Università di Pisa

# *Detailed modeling of sub-100-nm MOSFETs based on Schrodinger DD per subband and experiments and evaluation of the performance gap to ballistic transpor*

**Gilberto Curatola**
Philips Research Leuven
**Gerben Doornbos**
Philips Research Leuven
**Josine Loo**
Philips Research Leuven
**Youri V. Ponomarev**
Philips Research Leuven
**Giuseppe Iannaccone**
Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni,
Università di Pisa

G. Curatola, G. Doornbos, J. Loo, Y.V. Ponomarev, G. Iannaccone, *Detailed modeling of sub-100-nm MOSFETs based on Schrodinger DD per subband and experiments and evaluation of the performance gap to ballistic transport*, IEEE Transactions on Electron Devices, **52**, pp.1851-1858 (2005).

# Detailed Modeling of Sub-100-nm MOSFETs Based on Schrödinger DD Per Subband and Experiments and Evaluation of the Performance Gap to Ballistic Transport

Gilberto Curatola, Gerben Doornbos, Josine Loo, Youri V. Ponomarev, and Giuseppe Iannaccone, *Member, IEEE*

***Abstract*—We analyze in detail the requirements for the detailed physical modeling of nanoscale MOSFETs and show that Schrödinger drift—diffusion per subband simulations are adequate for the inverse modeling of bulk-Si MOSFETs with gate length down to 40 nm (channel length down to 26 nm) from their dc electrical characterization. We show that a proper treatment of quantum effects both in the channel and in the polysilicon gate through the direct solution of Schrödinger equation, and a transport model based on two-dimensional subbands are required for accurate and—after calibration—predictive modeling. The model is included in the NANOTCAD2D code (Curatola and Iannaccone, 2003). We also evaluate the performance gap to ballistic transport, by comparing the experiments with simulations based on a fully ballistic transport model on the devices structures extracted with the inverse modeling procedure.**

***Index Terms*—Ballistic transport, drift-diffusion (DD), MOS devices, Schrödinger equation.**

## I. Introduction

THE continuous scaling of semiconductor devices for achieving performance improvement and higher density of integration imposes severe challenges to device fabrication and modeling [2]. Since conventional scaling has already shown to be inadequate for keeping the pace demanded by the ITRS [2], a concurrent use of experiments and detailed TCAD simulations is, today, a fundamental tool for driving technology toward device structures that will dominate the Semiconductor Industry in the next decade. Unfortunately, conventional semiclassical modeling approaches based on drift-diffusion (DD) transport in the bulk have revealed their insufficiency in predicting the electrical characteristics of ultrascaled devices and more accurate models need to be implemented in TCAD simulators.

The main issues that need to be properly considered are represented by the strong quantum confinement in the vertical direction, due in bulk devices to the high vertical electric field and the large impurity concentration, and by non local transport effects, due to the large longitudinal fields and by partially ballistic transport.

The latter aspect can be addressed by the use of more sophisticated transport models, such as the energy-balance (EB) [3] or Hydrodynamic [4] model, at the cost of increased computational load and time consuming simulations. Nevertheless, even if the underlying models are somewhat out of their region of validity, the majority of the routinely performed simulations are still based on the DD approach due to its simplicity and to its stable implementation, and interesting simulations of very short devices have also been reported [5] by means of *ad hoc* adjustments to the original models.

On the other hand, we believe that the issue of quantum confinement in the vertical direction is much more critical for accurate device simulation, and, after proper calibration, for predictive evaluation of scaling perspectives. Some corrections to the standard DD (or EB) approach in the bulk in order to take into account quantum confinement have been suggested in the literature [6]–[10] and are today implemented in state-of-the-art commercial TCAD tools [11], [12].

The so called "density gradient" or "Bohm quantum potential" approaches [8]–[12] used in conjunction with bulk DD transport, have often been called with emphasis "quantum DD," and have shown to provide pretty accurate electrostatics in particular in the strong inversion regime.

While we think that such approaches are interesting in the sense that they prolong the "life" and the range of applicability of standard DD, they fall short as long as two important issues are concerned: catching the physical "fact" of transport occurring in two-dimensional (2-D) subbands, with related possible modifications of scattering rates and mobility, and properly addressing quantum confinement simultaneously occurring both in the polysilicon gate and in the channel.

Here, we prefer to use a modeling approach based on the direct solution of the Schrödinger equation, coupled with drift-diffusion transport in each 2-D subband. We have called such approach "Schrödinger DD per subband" (SDDS). Our model also enables us to switch off the effect of scattering and to easily evaluate device characteristics in a purely ballistic regime [13], [14]. This is particularly useful in deep submicrometer devices in order to evaluate the effects of scaling on device performance and to estimate how far nanoscale

Manuscript received January 10, 2005; revised May 31, 2005. The review of this paper was arranged by Editor S. Datta.

G. Curatola, G. Doornbos, J. Loo, and Y. V. Ponomarev are with Philips Research Leuven, B-3001 Leuven, Belgium.

G. Iannaccone is with the Dipartimento di Ingegneria dell'Informazione, Università degli Studi di Pisa, 56122 Pisa, Italy.

Digital Object Identifier 10.1109/TED.2005.852722

0018-9383/$20.00 © 2005 IEEE

MOSFETs are from the maximum achievable current drive capability. Such issue is particularly relevant to the need, also stressed by the ITRS, to understand ballistic transport and the attainable performance improvements.

We apply our model to perform the inverse modeling of a series of fabricated bulk-Si MOSFETs with channel length down to 26 nm, and to the evaluation of the existing gap with respect to ballistic transport. Results will show that the proposed approach is effective for both tasks.

The paper is organized as follows: In Section II we briefly describe the two transport models implemented in the simulator, DD and fully ballistic per subband.

In Section III the SDDS model is applied to simulate fabricated bulk-Si MOSFETs with different channel lengths down to 26 nm. The importance of a correct treatment of quantum effects both in the polysilicon layer and in the channel is highlighted. Finally, in Section IV we evaluate by means of the ballistic model how far the investigated devices are from the ballistic limit by means of the direct evaluation of the ballistic efficiency of such devices. Conclusion and discussion end the paper.

## II. Model

In the following, we will give an overview of the two different transport models used for the treatment of out of equilibrium problems, i.e., a DD per subband coupled to the Schrödinger equation and a ballistic approach. More detailed information on the simulation tool can be found in [1].

