Thermal Simulation Techniques for Nanoscale Transistors

Jeremy Rowlette[1], Eric Pop[1], Sanjiv Sinha[2,3], Mathew Panzer[2], and Kenneth Goodson[2]
Dept. of Electrical[1] and Mechanical Engineering[2], Stanford University, Stanford, CA and [3]Intel Corporation, Hillsboro, OR
Contact: rowlette@stanford.edu, Bldg 530 Room 101, Stanford CA 94305-3030, tel 650.725.3314

ABSTRACT

Thermal simulations are important for advanced electronic systems at multiple length scales. A major challenge involves electrothermal phenomena within nanoscale transistors, which exhibit nearly ballistic transport both for electrons and phonons. The thermal device behavior can influence both the mobility and the leakage currents. We discuss recent advances in modeling coupled electron-phonon transport in future nanoscale transistors. The solution techniques involve solving the Boltzmann Transport Equation (BTE) for both electrons and phonons. We present a practical method for coupling an electron Monte Carlo simulation with an analytic "split-flux" form of the phonon BTE. We use this approach to model self-heating in a 20 nm quasi-ballistic n+/n/n+ silicon diode, and to investigate the role of hot electron and hot phonon transport.

1 INTRODUCTION

Detailed electrothermal modeling in silicon devices has evolved over the past two decades beginning with the work of Wachutka [1]. Classically, Joule heat generation rates are simulated using the dot product of the current density and the electric field across the simulation mesh, and the temperature profile is computed through a solution of the heat diffusion equation (Fourier Law) [2]. This classical model assumes that the electrons and phonons are in local equilibrium and is applicable when the desired resolution is within several mean-free-paths of the energy carrying particles (10-100 nm in silicon). In the time domain this corresponds to periods much longer than the relevant relaxation times (0.1-10 ps). More advanced approaches have included coupled hydrodynamic models [3, 4] which assume that the individual systems, i.e. electrons or phonons, are in thermal equilibrium with themselves but not with each other. This approach allows the assignment of a single effective temperature to each particle system which forces the particle distribution functions to retain their equilibrium shapes. Furthermore, a single effective relaxation time describes the rate of energy exchange between the two systems. Since relaxation times are in general strongly energy-dependent, the use of a single average relaxation time can be a poor approximation. This is particularly true for highly non-equilibrium transport conditions.

Fundamentally, the physical mechanism through which self-heating occurs in a semiconductor device is that of electrons scattering with phonons. Thus only a simulation approach which deliberately incorporates all such scattering events, and the energy dispersion characteristics of each particle type, will capture the full microscopic, detailed picture. This can be achieved by solving the Boltzmann Transport Equation (BTE) for both electrons and phonons:

$$(1)\quad \frac{\partial N_q(r,t)}{\partial t} + v_g \cdot \nabla_r N_q(r,t) + F \cdot \nabla_{\hbar q} N_q(r,t) = \left.\frac{\partial N_q(r,t)}{\partial t}\right|_{collision} + \dot{n}_q(r,t)$$

Here, N_q is the number of particles in a particular mode with wave vector q at position r. v_g is the group velocity for that mode and F is an external force acting on the particle. The two right-hand terms correspond to changes in the number of particles in that mode due to collisions and due to the creation of particles (e.g. during recombination-generation processes for electrons or phonon creation from electron-phonon scattering) respectively.

The BTE is valid for the semi-classical transport regime where charge and energy carriers can be treated as particles between scattering events but the frequency and nature of scattering is described using a quantum mechanical treatment (see for example Ref. 5). At sufficiently high power densities (~10^{12} W/cm^3) and length (temporal) scales well below the relaxation length (time), i.e. < 100 nm (< 10 ps), of the energy carriers, both hot electron and hot phonon effects are to be anticipated [6, 7]. Such conditions could be encountered in highly scaled transistors. Under these conditions, the distribution functions for electrons and phonons are often highly skewed with respect to their near-equilibrium distributions and do not have simple analytic relationships. Solving the BTE (or one of its limiting forms) is necessary for capturing the details of hot carrier transport physics near the peak power generation point often referred to as the "hotspot."

