

Review

Quantum Transport in a Silicon Nanowire FET Transistor: Hot Electrons and Local Power Dissipation

Antonio Martinez ^{1,*} and John R. Barker ²

¹ College of Engineering, Swansea University, Engineering East, Fabian Way, Crymlyn Burroughs, Swansea SA1 8EN, UK

² James Watt School of Engineering, College of Science and Engineering, University of Glasgow, Glasgow G12 8LT, Scotland, UK; john.barker@glasgow.ac.uk

* Correspondence: a.e.martinez@swansea.ac.uk

Received: 12 June 2020; Accepted: 23 July 2020; Published: 26 July 2020



Abstract: A review and perspective is presented of the classical, semi-classical and fully quantum routes to the simulation of electro-thermal phenomena in ultra-scaled silicon nanowire field-effect transistors. It is shown that the physics of ultra-scaled devices requires at least a coupled electron quantum transport semi-classical heat equation model outlined here. The importance of the local density of states (LDOS) is discussed from classical to fully quantum versions. It is shown that the minimal quantum approach requires self-consistency with the Poisson equation and that the electronic LDOS must be determined within at least the self-consistent Born approximation. To bring in this description and to provide the energy resolved local carrier distributions it is necessary to adopt the non-equilibrium Green function (NEGF) formalism, briefly surveyed here. The NEGF approach describes quantum coherent and dissipative transport, Pauli exclusion and non-equilibrium conditions inside the device. There are two extremes of NEGF used in the community. The most fundamental is based on coupled equations for the Green functions electrons and phonons that are computed at the atomically resolved level within the nanowire channel and into the surrounding device structure using a tight binding Hamiltonian. It has the advantage of treating both the non-equilibrium heat flow within the electron and phonon systems even when the phonon energy distributions are not described by a temperature model. The disadvantage is the grand challenge level of computational complexity. The second approach, that we focus on here, is more useful for fast multiple simulations of devices important for TCAD (Technology Computer Aided Design). It retains the fundamental quantum transport model for the electrons but subsumes the description of the energy distribution of the local phonon sub-system statistics into a semi-classical Fourier heat equation that is sourced by the local heat dissipation from the electron system. It is shown that this self-consistent approach retains the salient features of the full-scale approach. For focus, we outline our electro-thermal simulations for a typical narrow Si nanowire gate all-around field-effect transistor. The self-consistent Born approximation is used to describe electron-phonon scattering as the source of heat dissipation to the lattice. We calculated the effect of the device self-heating on the current voltage characteristics. Our fast and simpler methodology closely reproduces the results of a more fundamental compute-intensive calculations in which the phonon system is treated on the same footing as the electron system. We computed the local power dissipation and “local lattice temperature” profiles. We compared the self-heating using hot electron heating and the Joule heating, i.e., assuming the electron system was in local equilibrium with the potential. Our simulations show that at low bias the source region of the device has a tendency to cool down for the case of the hot electron heating but not for the case of Joule heating. Our methodology opens the possibility of studying thermoelectricity at nano-scales in an accurate and computationally efficient way. At nano-scales, coherence and hot electrons play a major role. It was found that the overall behaviour of the electron system is dominated by the local density of states and the scattering rate. Electrons leaving the simulated drain region were found to be far from equilibrium.

Keywords: silicon nanowires; nano-transistors; quantum transport; hot electrons; self-cooling; nano-cooling; thermoelectricity; heat equation; non-equilibrium Green functions; power dissipation

1. Introduction

1.1. Context

Silicon metal-oxide semiconductor-field effect transistors (MOSFETs) have underpinned the last fifty years of the microelectronics revolution. The scalability of the basic planar single gate MOSFET has already approached the long-anticipated physical limits to device performance that require considerable architectural ingenuity to overcome. These limits include short-channel effects (SCE) such as: quantum mechanical tunnelling between source and drain and between drain and substrate; drain-induced barrier lowering (DIBL); threshold voltage roll-off. However, since the supply voltage has not been downscaled as smaller feature sizes have been achieved, the power densities have increased close to the limit of $150 \text{ W}\cdot\text{cm}^{-2}$ beyond which air cannot cool the device temperature [1]. The move to a 3D FinFET architecture has given some respite [2,3]. Electronic devices at nanometre scales have already been fabricated and tested [4–7]: effective gate lengths of some of the recently produced transistors are well under 20 nm [8]. There is a consensus emerging that the most promising successor to the planar MOSFET is the gate-all-around nanowire-FET (NWFET) (a cross-sectional model is displayed in Figure 1 of Section 3); it has far superior electrostatic control to the FinFET [9]. Silicon NWFETs have been fabricated with circular cross sections with diameters in a range of 9 to 1 nm [7]. These devices show very good performances and ideal sub-threshold slopes (60 mV/dec). They also showed a substantial decrease in the on-current as the cross section decreases that cannot be explained by the reduction in area. Instead, it was suggested that increases in scattering rates or the enhancement of other scattering mechanism such as surface scattering or stresses in the fabrication might be responsible. In a recent study [10], we have shown that NWFETs with ultra-scaled square cross section and circular cross-section have very similar properties.

1.2. Electro Thermal Properties of NWFET

The gate-all-around NWFET has excellent SCE reduction at very small gate lengths (~10 nm) but requires small diameters (~5 nm) for best performance. At these scales, however, there are issues with atomistic effects (the effect of the random discrete nature of dopants and defects [11]) but the major future issue has been long-known: any conventional electronic logic is ultimately limited by heat dissipation [12]. Comparatively few studies have been made of the electro-thermal properties of NWFETs. It is known that silicon nanowire structures have a reduced thermal conductivity compared to bulk [13,14] due to boundary scattering of phonons at the oxide sheath-silicon channel interface, which further exacerbates self-heating and hence device performance. Such devices are expected to be prone to the appearance of self-heating effects and localised hot spots. Unfortunately, the small dimensions of ultra-scaled nanowires provides a serious challenge to measurements of self-heating effects and local hot spots (although recent progress in scanning thermometry is promising [15,16]).

1.3. Need for Efficient Modelling

Fabrication of nano-devices is of high lithographic precision and it is expensive. These factors make a comprehensive experimental investigation of a wide range of prototype devices prohibitive. From a design point of view, there is thus scope for good fast physics-based device modelling tools for NWFETs that assists the control of adverse electro-thermal effects. Good theoretical analysis of the electrical and carrier transport properties of proposed nanowire devices is required prior to optimised design and fabrication; in particular, the understanding of the interaction between the carrier flow and the electrostatic control.

The simulation of the thermal performance, power dissipation and carrier lattice interaction are also important in the design of self-cooling and heat management of nanostructures. High channel temperatures lead to raised temperatures in the interconnects at the silicon source extension-metal contact [17]. The scope for conduction cooling via the source/drain/interconnects [18–20] has been appreciated for some time [21]. The number of research programmes on nano-cooling is rising [22]. In addition, research into the impact of quantum effects on the thermodynamic efficiency of nanostructures [22] has been a long-standing concern.

1.4. Review of Modelling Approaches

Electronic transport and heat transport at nano scales have been studied by a variety of methodologies. Semi-classical techniques complemented with quantum corrections such as the Drift-Diffusion (DD) approach [23–27] and the Monte Carlo method [28–37] have been widely used.

The Drift-Diffusion approach is particularly computationally efficient but assumes that the electron charge fluid is both in local equilibrium [38] and moves due to electric fields and gradients of concentration. The equation to be solved is the charge density conservation law usually called the continuity equation. Generalizations of the DD approach and energy balance in concomitance with the heat equation have been carried out, a thorough investigation is presented in [39]. An extra equation can be added to the Drift-Diffusion model to treat the local changes in momentum or velocity of the charge fluid; the model that results is termed the hydrodynamic model. However, quantum effects such as confinement and tunneling need to be artificially added.

