be written in terms of $V_g - V_{Fn} - V_T$ as

$$V_g - V_{Fn} - V_T = \frac{Q_m}{2C_g} + \phi_t \log \frac{Q_m}{qN_c \cosh\left(\frac{\lambda}{2\phi_t}\frac{dV_{Fn}}{dx}\right)}. \quad (27)$$

Equation (27) can be interpreted as a generalization of electrostatics of EKV-like compact models in the presence of the saturation velocity effect. Indeed, if the term $(\lambda/2\phi_t)(dV_{Fn}/dx)$ tends to zero, (27) becomes the well known electrostatics of EKV-like models (21). A closed-form solution of (27) can be obtained to eliminate the smoothing functions, but would be beyond the scope of the present paper.

In order to build a model of the finite chain of B MOSFETs that can be handled more easily, we can observe in the example in Fig. 5, that when nonlinear transport emerges, it manifests its effects almost exclusively on the last B transistor of the chain. This fact suggests that we can aggregate the first $N-1$ B transistors in an approximate equivalent DD transistor with ratio $L/\lambda = N-1$, as it is represented in Fig. 6. The feasibility of this approach can be evaluated by observing Figs. 7 and 8, where the output characteristics are plotted for the two cases $N = 3$ and

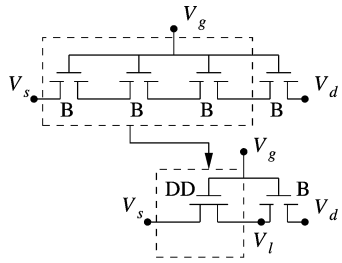

Fig. 6. Approximate aggregation of first $N-1$ B transistors in an equivalent DD transistor. The global circuit can be seen as a macromodel for a device in intermediate transport.

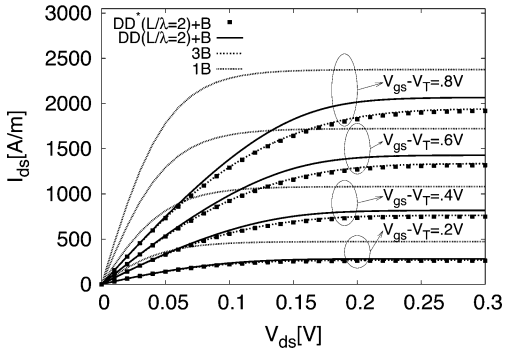

Fig. 7. Output characteristics of a chain of three B MOSFETs with different models and approximations. 2DD+1B denotes the series of a DD with $L/\lambda = 2$ and a B transistor, 3DD denotes a DD transistor with $L/\lambda = 3$, 2DD*+1B denotes reduced mobility model. $t_{\rm ox} = 2.5$ nm, $V_g - V_T = 0.2, 0.4, 0.6, 0.8$ V.

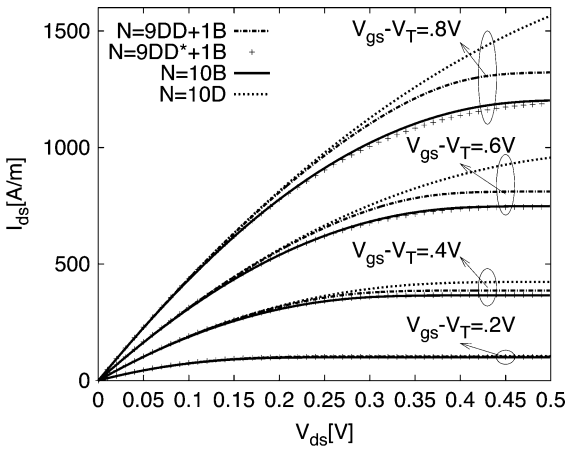

Fig. 8. Output characteristics of a chain of ten B MOSFETs with different models and approximations. $9\text{DD} + 1\text{B}$ denotes the series of a DD transistor with $L/\lambda = 9$ and a B transistor, 10DD denotes a DD transistor with $L/\lambda = 10$, 9DD*+1B denotes reduced mobility model (28) $t_{\rm ox} = 2.5$ nm, V_g-$V_T = 0.2, 0.4, 0.6, 0.8$ V.

