Accurate Three-Dimensional Simulation of Electron Mobility Including Electron-Electron and Electron-Dopant Interactions

C. Heitzinger,[1] C. Ringhofer,[1] S. Ahmed,[2] D. Vasileska,[2]
[1]Department of Mathematics and Statistics
Arizona State University, Tempe, AZ 85287, USA
[2]Department of Electrical Engineering
Arizona State University, Tempe, AZ 85287, USA

ABSTRACT

Electron-electron and electron-impurity interactions play an important role in ultra-small MOSFETs. Previous simulation approaches (particle-mesh and particle-particle particle-mesh methods) were found to be very time consuming. Here a fast multi-pole method is used instead which was found to yield physically correct results in bulk mobility and device simulations within significantly decreased simulation times.

We describe the simulation method and compare the electron mobility for five doping concentrations as obtained experimentally and by five simulation methods.

I. INTRODUCTION

The constant need for integration of more functions on a chip necessitates continuous scaling of the semiconductor devices. According to the SIA Road-map for Semiconductors it is expected that by 2014 production devices will have a gate-length of 35 nm and a oxide thickness of less than 1 nm. There are, however, a number of barriers that have to be overcome in accomplishing this goal.

One problem is finding the proper oxide that will satisfy the requirements for low-power dissipation. An attempt in solving this problem has been made by the introduction of the high-k dielectrics, but a universal solution to this problem has not been found yet.

Another issue that needs to be solved for proper operation of nanoscale MOSFETs is the shallow source and drain implantation process. Yet another problem that will pose serious problems in circuits operation are the transistor mismatches due to the different number and different patterns of the impurity atoms underneath the gate which will affect the off-state current and the device threshold voltage. There are significant fluctuations in the threshold voltage from device to device fabricated on the same chip. The solution to this

problem is to use undoped silicon films. However unintentional doping, i.e., the presence of even a single impurity atom, can significantly alter the performance of the device under consideration.

Now the continuous scaling of MOSFET devices into the nanoscale regime requires refined models for electron transport. Coulomb interactions must be considered because of two reasons. First the Coulomb force is a long range force implying that both the short and the long range interactions must be included in a particle ensemble. Second the hot-carrier and short-channel effects will have a significant impact on device performance due to the small number of carriers and impurities in the active region of nanoscale devices.

Although ensemble Monte Carlo (EMC) simulators usually do not include the full Coulomb interactions due to the necessary computational time and resources, two techniques have been used in the past for properly including the short and the long range electron-electron and electron-ion interactions in particle based simulations: P^3M (particle-particle particle-mesh) methods [1], adapted to the boundary conditions necessary for semiconductor devices, and the corrected Coulomb approach [2, 3]. However both approaches suffer from drawbacks when used in the context of transport simulations in semiconductors.

To overcome these problemes we propose to use a FMM (fast multi-pole method) [4, 5, 6, 7], as summarized in the next section, instead. For our applications this means that all Coulomb field effects can be included while avoiding the long computation time of the P^3M method.

After describing the simulation methods in more detail, we calculate the electron mobility in silicon using five different methods in the examples section and compare the results to experimental values.

II. SIMULATION METHOD

In this section we discuss the three main simulation methods and the EMC simulator.

The PPPM Method

P^3M methods work by splitting the potential into two parts: the short range interactions, which are obtained by direct evaluation, and the long range interactions, which are handled by distributing the sources onto a grid and solving the Poisson equation on this grid. For a given particle the potential is evaluated by summing the short range and long range part and subtracting the influence of nearby particles in the long range part in order to prevent taking nearby particles into account twice.

P^3M methods are mainly used in astrophysical simulations and quantum chemistry. However, when they are applied to simulations of semiconductor devices, there are some drawbacks. Firstly DFT cannot be used to solve the Poisson equation since the boundary conditions are non-periodic, whereas periodic boundary conditions are commonly used for astrophysical problems.

Secondly the inverse of the Laplace operator must be calculated for each particle in order to prevent counting nearby particles twice which is, of course, very time consuming. In order to alleviate this problem the inverse can be precomputed (which takes several

hours) and stored (which requires lots of memory), but this must be repeated for every new device structure.