The Monte Carlo method [28] which is founded on a particle representation of the semi-classical Boltzmann equation is ideal to treat non-equilibrium processes semi-classically; the flows are energy-resolved but the carrier or electron transport is treated classically. It is assumed that the electron de Broglie wavelength is much smaller than the scale of variation of the electrical potential. This last assumption is violated in small nanostructures. Any quantum coherent transport is neglected in the semi-classical formalism. In addition, the scale of the thermalisation (as will be shown in this paper) is much larger than the dimension of the nanostructure.

A powerful quantum transport methodology, the Keldysh non-equilibrium Green function (NEGF) formalism, was applied from the 70's in order to describe non-equilibrium quantum transport process beyond the linear response in nanostructures [40]. The formalism was used to study high electric field effects [41]. The original formalism was developed in the 60's by Schwinger [42], Keldysh [43], Kadanov and Baym [44] to address highly non-equilibrium processes such as those in Masers. The formalism uses matrix equations containing a mix of dynamical and kinetic Green functions [43].

The transport methodology satisfies the second law of thermodynamic as the entropy generated by the transfer of electrons from one reservoir to another reservoir is always positive [22]. The non-equilibrium scattering rates are self-consistently dependent on the local electric field (or equivalently the local electrostatic potential), which is not the case for Boltzmann transport. This feature was pointed out using a super-resolvent non-linear time-dependent quantum kinetic equation based on the density matrix formalism [45–47], and confirmed in later studies by the NEGF formalism [41]. Electro-thermal simulations have been carried out using a variety of methodologies ranging from semiclassical [48], a blend of quantum transport with classical heat equation to a pure quantum mechanical [49] and atomistic methodologies [50]. All these methodologies have their weaknesses and strengths. The semi-classical methodologies are very efficient and handle micrometer size devices however they ignore coherence, tunneling and renormalization of the density of states and therefore lose accuracy for devices of few tens of nanometers. The pure atomistic methodologies (mostly based in the NEGF formalism) consider the wave nature of electrons, coherence and quantum effects that require substantial computational power, and it is computationally feasible only for very small dimensions. However, effects and boundary conditions that are far from the active device region could affect the electrical and heat performance of the device, for example, remote optical phonon scattering and remote coulomb interactions [51–54], phonon reflections or transmission through interfaces [55,56].

1.5. Content of This Paper

In the present paper, Section 2 provides a brief review and perspective on the key concepts and the transition from semi-classical to fully quantum mechanical modelling of electro-thermal effects applicable to quasi-one dimensional NWFETs. Section 2.1 discusses the basic concepts of electro-thermal modelling. Section 2.2 describes the crucial local density of states in energy and the need for a transition from semi-classical to full quantum transport for electro-thermal device modelling. Section 2.3 gives a brief overview of the non-equilibrium Green function (NEGF) methodology augmented by explicit technical details in an Appendix (Appendix A discusses the local density of states for nanowires and the multi-valley picture; Appendix B describes how contact self-energies are used to project the open many-body problem into a finite device simulation domain; Appendix C discusses details of the Keldysh NEGF formalism; Appendix D gives explicit details of the phonon Green functions and the generic electron-phonon self-energies; Appendix E demonstrates explicitly that the NEGF formalism shows no heat-dissipation for elastic acoustic phonon scattering; Appendix F describes detailed balance within the NEGF formalism; Appendix G describes the importance of the Kramers-Kronig relations for self-energies).

In Section 3, we present our simulation methodology for the silicon gate-all-around nanowire FET that uses the NEGF picture for the electrons coupled to a semi-classical Fourier heat equation. The device structure and the quasi-1D representation is introduced in Section 3.1. The NEGF methodology for the density and current density is outlined in Section 3.2. Section 3.3 discusses energy balance and power. Section 3.4 presents the results and discussion including comparison with other simulated cases. Section 3.5 discusses perspectives and challenges. Finally, the conclusions are presented in Section 4. In the following we exclude radiative emission and absorption as this is not relevant for conventional integrated circuits, although this is relevant to photo-excited hot electron phenomena.

2. Methodology and Models

2.1. Basic Concepts of Electro-Thermal Modelling

The key quantity in electro-thermal phenomena is the local energy current density $J_{Energy}(\mathbf{r}, t)$ synonymous with the heat current density. If we consider a finite volume Ω bounded by an enclosing surface S then the total power generated within the volume Ω is balanced by the net integrated energy current density flowing out of the surface S :

$$\iiint_{\Omega} P(\mathbf{r}, t) d^3\mathbf{r} = - \iint_S J_{Energy}(\mathbf{r}, t) \cdot d\mathbf{S} \quad (1)$$

where $P(\mathbf{r}, t)$ is the local power density, synonymous with the local rate of heat generation. Applying the divergence theorem we obtain:

$$P(\mathbf{r}, t) = -\nabla \cdot J_{Energy}(\mathbf{r}, t) \quad (2)$$

If Ω represents the active volume of an FET device there is no net heat generation in the device when it is switched off as it is in thermal equilibrium with the lattice and its environment. If the device is switched on (gate voltage applied) and a (chemical potential difference) voltage difference established between the source and drain contacts then an electrical current flows characterised by the current density $J(\mathbf{r}, t) = e j(\mathbf{r}, t)$ where $j(\mathbf{r}, t)$ is the particle current density.

In a purely ballistic channel [57,58] for which the electron flow is quantum coherent (no dissipative scattering), the electron transport involves a net injection of electrons at high energy and momentum from the source followed by net extraction at the drain where the excess energy and momentum are dissipated: the flow is quantum transmission and the conductance $G = 1/R$ (R : The electrical resistance) of the channel is quantised in units of $e^2/\pi\hbar$. This extreme case (best studied for low temperatures) illustrates that quantum transport can lead to hot spots, in this case strong heating occurs in the

drain contact. If a defect or impurity occurs in a ballistic channel such that its interaction with an electron is representable as a simple scalar potential perturbation, the flow then involves reflection and tunnelling transmission that may be described by T_c the transmission coefficient for the channel. Then $G_{min} = T_c e^2 / \pi \hbar$ which is Landauer's formula [59]. If $T_c < 1$ there is elastic (potential) scattering in the channel. Again, the dissipation of energy is in the contact region; the dissipation of momentum is partly within the channel via the defect/impurity. This simple picture ignores the electrical field profile $E(r,t) = -\nabla\phi(r,t)$ in the channel. In practice, it is neither abrupt nor uniform; instead it may be determined self-consistently from Poisson's Equation:

$$\nabla \cdot \epsilon_d \nabla \phi(r,t) = -\rho(r,t) \quad (3)$$

where ρ is the charge density in the channel and ϵ_d is the dielectric permittivity.

Silicon nanowire transistors do not operate in the purely ballistic regime although elements of the simple transmission/reflection flows may occur as we discuss later. In general, the applied electric field supplies energy and momentum to the carriers throughout the device and the electrostatic potential varies smoothly between source and drain. For dissipative transport the energy and momentum gained locally from the field will be transferred to the phonon system by scattering within the device involving a loss-gain process in which phonons may be emitted or absorbed by the electron assembly. Inelastic scattering gives rise to energy dissipation whereas the momentum dissipation occurs for elastic and inelastic processes.