The highly non-equilibrium carrier transport can lead to measurable effects on the macroscopic behavior of devices, e.g. through the reduction in transconductance (or drive current) and increase in subthreshold leakage power. For example, it has been postulated that heat transport can be impeded strongly near the hotspot which can lead to anomalously high temperatures inside the device [8, 9]. This is partly explained by the fact that hot electrons preferentially impart their energy to optical phonon modes (via scattering) which have relatively low group velocities (~ 1000 m/s). Therefore, a significant portion of the heat energy tends to stagnate near the hot spot until the optical phonons can decay to faster moving acoustic modes (~10 ps). The accumulation of optical phonon modes in the vicinity of the hotspot will cause electrons to experience an increase in momentum relaxing scattering events, which negatively impacts electron transport. Furthermore, these hot optical phonons may contribute strongly to subthreshold conduction giving rise to anomalously high leakage currents.

Recently, we have achieved the coupling of the electron and phonon systems inside realistic device structures by solving the BTE for electrons and phonons in an iterative fashion. In this work, we discuss a particular computational technique for solving the coupled BTE for the electron-phonon system. We present recent simulation results where this combined technique was used to study self-heating in a one-dimensional (1-D) 20 nm long quasi-ballistic n+/n/n+ silicon diode.

2 COUPLING ELECTRONS AND PHONONS

To model the effects of self-heating, the electron and phonon systems must be coupled together. It is not sufficient for electron-phonon scattering to be simply included in an electron transport model since the phonons that are generated during the simulation are not "sensed" by the simulated electrons. There needs to be a way to feed the updated information about the phonon populations (modal occupation) back into the scattering rates for the electron system. Furthermore, the phonons generated during the simulation must be allowed to propagate and decay as they would in a real device. The complexity and magnitude of such a task, however, has prevented truly rigorous solutions of the coupled transport physics at such length and time scales. Various approximations in either the electron or phonon models are typically necessary [3,10,11]. In this work, we take a novel approach to coupling the two carrier populations by combining an electron Monte Carlo (EMC) technique, with a simplified "split-flux" form of the phonon BTE (SF-BTE) and solving them in iteration in a realistic device geometry.

The EMC simulator used in this work was developed by Pop *et*

0-7803-9254-X/05/$20.00 ©2005 IEEE.

al. [12] and is optimized for calculating detailed electron-phonon scattering and thereby heat generation in low voltage (< 1.5 Volt) silicon devices. This approach includes a full description of the phonon dispersion relationship in silicon. The SF-BTE, developed by Sinha *et al*. [13], is a computationally efficient solution to the phonon BTE, which captures ballistic phonon conduction near the hotspot and also yields a convenient interface to continuum calculations far from the hotspot. In effect, the phonon distribution is split into a near-equilibrium component obeying Fourier's heat conduction law and a far-from-equilibrium component which dominates the transport near the hotspot. For detailed descriptions of these two techniques we refer the reader to Ref.'s 12 and 13.

For each iteration, two independent simulations are performed: one for the electron system and the other for the phonon system. Outputs from each simulation are fed back into the other and the simulation proceeds until satisfactory convergence is achieved. We find that this method achieves rapid convergence within three iterations. The coupled approach begins with an isothermal (300 K) EMC simulation whose initial conditions are given by a drift-diffusion device simulator such as MEDICI. The EMC computes electron transport self-consistently with the electric field (Poisson equation) across the device grid, taking into account the doping profile and impurity scattering. Net phonon generation rates as a function of position and phonon frequency are gathered from the EMC and fed into the SF-BTE. In the latter, the phonons are allowed to propagate in the absence of the electron system and only phonon-phonon scattering is included. Phonon-phonon scattering rates are determined at the beginning of the SF-BTE simulation for acoustic modes using formulae derived from first-order anharmonic perturbation theory [14] and an updated temperature profile inferred from the EMC calculation. Because of a lack of experimental data, a best-known value of 10 ps is assigned to the lifetime for all optical modes [15, 16]. At the end of the SF-BTE calculation, an updated distribution of phonons as a function of position is computed. This distribution of phonons is then used to compute the electron-phonon scattering rates to be used during the subsequent EMC simulation. We now discuss aspects of the feedback pathways which ensure a closed-loop process between the two systems.

