Transient Simulation of Semiconductor Devices Using a Deterministic Boltzmann Equation Solver

SUNG-MIN HONG (Member, IEEE), AND JAE-HYUNG JANG (Member, IEEE)

School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju 61005, South Korea

CORRESPONDING AUTHOR: S.-M. HONG (e-mail: smhong@gist.ac.kr)

ABSTRACT In this paper, the transient simulation of semiconductor devices using a deterministic Boltzmann equation solver is presented. Transient simulation capability is implemented in a deterministic Boltzmann equation solver for the 3-D momentum space based on the spherical harmonics expansion. The numerical simulation results with implicit time marching methods demonstrate that the transient simulation using a deterministic Boltzmann equation solver can be performed. The impact of the quasi-static approximation for the current density, which is widely adopted in the momentum-based equations, is tested for various devices such as homogeneous samples, an N^+NN^+ structure and an MOSFET.

INDEX TERMS Semiconductor device modeling, transient simulation, Boltzmann transport equation, deterministic Boltzmann equation solver, spherical harmonics expansion.

I. INTRODUCTION

Deterministic Bolzmann equation solvers [1]–[7], which solve the semiclassical Boltzmann transport equation for carrier transport in a deterministic manner without relying on the Monte Carlo method, recently have gained popularity in the modeling and simulation society. Among various variations of these solvers, for three-dimensional electron gases, the spherical harmonics expansion method is frequently used.

Starting from the steady-state simulation using the Gummel iteration method [1], the full Newton-Raphson method has been introduced [5], [7]. Introduction of the full Newton-Raphson method enables small-signal and noise analyses [5], [7]. An electrothermal simulation [8] can be easily performed, since all variables are always evaluated. It can be used as a valuable tool to simulate the hot-carrier-related problems such as the oxide reliability [9], [10]. It is expected that deterministic Boltzmann equation solvers will play an increasingly important role in the semiconductor device simulation.

In spite of the recent progress introduced above, a fundamental simulation capability – transient simulation with an implicit time marching scheme – is still lacking [11]. It is in contrast to the Monte Carlo method. The Monte Carlo method is inherently transient, therefore, transient responses of the simulated system can be easily obtained. For example, the transient simulation via full-band Monte Carlo

simulations for bipolar transistors, inverters, and ring oscillators can be found in [12] and [13]. However, in the deterministic Boltzmann equation solvers, additional effort for the transient simulation is required, because the time derivative term should be separately implemented.

Although some reports on transient simulation results can be found in [14] and [15], those works adopt the explicit time marching scheme, which is very different from the implicit time marching scheme used in the conventional device simulators. Since the maximum time step for the explicit time marching technique is restricted due to the stability issue, it can be applied only to the very short time duration. Therefore, transient simulation capability using an implicit time marching technique is desirable for practical purposes.

One application area of particular interest is the simulation of plasma modes in the gated electron gas [16]. Recently, the plasma instability has been simulated by solving the balance equations numerically [17]. In [18], it is reported that the higher-order transport model (for example, the Boltzmann transport equation) is more desirable. Therefore, it is very interesting to check whether the plasma effects can be obtained by solving the Boltzmann transport equation directly.

In this work, transient simulation results using the deterministic Boltzmann equation solver are reported.

The transient Boltzmann equation solver is applied to the simulation of the plasma effects. The organization of the paper is as follows. In Section II, our simulation framework is briefly introduced. In Section III, numerical results for homogeneous samples and an N^+NN^+ structure are presented. In Section IV, the plasma effects in a MOSFET are tested. Our conclusions are presented in Section V.

II. SIMULATION FRAMEWORK
A. STABILIZATION SCHEME

A deterministic Boltzmann equation solver based on the spherical harmonics expansion for the three-dimensional momentum space has been newly implemented into our in-house device simulation framework, G-Device [17]. Three non-parabolic valleys in the conduction band of silicon are considered. The phonon scattering mechanisms and their parameters are taken from [6]. The isotropic-elastic approximation for the impurity scattering is adopted with the Brooks-Herring model. Injection boundary conditions [5], [6] are used for ohmic contacts. Band gap narrowing and the Pauli principle are not considered.

In the spherical harmonics expansion, the distribution function is expanded up to a predefined maximum order, l_{max}. When l_{max} is too low, considerable errors can be obtained. For example, it has been demonstrated that the first-order expansion is not sufficient for small devices [3], [19]. Since an appropriate value of l_{max} depends on the device structure and the bias point, it is important to check the convergence of the simulated quantities carefully [20]. In this work, the convergence of the simulated quantities (such as terminal currents and voltages) is verified by simulations with various l_{max} values.

The strong electric field inside the device enforces introduction of a special stabilization scheme, such as the H-transformation [1] or the maximum entropy dissipation scheme [3], [21]. In particular, by virtue of its construction, the H-transformation requires interpolation of the previous solution variables under the present band profile, which inevitably results in interpolation errors. Therefore, the transient simulation capability that is compatible with the H-transformation is not currently available. In this work, following the suggestion in [11], the kinetic-energy-based scheme [3] instead of the H-transformation has been adopted. Developing a new stabilization scheme, which allows the transient simulation while taking advantages of the H-transformation, would be a formidable task. Progress on this issue will be reported elsewhere.

B. TRANSIENT CAPABILITY

With adoption of the kinetic-energy-based stabilization scheme, implementing the transient capability is straightforward.