More detailed non-phenomenological analysis requires an electron transport formalism and ultimately a coupled phonon transport picture. The deepest semi-classical picture is provided by the well-known Boltzmann transport equation for the electron probability distribution $f[r,k,t]$:

$$\frac{\partial f[r,k,t]}{\partial t} + v(k) \cdot \frac{\partial f[r,k,t]}{\partial r} + eE(r,t) \cdot \frac{\partial f[r,k,t]}{\partial \hbar k} = -\frac{\partial f[r,k,t]}{\partial t} \Big|_s \quad (4)$$

Here the velocity v is determined from the k -dependence of the conduction band energy $\epsilon(k)$:

$$v(k) = \frac{1}{\hbar} \frac{\partial \epsilon(k)}{\partial k} \quad (5)$$

The second term in Equation (4) represents diffusive effects, the third term represents the drift of the distribution induced by an applied electric field, the RHS represents the change in f due to scattering processes. The scattering integral (actually a sum over k -states) is given by

$$\frac{\partial f[r,k,t]}{\partial t} \Big|_s = \sum_{k'} \{ f[r,k',t](1-f[r,k,t])R[r,k',k,t] - f[r,k,t](1-f[r,k',t])R[r,k,k',t] \} \quad (6)$$

We shall be mostly concerned with phonon scattering for which the scattering rate from k to k' may be written as a sum over the phonon modes b as:

$$R[r,k,k',t] = \sum_{bK} \Lambda_b[k,k'] \left\{ \delta_{k-q-k',K} (f_{\text{phonon } b}[r,q,t] + 1) \delta(\epsilon(k') - \epsilon(k) + \hbar\omega_{bq}) \right. \\ \left. + \delta_{k+q-k',K} f_{\text{phonon } b}[r,q,t] \delta(\epsilon(k') - \epsilon(k) - \hbar\omega_{bq}) \right\} \quad (7)$$

Here Λ_b is the scattering kernel, K represents a reciprocal lattice vector for Umklapp processes, $\hbar\omega_{bq}$ is the energy of the b mode phonon at wave vector q . The delta-functions in energy and momentum represent the conservation of energy and momentum within a scattering event. The function $f_{\text{phonon } b}[r,q,t]$ is the probability distribution function for a phonon in mode b with wave vector q . The scattering rates are local in space and time, and evaluated in the Born approximation and do not depend on the applied electrostatic field. The first term in the brackets $\{ \dots \}$ of Equation (7) represents electron scattering with the emission of a phonon; the second term describes absorption of a phonon. If the phonon

distributions are in internal thermal equilibrium at a well-defined lattice temperature T they assume the form of a Bose-Einstein distribution:

$$f_{\text{phonon } b}[\mathbf{r}, \mathbf{q}, t] \rightarrow \left(e^{\frac{\hbar\omega_{bq}}{k_B T}} - 1 \right) \quad (8)$$

The Boltzmann transport equation is usually solved numerically by using the Monte-Carlo method [28]. Let us now specialise to the steady-state electro-thermal problem for which $\frac{\partial f}{\partial t} = 0$.

The local kinetic energy current density is given by:

$$\mathbf{J}_{\text{kinetic}}(\mathbf{r}) = \sum_{\mathbf{k}} \varepsilon(\mathbf{k}) \mathbf{j}(\mathbf{r}, \mathbf{k}) \rightarrow \int d^3k \rho_{\mathbf{k}} \varepsilon(\mathbf{k}) \mathbf{j}(\mathbf{r}, \mathbf{k}) = \int d^3k \rho_{\mathbf{k}} \varepsilon(\mathbf{k}) \mathbf{v}(\mathbf{k}) f[\mathbf{r}, \mathbf{k}] \quad (9)$$

where the summation over \mathbf{k} is replaced by integration by introducing $\rho_{\mathbf{k}}$ the density of \mathbf{k} -states.

We may obtain an expression for $\mathbf{J}_{\text{kinetic}}(\mathbf{r})$ by multiplying the steady-state version of Equation (4) by $\varepsilon(\mathbf{k})$ and summing over all \mathbf{k} to obtain after an integration by parts and re-arranging the collision integral:

$$\nabla \cdot \mathbf{J}_{\text{kinetic}}(\mathbf{r}) - \mathbf{E}(\mathbf{r}) \cdot \mathbf{J}(\mathbf{r}) = - \sum_{\mathbf{k}} \varepsilon(\mathbf{k}) \frac{\partial f[\mathbf{r}, \mathbf{k}]}{\partial t} \Big|_{\text{sc}} \quad (10)$$

$$= \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \{ \varepsilon(\mathbf{k}) - \varepsilon(\mathbf{k}') \} f[\mathbf{r}, \mathbf{k}] (1 - f[\mathbf{r}, \mathbf{k}']) R[\mathbf{r}, \mathbf{k}, \mathbf{k}'] = -P(\mathbf{r}) \quad (11)$$

$$P(\mathbf{r}) = -\nabla \cdot \mathbf{J}_{\text{Energy}}(\mathbf{r}) = -\nabla \cdot \mathbf{J}_{\text{kinetic}}(\mathbf{r}) + \mathbf{E}(\mathbf{r}) \cdot \mathbf{J}(\mathbf{r}) \quad (12)$$

where from Equation (7) $\varepsilon(\mathbf{k}') - \varepsilon(\mathbf{k})$ is the energy transferred to or absorbed from the lattice per unit time when an electron scatters from state \mathbf{k}' to state \mathbf{k} by emitting or absorbing a phonon. The explicit expression for the local power density $P(\mathbf{r})$ is

$$P(\mathbf{r}) = \sum_{\mathbf{k}} \sum_{\mathbf{k}'} \{ \varepsilon(\mathbf{k}') - \varepsilon(\mathbf{k}) \} f[\mathbf{r}, \mathbf{k}] (1 - f[\mathbf{r}, \mathbf{k}']) R[\mathbf{r}, \mathbf{k}, \mathbf{k}'] \quad (13)$$

showing that for *elastic* scattering processes $\varepsilon(\mathbf{k}') = \varepsilon(\mathbf{k})$, the contribution to the power density vanishes: no dissipation. Re-calling our discussion of ballistic transport it is evident from Equation (13) that in semi-classical transport the heat dissipation must occur where the scattering rate $R[\mathbf{r}, \mathbf{k}, \mathbf{k}']$ is non-zero.

Equation (12) makes the Joule heating explicit, where $\mathbf{E}(\mathbf{r})$ is the local electrical field and $\mathbf{J}(\mathbf{r}) = e\mathbf{j}(\mathbf{r})$ is the electrical current density. The Joule power density $\mathbf{E} \cdot \mathbf{J}$ when integrated throughout the device produces the familiar Joule power $I V$. The first term on the RHS of Equation (12) tracks the spatial variation of the “kinetic energy” current density of the electron ensemble. Crucially, if the kinetic energy current density $\mathbf{J}_{\text{kinetic}}(\mathbf{r})$ does not change with position \mathbf{r} the electron energy distribution will follow the bending of the electrostatic potential and the electron mean kinetic energy will not increase relative to the conduction band edge; therefore the energy distribution of the electrons will remain in local equilibrium, there will be no hot electron phenomena. On the other hand, if there is a large change in the spatial dependence of the kinetic energy current density there will be heating or cooling of the electronic sub-system.

Equations (9) and (10) lead directly to the thermodynamically significant detailed balance equations and have much wider generality than the classical picture.

The recognised route to simulating devices operating under non-isothermal conditions is to add a transport model to a standard heat equation. Using the local heat generation rate $P(\mathbf{r}, t)$ as the source of excess heat we obtain the heat equation (Fourier’s Equation) for the local lattice temperature $T(\mathbf{r}, t)$:

$$C_h \frac{\partial T(\mathbf{r}, t)}{\partial t} = \nabla \cdot \kappa \nabla T(\mathbf{r}, t) + P(\mathbf{r}, t) \quad (14)$$

where C_h is the total heat capacity and $\kappa(\mathbf{r}, t)$ is the local thermal conductivity.

For steady-state simulations Equation (8) reduces to:

$$\nabla \cdot \kappa \nabla T(\mathbf{r}) = -P(\mathbf{r}) \quad (15)$$

There are three main approaches to specify the heat source term P : (i) the Joule-heating model, exemplified by local equilibrium carriers commonly used in drift-diffusion modelling; (ii) assuming non-equilibrium electrons (hot electrons) interacting with equilibrium phonons; (iii) non-equilibrium electrons interacting with local equilibrium phonons.