Thirdly the separation of the Coulomb interaction in a short and a long range component is inaccurate and the long range part may depend on the mesh size. Non-uniformity in the source distributions can cause degradation of performance and errors may originate from the numerical differentiation used to calculate the field from the mesh potential depending on the scheme to couple a particle to the mesh. Secondly the k-space treatment of the short range part is inaccurate, since it is necessary to know the distribution function to accurately describe the screening. The evaluation of the distribution function is, in turn, rather CPU intensive and noisy because of the poor statistics usually found in realistic device structures.

Although real space treatments eliminate the problem of double counting the force, a drawback is that the Poisson equation must be solved repeatedly to properly describe the self-consistent fields which consumes over 80% of the total time spent in 3D simulations.

The Corrected Coulomb Approach

The corrected Coulomb approach [2, 3] is a purely numerical scheme that generates a corrected Coulomb force look-up table for the individual electron-electron and electron-ion interaction terms. To calculate the proper short range force one defines a 3D box with uniform mesh spacing in each direction. A single (fixed) electron is then placed at a known position within the 3D domain, while a second (target) electron is swept along the domain in small increments so that it passes through the fixed electron.

The 3D box is usually made sufficiently large so that the boundary conditions do not influence the potential solution. The electron charges are assigned to the nodes using one of the charge assignment schemes. A 3D Poisson equation solver is then used to solve for the node or mesh potentials. At self-consistency the force on the swept electron $F = F_{\text{mesh}}$ is interpolated from the mesh or node potential.

In a separate experiment the Coulomb force $F_{\text{tot}} = F_{\text{coul}}$ is calculated using the standard Coulomb law. For each electron separation one then tabulates F_{mesh} ,F_{coul} and the difference between the two $F_{sr} = F_{\text{coul}} - F_{\text{mesh}}$, which is called the corrected Coulomb force or the short range force. The latter is stored in a separate look-up table.

It is clear that the mesh force and the Coulomb force are identical when the two electrons are separated by several mesh points. Therefore adding the two forces in this region would result in double counting of the force. Within three to five mesh points F_{mesh} starts to deviate from F_{coul}. When the electrons are within the same mesh cell, the mesh force approaches zero due to the smoothing of the electron charge when divided amongst the nearest node points. The generated look-up table for F_{sr} also provides important information concerning the determination of the minimum cut-off range based upon the point where F_{coul} and F_{mesh} begin to intersect, i.e., where F_{sr} goes to zero.

The FMM Approach

The FMM is based on the idea of condensing the information of the potential generated by point sources in several truncated series expansions. After calculating suitable expansions,

the long range part of the potential is obtained by evaluating the truncated series at the point in question and the short range part is calculated by direct summation. Its computational effort is only $O(n)$ where n is the number of particles, instead of proportional to $\binom{n}{2}$ for direct summation.

For 3D simulations we used the standard Coulomb potential. Note that in two dimensions the Coulomb potential at a point $x \in \mathbb{C}$ produced by a point electric charge q_0 at $x_0 \in \mathbb{C}$ is given by

$$\Phi(x) = -\frac{q_0}{2\pi\varepsilon}\mathrm{Re}\big(\log(x - x_0)\big).$$

This two-dimensional potential is obtained by postulating that (1) the Coulomb force is a long range force and that (2) Gauss' law holds. The approximation to the field can be obtained by numeric differentiation.

Using the PM (particle-mesh) approach, the electric field due to the dopants and the applied boundary biases was obtained at the start of the simulation by solving the Poisson equation. Hence the total field acting on each electron is the sum of this constant field and the contribution from the electron-electron interaction handled by the FMM calculations.

The 3D Ensemble Monte Carlo Simulator

We close this section with a short description of our 3D EMC device simulator. The Monte Carlo transport kernel is based on 3D electrons distributed amongst three equivalent valleys with energy dependent effective masses. In the k-space portion of the MC transport kernel we have included acoustic and intervalley (g- and f-phonons) scattering processes. Surface roughness is included as a real-space scattering process. We assume 50% to be specular and 50% to be diffusive reflections from an ideal atomically flat interface.

III. RESULTS

As the first example we simulate the electron mobility using five different approaches. This includes three-dimensional simulations by the 3D EMC simulator and the FMM method as described in the previous section. The second application are two-dimensional simulations of a 25 nm n-channel MOSFET.