### A. SDDS

In Fig. 1 the flow diagram of the ballistic model and of the SDDS approach is represented. The 2-D Poisson equation is solved with the Newton–Raphson algorithm after being discretized using the box integration method. Available energy states are obtained by solving the Schrödinger equation with the effective mass approximation, assuming parabolic energy bands. In particular, we assume that quantum confinement is strong only along the growth direction, let say $x$ (perpendicular to $Si/SiO_2$ interface), and the 1-D Schrödinger equation is solved along the confined direction. Repeating the calculation for each mesh point $y_j$ along the longitudinal direction, the subband profile $E^l(y)$ can be therefore extracted.

If we consider the SDDS approach, the electron density for each minimum of the conduction band, in equilibrium ($V_{\text{DS}} = 0$) can be written as

$$n(x,y) = \rho_{2D} k_B T_c \sum_t |\Psi^l(x,y)|^2 F_0 \left( \frac{E_F(y) - E^l(y)}{k_B T} \right) \tag{1}$$

where $F_0$ is the Fermi–Dirac integral of order zero, $\Psi^l(x,y)$ represents the $l$th eigenfunction with eigenvalue $E^l(y)$, $\rho_{2D}$ represents the 2-D electron gas (2DEG) density of states and $E_F(y)$ is the Fermi level in the current flow direction. The total electron density is obtained by summing up the contributions of all minima, considering mass anisotropy. After integration over the

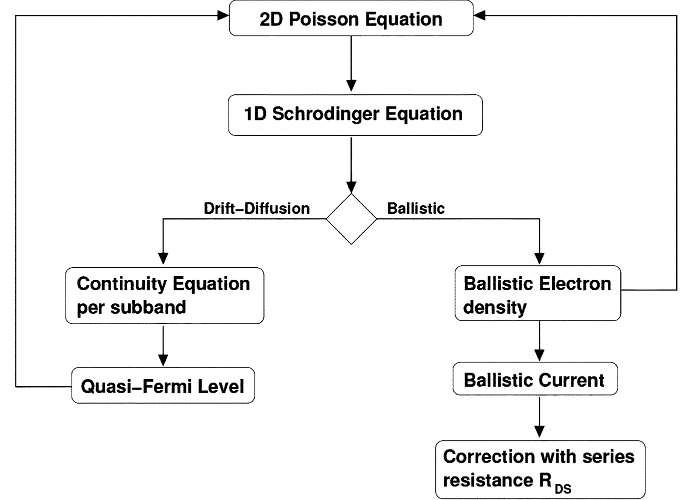


Fig. 1. Flow diagram of an out-of-equilibrium simulation with NANOTCAD2D.

confined direction, the electron density per unit area in the $l$th subband can be expressed as

$$n^l(y) = \rho_{2D} k_B T_c F_0 \left[ \frac{E_F(y) - E^l(y)}{k_B T} \right]. \tag{2}$$

If we assume to be in equilibrium and then apply a voltage $V_{\text{DS}}$ to the device, (2) is used as the initial guess of the carrier density in each subband and also to impose the Dirichlet boundary conditions (Ohmic contacts) to the continuity equation in each subband.

The continuity equation is solved in the DD limit, assuming that carriers are in equilibrium with the lattice and that the kinetic part of the energy density is negligible with respect to the thermal contribution due to the random motion of carriers. To avoid numerical instabilities, the continuity equation per subband has been discretized using the Scharfetter-Gummel [15] approach which yields

$$J^l_{i+1/2} = \frac{q D^l_n}{h_i} \left[ B\left(\Delta^l_i\right) n^l_{i+1} - B\left(-\Delta^l_i\right) n^l_i \right] \tag{3}$$

where superscript $l$ identifies the selected subband, subscript $i + 1/2$ means that the current is evaluated at an intermediate point between $y_i$ and $y_{i+1}$, $n$ represents the electron density per unit area and $h_i = y_{i+1} - y_i$ is the grid spacing. The function $B(\Delta) = (\Delta)/(\exp(\Delta) - 1)$ represents the Bernoulli function with argument $\Delta$.

Equation (3) has been derived assuming that the electrostatic potential varies linearly between two adjacent grid points. Particular attention must be given to the inclusion of Fermi–Dirac statistics in deriving (3) in order to properly account for degeneracy effects. Therefore, we have to define a modified diffusion coefficient in each considered subband and to correct the argument of the Bernoulli function as follows:

$$D^l_{n_i} \equiv \mu^l_{n_i} \frac{k_B T_c}{q} \lambda^l_{F_i} \tag{4}$$

$$\Delta^l_i \equiv \frac{q(\varphi_{i+1} - \varphi_i)}{k_B T_c \lambda^l_{F_i}} \tag{5}$$

where $\varphi$ represents the electrostatic potential and $\lambda_{F_i}^l$ is a dimensionless corrector term, corresponding to the selected subband $l$ and evaluated at the grid point $i$, which allows us to implement Fermi–Dirac statistics and is given by

$$\lambda_{F_i}^l = \frac{q}{k_B T_c} \frac{n_i^l}{\partial n_i^l / \partial E_F} = \frac{F_0\left(E_{F_i} - E_i^l\right)}{F_{-1}\left(E_{F_i} - E_i^l\right)} \tag{6}$$

As far as the mobility model is concerned, we use a physically based local semi-empirical model valid in the bulk [16], which expresses the global mobility, by means of the Matthiessens rule, as a function of the carrier mobility in the bulk, the mobility limited by surface acoustic phonon scattering and the mobility limited by surface roughness. The effect of Coulomb screening of impurities by charge carriers in the bulk mobility term is also included in the model. The local mobility in the $l$th subband, where we solve the continuity equation, is obtained as a weighted average over the corresponding eigenfunctions

$$\mu^l(y) = \frac{\int \Psi^l(x,y)^* \mu(x,y) \Psi^l(x,y) dx}{\int \Psi^l(x,y)^* \Psi^l(x,y) dx}. \tag{7}$$

Equation (7) expresses, as a first-order approximation, the mobility in each 2-D subband as the weighted average of the local mobility with the local density of states along the vertical direction on the considered subband. We are aware that this approach represents a brutal approximation, that has the advantage of tending to bulk mobility [16] when confinement tends to vanish. An appropriate mobility model would require the computation of the scattering rates in each subband, based on the evaluation of the overlap integrals between different eigenfunctions, but would be out of the scope of the present paper. Let us stress here the fact that we neglect here intersubband scattering, i.e., we enforce current conservation *in each subband*.