Rigorously, a detailed phonon generation spectrum needs to be calculated for each grid point in the device during the EMC simulation and then fed into the subsequent BTE simulation as a source term. However, to compute such a spectrum at each grid point would require impractically long simulations. In this work, we significantly reduce the computation time while maintaining the critical elements of the coupled transport physics by developing a phonon generation spectra lookup table which is cross-referenced to a local power generation rate. The latter quantity is more easily tracked during the EMC routine. The local power density, determined as the sum of net phonon emission events at each grid location, is then used to determine an appropriate phonon generation spectra by using a linear interpolation scheme. The major assumption employed here is that the phonon generation spectrum for the spatially nonuniform (i.e. nonuniform electric field and doping) condition is close to that for the homogeneous condition (for which the lookup table is generated) provided that the power densities for the two scenarios are equivalent.

Once the SF-BTE has been run using the computed phonon generation spectra, we use the output phonon distributions to recompute the electron-phonon scattering rates $S_{e\text{-}p}(k, k')$ in the subsequent EMC simulation, which scale in proportion to $N_{q,s}$ (the phonon occupation number) as follows

$$S_{e\text{-}p}(k,k') = \frac{1}{\tau_{k,k'}} \propto \left(N_{q,s} + \frac{1}{2} \pm \frac{1}{2} \right) \tag{2}$$

where the +/- signs correspond to emission and absorption of phonons during an electron transition from state |k> to |k'>. At equilibrium, the phonon occupation number is given by the Bose-Einstein distribution as a function of the temperature T:

$$N_{q,s} = \left[\exp\left(\frac{\hbar\omega_q}{k_B T} \right) - 1 \right]^{-1} \tag{3}$$

Although the occupation number is the fundamental parameter needed to compute scattering rates in the EMC, it is impractical to employ phonon occupation numbers for all modes and branches and for all grid points when computing the scattering rates. A second approximation is then made to dramatically simplify the calculation. Using the actual distributions calculated in the SF-BTE simulation, an effective temperature $T_{eff,s}(x)$ is computed for each phonon branch, s, and then passed to the EMC in the form of four separate temperature vectors, one for each phonon branch. The temperatures are then used to adjust the scattering rates in a manner we will discuss shortly. Aside from the added complexity and the reduction in computational speed, there is nothing fundamentally preventing the use of additional temperatures to account for the occupation of individual phonon wave vectors or ranges of phonon wave vectors during scattering rate calculations. One temperature per phonon branch is just one approximation. We discuss the use of additional temperatures in the following section.

The effective temperature for branch s ($T_{eff,s}$) is calculated using the following expression derived in Ref. 11

$$\frac{1}{8\pi^3}\int N_{q,s}(T_{eff,s})\hbar\omega_{q,s}d\vec{q} = \frac{1}{8\pi^3}\int \left(N_{q,s}(T_F) + n_{q,s}\right)\hbar\omega_{q,s}d\vec{q} \tag{4}$$

with the integrations performed over the first Brillouin zone (B.Z.). In this expression, $N_{q,s}(T_F)$ is the near-equilibrium occupation of phonons and $n_{q,s}$ is the far-from equilibrium occupation having wave vectors q of branch s and frequency $\omega_{q,s}$. T_F is the temperature of the near-equilibrium phonons which obey Fourier's Law. To arrive at $T_{eff,s}$, we first find $n_{q,s}$ from the far-from equilibrium phonon BTE derived under the relaxation time approximation:

$$v_{q,s} \cdot \nabla_r n_{q,s} = -\frac{n_{q,s}}{\tau_{q,s}} + \dot{n}_{q,s} \tag{5}$$