Resistor

To check the physical validity and speed of the FMM approach, we first simulated an n^+–n–n^+ resistor using a 3D EMC device simulator [3, 9]. The theoretically calculated results for the low-field electron mobility using the P^3M, the corrected Coulomb, the FMM, the PM (particle-mesh), and the Brooks–Herring k-space approach (bulk MC) are compared with available experimental data [10] in Figure 1. It is found that the low-field mobility calculated with the FMM approach agrees very well with the experimental values for both low and high doping concentrations. The k-space Brooks–Herring approach significantly overestimates the low-field electron mobility for high substrate doping densities.

The corresponding simulation times are shown in Table 1. It is found that for higher doping concentrations the FMM result agrees best with experimental data and significantly better than the previous approaches.

Method	Average time per iteration	Total Time (400 iterations)
P^3M	35 s	233 min
FMM	17 s	113 min

Table 1: The simulation times on a 600 MHz Pentium III computer. The total number of charges is $\approx 22\,000$, and the MC mesh size is $50 \cdot 20 \cdot 20$.

Device simulation: 25nm MOSFET

The device under investigation is a 25 nm n-channel MOSFET with an oxide thickness of 1.2 nm, source and drain region doping of $10^{19}\,\mathrm{cm}^{-3}$ and channel doping of $5 \cdot 10^{19}\,\mathrm{cm}^{-3}$. The electric field profile along the depth of the device due to impurities and the applied boundary bias at the ohmic contacts and at the gate contact is shown in Figure 4. One can clearly see the extensions of the depletion regions and of the channel region underneath the gate oxide.

The output characteristics of the device for a gate voltage of 1.5 V are shown in Figure 5. We see that the use of the FMM approach gives rise to slightly degraded device transconductance at low drain voltages, but at high drain voltages, when there is a large number of electrons in the channel region of the device, the FMM and the PM approach lead to almost the same amount of current but different output conductance.

The electron distributions in the device obtained by the PM and the FMM approach are shown in Figure 6. We see that the inclusion of the short range interactions leads to more scattered electrons. We also have to note that the artificially large force that results when two electrons are very close to each other was eliminated by introducing a proper cut-off range. Finally we show the energy and the velocity characteristics along the channel region in Figures 7 and 8.

There are several noteworthy features that can be observed when analyzing this data. First, the inclusion of the electron-electron interactions leads to fast thermalization of the carriers at the drain end of the device. Second, the average electron energy and the average electron velocity increase when using the FMM approach, i.e., when including the short range electron-electron interactions. In summary the results presented in this work suggest that electron-electron interactions, namely both the short and the long range components, must be properly included when modeling nanoscale MOSFETs.

IV. CONCLUSION

Since FMM implementations are independent of any grids possibly used during the MC part of the simulation, it can be included into existing MC device simulation codes in a straightforward manner. In summary simulation times are significantly reduced and the simulation results agree better with experimental data than previously used methods while accounting rigorously for short range and long range electron-electron and electron-ion interactions.

The careful treatment of electron-ion and electron-ion interactions is a premise for the realistic simulation of nanoscale MOSFETs. Unintentional doping leads to considerable

fluctuation in the device parameters. Furthermore the electron-electron interactions affect the thermalization of the carriers at the drain end of the device. Impurities near the middle portion of the source end of the channel have most significant impact on the device drive current. Regarding the electron-electron interactions it was found that they affect the carrier velocity near the drain end of the channel.

Simulation times are an important issue for MC simulations. The simulation of the characteristics of a device is especially tedious considering that previously one simulation took about a day. The speed-up observed in the FMM simulations will facilitate the exploration of future devices and new device structures.

ACKNOWLEDGMENT

The first author acknowledges support by the Austrian Science Fund (Fonds zur Förderung der wissenschaftlichen Forschung, FWF) via an Erwin Schrödinger Fellowship. The authors also acknowledge support from NSF under contract No. ECS 021-8008.

REFERENCES

[1] A. Sangani and G. Mo, "An $O(N)$ algorithm for Stokes and Laplace interactions of particles," *Phys. Fluids*, vol. 8, pp. 1990–2010, Aug. 1996.

[2] W. Gross, D. Vasileska, and D. Ferry, "A novel approach for introducing the electron-electron and electron-impurity interactions in particle-based simulations," *IEEE Electron Device Lett.*, vol. 20, pp. 463–465, Sept. 1999.

[3] W. Gross, D. Vasileska, and D. Ferry, "Three-dimensional simulations of ultra-small MOSFETs with real-space treatment of the electron-electron and electron-ion interactions," *VLSI Design*, vol. 10, pp. 437–452, 2000.