When the entire Boltzmann equation is projected, the resultant Boltzmann equation reads [5], [6]

$$\sum_{l'} \frac{\partial}{\partial t}\left[Z_{l,l'}f_{l'}\right] + \frac{\partial}{\partial \varepsilon}\left[\mathbf{F} \cdot \mathbf{A}_{l,l'}f_{l'}\right] + \nabla_{\mathbf{r}} \cdot \left[\mathbf{A}_{l,l'}f_{l'}\right]$$
$$- B_{l,l'}f_{l'} = \hat{S}_l^{in} - \hat{S}_l^{out} \tag{1}$$

where l and l' represent the index pairs for the spherical harmonics expansion, t is the time, $Z_{l,l'}$ is the density-of-states, f_l is the projected distribution function, ε is the kinetic energy, $\mathbf{F}$ is the force, $\mathbf{A}_{l,l'}$ and $B_{l,l'}$ are transport parameters [6], and $\hat{S}_l^{in}$ and $\hat{S}_l^{out}$ are scattering integrals. Since the discretization of all other terms except for the time derivation has been investigated extensively in previous works, the discretization of the time derivative term, $\frac{\partial}{\partial t}[Z_{l,l'}f_{l'}]$, is the remaining task.

An implicit time marching scheme is adopted to discretize the time derivative term. In an implicit time marching scheme, the time derivative of $Z_{l,l'}f_{l'}$ at a time instance t_i can be written as

$$\frac{\partial[Z_{l,l'}f_{l'}]}{\partial t}\bigg|_{t=t_i} \approx Z_{l,l'}(\mathbf{r}, \varepsilon) \sum_{j=0}^{N-1} a_{i,j}f_{l'}(\mathbf{r}, \varepsilon, t_{i-j}), \tag{2}$$

where N is the order of the time-marching scheme and $a_{i,j}$ is a coefficient connecting t_i and t_{i-j}. Note that the density-of-states at given $(\mathbf{r}, \varepsilon)$ is time-invariant in the kinetic-energy-based scheme.

It has been found that the first-order method (the backward Euler method) introduces excessively strong over-damping even with a fairly fine time step. Therefore, two second-order methods are applied. When the periodic input signal is applied, the modified second-order method based on trigonometric functions [22], [23], which is efficient for the periodic case, is used. In this scheme, with a constant time step of Δt, the coefficients are given by

$$a_{i,0} = \frac{1}{\Delta t}\frac{z\cos(2z) - z\cos(z)}{\sin(2z) - 2\sin(z)}, \tag{3}$$

$$a_{i,1} = \frac{1}{\Delta t}\frac{z\sin(z)}{\cos(z) - 1}, \tag{4}$$

$$a_{i,2} = \frac{1}{\Delta t}\frac{z}{2\sin(z)}, \tag{5}$$

where $z = \frac{2\pi}{n_T}$ and n_T is the number of time steps per period. On the other hand, when the input signal is not periodic, following the second-order backward differentiation formula, the coefficients are given by

$$a_{i,0} = \frac{1.5}{\Delta t}, \quad a_{i,1} = -\frac{2}{\Delta t}, \quad a_{i,2} = \frac{0.5}{\Delta t}. \tag{6}$$

III. NUMERICAL RESULTS

In this section, electrical properties of homogeneous samples and an N^+NN^+ structure are simulated by using the transient Boltzmann transport equation solver. For some cases, the results of the drift-diffusion (DD) model are also shown for comparison. The standard DD model neglects the time

derivative of the current density, $\frac{\partial}{\partial t}\mathbf{J}_n$ [24]. The current density is given by

$$\mathbf{J}_n = q\mu_n n\mathbf{E} + qD_n\nabla_{\mathbf{r}}n, \tag{7}$$

where q is the absolute elementary charge, μ_n is the electron mobility, n is the electron density, $\mathbf{E}$ is the electric field, and D_n is the diffusivity. In this work, the standard DD model is denoted as the "QSJ" (quasi-static current density) DD model. It is noted that the QSJ DD model still considers the time derivative of the electron density by solving the continuity equation. On the other hand, the DD model which does not neglect $\frac{\partial}{\partial t}\mathbf{J}_n$ [23] is called the "full" DD model. The equation for the current density reads

$$\mathbf{J}_n + \tau\frac{\partial}{\partial t}\mathbf{J}_n = q\mu_n n\mathbf{E} + qD_n\nabla_{\mathbf{r}}n, \tag{8}$$

where τ is the momentum relaxation time.

Throughout this work, a uniform energy spacing of 10 meV is adopted. This value is found to be small enough from our numerical experience.

A. HOMOGENEOUS SAMPLES AT EQUILIBRIUM

The donor density is 2×10^{17} cm^{-3}. In the DD model, the mobility (μ_n) of 518 cm^2 V^{-1} sec^{-1} and the momentum relaxation time (τ) of 86 fsec are used. Since the DC bias voltage is zero, the DC electric field inside the sample vanishes. A small excitation voltage is applied across the sample. After the initial transient effect is diminished, the current is recorded. By taking the Fourier coefficients of the current, the admittance is calculated.