After the continuity equation has been solved and the new guess of the electron density per unit area has been obtained, the quasi-Fermi level is calculated from the comparison between (1) and electron profile extracted from the continuity equation, and a new cycle starts until convergence is reached.

### B. Ballistic Transport

If the device length is smaller than the mean free path, it is very probable for carriers not to undergo any scattering event during their motion within the device and transport is *ballistic*.

The degree of ballistic transport of the MOSFET is mainly controlled by a small region near the source [17], [18] instead of the whole channel region, which implies that the device is essentially ballistic or quasi-ballistic even if the channel length is larger than the mean free path. If we consider the subband profile along the channel, electrons travelling from source to drain must overcome an energy barrier modulated by the gate voltage and by the drain voltage (DIBL). If the length, measured from the peak of the barrier, over which the potential drops by $k_B T/q$, is lower than the mean-free path, carriers that pass the barrier peak cannot be reflected back to the source even if they undergo a scattering event. Therefore, scattering near the source limits the steady-state drain current while scattering near the drain can be considered a second order effect. In our model, carriers are injected into the channel from a thermal equilibrium reservoir (source/drain) and contribute to the current only if they have a longitudinal energy higher that the barrier. We simply assume that electrons with injected longitudinal energy lower than the subband maximum are reflected back to their originating contact, while the others are transmitted over the barrier and contribute to the current [19], [20]. In our approach we do not restrict the calculation to the quantum limit but several subbands are taken into account. For each subband $l$ we evaluate the subband maximum $E_{\max}^l$ and its corresponding longitudinal position $y_{\max}^l$. All electrons with longitudinal energy lower than $E_{\max}^l$ are in equilibrium with the originating contact, while electrons with longitudinal energy higher than $E_{\max}^l$ conserve the chemical potential of the injecting reservoir. The occupation factor $f(y,E)$ is therefore given by a Fermi–Dirac distribution $f(E_{\rm FS},E)$ with the source fermi level $E_{\rm FS}$ if $y < y_{\max}^l$ and $E < E_{\max}^l$; by a Fermi–Dirac distribution $f(E_{\rm FD},E)$ with the drain fermi level $E_{\rm FD}$ if $y > y_{\max}^l$ and $E < E_{\max}^l$. Instead, if $E > E_{\max}^l$, then for each $y$ and $E$ there is a state travelling toward the right (coming from the source) and a state travelling toward the left (coming from the drain). In this case the occupation factor is given by $f(y,E) = (1/2)[f(E_{\rm FS},E) + f(E_{\rm FD},E)]$.

The electron current for each conduction band minimum can be evaluated assuming that there is no tunnel current through the barrier so that only electrons with longitudinal energy higher than $E_{\max}^l$ contribute

$$J_n = g_e \sum_l 2q \frac{\sqrt{2m_z}}{h^2} (k_B T)^{\frac{3}{2}} \times \left[ F_{\frac{1}{2}}\left(\frac{E_{\rm FS} - E_{\max}^l}{k_B T}\right) - F_{\frac{1}{2}}\left(\frac{E_{\rm FD} - E_{\max}^l}{k_B T}\right)\right] \tag{8}$$

where $F_{1/2}$ represents the Fermi–Dirac integral of order $1/2$.

## III. Results and Discussion

### A. Devices

We have focused our attention on planar silicon $n$-channel MOSFETs fabricated with a low thermal budget process [21]. A super-steep retrograde (SSR) boron profile is used in order to minimize short channel effects where implantation is performed with the gate, source/drain already in place [22]. In order to avoid undesired boron diffusion, the only thermal step experienced by the pocket implant is the final high-temperature RTA spike annealing (900 ° C–1100 °C) and optimized PECVD of oxide was used to reduce the thermal budget while preserving the quality of the deposited spacers. A single activation anneal and the back-end processing are, therefore, the only thermal steps experienced by the SSR. The equivalent oxide thickness of the gate dielectric (heavily nitrided $SiO_2$) was EOT $= 1.5$ nm. Different gate lengths ranging from 10 $\mu$m down to 40 nm have been explored.

### B. Simulations and Experiments

Accurate modeling of the two dimensional doping distribution is crucial as the gate length is reduced and the effects of pockets and lateral extension diffusion become progressively more important. Starting from an initial doping profile obtained with the TSUPREM4 [23] process simulator, we have used both the $C$-$V$ and $I$-$V$ characteristics [24], [25] measured on

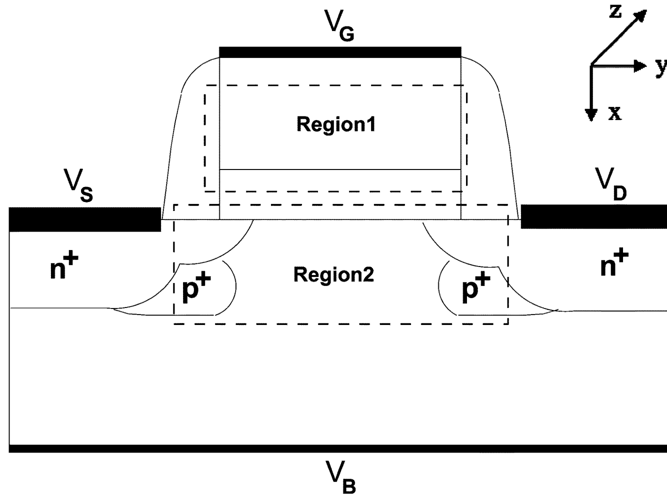


Fig. 2. Schematic representation of the simulated bulk MOSFET. Shown are the two quantum regions where we solve the 1-D Schrödinger equation (region1-region2) and the continuity equation per subband (region2) or the ballistic current equation (region2).