$\dot{n}_{q,s}(x)$ is the source term determined from the EMC output and the post-EMC lookup table routine (i.e. the phonon spectra), $\tau_{q,s}$ is the phonon-phonon scattering time (or lifetime) and $v_{q,s}$ is the group velocity for mode ω_q in branch, s. Then, T_F is found by

$$\frac{1}{8\pi^3}\int \frac{n_{q,s}}{\tau_{q,s}}\hbar\omega_{q,s}d\vec{q} + k_{Si} \cdot \nabla_r^2 T_F = 0 \tag{6}$$

where the second term is simply Fourier's Law for the near-equilibrium phonon heat flux, and k_{Si} is the effective thermal conductivity tensor for silicon. Once effective temperatures have been computed for each phonon branch, the maximum scattering rate for each electron-phonon scattering type (e.g. intravalley acoustic, intervalley g-type and f-type, emission/absorption, etc. [12]) is calculated using the maximum temperature for the appropriate phonon branch and the simulation begins. To include the dependence of the local phonon concentration, we then employ a temperature based rejection algorithm similar to a technique discussed in Ref. [17]. When an electron at a grid location (x_i) is chosen to scatter with a particular phonon type (of branch s), the effective temperature for the particular branch of phonon at that exact grid point is compared to the maximum effective temperature for that branch and the electron scattering rate is adjusted locally. The local scattering rate is derived from the maximum scattering rate for that phonon type

using a scaling factor Γ defined as

$$\Gamma = \frac{N_{q,s}\left(T_{eff,s}(x_i)\right) + \frac{1}{2} \pm \frac{1}{2}}{N_{q,s}\left(T_{eff,s,MAX}\right) + \frac{1}{2} \pm \frac{1}{2}} \quad (7)$$

where, again, the +/- signs correspond to phonon emission and absorption, respectively and $0 \le \Gamma \le 1$.The rejection algorithm is implemented by first generating a random number $0 \le X \le 1$ with uniform probability density. The scaling factor Γ is then computed using the effective phonon temperature at that grid point and compared to X. If $\Gamma < X$ then the scattering event chosen at that instant in time and position is allowed to take place. Otherwise, the scattering event is rejected and the electron continues on its initial trajectory unperturbed. With the phonon generation source term output from the EMC and the electron-phonon scattering rates being adjusted to account for phonon occupation (via the effective temperature and the rejection algorithm), the electrons and phonons form a closed loop system. We now discuss the application of this algorithm to the simulation of a simple 1-D silicon device.