[4] L. Greengard and V. Rokhlin, "A fast algorithm for particle simulations," *J. Comput. Phys.*, vol. 135, no. 2, pp. 280–292, 1997.

[5] R. Beatson and L. Greengard, "A short course on fast multipole methods," in *Wavelets, Multilevel Methods and Elliptic PDEs (Leicester, 1996)*, Numer. Math. Sci. Comput., pp. 1–37, New York: Oxford Univ. Press, 1997.

[6] H. Cheng, L. Greengard, and V. Rokhlin, "A fast adaptive multipole algorithm in three dimensions," *J. Comput. Phys.*, vol. 155, no. 2, pp. 468–498, 1999.

[7] MadMax Optics, Hamden, CT, USA, *FMMPART3D user's guide*, version 1.0 ed.

[8] C. Wordelman and U. Ravaioli, "Integration of a particle-particle-particle-mesh algorithm with the ensemble monte carlo method for the simulation of ultra-small semiconductor devices," *IEEE Trans. Electron Devices*, vol. 47, pp. 410–416, Feb. 2000.

[9] D. Vasileska and S. Goodnick, "Computational electronics," *Materials Science and Engineering Reports*, vol. R38, no. 5, pp. 181–236, 2002.

[10] C. Canali, G. Ottaviani, and A. Alberigi-Quaranta, "Drift velocity of electrons and holes and associated anisotropic effects in silicon," *J. Phys. Chem. Solids*, vol. 32, p. 1707, 1971.