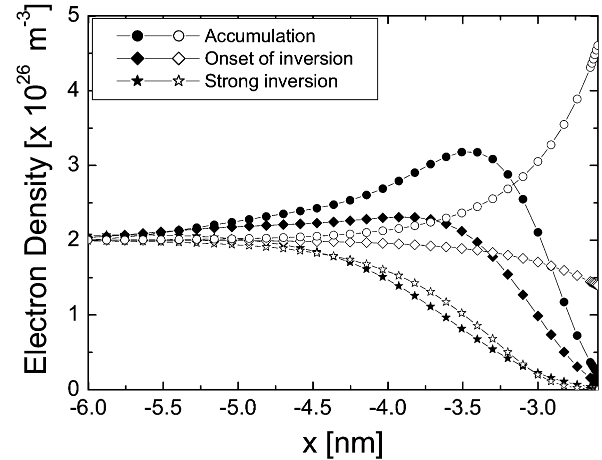


Fig. 3. Electron density in the polysilicon layer obtained with a semiclassical approach(open symbol) and with a full quantum approach (filled symbol) in conditions of accumulation, onset of inversion and strong inversion.

devices fabricated with the same process but with different gate lengths. For the capacitance–voltage ($C$–$V$) measurements a special structure consisting of two equal transistors in parallel has been used, which has been shown to be effective in reducing undesired leakage effects [26], and results of measurements have been carefully compared with results of simulations obtained with our Poisson/Schrödinger solver at equilibrium.

We initially focus on very long structures ($L_G = 880$ nm), where the process spread of the gate length and the effect of source/drain and pockets doping can be considered negligible, in order to use the measured current–voltage ($I$–$V$) and $C$–$V$ characteristics to extract the doping profiles in the well and in the polysilicon. The latter is very difficult to predict with a process simulation while its accurate determination is crucial because of the strong polysilicon depletion effects that are significant for EOTs lower than 2 nm. In particular, the accumulation and low inversion $C$-$V$ curve coupled with the transfer characteristics is used to determine the correct doping profile and the gate length, while the strong accumulation $C$–$V$ curve is used to determine the donor concentration in the poly.

With this aim, it is extremely important to account for polysilicon depletion effects through the direct solution the 1-D Schrödinger equation. Indeed, due to the abrupt energy barrier at the poly/$SiO_2$ interface, electrons are repelled from the interface and a dark space charge forms [27], [28] both when the transistor operates in accumulation and in inversion. In Fig. 2 a schematic representation of the simulated device is given. We have considered two different quantum regions, one in the poly(region1) and one in the channel (region2), and solved in both regions the 1-D Schrödinger equation in the vertical direction for each $y$, extending each region into the oxide in order to allow for wavefunction penetration. In Region 2 the continuity equation is solved for each subband, assuming for simplicity that intersubband scattering is negligible, and the 2-D Poisson equation is solved self-consistently in the whole structure in order to account for short-channel effects. In Fig. 3 the electron charge density in the polysilicon layer, obtained with a semiclassical approach and with a full Schrödinger approach, is shown for three different operating modes of the transistor: accumulation, onset of the inversion and strong inversion. At the onset of inversion, electrons are pushed away from the interface leaving uncompensated ionized donors. Since charge neutrality must be ensured and extra electrons can not be provided by the inversion layer which is not yet formed, we have a bump in the electron density, as clearly shown in the figure. The local peak in the electron density has a direct effect on the electric field which does not increase gradually as we move from the polysilicon toward the oxide, but reveals a local minimum located in correspondence of the charge peak and induces a corresponding increase of the electrostatic potential toward the interface and inside the oxide. The consequences of quantum effects in the poly are, mainly, a negative shift of the threshold voltage and a degradation of the gate capacitance due to the presence of the depletion capacitance which adds in series.

Taking into account for the quantum effects in the poly and in the channel, it is possible to reproduce very well the experimental $C$–$V$ curves both in accumulation and in inversion, as shown in Fig. 4(a) and to extract the donor distribution in the polysilicon gate. In particular, a uniform donor concentration of $N_D = 1.2 \times 10^{20}$ $\text{cm}^{-3}$ has been considered. Also shown in the Fig. 4(b) is the extracted boron profile in the channel.

Devices with shorter gatelengths may be used to extract source/drain and pocket doping with 2-D inverse modeling. Fixing the extracted SSR profile, we repeated the previously described procedure for a MOSFET with gatelength of 64 nm, in order to correctly extract the doping profiles of pockets, deep HDDs and LDD extensions. Starting profiles have been obtained with TSUPREM4 and fitting of the $C$–$V$ and $I$–$V$ characteristics has been obtained using analytical gaussian doping profiles.

To summarize, a correct treatment of quantum effects both in the channel and in the polysilicon layer is fundamental in order to properly extract the 2-D doping profile in the structure and, therefore, to reproduce with simulations the electrical

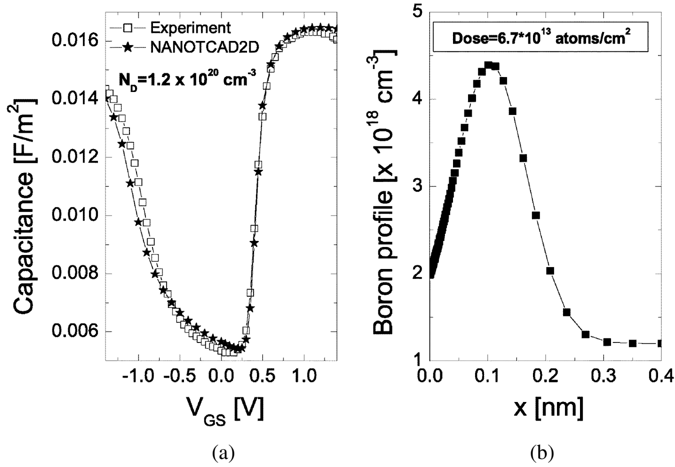


Fig. 4. (a) Experimental and simulated $C$–$V$ curve of the large nMOS device with $W = L = 10\ \mu$m. The extracted value of the donor concentration in the poly is shown in the inset. (b) Extracted boron profile as a function of the confined direction.