$N = 10$, respectively. We can see in Fig. 7 that the series of a simple DD and a B transistor has a little but not negligible error with respect to the B chain. The reason of this behavior is caused by the approximation of a fully linear transport in the first $N-1$ transistors. In order to improve this approximation, among other possibilities,in currents, we choose to introduce the nonconstant mobility (26) in the DD equivalent transistor. It can be seen that this improvement (indicated with DD* + B in Figs. 7 and 8), gives a much better behavior than the simple DD + B.

A remarkable point is that the introduction of the variable mobility (26) in the DD section of DD+B transistor, eliminates the need to introduce unphysical smoothed clamping functions which always appear when saturation velocity effect is introduced in compact models, in order to eliminate the artificial negative conductance behavior. In our case, the B transistor covers the role of the conventional smoothed clamping functions. Therefore we have obtained a good approximation of an arbitrary B chain, which consists by the series of a DD and a B transistor. In summary, the most appropriate set of equations is

$$\begin{cases} I_{\rm ds} = \mu_n \frac{Q_n \phi_t}{L}\left(\frac{Q_{\rm ms}^2 - Q_{\rm ml}^2}{2Q_n^2} + \frac{Q_{\rm ms} - Q_{\rm ml}}{Q_n}\right) \\ I_{\rm ds} = Q_n v_{\rm th} \mathcal{W}\left(\frac{e^{\frac{V_g - V_d - V_T}{\phi_t}} + e^{\frac{V_g - V_l - V_T}{\phi_t}}}{2}\right) \tanh\left(\frac{V_{dl}}{2\phi_t}\right) \end{cases} \tag{28}$$

where the subscript l denotes the internal node and $(\partial V_{Fn}/\partial x)$, that is contained in μ_n is substituted by its mean value in the first $N-1$ MOSFETs, that is $(V_l - V_s)/(L - \lambda)$. A very significant aspect is that no smoothed clamping functions are needed, because their role is covered by the last B MOSFET, intrinsically preventing the unphysical negative output conductance that appears when mobility longitudinal degradation is plainly introduced in compact models [30].

V. Compact Model for Mosfets

The previous discussion can be used to build the basis for a compact model. A more realistic description of the transport must include the effect of mobility degradation at high vertical fields. We can use the simple expression [8], [11]

$$\mu_{n\perp} = \frac{\mu_{n0}}{1 + \theta(Q_{\rm ms} + Q_{\rm md})} \tag{29}$$

where θ is a suitable fitting parameter. Equation (29) is inherently continuous and symmetrical and captures the essential behavior of mobility degradation at high vertical fields. In order to describe the two-dimensional effects, similarly to [31], we introduce two geometrical capacitances that model the electrostatics coupling of source and drain on the peak of B transistor (drain-induced barrier lowering): C_s and C_d. It is possible to introduce the polysilicon depletion effect, similar to [32], with the additional parameter N_p, or the doping of the gate polysilicon. Therefore, we end up with the following parameters: $(L/\lambda), t_{\rm ox}, N_A, N_p,\ \Phi_m - \chi, C_s, C_d, R_s, R_d$, and θ. In Fig. 9, the compact model is fitted to experimental curves from a FinFET with $L = 80$ nm taken from the literature [33]. The model parameters have the following values: $C_g = 0.021$ F/m^2, $N_p = 5 \cdot 10^{19}$ cm^{-3}, $C_s = 0.36C_g$, $C_d = 0.42C_g$, $L/\lambda = 8.5$, $R_s = R_d = 34\ \Omega$m, $\theta^{-1} = 9.6\ \mu$Cm^{-2}. Although our proposed model is not yet complete, we believe that it can represent the basis of a very detailed physics-based compact model that, as we have seen, in limiting cases reduces to other well known compact models.