3 SELF-HEATING IN NANO-TRANSISTORS

Fig. 1 shows the 1-D n+/n/n+ silicon device simulated in this work. The 150 nm "source/drain" regions are doped to 10^{20} cm^{-3} and are separated by a 20 nm lightly doped (10^{16} cm^{-3}) "channel." The electrical characteristics of the device are shown in Fig. 2. Although, infinite in extent in the transverse plane and lacking a gate terminal, such a device structure resembles the core of future transistor structures, e.g. FinFETs. The band diagram along the channel is similar to that along the channel of typical CMOS devices, especially double- (surround-) gate devices where vertical (transverse) symmetry exists. Fig. 3 shows the steady-state power density generated within the device at three different bias conditions: 0.6, 0.8, and 1.0V. These values are computed as the net sum of all phonon emission minus all phonon absorption events at every grid point. Notice that nearly all of the power generation occurs within the drain and that the generation profile is exponentially decaying with a characteristic length of about 20 nm. This characteristic length is comparable to the electron energy relaxation length and is nearly independent of bias voltage [6]. Fig. 4 shows the phonon generation spectra computed at four positions within the drain region using the lookup table and linear interpolation method. Fig. 5 shows the effective temperature profiles for each of the four phonon branches for the 1V bias condition and fixed temperature (300 K) boundary conditions. Near the hotspot, the temperature is dominated by the effects of ballistic transport and a temperature slip is observed [13]. Farther from the hotspot, the temperature profiles resemble that which would be predicted using the classical diffusion equation (dashed curves). The longitudinal optical (LO) phonon branch temperature peaks at greater than 365 K near the peak power generation point or hotspot. The impact on average electron energy is shown in Fig. 6. It is interesting to note here that the average energy increases in both the source and drain but there is little effect in the channel. This indicates near-ballistic conditions within the channel. Despite an appreciable temperature rise within the device, it is found that the overall impact on device current is only about -2.2 % for the highest bias condition (1 V). This effect may be higher in a real FinFET where the thermal conductivity is significantly reduced from phonon boundary scattering [18]. The impact on leakage current could be much larger due to the strong (~ exponential) dependence of the electron occupation on temperature and is under investigation. The hot electrons injected into the channel under near ballistic conditions scatter strongest with the LO phonons and in particular the g-type LO (g-LO) phonons with wave vector near 0.3 of the B.Z. edge [19]. Since the g-LO is one of the strongest scatterers with electrons while also having a relatively low density of states, its occupation value should be a good indication of hot phonon effects [6, 7]. To understand how a single LO branch temperature is able to capture the g-type LO phonon occupation, it is instructive to review the results of Fig. 7. The effective temperature of the g-LO phonon, and consequently its occupation number, is seen to be significantly underrepresented (~200 K discrepancy) by using a single temperature assigned to the LO branch. Therefore, additional phonon temperatures may be necessary to model hot phonon effects.

4. CONCLUSIONS

We have demonstrated the coupling of electron Monte Carlo simulations with an analytic split-flux form of the phonon BTE. We used this technique to evaluate self-heating in a quasi-ballistic Si diode where hot electron and phonon conditions prevail. Despite appreciable temperature rises within the device, a negligible impact on the macroscopic device behavior was observed. The effects on subthreshold leakage could, however, be substantial. This study is ongoing, but preliminary results indicate that four effective temperatures (one per phonon branch) may not be sufficient to capture the hot phonon effects within the device. The next logical step will be to incorporate additional phonon temperatures into the EMC simulation beginning with a temperature that will effectively model the high occupation of the g-type LO phonons.

ACKNOWLEDGMENTS

This work was supported by the Semiconductor Research Corporation (SRC) Task 1043.

REFERENCES

[1] G. Wachutka. "Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling," *IEEE Trans. CAD* **9**, 1141 (1990)

[2] G. Workman *et al.* "Physical modeling of temperature dependences of SOI CMOS devices and circuits including self-heating," *IEEE TED* **9**, 1141 (1990)

[3] J. Lai and A. Majumdar. "Concurrent thermal and electrical modeling of submicrometer silicon device,." *J. Appl. Phys*. **74,** 3005 (1999)

[4] J-H Chun *et al*. "Electrothermal simulation of nanoscale transistors with optical and acoustic phonon heat conduction," to appear in *IEEE SISPAD* (2005)

[5] M. Lundstrom, "Fundamentals of carrier transport," *Cambridge U. Press* (2000)

[6] E. Pop *et al*. "Joule heating under quasi-ballistic transport conditions in bulk and strained silicon devices," to appear in *IEEE SISPAD* (2005)

[7] M. Artaki and P. Price "Hot phonon effects in silicon field-effect transistors," *J. Appl. Phys*. **65**, 1317 (1989)

[8] G. D. Mahan and F. Claro. "Nonlocal theory of thermal conductivity," *Phys. Rev. B* **38**, 1963 (1988)