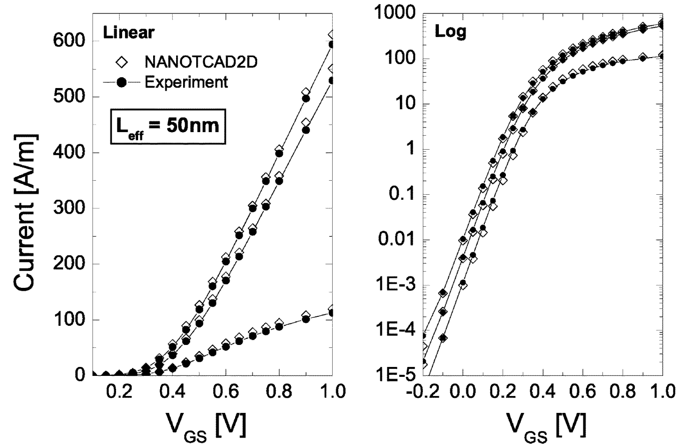


Fig. 5. Transcharacteristics for the $L_G = 64$ nm ($L_{\rm eff} = 50$-nm) device, represented in linear (left) and logarithmic scale (right) for three different value of the drain-to-source voltage: $V_{\rm DS} = 0.05, 0.55, 1.05$ V. The extracted effective channel length is shown in the inset.

characteristics of the considered devices. Indeed, we have experienced, as expected, a very strong dependence of the simulated $I$–$V$ curves on the extracted doping profile in the well, in the extension regions, and the doping of pocket implants. It is also worth noticing that the threshold voltage of the considered devices is strongly affected by the donor concentration in the polysilicon layer, which has been extracted by comparison between the measured and simulated $C$–$V$ characteristics.

Particular attention must be given to the parasitic S/D series resistances which reduce the on-current and cause a degradation of the electrical performance of the device. The total parasitic resistance of each device can be modeled as the sum of four different contributions, i.e., 1) silicide contact resistance $R_c$; 2) deep resistance $R_{\rm dp}$ due to the HDD region; 3) extension resistance $R_{\rm ext}$ due to the LDD formation; 4) overlap resistance $R_{\rm ov}$ due to the underdiffusions. In order to account for the parasitic resistances in our SDDS approach, different simulations have been performed by progressively increasing the lateral dimension of the quantum region (Region 2 in Fig. 2) where the continuity equation is solved, until the on-current remained practically unaffected by further enlargement of this region. The optimal choice which minimizes computing time and gives an accurate prediction of the on current consists in extending the quantum region in order to include the entire LDD extensions so that the overlap and extension resistance effect is fully taken into account in the simulation. For the specific process we consider, the deep resistance and the silicide contact resistance have been estimated as negligible.

Comparison of the transfer characteristics obtained from experiments and SDDS simulations, is shown in Fig. 5 and allows us to conclude that the fitting has been very good. More importantly, in Fig. 6 we show the same comparison for a MOSFET with gatelength of 40 nm (channel length of 26 nm) for which no fitting has been performed, and the doping profiles previously obtained have been used. The agreement in this case is excellent, and shows how the SDDS approach is a adequate tool for inverse device modeling, and can be used for predictive simulations, after proper calibration, and evaluation of scaling perspectives of a technology at hand.

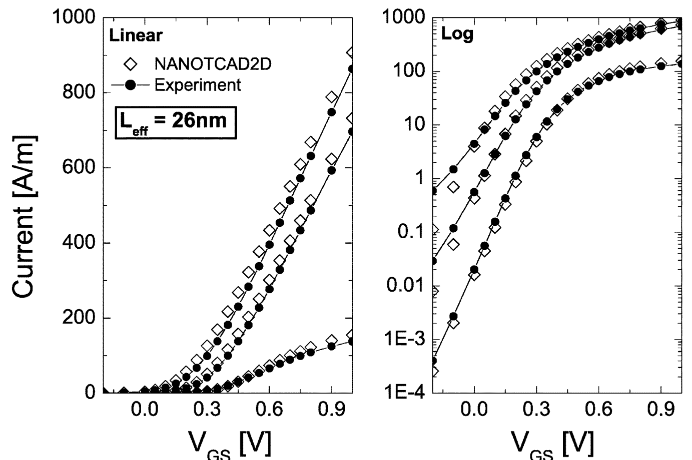


Fig. 6. Transcharacteristics for the $L_G = 40$-nm device ($L_{\rm eff} = 26$ nm), represented in linear (left) and logarithmic scale (right) for three different value of the drain-to-source voltage: $V_{\rm DS} = 0.05, 0.55, 1.05$ V. The extracted effective channel length is shown in the inset.

Our results also show that the accurate extraction of doping profiles and the proper treatment of quantum confinement are sufficient for reproducing characteristics of MOSFETs down to channel length of 26 nm, even if a basic transport model such as DD is adopted.

## IV. Evaluating the Performance Gap to Ballistic Transport

If we apply the ballistic model for the evaluation of the current, the difference with respect to the previously described procedure is that, instead of solving the continuity equation per subband inside region two of Fig. 2, ballistic current is evaluated from (8). Poly depletion effects are again taken into account as well as quantum confinement in the channel.

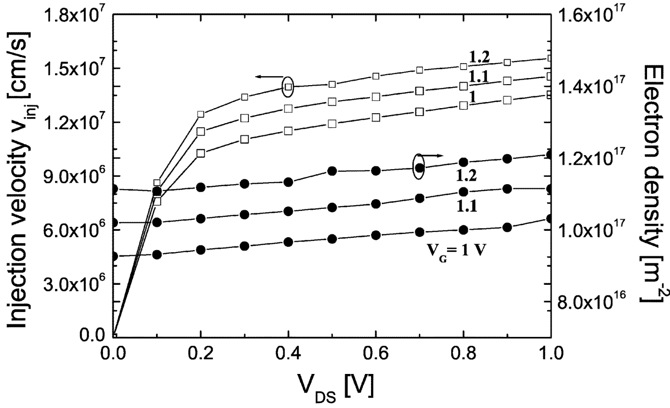


Fig. 7. (Left) Injection velocity at the top of the barrier as a function of $V_{\mathrm{DS}}$ for different $V_{\mathrm{GS}}$. (Right) Inversion charge at the top of the barrier. In a well-tempered MOSFET in strong inversion, the electron density at the barrier peak is roughly only proportional to the gate overdrive $(V_G - V_T)$ and it is independent of the drain-to-source voltage.