In this example, an analytic solution for the admittance is readily available by solving the DD model. In conjunction with the equation for the current density, the Poisson equation and the continuity equation should be solved. At equilibrium, the linearized equations at the angular frequency ω are given by

$$-\epsilon\frac{\partial^2}{\partial x^2}\delta\phi = -q\delta n, \tag{9}$$

$$j\omega q\delta n = \frac{\partial}{\partial x}\delta J_n, \tag{10}$$

where ϵ is the permittivity, x is the position, $\delta\phi(x)$ is the AC component of the electrostatic potential, $\delta n(x)$ is the AC component of the electron density, j is the imaginary unit, $\delta J_n(x)$ is the AC component of the current density. In the case of the full DD model, all those equations are combined into

$$\frac{\partial^2}{\partial x^2}\delta n = \left[\frac{qn}{\epsilon V_T} + \frac{j\omega(1+j\omega\tau)}{D_n}\right]\delta n, \tag{11}$$

where V_T is the thermal voltage and the Einstein relation is employed. It is noted that the term $\omega\tau$ should be neglected in the QSJ DD model. By integrating the total current density (which is a sum of δJ_n and the displacement current
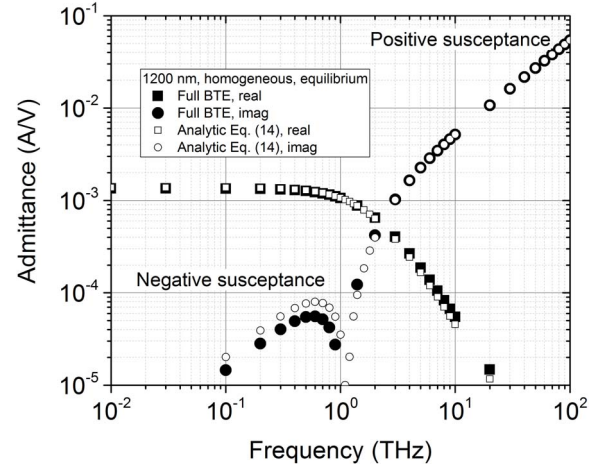

FIGURE 1. Admittance of a homogeneous sample at equilibrium. The sample length is 1200 nm. It is obtained by solving the Boltzmann transport equation (the full BTE model). The results of an analytic equation, (14), are shown for comparison.

density) over the sample, the AC current, δI, is readily obtained as

$$\frac{\delta I}{A} = \frac{1}{1+j\omega\tau}q\mu_n\left[n\frac{\delta V}{L} - V_T\frac{\delta n(L)-\delta n(0)}{L}\right] + j\omega\epsilon\frac{\delta V}{L}, \tag{12}$$

where A is the device area (1.0 μm^2 in this example) and δV is the AC voltage.

When the ideal ohmic boundary condition is considered,

$$\delta n(0) = 0, \quad \delta n(L) = 0, \tag{13}$$

only the trivial solution of (11) exists. The admittance is then analytically given as

$$Y(\omega) = \lim_{\delta V\to 0}\frac{\delta I}{\delta V} = \frac{1}{1+j\omega\tau}q\mu_n n\frac{A}{L} + j\omega\epsilon\frac{A}{L}. \tag{14}$$

At high frequencies, the capacitive contribution from the second term dominates. At low frequencies, in the case of the full DD model, the susceptance (the imaginary part of the admittance) may or may not become negative, depending on the sign of $\tau q\mu_n n - \epsilon$. The negative susceptance is originated from the time derivative of the current density. For a relatively long sample, a reasonable agreement between the results from the Boltzmann transport equation and (14) is achieved as shown in Fig. 1. The sample length is 1200 nm.

However, for a relatively short sample, the boundary condition has more significant impact on the simulation results. Fig. 2 shows the admittance of a 120-nm-long sample at equilibrium. The result from the Boltzmann transport equation shows that the simple analytic expression with the ideal boundary condition, (14), fails to predict the admittance even at low frequencies. In order to see the effect of the boundary condition, the admittance is also calculated with the equilibrium distribution applied to ohmic contacts (empty symbols in the figure). It corresponds to the ideal ohmic boundary
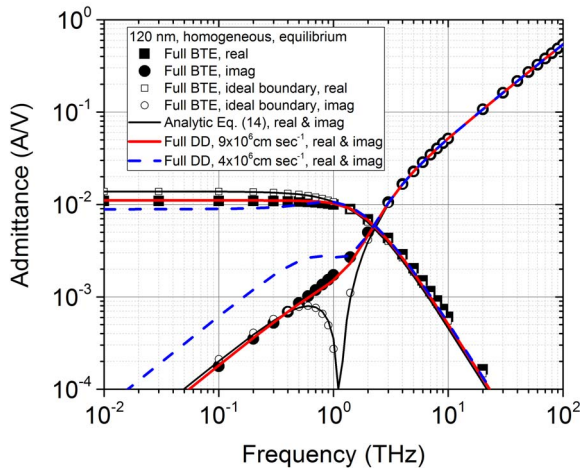
FIGURE 2. Admittance of a homogeneous sample at equilibrium. The sample length is 120 nm. It is obtained by solving the Boltzmann transport equation (the full BTE model). Impact of the ideal ohmic boundary condition on the admittance is clearly shown. The full drift-diffusion (DD) model results with two different recombination velocities are compared.

condition, (13). As shown in Fig. 2, the admittance with the ideal ohmic boundary condition agrees well with (14).

Using the DD model, the reason of discrepancy can be explained as follows. With the recombination velocity of contacts, v_{rec}, more realistic boundary conditions for two contacts are written as

$$\delta J_n(0) = q\delta n(0)v_{rec}, \qquad (15)$$

$$\delta J_n(L) = -q\delta n(L)v_{rec}. \qquad (16)$$

Then, a non-trivial solution, δn, of (11) can be found. Therefore, because of non-vanishing $\delta n(L) - \delta n(0)$, the admittance can be significantly deviated from (14). In Fig. 2, the full DD results with two different values of v_{rec} are also shown. With a certain value of the recombination velocity (9×10^6 cm sec^{-1} in this example), excellent agreement with the full BTE model is obtained. When the recombination velocity is much lower than the optimal value (for example, 4×10^6 cm sec^{-1} in the figure), the low-frequency conductance is reduced and the low-frequency susceptance is overestimated. This reveals that not only the correct transport equation but also the correct boundary condition are critically important to gain the correct admittance.