(i) If we neglect electron-hole recombination and generation processes the Joule heating model [60] assumes:

$$P(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) \cdot \mathbf{J}(\mathbf{r}, t) \quad (16)$$

where \mathbf{E} is the local electric field and \mathbf{J} is the local electrical current density.

Evidently, the highest heating rate will occur where the scalar product is at a maximum. In conventional FETs most studies indicate that the maximum heating occurs under the gate where most of the potential drop takes place and where the electrical current density is the largest.

(ii) Here, we require an Equations (12)–(14) which picks up the transfer of energy from the electron system to the phonons in the local lattice. If the phonons are assumed to be in internal thermal equilibrium at temperature T then we can evaluate the net energy loss via the RHS of Equation (14) by using the matrix-elements and phonon distribution functions within the scattering rates R of Equation (7) by their equilibrium values [12,61]. More generally, we may evaluate a local lattice temperature $T(\mathbf{r})$ (from self-consistent solution of the coupled heat balance equation [62,63]).

(iii) To obtain full generality [34–36,64–69], we must introduce and solve self-consistently the full set of phonon transport equations for each phonon type at each location in the system including internal equilibrium processes such as phonon-phonon scattering and external processes such as boundary scattering. If the phonon distributions are parameterised by a different non-equilibrium value $T_b(\mathbf{r})$ for each mode b then it is still relatively easy to proceed; if not, the problem becomes a major challenge.

Further progress is only possible by resorting to a full space-dependent, energy-resolved quantum statistical approach to which we now turn.

2.2. LDOS and the Transition from Classical to Quantum Modelling

Consideration of the local energy density of states (LDOS) provides a clear example of the need for a transition to a quantum description of electro-thermal transport. Semi-classically, one assigns the particle momentum \mathbf{p} to a wave vector \mathbf{k} by $\mathbf{p} = \hbar \mathbf{k}$, where the underlying implication of a plane wave state $\exp[i\mathbf{k} \cdot \mathbf{r}]$ only has meaning in the evaluation of the scattering rate matrix elements in the collision integral. Its amplitude is normalised to unity. The local density of states is then defined as:

$$\rho_{LDOS}(\mathbf{r}, \varepsilon) = \sum_{\mathbf{k}} \delta(\varepsilon - \varepsilon(\mathbf{k}) - V(\mathbf{r})) = \int d^3k \rho_{\mathbf{k}} \delta(\varepsilon - \varepsilon(\mathbf{k}) - V(\mathbf{r})) \quad (17)$$

Noting that $\rho_{\mathbf{k}} = 2L^N/(2\pi)^N$ where N is the number of dimensions and L is the normalisation length, Equation (14) is easily evaluated for 3D, 2D and 1D systems as:

$$\rho_{LDOS}^{3D}(x, y, z, \varepsilon) = \left(\frac{m}{\pi^2 \hbar^2} \right) \sqrt{\frac{2m(\varepsilon - V(x, y, z) - \varepsilon_c)}{\hbar^2}} \quad (18)$$

$$\rho_{LDOS}^{2D}(x, y, \varepsilon) = \left(\frac{m}{\pi \hbar^2} \right) \theta(\varepsilon - V(x, y) - \varepsilon_c) \quad (19)$$

$$\rho_{LDOS}^{1D}(x, \varepsilon) = \left(\frac{2m}{\pi \hbar^2} \right) / \sqrt{\frac{2m(\varepsilon - V(x) - \varepsilon_c)}{\hbar^2}} \quad (20)$$

where m is the electron mass for a simple isotropic band structure, ϵ_c is the conduction band edge; it is assumed that the dimensionality factor $L = 1$ and θ is the unit step function. In the Appendix A, Figure A1, Equation (20) is illustrated classically and for a simple quantum generalisation for a typical band edge profile for a 1D NWFET.

The quantum origins of Equation (17) are very simple; consider for example, a general quantum Hamiltonian H that has a complete set of eigenvalues and eigenstates $\{ \epsilon_\alpha, | \alpha \rangle \}$ with completeness condition:

$$\sum_{\alpha} | \alpha \rangle \langle \alpha | = I \tag{21}$$

where I is a unit operator. The density of states in energy per unit volume follows as

$$\rho_{DOS}(\epsilon) = \sum_{\alpha} \delta(\epsilon - \epsilon_{\alpha}) = \sum_{\alpha} \langle \alpha | \delta(\epsilon - H) | \alpha \rangle \tag{22}$$

$$\rho_{DOS}(\epsilon) = Tr[\delta(\epsilon - H)] \tag{23}$$

$$\rho_{DOS}(\epsilon) = \sum_{\beta} \langle \beta | \delta(\epsilon - H) | \beta \rangle = \sum_{\beta, \alpha} | \langle \beta | \alpha \rangle |^2 \delta(\epsilon - \epsilon_{\alpha}) \tag{24}$$

where Equation (23) shows the representation independence of the trace operation. Equation (24) re-expresses the result in terms of any basis $\{ \epsilon_{\beta}, | \beta \rangle \}$ in particular if we choose the position representation $\{ | r \rangle \}$, Equation (24) leads to the local density of states

$$\rho_{LDOS}(r, \epsilon) = \sum_{\alpha} | \langle r | \alpha \rangle |^2 \delta(\epsilon - \epsilon_{\alpha}) = \langle r | \delta(\epsilon - H) | r \rangle \tag{25}$$

For a simple free electron model the eigenstates α are just k -states : $\langle r | k \rangle = \exp[ik \cdot r]$ with eigenvalues $\epsilon(k)$ and we recover Equation (14) for $V = 0$.

By using the operator identity

$$\delta(\epsilon - H) = \lim_{\eta \rightarrow 0^+} \frac{1}{2\pi i} \{ (\epsilon - i\eta - H)^{-1} - (\epsilon + i\eta - H)^{-1} \} \tag{26}$$

Equation (25) may be formulated in terms of the advanced and retarded Green's operators for the steady-state operator equation

$$(\epsilon + i\eta - H)G^R = I; (\epsilon - i\eta - H)G^A = I \tag{27}$$

$$\begin{aligned} \rho_{LDOS}(r, \epsilon) &= \langle r | \delta(\epsilon - H) | r \rangle = \frac{-1}{2\pi i} \left\{ G^R(r, r', \epsilon) - G^A(r, r', \epsilon) \right\} \Big|_{r' \rightarrow r} \equiv A(r, \epsilon) \\ &= \left(\frac{-1}{\pi} \right) \text{Im} G^R(r, r', \epsilon)_{\infty} \end{aligned} \tag{28}$$

In the position representation Equations (27) are matrix equations for the Green functions $G^R(r, r', \epsilon)$, $G^A(r, r', \epsilon)$; they determine the description of the energy structure of the system.

In thermal equilibrium, the energy distribution of electronic states will follow a Fermi-Dirac distribution

$$f_0(\epsilon - \mu) = \left\{ 1 + \exp \left[\frac{(\epsilon - \mu)}{k_B T} \right] \right\}^{-1} \tag{29}$$

so that the equilibrium particle density will be

$$\begin{aligned} n_0(r, \epsilon) &= \sum_{\alpha} | \langle r | \alpha \rangle |^2 \delta(\epsilon - \epsilon_{\alpha}) f_0(\epsilon - \epsilon_{\alpha} - \mu) \\ &= \langle r | \delta(\epsilon - H) f_0(\epsilon - H - \mu N) | r \rangle \equiv \frac{-i}{2\pi} G_0^<(r, r', \epsilon) \Big|_{r' \rightarrow r} \end{aligned} \tag{30}$$

Here we introduce the statistical (superscript “<” denotes “lesser than”) equilibrium Green function. This is easy to evaluate if the eigenstates of H are known. Equations (21)–(30) provide a simple account of how the NEGF formalism arises.