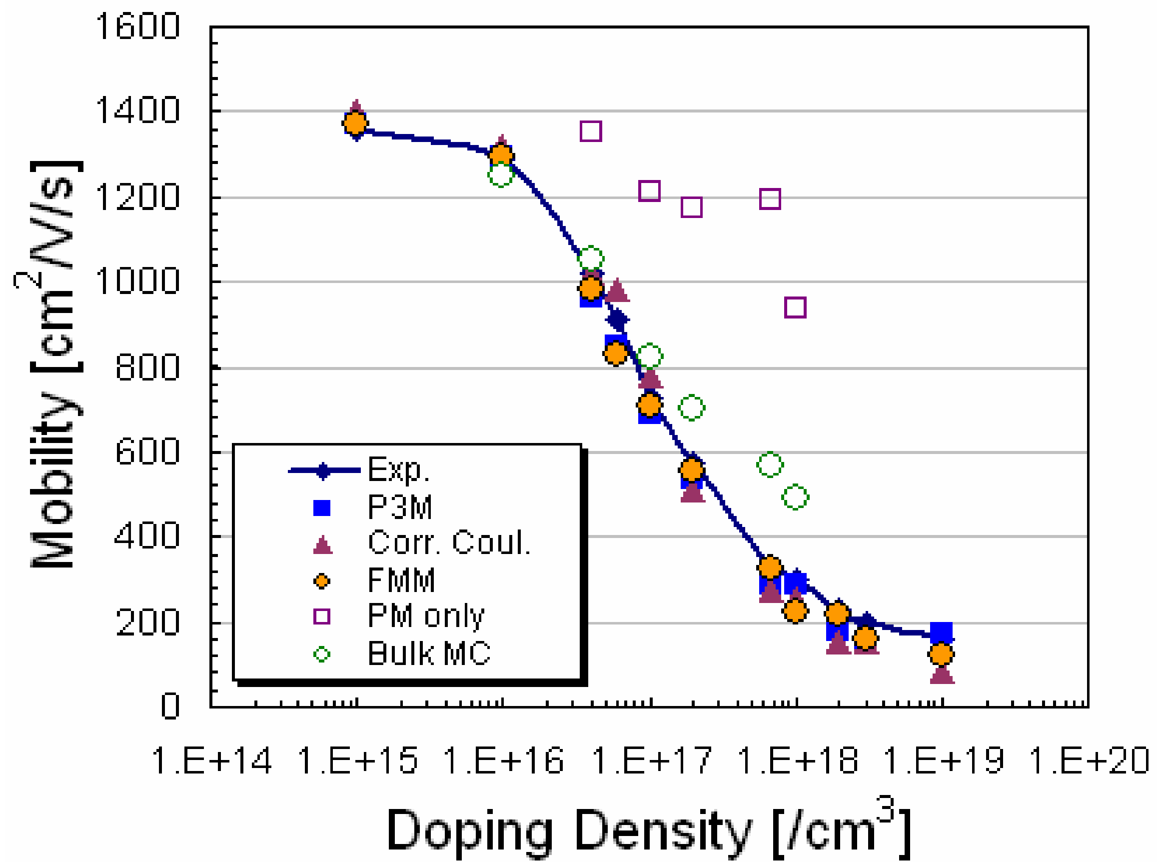


Figure 1: The electron mobility for five doping concentrations (from $10^{15}\,\text{cm}^{-3}$ to $10^{19}\,\text{cm}^{-3}$) as obtained experimentally and by five simulation methods.

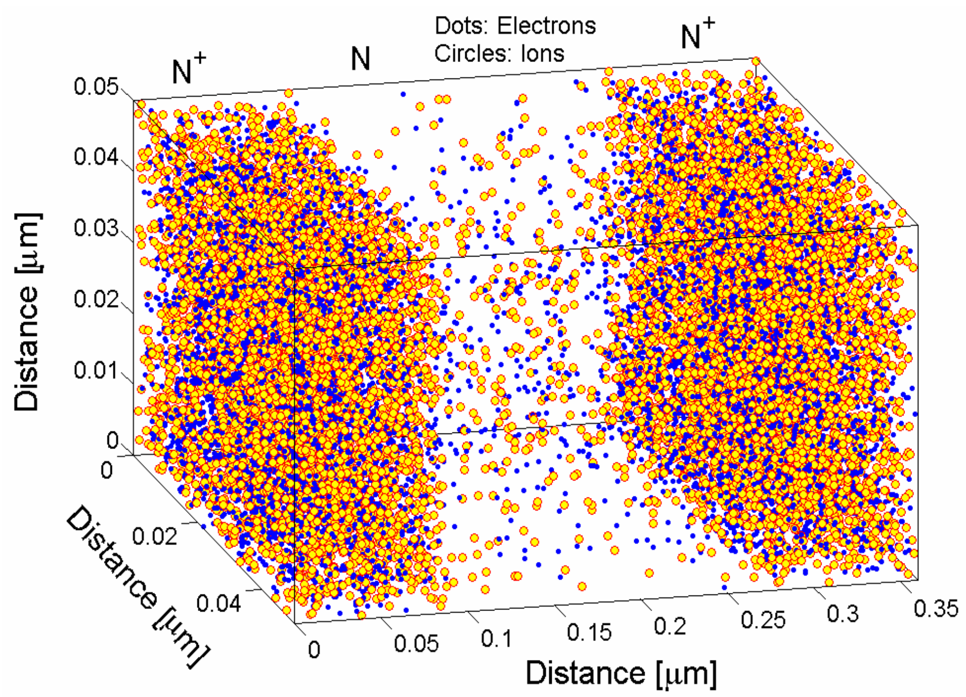


Figure 2: The electron distribution within the resistor ($360\,\text{nm} \cdot 50\,\text{nm} \cdot 50\,\text{nm}$).