[9] G. Chen. "Nonlocal and nonequilibrium heat conduction in the vicinity of nanoparticles," *ASME J. Heat Transfer* **118**, 539 (1996)

[10] R. Lake and S. Datta,. "Energy balance and heat exchange in mesoscopic systems," *Phys. Rev. B.* **46**, 44763 (1992)

[11] J.C. Vaissiere, *et al*. "Numerical solution of coupled steady-state hot-phonon-hot-electron Boltzmann equations in InP," *Phys. Rev. B*. **46**, 13082 (1992)

[12] E. Pop. *et al*. "Analytic band Monte Carlo model for electron transport in Si including acoustic and optical phonon dispersion," *J. Appl. Phys*. **96**, 4998 (2004)

[13] S. Sinha *et al*. "A split-flux model for phonon transport near hotspots," Int. Mech. Eng. Congress and Expo., Anaheim, CA, 2004; and in press, *ASME J. Heat Transfer* (2005)

[14] M. G. Holland, "Analysis of lattice thermal conductivity," *Phys. Rev*. **132**, 2461 (1963)

[15] J. Menendez and M. Cardona, "Temperature dependence of the first-order Raman scattering by phonons in Si, Ge, and a–Sn: Anharmonic effects," *Phys. Rev. B* **29**, 2051 (1984)

[16] S. Sinha *et al*., "Scattering of g-process longitudinal optical phonons at hotspots in silicon," *J. Appl. Phys*. **97**, 023702 (2005)

[17] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Rev. Mod. Phys*. **55**, 645 (1983)

[18] Y. S. Ju and K. E. Goodson. "Phonon scattering in silicon films with thickness of order 100 nm," *App. Phys. Lett*. **74**, 3005 (1999)

[19] E. Pop *et al*. "Monte Carlo simulation of Joule heating in bulk and strained silicon," *Appl. Phys. Lett*. **86**, 82101 (2005)

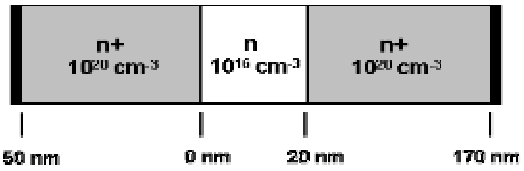

Figure 1. n+/n/n+ device structure analyzed in this work. The doping profile between regions is Gaussian having a 1.25 nm/decade slope.

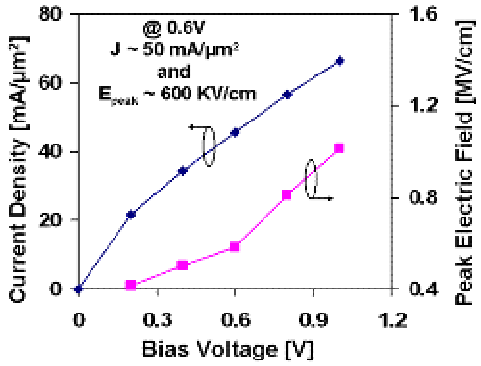

Figure 2. Electrical characteristics of the 20 nm n+/n/n+ diode as a function of applied bias.

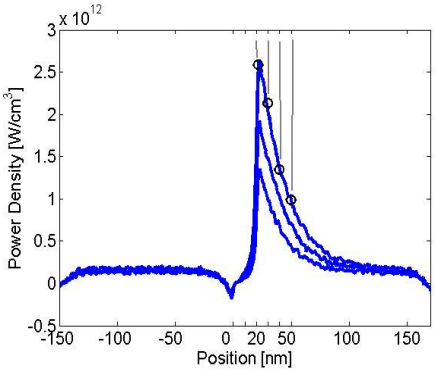

Figure 3. Computed heat dissipation rate along the 20 nm device for 0.6, 0.8 and 1.0V operation (from bottom to top). The circles correspond to positions in the device where the phonon generation spectra are shown in Fig. 4.