However, if H is split into an interaction free Hamiltonian H_0 (which may include the Hamiltonian for non-interacting phonons) and an interaction term H_{int} describing the electron-phonon interactions, then the trace operation in Equations (28) and (30) must be widened to include a trace over the phonon states to maintain completeness. A self-consistent perturbation theory therefore requires a simultaneous treatment of the states within thermal equilibrium and the dynamical evolution of the states in non-equilibrium.

The route to doing this was laid out in [41–44] in the 1960s, but was difficult to apply, partly because of computational resources. Important technical issues and improvements were made much later [40,70–74]

2.3. The NEGF Formalism

The NEGF formalism is ultimately based on statistical quantum field theory; in the following we will draw from references [41–44,70–74], but we will only review the parts of that theory necessary to consider electro-thermal transport and we avoid the very technical derivations and issues. An excellent practical introduction to the NEGF as applied to devices will be found in works and references therein by Datta [57,58], background many-body perturbation theory is discussed in detail in reference [71]. The Keldysh approach is outlined in 2.3.1 but may be skipped to proceed directly to the required steady-state NEGF equations that are described in Section 2.3.2. More technical details will be found in the Appendix, Sections Appendices B–G.

2.3.1. The Keldysh Picture

Let us start with a generic second-quantised many-body Hamiltonian $H = H_0 + H_{int}$ describing a system of interacting (via H_{int}) electrons and phonons. Any property of the system may be represented by an operator Ξ and its statistical expectation value. $\langle \Xi \rangle$ may be computed in thermal equilibrium from the grand canonical ensemble ρ_{eq} of statistical mechanics giving:

$$\langle \Xi(t) \rangle_{eq} = \frac{\text{Tr}[\Xi \rho_{eq}]}{\text{Tr}[\rho_{eq}]}; \rho_{eq} = \exp\left[-\frac{H - \mu N}{k_B T}\right] \rho_{eq} = \exp\left[-\frac{H - \mu N}{k_B T}\right] \quad (31)$$

where T is the absolute temperature and μ is the chemical potential. Here N represent the second-quantised number operator for excitations (electrons and phonons). The trace operation $\text{Tr}[\dots]$ means sum the diagonal matrix elements of the operand in any complete set of many-body states.

An important example of interest is the single electron equilibrium Green function that has a two-point space-time structure:

$$\langle \Xi(\mathbf{r}, \mathbf{r}', t, t') \rangle_{eq} \equiv G_{eq}(\mathbf{r}, \mathbf{r}', t, t') \equiv \langle T \psi(\mathbf{r}, t) \psi^\dagger(\mathbf{r}', t') \rangle_{eq} \quad (32)$$

Equation (32) represents the probability amplitude for the insertion of an electron at (\mathbf{r}', t') and the removal of an electron at (\mathbf{r}, t) . The operators $\psi(\mathbf{r}, t)$, $\psi^\dagger(\mathbf{r}, t)$ are anti-commuting electron annihilation and creation field operators in the Heisenberg picture (state vectors constant, operators time-dependent.) T is the Dyson time-ordering operator which for fermions re-orders the field operators so that the term with earliest time is on the right and also multiplies the result by +1 or –1 depending on whether the new order is an even or odd permutation of the original product.

Different time-orderings of Equation (32) give Green functions with different behaviour. The Green functions $G_{eq}^<$, $G_{eq}^>$ (cf. Section 2.2) are the specific time-orderings:

$$G_{eq}^<(\mathbf{r}, \mathbf{r}', t, t') = G_{eq}(\mathbf{r}, \mathbf{r}', t, t') \quad (t < t') \quad (33)$$

$$G_{eq}^>(r, r', t, t') = G_{eq}(r, r', t, t')(t > t') \tag{34}$$

$$G_{eq}^R(r, r', t, t') = -i\theta(t - t') < \{ \psi^\dagger(r', t'), \psi(r, t) \} >_{eq} \tag{35}$$

$$G_{eq}^A(r, r', t, t') = i\theta(t' - t) < \{ \psi(r, t), \psi^\dagger(r', t') \} >_{eq} \tag{36}$$

The “lesser/greater” Green functions $G^</math>/ $G^>$ are so named by the conditions $t < t'$ or $t > t'$). Here, the brackets $\{ , \}$ represent the anti-commutator.$

Using the above Equations (31) and (32) it is relatively straightforward to derive Equations (28) and (30). Equation (31) plus considerable manipulation yields an important boundary condition that relates $G^<$ to $G^>$ in thermal equilibrium:

$$G_{eq}^<(r, r', t, t')|_{t=0} = -\exp[\beta\mu]G_{eq}^>(r, r', t, t')|_{t=-i\hbar\beta} \tag{37}$$

The latter is the basis for developing finite temperature thermal equilibrium perturbation theory in complex time, see reference [71]

The non-equilibrium problem is to start at some initial time t_0 with a system (such as a device) in thermal equilibrium and then initiate an external perturbation H_{ext} (applied fields or temperature gradients) that drives the system into non-equilibrium. To establish an infinite order perturbation theory for interacting systems we write the full Hamiltonian as

$$H = H_0 + H_{ext}(t) + H_{int} \tag{38}$$

where H_0 describes free excitations (electrons, phonons in our case); $H_{ext}(t)$ describes coupling to external fields switched on *at some time* $> t_0$; and H_{int} represents interaction between excitations (electron-phonon coupling in our case). The system density matrix satisfies the equation of motion:

$$i\hbar \frac{\partial \rho}{\partial t} = [H, \rho] \tag{39}$$

Assuming adiabatic switching-on of the interactions, the density matrix at time t is in the interaction picture given by the S -matrix (evolution operator):

$$\rho(t) = S(t, -\infty)\rho(-\infty)S^\dagger(t, -\infty) \tag{40}$$

$$S(t, t') = Texp\left[-i/\hbar \int_{t'}^t dt'' H_{int}(t'')\right] \tag{41}$$

$$S^\dagger(t, t') = Texp\left[-i/\hbar \int_t^{t'} dt'' H_{int}(t'')\right] \tag{42}$$

The time ordering is along the real-time integration paths.

$$\langle \Xi(t) \rangle = Tr[\Xi(t)\rho(t)] = Tr[S^\dagger(t, -\infty)\Xi(t)S(t, \infty)\rho(-\infty)] \tag{43}$$

In a zero-temperature many-body systems with a stable ground state $|0\rangle$ (vacuum state) it is possible to reduce Equation (32) to:

$$\langle \Xi(t) \rangle = \frac{\langle 0|T[S(-\infty, \infty)\Xi(t)]|0 \rangle}{\langle 0|S(-\infty, \infty)|0 \rangle} \tag{44}$$

where the time-integrals are along the real time axis from $-\infty$ to ∞ . The familiar many-body perturbation theory then follows using Wick’s theorem [71] that provides cancellation of divergent terms in the

perturbation expansion of the numerator of Equation (33). This simplification does not hold for non-equilibrium systems. Instead, in the Keldysh formalism Equation (33) is written as

$$\langle \Xi(t) \rangle_{0eq} = Tr[T_{\text{contour}} S_{\text{contour}} \Xi(t) \rho_{0eq}] \tag{45}$$

where the S -matrix, S_{contour} , is evaluated along a time contour C_{Keldysh} that runs from $-\infty$ to t (C^+) and then back from t to $-\infty$ (C^-).