The time derivative of the current density is usually neglected in the conventional device simulation. In order to examine the impact of the time derivative term, two different simulation sets are generated. In addition to the simulation with correct implementation of the time derivative term (the "full" BTE model), an approximated implementation (the "QSJ" BTE model) is tried. In the QSJ BTE model, the time derivative term is implemented only for the zeroth-order distribution function, while it is completely ignored for the higher-order distribution functions. The results shown in Fig. 3 show that the approximation yields significant errors in the admittance. Even in terms of the overall magnitude,
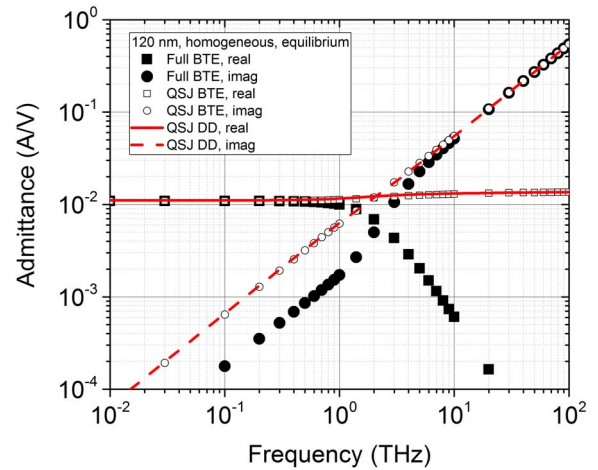

FIGURE 3. Admittance of a homogeneous sample at equilibrium. In the QSJ BTE model, the time derivative of the distribution function is considered only at the zeroth-order.
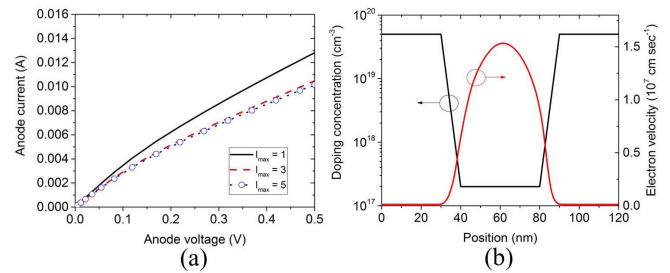

FIGURE 4. (a) DC IV curve of a 120-nm-long structure. l_{max} = 1, 3, and 5. (b) Doping profile and electron velocity as functions of the position. The anode voltage is 0.5 V.

a significant error is observed at frequencies of around two or three THz.

B. N$^+$NN$^+$ STRUCTURE
An N$^+$NN$^+$ structure whose length is 120 nm is simulated. The device area is 1.0 μm^2. The 40-nm-long lowly doped region is surrounded by two highly doped regions. The DC IV curve is shown in Fig. 4(a). As the maximum order of the spherical harmonics expansion, l_{max}, increases, the anode current is converged. In Fig. 4(b), the doping profile and the electron velocity at the anode voltage of 0.5 V are shown as functions of the position. The maximum velocity is much larger than the saturation velocity, due to the strong quasi-ballistic transport. Since the anode current is already converged with $l_{max} = 5$, the transient simulation also has been performed with $l_{max} = 5$.

Small-signal anode currents of the 120-nm-long structure for the first five periods are shown in Fig. 5. The DC voltage is 0.5 V and various frequencies are tested. At low frequencies, the conductive components dominate and the phase of the current is very similar to that of the applied voltage. As the frequency increases (for example at 10 THz), the capacitive contribution becomes significant.
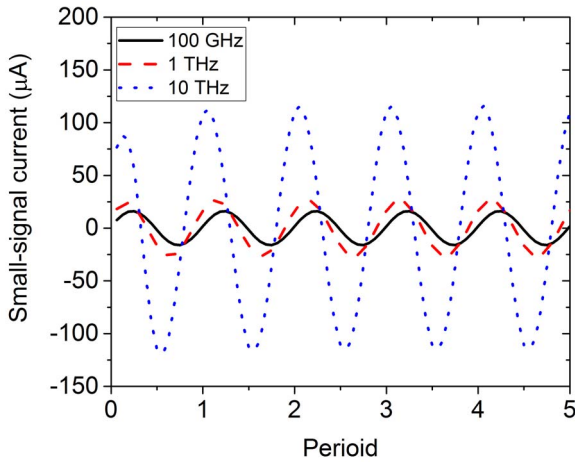
FIGURE 5. Transient simulation results for the anode current. The small-signal response without the DC current is drawn. The amplitude of the input signal is 1 mV. The 120-nm-long structure is considered. The DC voltage is 0.5 V and various frequencies are tested. Time is normalized to the period.
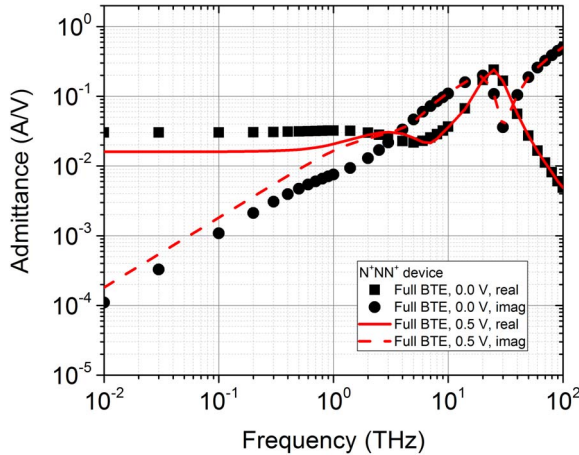

FIGURE 6. Frequency dependence of the small-signal admittance calculated with the full BTE model. Two DC bias voltages of 0.0 V and 0.5 V are considered.