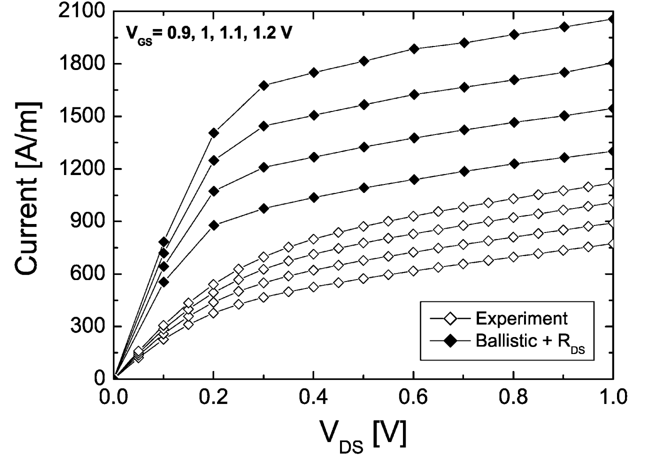


Fig. 8. Output characteristics for the 26-nm channel length MOSFET obtained with the ballistic simulation (filled symbols) and from experiments (open symbols).

Ballistic transport represents, for a given device structure, the upper performance limit. Of course, even considering the most optimistic case, one has to take into account that for a given device structure and doping profile, the source and drain series resistances must be included. Since the ballistic model does not account for parasitic series resistance $R_{\mathrm{DS}}$, the simulated current has been therefore corrected with the measured value of $R_{\mathrm{DS}}$. The series resistance has been extracted from the measured $I$–$V$ characteristics of different gate length devices by means of the Shift and Ratio method [29] and the obtained value is $\sim$110 Ohm-$\mu$m.

As a measure of the performance improvements that would be available if transport was ballistic, it would be interesting to make a comparison between the experimental characteristics of the nanoscale MOSFETs and simulations of ballistic MOSFETs, with the same structures and doping profile obtained as shown in Section III.

Even if no consensus has been reached up to now on whether the ballistic limit will be reached as we proceed with scaling, comparisons between experiments and Monte Carlo simulations [30], [31] show that MOSFETs which channel length down to 45 nm operate roughly at 50% of their ballistic limit and that such level can *not* be increased with scaling.

We believe that two parameters can be used as indexes of performance gap to ballistic transport: the "ballistic efficiency" defined in [32] as

$$\theta = \frac{v_{\mathrm{eff}}}{v_{\mathrm{inj}}} = \frac{J_{\mathrm{eff}}}{J_{\mathrm{Ball}}} \frac{Q_{\mathrm{inv}}^{\mathrm{Ball}}}{Q_{\mathrm{inv}}^{\mathrm{eff}}} \tag{9}$$

where $v_{\mathrm{eff}}$ represents the measured injection velocity at the barrier peak, $v_{\mathrm{inj}}$ represents the thermal ballistic velocity, $Q_{\mathrm{inv}}^{\mathrm{Ball}}, Q_{\mathrm{inv}}^{\mathrm{eff}}$ represent the mobile charge densities evaluated at the peak of the barrier for the ballistic and the real devices, respectively. $J_{\mathrm{eff}}$ and $J_{\mathrm{Ball}}$ represent the measured and the simulated ballistic current, respectively. An alternative parameter could be simply the ratio of the two currents $J_{\mathrm{eff}}/J_{\mathrm{Ball}}$. The latter approach is preferable since it allows to evaluate the performance gap to the ballistic limit directly from simulations, avoiding, consequently, the problematic extraction of the injection velocity from measurements performed on nanoscale devices. In the following, results obtained with the two different approaches are examined.

The ensemble ballistic velocity $v_{\mathrm{inj}}$ can be easily obtained from simulations as: $v_{\mathrm{inj}} = J_{\mathrm{ball}}/Q_{\mathrm{inv}}^{\mathrm{Ball}}$.

Fig. 7 shows the ballistic injection velocity $v_{\mathrm{inj}}$ and the electron density per unit area at the peak barrier as a function of $V_{\mathrm{DS}}$ and $V_{\mathrm{GS}}$. The electron density is almost constant with $V_{\mathrm{DS}}$ at a value given, to first order, by $C_{\mathrm{eff}}(V_{\mathrm{GS}} - V_T)$ (Deviations can be considered a consequence of short-channel effects). As far as the velocity is concerned, it is interesting to note that in the SDDS approach the velocity saturation occurs because of a reduction of the mobility in the high field regime due to an increased number of scattering events. Carriers accelerated by the high electric field scatter more during their travel toward the drain and hence the ensemble momentum relaxation time is reduced. The ensemble average velocity $v_d \approx \mu E$ saturates in the proximity of the drain.

In the ballistic regime, where the concept of mobility is meaningless, carriers with positive momentum $(+k_y)$ are injected from the source contact and carriers with negative momentum $(-k_y)$ are injected from the drain contact and the two different populations do not mix if we suppose that contacts are reflectionless. In equilibrium, the net velocity is zero because there is an equal number of carriers with positive and negative velocity. As the drain voltage is increased the negative half is unable to overcome the energy peak and only carriers with positive momentum contribute to the total current. The velocity therefore saturates in the proximity of the peak barrier, located near the source, at the thermal injection velocity $v_d \approx v_{\mathrm{th}}$ [17].

In Fig. 8, the output characteristics obtained with the ballistic simulation (considering the extracted series resistances) are compared with experiments for the 26-nm channel length MOSFET. In Fig. 9 the same comparison is shown for the transfer characteristics.

Due to the problematic extraction of injection velocity from experimental results, accurate determination of ballistic efficiency is not straightforward and different methods have been suggested to overcome the problem. In particular, as pointed out in [30], the conventional approach to obtain carrier velocity

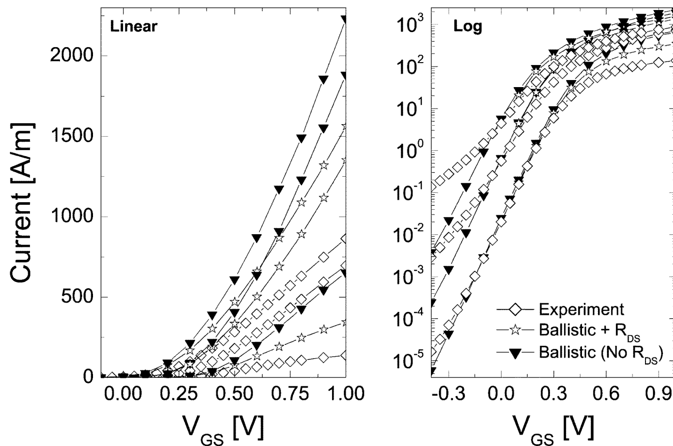


Fig. 9. Transcharacteristics for the $L_G = 40$-nm ($L_{\mathrm{eff}} = 26$-nm) device for three different source-to-drain voltages $V_{\mathrm{DS}} = 0.05, 0.55, 1.05$ V. Comparison between the experimental results and the results of simulation with the ballistic approach. The correction to the ballistic current due to series resistances is also shown in the plot.