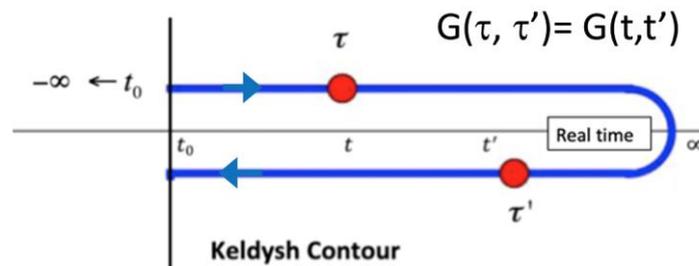


Figure 1. The Keldysh Contour. The general single particle Green function is the expectation value $G(r,r',\tau,\tau') \equiv -i\langle T_{\text{Keldysh}} \psi(r,\tau) \psi^\dagger(r',\tau') \rangle_{0eq}$ where the time-ordering operator orders the field operators according to the temporal positions τ, τ' on the Keldysh contour. The temporal values of τ, τ' are denoted by the corresponding real times t, t' . In the figure are shown the temporal positions and values for the “lesser” Green Function defined for all times τ on the upper branch (C^-), and for all times τ' on the lower branch (C^+), $G^<(r,r',\tau,\tau') \equiv -i\langle T_{\text{Keldysh}} \psi(r,\tau) \psi^\dagger(r',\tau') \rangle_{0eq}$.

Figure 1 displays the Keldysh contour showing the arguments (red) of the contour ordered Green function $G^<(t,t')$. T_{contour} is the time ordering operator on the contour $C_{\text{contour}} = C_{\text{Keldysh}}$. By inserting the identity operator $S(t,\infty)S(\infty,t)$ the Keldysh contour is extended from $(-\infty,\infty)$ and back again: $(\infty,-\infty)$.

In Equation (45) ρ_{0eq} is the equilibrium density matrix for non-interacting excitations. The subsequent time-dependent many-body infinite perturbation theory then follows a similar Feynman diagram expansion as for zero temperature theory but involves integrations along the Keldysh contour. The perturbation theory is most easily derived using equations of motion for a hierarchy of Green functions.

In the steady-state, neglecting initial state correlations and transients, and choosing a time-independent external perturbation H_{ext} , the basic single-particle electron Green functions may be expressed in the Heisenberg picture:

$$\Xi(t) = \exp\left[\frac{i(H_0 + H_{ext} + H_{int})(t - t_0)}{\hbar}\right] \Xi(t_0) \exp\left[\frac{-i(H_0 + H_{ext} + H_{int})(t - t_0)}{\hbar}\right] \tag{46}$$

where now the Hamiltonian H has been augmented to include the perturbing interaction H_{ext} .

In the case of the single particle electron Green function $G(r,r',t,t')$ we find:

$$G(r,r',t,t') \equiv -i\langle T_{\text{Keldysh}} \psi(r,t) \psi^\dagger(r',t') \rangle_{0eq} \tag{47}$$

where as before the terms $\psi(r,t), \psi^\dagger(r,t)$ are annihilation, creation operators in the Heisenberg picture Equation (46) for the electron quantum fields, time-ordered by T_{Keldysh} (the Dyson time-ordering operator on the Keldysh contour). Equation (47) may be partitioned into four piece-wise defined Green functions chosen according to the location of the time variables t,t' on the Keldysh contour. A common choice, which is adopted here, is closest to the Kadanoff-Baym definitions (a better choice for constructing equations of motion and self-energies is detailed by Lifshitz and Pitaevskii [70]):

$$G^<(r,r',t,t') = G(r,r',t,t') (t < t') = i\langle \psi^\dagger(r',t') \psi(r,t) \rangle_{0eq} \tag{48}$$

$$G^>(r, r', t, t') = G(r, r', t, t') \quad (t > t') \quad (49)$$

$$G^R(r, r', t, t') = -i\theta(t - t') < \{ \psi^\dagger(r', t'), \psi(r, t) \} >_{0eq} \quad (50)$$

$$G^A(r, r', t, t') = i\theta(t' - t) < \{ \psi(r, t), \psi^\dagger(r', t') \} >_{0eq} \quad (51)$$

The Heisenberg equations of motion for these Green functions are coupled differential—integral equations that may be obtained by differentiating with respect to the two variables t and t' (and r, r'). These equations immediately show up coupling to the two-electron Green functions that in turn are coupled to higher order Green functions to form a well-known hierarchy. Within systematic perturbation theory the hierarchy is curtailed by introducing self-energy terms that close off the hierarchy by simplifying the higher order Green functions. Under these conditions the equations of motion for the piece-wise time sub-domain Green functions are coupled but together form a closed system of equations. Using the Langreth theorem [40], which connects the Keldysh contour ordered Green functions to the real-time Green functions $G^<, G^>, G^R, G^A$, all the relevant Green functions satisfy a Dyson-like form with a suitable self-energy $\Sigma(r, r', t, t')$:

$$G^\lambda(r, r', t, t') = G_0^\lambda(r, r', t, t') + \iint \iint d^3r'' d^3r''' d t'' d t''' \{ G_0^\lambda(r, r'', t, t'') \Sigma(r'', r''', t'', t''') \} G^\lambda(r''', r', t''', t') \quad (52)$$

Here, G_0^λ is the relevant unperturbed Green function ($H_{int} = 0$).

Because the piece-wise Green functions are not independent, the self-energies are functionals of the other self-energies and Green functions.

Considering now a non-equilibrium steady-state, energy-resolved picture let us transform to new variables $\tau_{relative} = t - t'$; $\tau_{mean} = (t + t')/2$; then performing a Fourier transform of the Green functions over the relative time variable $\tau_{relative}$ yields the same forms described in elementary fashion in Section 2.2:

$$G^<(r, r', \varepsilon) = \left(\frac{1}{\hbar} \right) \int_{-\infty}^{\infty} d\tau_{relative} \exp \left[\frac{i\varepsilon\tau_{relative}}{\hbar} \right] G^<(r, r', \tau_{relative}, \tau_{mean}) \quad (53)$$

(where there is no dependence on τ_{mean} in the steady state) and similarly for $G^>, G^A, G^R$. The Dyson Equation (52) become matrix equations in the space variables and the self-energies become convolution integrals over intermediate energies.

2.3.2. The Steady State NEGF Equations

For the device Hamiltonian comprising an uncoupled term, an interaction term H_{int} coupling electrons to phonons, and an applied perturbation H_{ext} resulting from the applied electric fields and the confinement potential:

$$H = H_0 + H_{ext} + H_{int} \quad (54)$$

The resulting steady-state coupled Green function equations are:

$$G^R = \{ \varepsilon + i\eta - H_0 - H_{ext} - \Sigma^R \}^{-1} = G_0^R + G_0^R \Sigma^R G^R \quad (55)$$

$$G^A = G^{R\dagger} \quad (56)$$

$$G^< = G^R \Sigma^< G^A; \quad G^> = G^R \Sigma^> G^A \quad (57)$$

Equation (55) is essentially a Dyson equation involving the self-energy Σ^R which is a functional of the other Green functions and self-energies. Equation (57) is deceptively simple but it is the analogue of the classical Boltzmann equation as discussed briefly in Appendix C. Indeed from Equations (55) and (57) we have

$$\{ \varepsilon + i\eta - H_0 - H_{ext} \} G^< = \Sigma^R G^< + \Sigma^< G^A \quad (58)$$

2.3.3. The Projection Theorem

So far we have tacitly treated an open system. A key requirement in NEGF theory applied to device modelling is to project from the open system comprising a device coupled to its environment of contacts and thermal reservoirs into a finite simulation domain (see Appendix B, Figure A3). This non-trivial matter was first solved by Caroli et al in 1971 and their derivation of the theorem is sketched in the Appendix B. The main result is that the finite device domain may be modelled by the device Hamiltonian H_W (W for nanowire) provided that an additional set of non-Hermitian self-energies be added to H_W for each relevant Green function. These contact self-energies $\Sigma_{contacts}^<$, $\Sigma_{contacts}^>$, $\Sigma_{contacts}^R$, $\Sigma_{contacts}^A$ are localised interactions on each contact-device boundary surface that describe the injection or extraction of carriers (and consequently the inflow and outflow of charge, current and energy).