Such a capacitive contribution exhibits the phase difference of 90 degrees.

The small-signal admittance is shown for various excitation frequencies in Fig. 6. Below a few hundred GHz, the conductance (the real part of the admittance) does not change significantly. Reduction of the conductance at 0.5 V originates from the DC IV characteristics shown in Fig. 4. Beyond 5 THz, the admittance does not depend on the bias voltage.

Thus far, the small-signal response has been calculated by the transient simulation. However, the transient Boltzmann transport equation solver can be applied to general large-signal cases. The large-signal response is shown in Fig. 7. The amplitude of the sinusoidal excitation is as large as 0.5 V. The frequency is 3 THz. As shown in the figure, the large-signal excitation with the peak-to-peak amplitude of 1 V can be simulated without numerical problems. This demonstrates
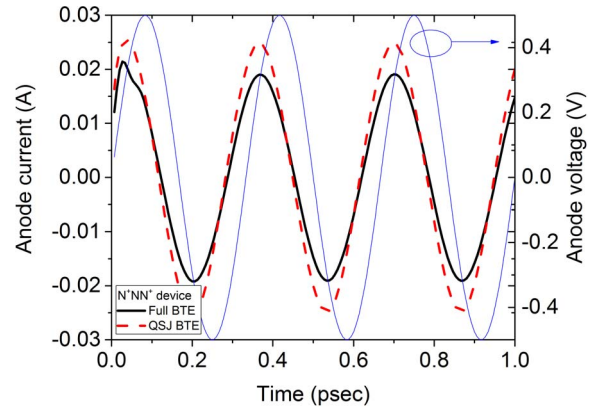

FIGURE 7. Large-signal simulation results of the anode current. The frequency of the voltage excitation is 3 THz. The peak-to-peak amplitude of the input signal is 1 V.

the robustness of the deterministic Boltzmann equation solver. The third order harmonic component (9 THz) of the anode current is about 0.8 % of the fundamental one (3 THz). Higher order harmonics are generated by the nonlinearity of the device, which cannot be captured by the small-signal analysis. Also, comparison with the QSJ simulation results reveals the importance of the time derivative term.

Next, the step response of the N^+NN^+ structure is simulated. The anode bias voltage is increased from 0 V to 0.5 V with a smooth waveform of $0.25 - 0.25 \cos(\frac{\pi t}{t_{ramp}})$ V, where t_{ramp} is the ramping time. Once after the voltage reaches 0.5 V, it does not change. Various ramping speeds have been tested. The duration of the ramping phase is used to normalize the time variable. For slow ramping rates, the results are similar with the quasi-static result as shown in Fig. 8. However, for faster ramping rates, the anode current is heavily deviated from the quasi-static result. For cases with large average ramping rates such as 5 V psec^{-1}, strong overshoot of the anode current is observed. In those cases, the displacement current contributes significantly to the total current as shown in Fig. 8. Even after the ramping period is finished, the anode current takes a considerable amount of time to be relaxed.

The time evolution of the electron distribution function is plotted in Fig. 9. The ramping rate is 5 V psec^{-1} and the elapsed times are $t = 0.1$ psec (just after the ramping period) and 0.2 psec (0.1 psec elapsed after the ramping period). At $t = 0.1$ psec, the electron distribution function (its zeroth-order expansion coefficient) is close to that in the equilibrium, because it takes a finite time to heat the electron gas. On the other hand, at $t = 0.2$ psec, it is clearly shown that high-energy electrons are generated near the anode terminal.

IV. APPLICATION TO PLASMA EFFECTS IN A MOSFET

Using the transient simulation capability, the detector operation of a scaled MOSFET is simulated. It is an ideal tool
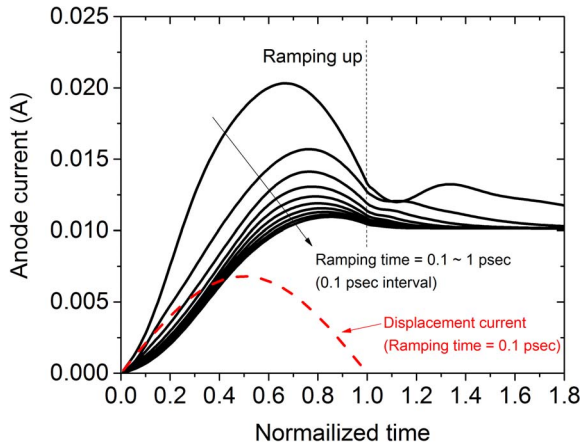
FIGURE 8. Step-response of the anode current. In addition to the total currents (solid lines), the displacement current with the ramping time oaf 0.1 psec (a dashed line) is also shown. The time variable is normalized with the ramping time. Various ramping rates are tested.
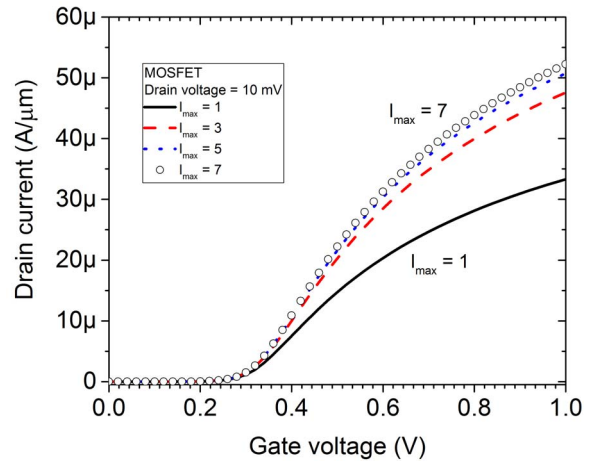