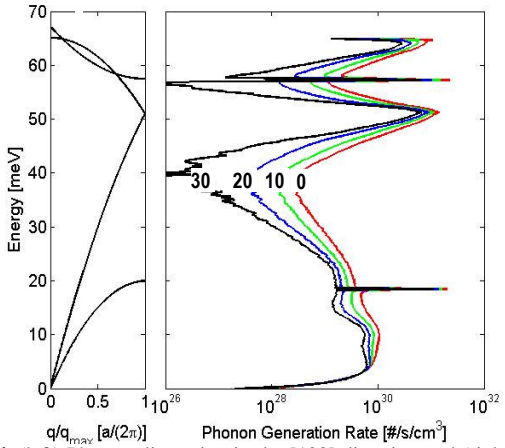

Figure 4. (left) Phonon dispersion in the [100] direction and (right) phonon generation spectra computed at 4 locations within the device for the 1V case (see Fig. 3). The red (right-most) curve corresponds to the location of the peak power dissipation. The remaining curves correspond to 10, 20, and 30 nm displaced from the peak generation point or hotspot.

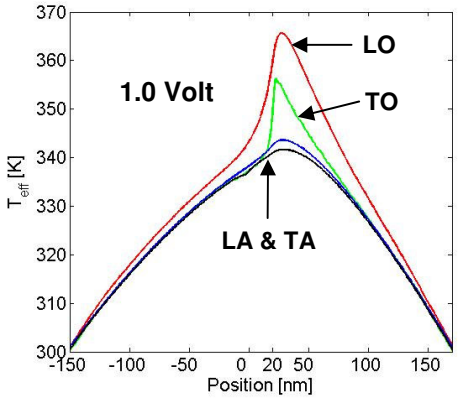

Figure 5. Effective temperature for each of the 4 branches: longitudinal optical (LO) and acoustic (LA), transverse optical (TO) and acoustic (TA), computed for the 1V condition. Note that the acoustic branch temperatures are similar to the diffusion temperature, as expected.

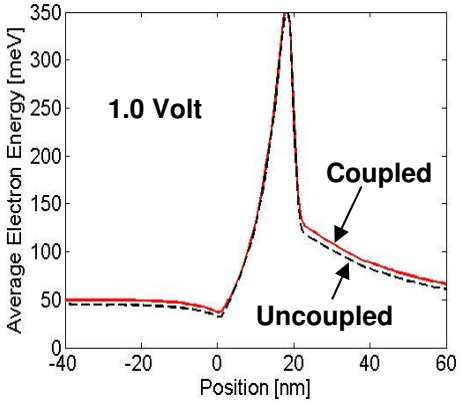

Figure 6. Average source-born electron energy vs. position for the coupled (solid red) and uncoupled (black dashed) simulations with 1V applied bias.

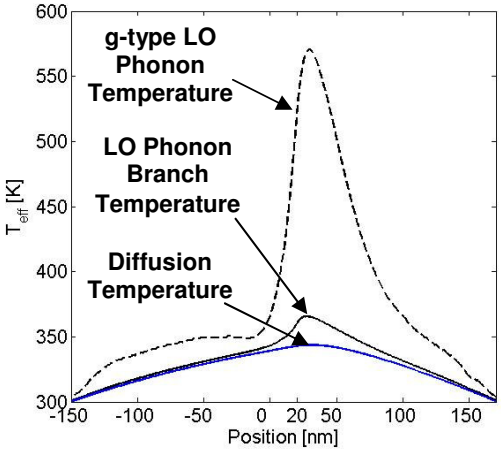

Figure 7. Effective temperature profiles for the longitudinal optical (LO) phonon branch (solid black) compared to the effective temperature of the g-type LO phonon alone (dashed black) for the 1V bias condition. The temperature computed from the heat diffusion equation (Fourier's Law) is also shown (solid blue line), as indicated.