2.3.4. Numerical Solution of NEGF Equations

Equations (55)–(57) may be evaluated in the position representation, by inserting complete sets of position eigenstates; if the position variables are discretised the result is a set of matrix equations. The problem is significantly simplified if the self-energies are diagonal. We further require expressions for H_{ext} and the self-energies Σ^λ , $\Sigma_{contacts}^\lambda$.

In solving the NEGF equations, there are three other equations to be solved self-consistently:

(i) The scattering self-energies must be evaluated self-consistently in at least the self-consistent Born approximation (fast algorithms for this and higher order perturbation are described in [75]. Appendix D gives an example of $\Sigma^<$, Σ^R in the SCBA for a generic electron-phonon interaction.

(ii) Poisson's equation for the self-consistent applied potential and confinement potential (necessary for image charge induced electrostatic self-energies [10]) must be evaluated.

(iii) If the lattice is considered to be held at constant ambient temperature, there is no further issue; if not, the heat equation(s) driven by the heat dissipation into the phonon system must be considered self-consistently. In the most advanced modelling, this latter step has been replaced by coupling the above equations self-consistently to the phonon Green functions [50,75]. A few groups use local atomic or molecular orbitals as the basis for tight binding style Hamiltonian models, but these techniques are restricted to a few hundred atoms [76]. The effective mass Hamiltonian method [76,77] is derived from tight binding and is discussed further in Section 3.

2.3.5. NEGF Electro-Thermal Modelling

The NEGF equations are very powerful compared to Boltzmann transport theory because they are able to describe quantum phenomena such as tunnelling, resonances, bound states, fully quantised scattering theory, coherent and non-coherent flows. The density of states particularly LDOS can then reveal shift and broadening in the energy levels) due to the scattering self-energies/ contact self-energies) and hot spots in the spatially-resolved energy current densities. In the latter case we may immediately recover the quantum equivalent of the classical Equations (11) and (12):

The NEGF equivalents of the electro-thermally important quantities described in Section 2.1 are as follows.

The energy- and space- resolved charge density involves only the diagonal part of $G^<(r, r', \epsilon)$ (following from the limit ($r \rightarrow r'$, $t \rightarrow t'$), of Equations (47) and (48))

$$n(r, \epsilon) = -i \left(\frac{e}{2\pi} \right) G^<(r, r', \epsilon) \Big|_{r' \rightarrow r} \quad (59)$$

For a general Hamiltonian, H the velocity operator is defined in the Heisenberg picture by $v \equiv dr/dt = \left(\frac{i}{\hbar} \right) [H, r]$.

For a simple effective mass Hamiltonian, the (electrical) energy-resolved current density is:

$$J(r, \epsilon) = ej(r, \epsilon) = (-e/4\pi m^*) (\nabla - \nabla') G^<(r, r', \epsilon) \Big|_{r' \rightarrow r} \quad (60)$$

$$J_{energy}(\mathbf{r}) = \int \varepsilon \mathbf{j}(\mathbf{r}, \varepsilon) d\varepsilon \quad (61)$$

The local power transfer to the lattice is

$$P(\mathbf{r}) = \nabla \cdot J_{energy}(\mathbf{r}) \quad (62)$$

or, separating out the Joule power,

$$P(\mathbf{r}) = -\nabla \cdot J_{kinetic}(\mathbf{r}) + E(\mathbf{r}) \cdot J(\mathbf{r}) \quad (63)$$

where the first term on the RHS of Equation (63) is the divergence of the electron kinetic energy current density, while the second term is the product of the local electric field and the current density and represents the local Joule heat. The LHS is the net energy flow from the electron gas to the phonon assembly and is the source term for a heat equation (s). Equation (63) is derived by writing the kinetic energy $\varepsilon_{kinetic}(\mathbf{r}, \varepsilon) = \varepsilon - e\phi(\mathbf{r})$. The kinetic energy current density is then

$$J_{kinetic}(\mathbf{r}) = \int \varepsilon_{kinetic} \mathbf{j}(\mathbf{r}, \varepsilon) d\varepsilon \quad (64)$$

and

$$J_{Energy}(\mathbf{r}) = J_{kinetic}(\mathbf{r}) + \phi(\mathbf{r})J(\mathbf{r}) \quad (65)$$

Taking the divergence of Equation (65) and using $E(\mathbf{r}) = -\nabla\phi(\mathbf{r})$ plus the continuity equation result $\nabla \cdot J(\mathbf{r}) = 0$, recovers Equation (63).

Using Equations (55)–(57), we obtain an expression for the divergence of the total energy current density

$$\nabla \cdot J_{Energy}(\mathbf{r}) = \int d\varepsilon \varepsilon \int d^3r' \left\{ \left[G^R(\mathbf{r}, \mathbf{r}', \varepsilon) - G^A(\mathbf{r}, \mathbf{r}', \varepsilon) \right] \Sigma^<(\mathbf{r}', \mathbf{r}, \varepsilon) + \right. \\ \left. G^<(\mathbf{r}, \mathbf{r}', \varepsilon) \Sigma^A(\mathbf{r}', \mathbf{r}, \varepsilon) - \Sigma^R(\mathbf{r}, \mathbf{r}', \varepsilon) G^<(\mathbf{r}', \mathbf{r}, \varepsilon) \right\} \quad (66)$$

$$\nabla \cdot J_{Energy}(\mathbf{r}) = \int d\varepsilon \varepsilon \langle \mathbf{r} | \left\{ G^R(\varepsilon) - G^A(\varepsilon) \right\} \Sigma^<(\varepsilon) + G^<(\varepsilon) \Sigma^A(\varepsilon) - \Sigma^R(\varepsilon) G^<(\varepsilon) | \mathbf{r} \rangle \quad (67)$$

The operator within the position matrix element of Equation (67) is related to the current operator as discussed in the Appendices D–F.

Since the self-energies here include coupling to the contacts it is possible to use Equation (66) to locate precisely where the energy dissipation occurs (for example, within particular device regions or at the contacts).

The above equations plus explicit forms for the electron-phonon interaction self-energies (see Appendix D) give the quantum equivalent of the classical electro-thermal formulation of Section 2.1, i.e., Equations (11) and (13). It is important to demonstrate that expression $\nabla \cdot J_{Energy}(\mathbf{r})$ vanishes for elastic scattering. This is demonstrated in Appendix E, where $\nabla \cdot J_{Energy}(\mathbf{r})$ is shown explicitly to vanish everywhere within the device volume using the self-energies for high temperature elastic acoustic phonon scattering in the self-consistent Born approximation. For the elastic case, there is no energy dissipation within the channel but when the contact self-energies are introduced there may still be energy dissipation in the drain contact region.

3. Simulation Methodology for the Silicon Gate-All-Around Nanowire FET

In this section we describe the carrier transport model and our electrothermal model used in the simulation of a gate-all-around nanowire transistor. In a nutshell, our method simultaneously solves the NEGF Equations (55)–(57), the Poisson Equation (3) and heat Equation (15) with the appropriate boundary conditions.

We start by estimating the electrostatic potential and use it to solve the NEGF Equations (55)–(57) for a range of energies; these two equations need to be solved self-consistently as electron-phonon scattering self-energies depend on the electron density. From the calculated Green functions, the electron density is obtained and inserted into the Poisson equation to calculate the new electrostatic potential. At the same time, the current and Green functions are used to calculate the local power dissipation that is the input into the heat equation in order to find the temperature profile. This process continues until density, current and temperature do not change. In what follows, we will describe in detail the different physical models, boundary conditions and simplifications used in solving the equations.