FIGURE 10. Input characteristics of the 60-nm MOSFET at the steady-state condition. The drain voltage is fixed to 0.01 V. The surface-roughness scattering is neglected.
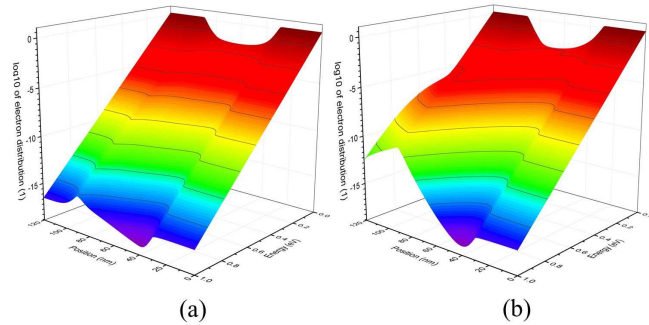

FIGURE 9. Electron distribution function for time points, (a) t = 0.1 psec and (b) t = 0.2 psec. The ramping rate is 5 V psec^{-1}.

to estimate the plasma modes in the two-dimensional electron gas. Previous studies on plasma effects in the scaled devices via full-band Monte Carlo simulations can be found in [25] and [26].

A double-gate MOSFET, whose oxide thickness is 1 nm, is considered. The silicon substrate is 10 nm. Only the upper half of the entire structure has been simulated. The doping profile along the channel direction is identical to that of the N^+NN^+ structure shown in the previous section. The gate length is 60 nm and the gate workfunction is 4.5 eV.

The one-dimensional electron transport is considered, while the two-dimensional Poisson equation is solved together [17]. No quantum confinement effect is included in this simulation. Therefore, the finite thickness of the electron gas [27], [28] is not considered. In order to evaluate the impact of this effect, the multisubband approach [4], [7] would be required, which is an interesting future research topic. The surface-roughness scattering mechanism is neglected. Only the bulk phonon and impurity scattering mechanisms are considered. The input characteristics at the steady-state condition are shown in Fig. 10. The

threshold voltage is about 0.3 V. It is also confirmed that the simulation with $l_{max} = 5$ gives reasonably converged values.

The detector operation [29] is simulated. The linear operation mode (so called "cold FET") has been exclusively considered in the simulation. The source terminal is directly connected to the ground and the drain terminal is electrically open. Therefore, no current is allowed at the drain terminal. However, it does not necessarily mean that the drain voltage also vanishes. Actually, when an input signal is applied to the gate terminal, a non-vanishing drain voltage is induced so that the drain current vanishes. The DC component of the induced drain voltage is often used to quantify the detector response [29].

Once after the steady-state solution is obtained, the input signal, whose voltage amplitude is 1 mV, is applied to the gate terminal. A figure-of-merit, the voltage responsivity, R_V, is defined as

$$R_V = \frac{|V_D|}{P_{AC}}, \qquad (17)$$

where V_D is the induced DC drain voltage and P_{AC} is the dissipated AC power of the MOSFET. The amplitude of the AC gate voltage does not change the responsivity because the responsivity is normalized with P_{AC}. It has been also numerically verified by the simulation results with different amplitudes. Since the two-dimensional cross section is simulated, R_V has to be divided by the width of the device in the third dimension [23].

As in the previous section, the transient simulation has been performed. After the initial transient effect is diminished, the DC drain voltage is extracted. Two simulation sets (the full BTE and QSJ BTE models) are tested. Only when the time derivative of the distribution function is considered (the full BTE model), the plasma mode is obtained by the simulation. In [23], it has been shown that even the small-signal admittance obtained by the full model can significantly differ from the one obtained by the QSJ model.
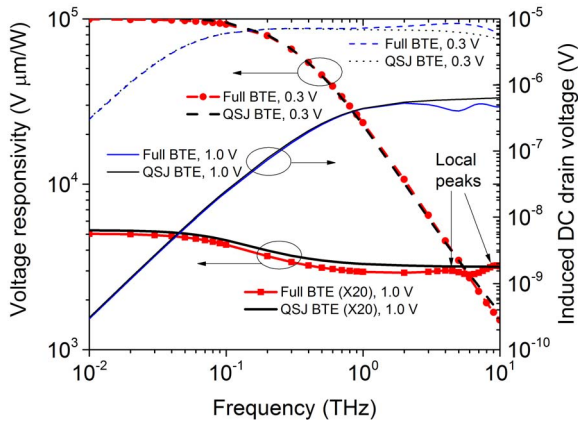
FIGURE 11. Absolute value of the responsivity, R_V, as a function of the frequency. Two gate bias voltages (0.3 V and 1.0 V) are tested. The responsivities at 1.0 V are scaled for clarity. Local peaks are pointed out by arrows. The induced DC drain voltage, $|V_D|$, is also shown in the right y-axis. The amplitude of the AC gate voltage is 1 mV.