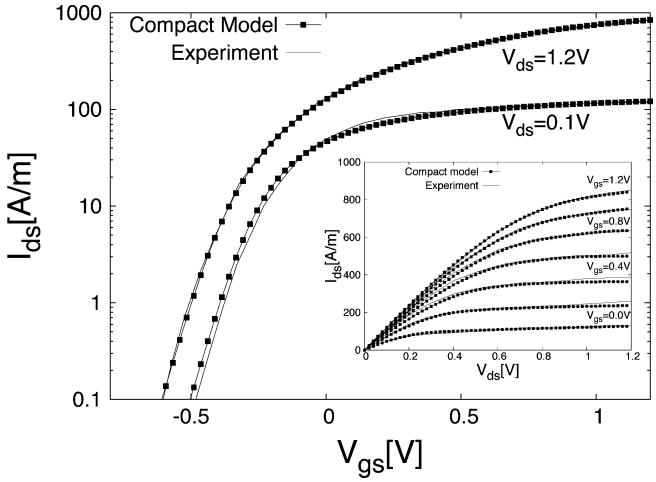

Fig. 9. Comparison between the transfer characteristics obtained from experiments on a FinFET with $L = 80$ nm reported in [33] and from a compact model fitted to the data. The comparison of output characteristics is illustrated in the inset.

VI. Conclusion

In this paper, we have proposed an analytical description of the intermediate regime between DD and B transport in nanoscale MOSFETs, that is suitable for inclusion in compact models of devices at the nanometer scale, in which quantum confinement is relevant. In our view, we have derived an elegant analytical representation for the dc characteristics of nanoscale DGMOSFETs.

Moreover we have derived a novel closed-form solution for B MOSFETs, analogous to the Natori model [34], that is analytical, explicit and inherently symmetrical. We have shown that, according to Büttiker approach [14], a long enough series of B MOSFETs can be interpreted as a DD transistor with constant mobility, while if the device is shorter, a behavior like saturation velocity effect appears. We have derived an alternative mobility model in which, remarkably, a "saturation velocity effect" seems to appear independently of optical phonon emission, that lead us to view B transport and saturation velocity effect as intimately linked. Our mobility model resembles a previous result by Arora [25], but is obtained with simpler physical hypotheses. We want to stress the fact that starting from the B chain interpretation, a physics-based macromodel for short MOSFETs is proposed, that can be also used in order to build a novel model of existing devices which are subject to the velocity saturation *without the introduction of unphysical smoothing functions*. Because of its structure, this model can be thought as an extension of EKV-like models [6]–[11] and is suitable for circuit simulation.

References

[1] G. Gildenblat, T.-L. Chen, X. Gum, H. Wang, and X. Cai, "SP: An advanced surface-potential-based compact MOSFET model," in *Proc. Custom Integrated Circuits Conf.*, Sep. 2003, pp. 233–240.

[2] Y. Cheng, M.-C. Jeng, Z. Liu, J. Huang, M. Chan, K. Chen, P. K. Ko, and C. Hu, "A physical and scalable *I–V* model in BSIM3v3 for analog/digital circuit simulation," *IEEE Trans. Electron Devices*, vol. 45, no. 2, pp. 277–287, Feb. 1997.

[3] P. Bendix, "Detailed comparison of the SP2001, EKV, and BSIM3 models," in *Proc. Int. Conf. Modeling Simulation Microsystems*, Apr. 2002, pp. 649–652.