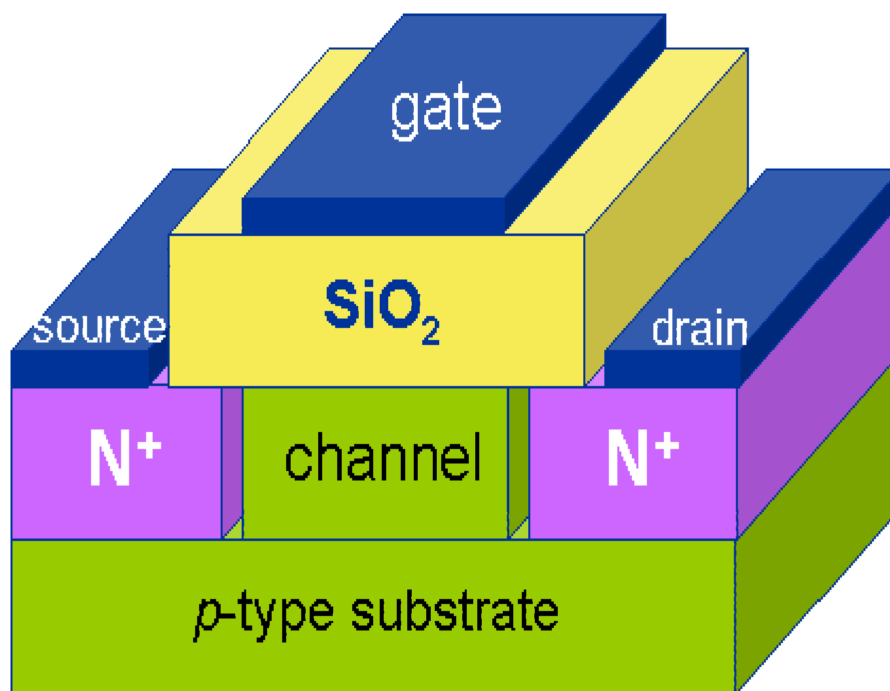


Figure 3: The simulated device has an oxide thickness of 1.0 nm, a channel length of 15 nm, a channel width of 0.5 μm, a source-drain junction depth of 15 nm, and the substrate doping is $N_A = 3 \cdot 10^{19}\,\mathrm{cm}^{-3}$ and the doping of the source-drain regions is $N_D = 10^{19}\,\mathrm{cm}^{-3}$.

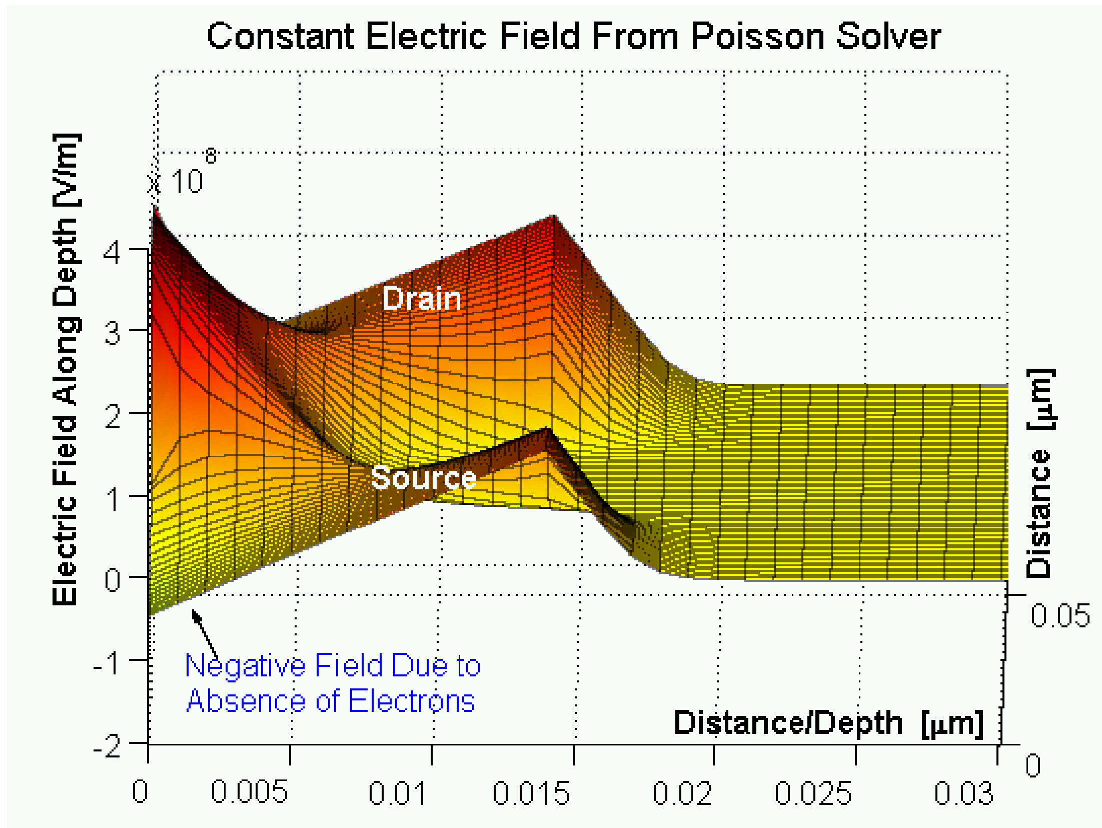


Figure 4: The electric field along the depth of the device due to the impurities and the applied boundary biases as calculated by a two-dimensional Poisson solver.