The difference of the small-signal response also affects the induced drain voltage. Therefore, the values of R_V obtained from the full and QSJ models may differ. The difference between two models is interpreted as an indicator for the impact of the plasma mode.

The voltage responsivity is shown as a function of the frequency in Fig. 11. Two gate bias voltages are tested. The responsivity shows a maximum value near the threshold voltage (0.3 V). At this bias condition, the two models give almost identical results. Therefore, significant impact of the plasma effect cannot be observed. However, above the threshold voltage (1.0 V), the full BTE results slightly deviate from the QSJ ones. Local peaks are observed at certain frequencies only in the full BTE result. Such differences are attributed to the plasma modes in the channel. It is noted that, depending on the frequency, the full BTE results can be larger or smaller than the QSJ results [23].

It has been found that the voltage responsivity of the 60-nm MOSFET is only slightly affected by the plasma effect. It is expected that the plasma modes in the electron gas will have a larger impact in aggressively scaled devices. Systematic investigation of the detector performance is of great research interest, and will be reported elsewhere.

V. CONCLUSION

In this work, the transient simulation has been performed using a deterministic Boltzmann equation solver. A kinetic-energy-based scheme is adopted with sufficiently dense grids. Implicit time marching methods are used.

It has been demonstrated that the transient simulation using a deterministic Boltzmann equation solver is possible. Examples for homogeneous samples, an N^+NN^+ structure, and a MOSFET have been shown. The importance of the boundary condition is also shown. Furthermore, considering the time derivative of the distribution completely (the full BTE model in this work) is mandatory for accurate calculation in high frequencies.

REFERENCES

[1] A. Gnudi, D. Ventura, G. Baccarani, and F. Odeh, "Two-dimensional MOSFET simulation by means of a multidimensional spherical harmonics expansion of the Boltzmann transport equation," *Solid-State Electron.*, vol. 36, no. 4, pp. 575–581, Apr. 1993, doi: 10.1016/0038-1101(93)90269-V.

[2] K. A. Hennacy, Y.-J. Yu, N. Goldsman, and I. D. Mayergoyz, "Deterministic MOSFET simulation using a generalized spherical harmonic expansion of the Boltzmann equation," *Solid-State Electron.*, vol. 38, no. 8, pp. 1485–1495, Aug. 1995, doi: 10.1016/0038-1101(94)00280-S.

[3] C. Jungemann, A. T. Pham, B. Meinerzhagen, C. Ringhofer, and M. Bollhöfer, "Stable discretization of the Boltzmann equation based on spherical harmonics, box integration, and a maximum entropy dissipation principle," *J. Appl. Phys.*, vol. 100, no. 2, Jul. 2006, Art. no. 024502, doi: 10.1063/1.2212207.

[4] S. Jin, M. V. Fischetti, and T.-W. Tang, "Theoretical study of carrier transport in silicon nanowire transistors based on the multisubband Boltzmann transport equation," *IEEE Trans. Electron Devices*, vol. 55, no. 11, pp. 2886–2897, Nov. 2008, doi: 10.1109/TED.2008.2005172.

[5] S.-M. Hong and C. Jungemann, "A fully coupled scheme for a Boltzmann–Poisson equation solver based on a spherical harmonics expansion," *J. Comput. Electron.*, vol. 8, pp. 225–241, Oct. 2009, doi: 10.1007/s10825-009-0294-y.

[6] S.-M. Hong, A.-T. Pham, and C. Jungemann, *Deterministic Solvers for the Boltzmann Transport Equation*. Vienna, Austria: Springer-Verlag, 2011, doi: 10.1007/978-3-7091-0778-2.

[7] D. Ruić and C. Jungemann, "Numerical aspects of noise simulation in MOSFETs by a Langevin–Boltzmann solver," *J. Comput. Electron.*, vol. 14, no. 1, pp. 21–36, Mar. 2015, doi: 10.1007/s10825-014-0642-4.

[8] H. Kamrani *et al.*, "A deterministic and self-consistent solver for the coupled carrier-phonon system in SiGe HBTs," *IEEE Trans. Electron Devices*, vol. 64, no. 2, pp. 361–367, Feb. 2017, doi: 10.1109/TED.2016.2640343.

[9] M. Bina *et al.*, "Predictive hot-carrier modeling of n-channel MOSFETs," *IEEE Trans. Electron Devices*, vol. 61, no. 9, pp. 3103–3110, Sep. 2014, doi: 10.1109/TED.2014.2340575.

[10] H. Kamrani *et al.*, "Microscopic hot-carrier degradation modeling of SiGe HBTs under stress conditions close to the SOA limit," *IEEE Trans. Electron Devices*, vol. 64, no. 3, pp. 923–929, Mar. 2017, doi: 10.1109/TED.2017.2653197.

[11] K. Rupp *et al.*, "A review of recent advances in the spherical harmonics expansion method for semiconductor device simulation," *J. Comput. Electron.*, vol. 15, no. 3, pp. 939–958, Sep. 2016, doi: 10.1007/s10825-016-0828-z.

[12] S. Tiwari, M. V. Fischetti, and S. E. Laux, "Overshoot in transient and steady-state in GaAs, InP, Ga.47In.53As and InAs bipolar transistors," in *Proc. Int. Electron Device Meeting*, 1990, pp. 435–438, doi: 10.1109/IEDM.1990.237074.