[4] D. B. M. Klaassen *et al.*, *The MOS Model, Level 1101*: Philips Research Laboratories, Apr. 2001.
[5] M. Miura-Mattausch, H. Ueno, M. Tanaka, H. Mattausch, S. Kumashiro, T. Yamaguchi, K. Yamashita, and N. Nakayama, "HiSIM: A MOSFET model for circuit simulation connecting circuit performance with technology," in *IEDM Tech. Dig.*, Dec. 2002, pp. 109–112.
[6] C. C. Enz, F. Krummenacher, and E. Vittoz, "An analytical MOS transistor model valid in all regions of operation and dedicated to low-voltage and low-power applications," *Analog Integr. Circuits Signal Process. J.*, no. 8, pp. 83–114, Jul. 1995.
[7] C.-K. Park, C.-Y. Lee, K. Lee, B.-J. Moon, Y.-H. Byun, and M. Shur, "A unified current-voltage model for long-channel nMOSFETs," *Solid State Commun.*, vol. 38, no. 2, pp. 399–406, 1991.
[8] R. Howes, W. Redman-White, K. Nichols, P. Mole, M. Robinson, and S. Bird, "An SOS MOSFET model based on calculation of the surface potential," *IEEE Trans. Computer-Aided Des. Integr. Circuits Syst.*, vol. 13, no. 4, pp. 494–506, Apr. 1994.
[9] A. I. A. Cunha, M. C. Schneider, and C. Galup-Montoro, "An MOS transistor model for analog circuit design," *IEEE J. Solid-State Circuits*, vol. 33, no. 10, pp. 1510–1519, Oct. 1998.
[10] H. K. Gummel and K. Singhal, "Inversion charge modeling," *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1585–1593, Aug. 2001.
[11] B. Iniguez, L. Ferreira, B. Gentinne, and D. Flandre, "A physically-based C_∞-continuous fully-depleted SOI MOSFET model for analog applications," *IEEE Trans. Electron Devices*, vol. 43, no. 4, pp. 568–575, Apr. 1996.
[12] G. Baccarani and S. Reggiani, "A compact double-gate MOSFET model comprising quantum-mechanical and nonstatic effects," *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1656–1666, Aug. 1999.
[13] L. Ge, J. G. Possum, and B. Liu, "Physical compact modeling and analysis of velocity overshoot in extremely scaled CMOS devices and circuits," *IEEE Trans. Electron Devices*, vol. 48, no. 9, pp. 2074–2080, Sep. 2001.
[14] M. Buttiker, "Role of quantum coherence in series resistors," *Phys. Rev. B, Condens. Matter*, no. 33, pp. 3020–3026, 1986.
[15] S. Datta, "Nanoscale device modeling: The Green's function method," *Superlatt. Microstruct.*, vol. 28, no. 4, pp. 253–278, Oct. 2000.
[16] J. Lopez-Villanueva, P. Cartujo-Cassinello, F. Gamiz, J. Banqueri, and A. Palma, "Effects of the inversion layer centroid on the performance of double-gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 47, no. 1, pp. 141–146, Jan. 2000.
[17] C. R. Crowell and M. Hafizi, "Current transport over parabolic potential barriers in semiconductor devices," *IEEE Trans. Electron Devices*, vol. 35, no. 7, pp. 1087–1095, Jul. 1988.
[18] R. M. Corless and D. J. Jeffrey, "The Wright ω function," in *Proc. Int. Conf. Artificial Intelligence, Automated Reasoning, Symbolic Computation*, 2002, pp. 76–89.
[19] R. Lipperheide, T. Weis, and U. Wille, "Generalized drude model: Unification of ballistic and diffusive electron transport," *Phys. Rev. B, Condens. Matter*, no. 13, pp. 3347–3363, Apr. 2001.
[20] G. Gildenblat, "One-flux theory of a nonabsorbing barrier," *J. Appl. Phys.*, vol. 91, no. 12, pp. 9883–9886, Jun. 2002.
[21] H. Wang and G. Gildenblat, "Scattering matrix based compact MOSFET model," in *IEDM Tech. Dig.*, Dec. 2002, pp. 125–128.
[22] Z. Ren and M. Lundstrom, "Simulation of nanoscale MOSFETs: A scattering theory interpretation," *Superlatt. Microstruct.*, vol. 27, no. 2–3, pp. 177–189, Feb. 2000.