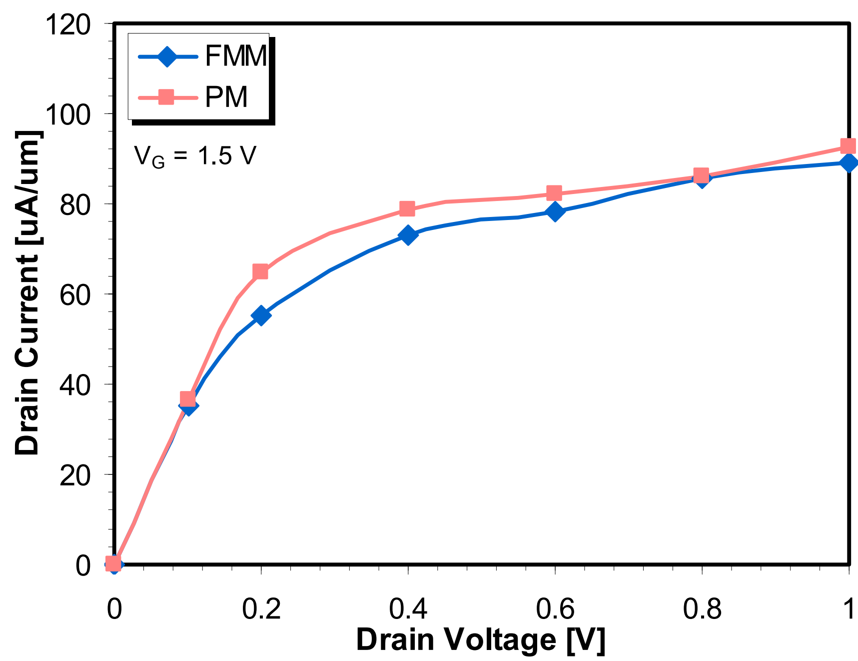
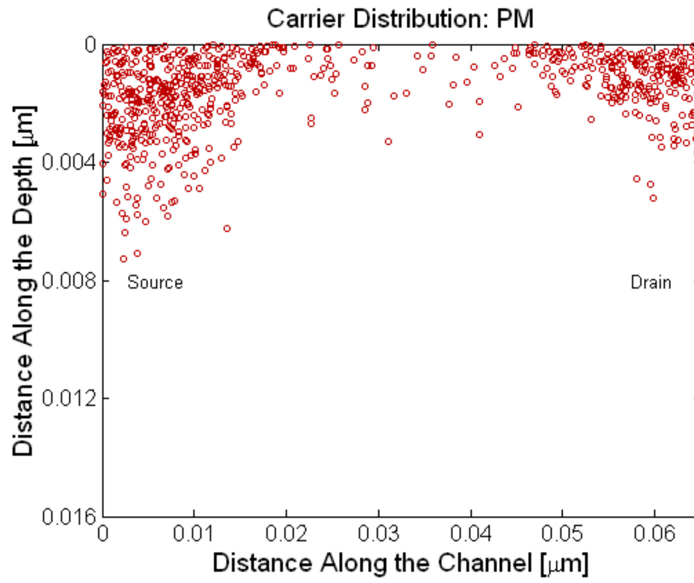
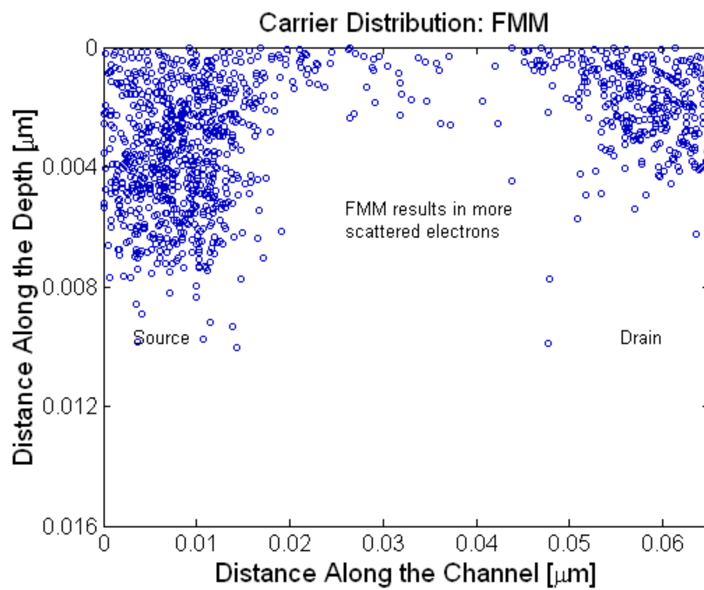


Figure 5: The output characteristics of the device for $V_G = 1.5\,\mathrm{V}$.