by evaluating the ratio $g_m/WC_{\mathrm{ox}}$ can induce large errors since the extracted barrier peak is erroneously shifted toward the drain with respect to the real position and, therefore, a higher injection velocity is predicted since electrons already started to be accelerated by the subband potential profile. Instead, the velocity at the top of the barrier can be better evaluated by extracting the term $Q_{\mathrm{inv}}^{\mathrm{eff}}$ and evaluating the ratio $J_{\mathrm{eff}}/Q_{\mathrm{inv}}^{\mathrm{eff}}$. Inversion carrier density can be obtained by integrating the effective long-channel capacitance $C_{\mathrm{eff}}$ and considering, in the integration limits, a correction term which accounts for short-channel effects and comprises the threshold voltage difference $\Delta V_T$ between the long-and short-channel device and the effect of series resistance $R_{\mathrm{DS}}$. Considering the capacitance $C_{\mathrm{eff}}$ measured on a long-channel device with $W = L = 10\ \mu$m, the extracted value for the effective velocity has been found to be $v_{\mathrm{eff}} = 7.22 \times 10^6$ cm/s, giving a ballistic efficiency of $\theta \approx 0.53$.

Alternatively, one could obtain a very similar results by looking directly at the $J_{\mathrm{eff}}/J_{\mathrm{ball}}$ ratio, that for the same bias point is 0.56. Such agreement is due to the fact that, due to the inverse modeling procedure, we have been able to extract in a very accurate way the doping profiles, so that the ratio $Q_{\mathrm{inv}}^{\mathrm{Ball}}/Q_{\mathrm{inv}}^{\mathrm{eff}}$ at the barrier peak is very close to one.

## V. Conclusion

We have presented a simulation approach based on a full evaluation of quantum confinement in the polysilicon and in the channel coupled with a DD transport model in each 2-D subband, which has been shown to be reliable and predictive for devices with channel length down to 26 nm. The presented approach has been applied first to the extraction of the doping profiles of bulk-Si MOSFETs realized with the same technology, with gate lengths from a few micron to 64 nm, and has been successfully used to simulate—without additional calibration—the $I$–$V$ characteristics of a MOSFET with channel length of 26 nm.

In addition, using the obtained doping profile in our simulator of ballistic MOSFETs, we have been able to estimate the performance gap to ballistic transport, in terms of the ratio of the experimental current to the ballistic current, for the same doping profiles and series S/D resistances.

As a final remark, it clearly appears from our study that an essential requirement for a reliable and predictive simulation is the capability to precisely extract (also through simulations on test devices) the doping profile of the device, which is necessary for an accurate description of the electrostatics. This aspect, also for nanoscale devices, appears to be more important than an accurate mobility model or a more sophisticated transport model.

## References

[1] G. Curatola and G. Iannaccone, “NANOTCAD2D: Two-dimensional code for the simulation of nanoelectronic devices and structures,” *Computational Materi. Sci.*, vol. 28, pp. 342–352, 2003.

[2] International Roaadmap for Semiconductors [Online]. Available: http://public.itrs.net/

[3] R. Stratton, “Semiconductor current-flow equation (diffusion and degeneracy),” *IEEE Trans. Electron Devices*, vol. ED–19, pp. 1288–1292, 1972.

[4] K. Blotekjaer, “Transport equations for electrons in two-valley semiconductors,” *IEEE. Trans. Electron Devices*, vol. ED–17, pp. 38–47, 1970.

[5] J. D. Bude, “MOSFET modeling in the ballistic regime,” in *Proc. Int. Conf. Simulation Semconductor Processes Devices*, 2000, pp. 23–26.

[6] M. J. VanDort, P. H. Woerlee, and A. J. Walker, “A simple model for quantization effects in heavily-doped silicon MOSFETs at inversion conditions,” *Solid State Electron.*, vol. 37, pp. 411–414, 1994.

[7] W. Hänsch, Th. Vogelsang, R. Kirchner, and M. Orlowski, “Carrier transport near the Si/SiC$_2$ interface of a MOSFET,” *Solid State Electron*, vol. 32, pp. 839–849, 1989.

[8] M. G. Ancona and H. F. Tiersten, “Macroscopic physics of the silicon inversion layer,” *Phys. Rev. B, Condens. Matter*, vol. 35, pp. 7959–7965, 1987.

[9] A. Wettstein, A. Schenk, and W. Fichtner, “Quantum device-simulation with the density-gradient model on unstructured grids,” *IEEE Trans. Electron Devices*, vol. 48, no. 3, pp. 279–284, Mar. 2001.

[10] G. Curatola, G. Iannaccone, and G. Fiori, “Effective Bohm quantum potential for device simulators based on drift-diffusion and energy transport,” in *Proc. SISPAD*, 2004, pp. 275–278.

[11] [Online]. Available: http://www.silvaco.com/

[12] [Online]. Available: http://www.synopsys.com/

[13] G. Fiori and G. Iannaccone, “Modeling of ballistic nanoscale metal-oxide-semiconductor field effect transistors,” *Appl. Phys. Lett.*, vol. 81, pp. 3672–3674, 2001.

[14] G. Curatola, G. Fiori, and G. Iannaccone, “Modeling and simulation challenges for nanoscale MOSFETs in the ballistic limit,” *Solid State Electron.*, vol. 48, pp. 581–587, 2004.

[15] D. L. Scharfetter and H. K. Gummel, “Large signal analysis of a silicon read diode oscillator,” *IEEE Trans. Electron Devices*, vol. ED–16, pp. 64–77, 1969.

[16] M. N. Darwish, J. L. Lentz, M. R. Pinto, P. M. Zeitzoff, T. J. Krutsick, and H. H. Vuong, “An improved electron and hole mobility model for general purpose device simulation,” *IEEE Trans. Electron Devices*, vol. 44, no. 12, pp. 1529–1538, Dec. 1997.