[13] S. E. Laux and M. V. Fischetti, "Monte Carlo study of velocity overshoot in switching a 0.1-micron CMOS inverter," in *Proc. Int. Electron Device Meeting*, 1997, pp. 877–890, doi: 10.1109/IEDM.1997.650521.

[14] Y. Koseki, V. Rhzhii, T. Otsuji, V. V. Popov, and A. Satou, "Giant plasmon instability in a dual-grating-gate graphene field-effect transistor," *Phys. Rev. B, Condens. Matter*, vol. 93, no. 24, Jun. 2016, Art. no. 245408, doi: 10.1103/PhysRevB.93.245408.

[15] S. Di, K. Zhao, T. Lu, G. Du, and X. Liu, "Investigation of transient responses of nanoscale transistors by deterministic solution of the time-dependent BTE," *J. Comput. Electron.*, vol. 15, no. 3, pp. 770–777, Sep. 2016, doi: 10.1007/s10825-016-0818-1.

[16] M. Dyakonov and M. Shur, "Shallow water analogy for a ballistic field effect transistor: New mechanism of plasma wave generation by DC current," *Phys. Rev. Lett.*, vol. 71, no. 15, pp. 2465–2468, Oct. 1993, doi: 10.1103/PhysRevLett.71.2465.

[17] S.-M. Hong and J.-H. Jang, "Numerical simulation of plasma oscillation in 2-D electron gas using a periodic steady-state solver," *IEEE Trans. Electron Devices*, vol. 62, no. 12, pp. 4192–4198, Dec. 2015, doi: 10.1109/TED.2015.2489220.

[18] Z. Kargar, T. Linn, D. Ruić, and C. Jungemann, "Investigation of transport modeling for plasma waves in THz devices," *IEEE Trans. Electron Devices*, vol. 63, no. 11, pp. 4402–4408, Nov. 2016, doi: 10.1109/TED.2016.2608422.

[19] K. Rahmat, J. White, and D. A. Antoniadis, "Simulation of semiconductor devices using a Galerkin/spherical harmonic expansion approach to solving the coupled Poisson–Boltzmann system," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 15, no. 10, pp. 1181–1196, Oct. 1996, doi: 10.1109/43.541439.

[20] M. V. Fischetti and W. G. Vandenberghe, *Advanced Physics of Electron Transport in Semiconductors and Nanostructures*. Cham, Switzerland: Springer Int., 2016, doi: 10.1007/978-3-319-01101-1.

[21] C. Ringhofer, "Numerical methods for the semiconductor Boltzmann equation based on spherical harmonics expansions and entropy discretizations," *Transp. Theory Stat. Phys.*, vol. 31, nos. 4–6, pp. 431–452, 2002, doi: 10.1081/TT-120015508.

[22] H. G. Brachtendorf and K. Bittner, "Grid size adapted multistep methods for high Q oscillators," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 32, no. 11, pp. 1682–1693, Nov. 2013, doi: 10.1109/TCAD.2013.2269800.

[23] C. Jungemann, T. Linn, K. Bittner, and H.-G. Brachtendorf, "Numerical investigation of plasma effects in silicon MOSFETs for THz-wave detection," *Solid-State Electron.*, vol. 128, pp. 129–134, Feb. 2017, doi: 10.1016/j.sse.2016.10.030.

[24] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, "A review of hydrodynamic and energy-transport models for semiconductor device simulation," *Proc. IEEE*, vol. 91, no. 2, pp. 251–274, Feb. 2003, doi: 10.1109/JPROC.2002.808150.

[25] M. V. Fischetti and S. E. Laux, "Long-range Coulomb interactions in small Si devices. Part I: Performance and reliability," *J. Appl. Phys.*, vol. 89, no. 2, pp. 1205–1231, 2001, doi: 10.1063/1.1332423.

[26] N. Sano, "Impact of the Coulomb interaction on nano-scale silicon device characteristics," *J. Comput. Electron.*, vol. 10, nos. 1–2, pp. 98–103, 2011, doi: 10.1007/s10825-010-0327-6.

[27] D. A. Dahl and L. J. Shan, "Electrodynamics of quasi-two-dimensional electrons," *Phys. Rev. B, Condens. Matter*, vol. 16, pp. 651–661, Jul. 1977, doi: 10.1103/PhysRevB.16.651.

[28] M. V. Fischetti, "Long-range Coulomb interactions in small Si devices. Part II. Effective electron mobility in thin-oxide structures," *J. Appl. Phys.*, vol. 89, no. 2, pp. 1232–1250, 2001, doi: 10.1063/1.1332424.

[29] M. Dyakonov and M. Shur, "Detection, mixing, and frequency multiplication of terahertz radiation by two-dimensional electronic fluid," *IEEE Trans. Electron Devices*, vol. 43, no. 3, pp. 380–387, Mar. 1996, doi: 10.1109/16.485650.
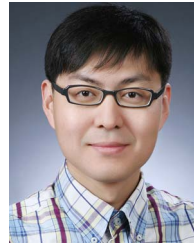
SUNG-MIN HONG (M'08) received the B.S. degree in electrical engineering and the Ph.D. degree in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2001 and 2007, respectively. He is currently an Assistant Professor with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea. His current research interests include deterministic Boltzmann equation solvers and simulation of terahertz devices.

JAE-HYUNG JANG (M'02) received the B.S. and M.S. degrees in electrical engineering from Seoul National University, in 1993 and 1995, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign, in 2002. He is a Professor with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea. His research interests include compound semiconductor devices and THz electronics.