[23] A. Rahman and M. Lundstrom, "A compact scattering model for the nanoscale double-gate MOSFET," *IEEE Trans. Electron Devices*, vol. 49, no. 3, pp. 481–489, Mar. 2002.
[24] D. Caughey and R. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proc. IEEE*, vol. 55, no. 12, pp. 2192–2193, Dec. 1967.
[25] V. K. Arora, "Drift diffusion and einstein relation for electrons in silicon sujected to a high electric field," *Appl. Phys. Lett.*, vol. 80, no. 20, pp. 3763–3765, May 2002.
[26] ——, "Quantum engineering of nanoeletronic devices: The role of quantum emission in limiting drift velocity and diffusion coefficient," *Microelectron. J.*, vol. 80, no. 31, pp. 853–859, 2000.
[27] S. Zukotynski and W. Howlett, "Carrier distribution function in semiconductors," *Solid State Electron.*, vol. 21, no. 1, pp. 35–41, Jan. 1978.
[28] W. R. Bandy and R. S. Winton, "A new approach for modeling the MOSFET using a simple, continuous analytical expression for drain conductance which includes velocity-saturation in a fundamental way," *IEEE Trans. Computer-Aided Des. Integr. Circuits Syst.*, vol. 15, no. 5, pp. 475–483, May 1996.
[29] M. A. Maher and C. Mead, "A physical charge-controlled model for MOS transistors," in *Advanced Research in VLSI*, P. Losleben, Ed. Cambridge, MA: MIT Press, 1987.
[30] K. Joardar, K. Gulallapalli, C. McAndrew, M. Burnham, and A. Wild, "An improved MOSFET model for circuit simulation," *IEEE Trans. Electron Devices*, vol. 45, no. 1, pp. 134–148, Jan. 1998.
[31] J. Wang and M. Lundstrom, "Ballistic transport in high electron mobility transistors," *IEEE Trans. Electron Devices*, vol. 50, no. 7, pp. 1604–1609, Jul. 2003.
[32] J.-M. Sallese, M. Bucher, and C. Lallement, "Improved analytical modeling of polysilicon depletion in MOSFETs for circuit simulation," *Solid State Electron.*, vol. 44, pp. 905–912, Jun. 2000.
[33] B. Yu, L. Chang, S. Ahmed, H. Wang, S. Bell, C.-Y. Yang, C. Tabery, C. Ho, Q. Xiang, T.-J. King, J. Bokor, C. Hu, M.-R. Lin, and D. Kyser, "FinFET scaling to 10 nm gate length," in *IEDM Tech. Dig.*, Dec. 2002, pp. 251–254.
[34] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor," *J. Appl. Phys.*, vol. 75, no. 8, pp. 4879–4890, Oct. 1994.

Giorgio Mugnaini received the M.S.E.E. degree in electrical engineering from the University of Pisa, Pisa, Italy, in 2003, where he is currently pursuing the Ph.D. degree.

His main field of activity is the development of compact models and numerical modeling tools for nanoscale MOSFETs.

Giuseppe Iannaccone (M'98) was born on April 28, 1968. He received the laurea degree *(cum laude)* in electrical engineering and the Ph.D. degree with a thesis on transport and noise phenomena in ultrasmall structures from the University of Pisa, Pisa, Italy, in 1992 and 1996, respectively.

In January 2001, he became an Associate Professor at the Information Engineering Department, University of Pisa. His interests include transport and noise modeling in nanoscale devices, devices and architectures for nanoelectronics, the design of passive RFID transponders, and the exploitation of quantum effects in conventional electron devices. He has participated in a series of European and national research projects as consortium coordinator or principal investigator, and has authored more than 90 papers in peer-reviewed journals, and 50 papers in proceedings of international conferences.