(a) Electron distribution (PM).



(b) Electron distribution (FMM).

Figure 6: The electron distribution along the channel as obtained **(a)** by the PM method and **(b)** using the FMM approach.

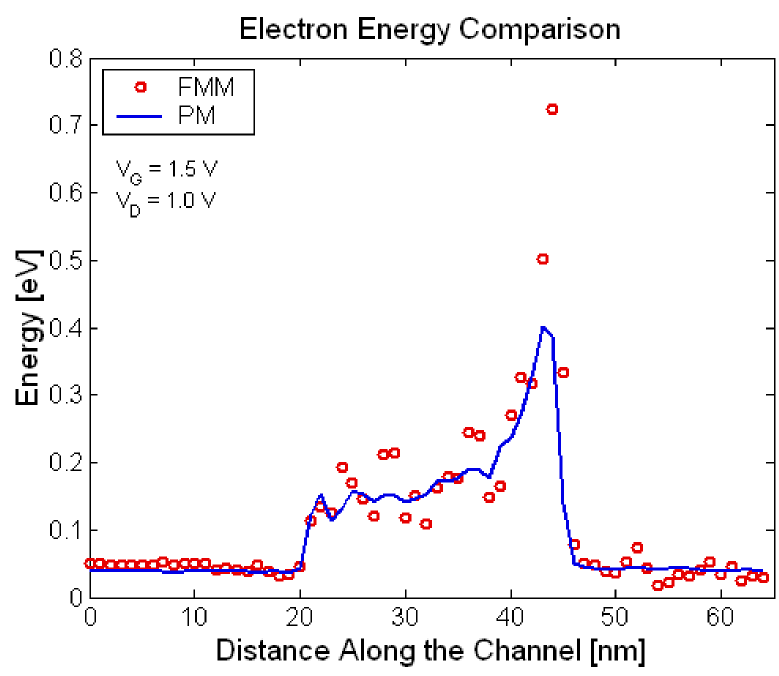


Figure 7: The average electron energies along the channel for $V_G = 1.5\,\mathrm{V}$ and $V_D = 1.0\,\mathrm{V}$.

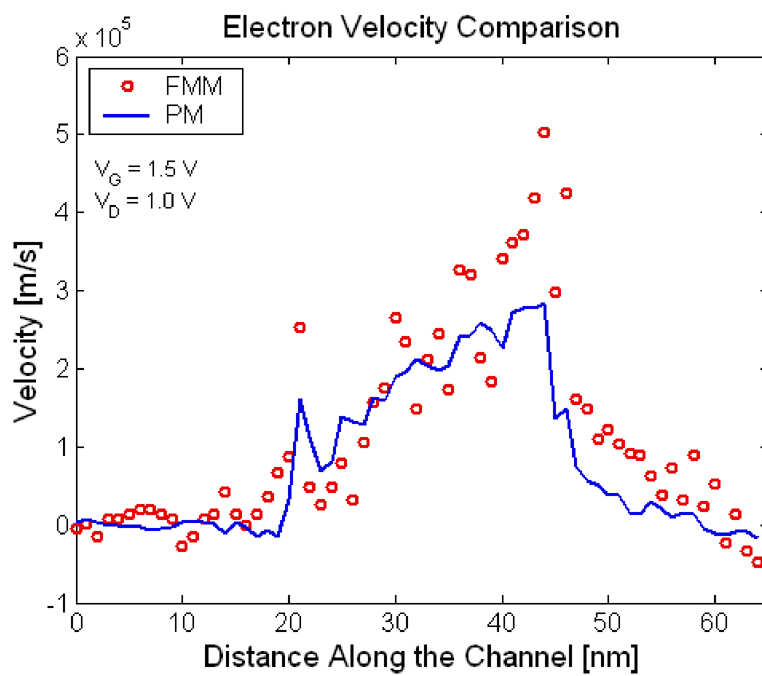


Figure 8: The average electron velocities along the channel for $V_G = 1.5\,\mathrm{V}$ and $V_D = 1.0\,\mathrm{V}$.