[17] M. Lundstrom and R. Zen, “Essential physics of carrier transport in nanoscale MOSFETs,” *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 133–141, Jan. 2002.

[18] A. Rahman, J. Guo, S. Datta, and M. Lundstrom, “Theory of ballistic nanotransistors,” *IEEE Trans. Electron Devices*, vol. 50, no. 11, pp. 1853–1864, Nov. 2003.

[19] K. Natori, “Ballistic metal-oxide-semiconductor field effect transistor,” *J. Appl. Phys.*, vol. 76, p. 4879, 1994.

[20] R. Venugopal, R. Zhibin, and M. S. Lundstrom, “Simulating quantum transport in nanoscale MOSFETs: Ballistic hole transport, subband engineering and boundary conditions,” *IEEE Trans. Electron Devices*, vol. 2, no. 1, pp. 135–143, Jan. 2003.

[21] Y. V. Ponomarev, J. J. G. P. Loo, C. J. J. Dachs, F. N. Cubaynes, M. A. Verheijen, M. Kaiser, J. G. M. van Berkum, S. Kubicek, J. Bolk, and M. Rovers, “A mnanufacturable 25-nm planar MOSFET technology,” in *Symp. VLSI Tech. Dig.*, 2001, pp. 33–34.

[22] Y. V. Ponomarev, P. A. Stolk, A. C. M. C. van Brandenburg, R. Roes, A. H. Montree, J. Schmitz, and P. H. Woerlee, “Channel profile engineering of 0.1 m-Si MOSFETs by throughthe-gate implantation,” in *IEDM Tech. Dig.*, 1998, pp. 635–638.

[23] *TSUPREM-4 Users Manual*. Sunnyvale, CA: TMA Inc., 1996.

[24] I. J. Djomehri and D. A. Antoniadis, "Inverse modeling of sub-100-nm MOSFETs using *I–V* and C-V," *IEEE Trans. Electron Devices*, vol. 49, no. 4, pp. 568–574, Apr. 2002.

[25] Z. K. Lee, M. B. McIlrath, and D. A. Antoniadis, "Two-dimensional doping profile characterization of MOSFETs by inverse modeling using *I–V* characteristics in the subthreshold region," *IEEE Trans. Electron Devices*, vol. 46, no. 10, pp. 1640–1649, Oct. 1999.

[26] F. Cubaynes, Ph.D. dissertation, Philips Research Leuven, Leuven, Belgium, 2004.

[27] A. S. Spinelli, A. Pacelli, and A. L. Lacaita, "Polysilicon quantization effects on the electrical properties of MOS transistors," *IEEE Trans. Electron Devices*, vol. 47, no. 12, pp. 2366–2371, Dec. 2000.

[28] A. Pacelli, A. S. Spinelli, and L. M. Perron, "Carrier quantization at flat bands in MOS devices," *IEEE Trans. Electron Devices*, vol. 46, no. 3, pp. 383–387, Mar. 1999.

[29] Y. Taur, D. S. Zicherman, D. R. Lombardi, P. J. Restle, C. H. Hsu, H. I. Hanafi, M. R. Wordeman, B. Davari, and G. G. Shahidi, "A new 'shift and ratio' method for MOSFET channel-length extraction," *IEEE Trans. Electron Devices*, vol. 13, no. 3, pp. 267–269, Mar. 1992.

[30] A. Lochtefeld and D. A. Antoniadis, *IEEE Electron Device Lett.*, vol. 22, no. 2, pp. 95–97, Feb. 2001.

[31] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M. Lundstrom, "On the performance limits for Si MOSFETs: A theoretical study," *IEEE Trans. Electron Devices*, vol. 47, no. 3, pp. 232–240, Mar. 2000.

[32] M. Lundstrom, "Elementary scattering theory of the Si MOSFET," *IEEE Electron Device Lett.*, vol. 18, no. 4, pp. 361–363, Apr. 1997.

**Gilberto Curatola** was born May 5, 1975 in Italy. He received his M.Sc. degree in electrical engineering and the Ph.D. degree for a thesis on the subject of quantum effects and transport in nanoscale field effect transistors, from the University of Pisa, Pisa, Italy in 2000 and 2005, respectively.

Since January 2005 he has been with Philips Research Leuven, working on device simulations for CMOS applications. His research interest includes characterization and simulation of deep-sub-micrometer silicon devices with emphasis on quantum effects and mobility.

**Gerben Doornbos** was born in 1969 in The Netherlands. He received the M.S. degree in experimental physics and the Ph.D. degree in physics from the Vrije Universiteit (VU), Amsterdam, Amsterdam, The Netherlands, in 1993 and 2001, respectively. His doctoral research focused on the dynamic behavior of magnetic vortices in Type II superconductors.

From 1999 to 2000, he was a Joined Researcher in the VU physics-applied computer science group, applying massively distributed computing to complex physics-based simulations. He joined Philips Research Leuven, Leuven, Belgium, in 2000, working on TCAD-based device modeling for advanced CMOS applications.

**Josine Loo** received the B.Sc. in physics from the Zuyd University, Netherlands in 1998.

From 1998 until 2005, she has been with Philips Research working on non-volatile memories and CMOS device technologies. She has coauthored more than 13 publications in journals and conferences proceedings and has several patents pending in the field of Si technology. Since May 2005 she has been with IMEC, Leuven, Belgium and is active in the field of RF and analog applications of CMOS technologies.

**Youri V. Ponomarev** received the B.Sc. in physics from the Moscow Institute of Physics and Technology (FizTech), Russia, in 1992, and Ph.D. degree in physics from Exeter University, U.K. in 1995 for his research on the Fractional Quantum Hall effect.

Since graduation, he has been with Philips Research Leuven, Leuven, Belgium working on advanced bipolar and CMOS device technologies. He has authored and co-authored more than 40 publications in scientific journals and conference proceedings and holds four U.S. patents in the field of Si technology. He is currently Principal Scientist working on the exploration of advanced Si technologies for biomedical applications.

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

Since January 2001 he has been an Associate Professor at the Information Engineering (Electrical 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 and research projects as consortium coordinator or principal investigator, and has authored more than 90 papers in peer-reviewed journals, and 60 papers in proceedings of international conferences.