3.1. Device Structure and the 1D Representation

In this work, the effective mass approximation [78–80] has been used to approximate the non-interacting Hamiltonian H_0 of the silicon crystal entering in Equation (54). The effective masses used have been extracted from tight-binding calculations, see reference [80]. In addition, our model for Si considers anisotropic bands and includes the six Δ -valleys, see Appendix A and Figure A2 for the k -space location of the valleys. The axis of the wire is oriented in the [100] direction, for a square cross section, 4 ellipsoids are equivalent as shown in Figure A2.

In devices with cylindrical or rectangular symmetry the spatial dependence of the Green functions may be reduced to 1D forms (in, x, x') by projecting out the transverse (y, z) dimensional dependence using self-consistent transverse eigenstates (discrete sub-bands n) computed at each point along the x -direction (see Section 2.3 and references therein). The full Green functions are then replaced by effective 1D Green functions: $G(x, x', \epsilon)$. This simplification is a key result for the efficient modelling of nanowire structures.

The device (typified by Figure 2) is divided in cross-sections perpendicular to the axis of the wire. The standard Schrödinger equation is solved in each cross-section by using the electrostatic potential calculated by the Poisson equation. This results in a set of eigenstates for each valley and cross-section considered. These eigenstates are commonly called sub-bands or modes and are associated with individual conduction band valleys. The longitudinal Hamiltonian is expanded in these sub-band states that are coupled to each other by the relevant overlap integrals. This method is called the coupled mode space method [81–84] and is very efficient for very narrow nanowires as the number of sub-bands considered can be kept to a minimum. In detail, the full volume of the device simulation domain is discretised in a rectangular mesh; the Schrödinger equation is then solved for each cross-section perpendicular to the wire axis. The potential energy entering in the equation is the electrostatic potential energy calculated from the Poisson equation and the confining potential at the SiO_2 interface. The kinetic energy is given by the anisotropic mass tensor corresponding to the particular valley. From the solution of the Schrödinger equation, a set of eigenvalues and eigenvectors (wavefunctions) are obtained for each cross section. These eigenvectors or modes are used to expand the 3D spatially discretized Green function [84] of size $(n_x \times n_y \times n_z, n_x \times n_y \times n_z)$, where n_y and n_z is the number of discretisation points in the cross-section direction and n_x in the transport direction. From these modes and by integration in the transversal spatial coordinates, an equation for the Green function is constructed that couples the modes in one cross section with the modes of the nearest neighbouring cross-sections [83,84]. The size of this Green function matrix is of order $(n_m \times n_x, n_m \times n_x)$, where n_m is the total number of modes used in the simulations. Note that the 3D spatially resolved Green function is substantially larger in size than the mode space Green function as mentioned before. In the discretised Hamiltonian, the matrix element (or coupling) between mode m in cross section i and mode l in cross section $i + 1$ is given by:

$$-\frac{\hbar^2}{2m_x(\Delta x)^2} \iint \psi_i^{*m}(y, z) \psi_{i+1}^l(y, z) dy dz \quad (68)$$

where $\psi_i^m(y, z)$ is the wavefunction for mode m in cross section i , the y and z are the spatial coordinates representing the cross section. Δx is the discretisation step in the longitudinal direction or the transport direction. The asterisk $*$ in the above expression stands for the complex conjugate. The double integral is over the whole cross-section. Equation (68) shows that the probability of an electron going from mode m in cross section i to mode l in cross-section $i + 1$ depends on the overlap between wavefunctions. It should be noted that if the potential energy in the cross-sections are similar the above integral between two different modes is zero. This means that there is a very small probability (or amplitude) for the electrons during their motion along the wire to switch modes. When the integral between different modes is assumed zero, the simulation method is called the uncoupled mode space approach. In this case the Green function in the mode/longitudinal space reduces to n_m Green functions of size $(n_x \times n_x)$. This reduces the computational load substantially. After the longitudinal Green function is obtained for each valley, mode and energy, the electron density and current is calculated from them and the mode wavefunctions. For details of the coupled mode space as applied to NEGF, we recommend the following references [83,84].

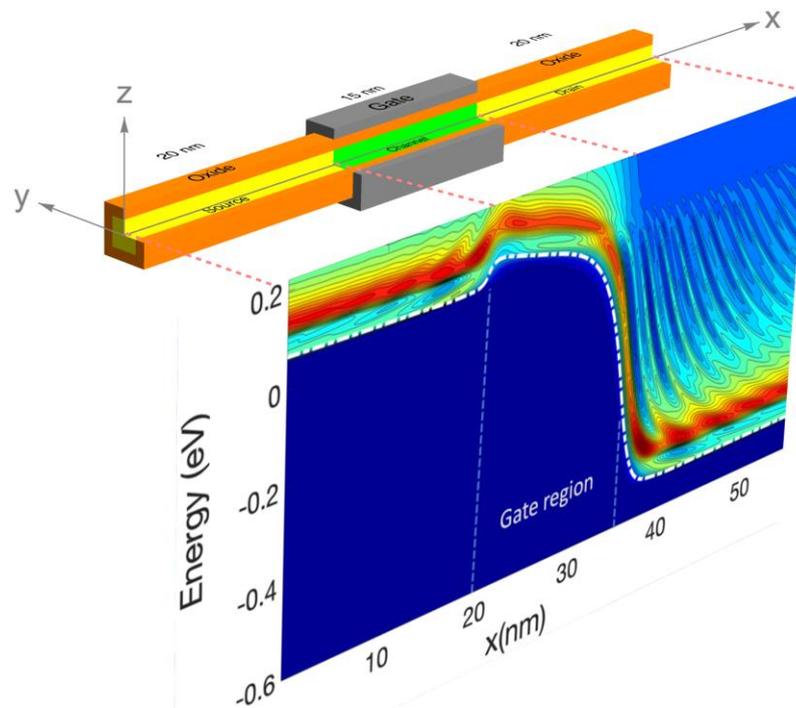


Figure 2. Schematics of a wrap-round gate silicon nanowire field-effect transistor of length 55 nm cut away to expose the channel region. The oxide thickness is 0.8 nm and the channel width is 2.2 nm. The channel length (gate length) is 15 nm, the source and drain extensions are 20 nm. The work function of the metal gate is 4.5 eV. The projected image shows a typical non-equilibrium Green function (NEGF)-computed local density of states (LDOS) for the device assuming a drain bias of 0.6V and a gate bias at 0.6 V. The location of the first 1D sub-band edge is shown along the central nanowire x-axis as a function of energy and position (dashed white line); it mimics the self-consistent potential along the same axis. In the projected image, the vertical energy scale is from -0.6 eV to 0.23 eV.

We have considered the acoustic and optical (for f-type and g-type) intervalley scattering mechanisms, as well as the intra-valley elastic acoustic phonons scattering mechanism. The specific parameters and models used can be found in references [85,86]. The electrostatic potential is calculated from the Poisson equation in the mean field approximation. We have used Neumann boundary conditions in the source and drain contacts; and Dirichlet boundary conditions at the gate contact.

Figure 2 shows the schematics of a typical wrap-round gate silicon nanowire with a silicon core comprising source, channel and drain regions. There is a high degree of symmetry. Therefore, we may use the sub-band projection methodology to obtain the 1D density of states throughout the silicon core as a function of energy and position as indicated.

Figure 3 shows the detailed 1D local density of states (LDOS) computed along the nanowire axis for each sub-band energy: it is proportional to $(\varepsilon - \varepsilon_n)^{-1/2}$ where ε_n is the local sub-band energy. Notably, the multiple reflection patterns (a simple example is in Appendix A) that are shown for energies lower than the top of the maximum sub-band energy around the middle of the channel. This is an indication of coherent transport. The wavelength of the electrons increases as the energy increases; this can be confirmed by the shorter separation of the maxima and minima in the figure.

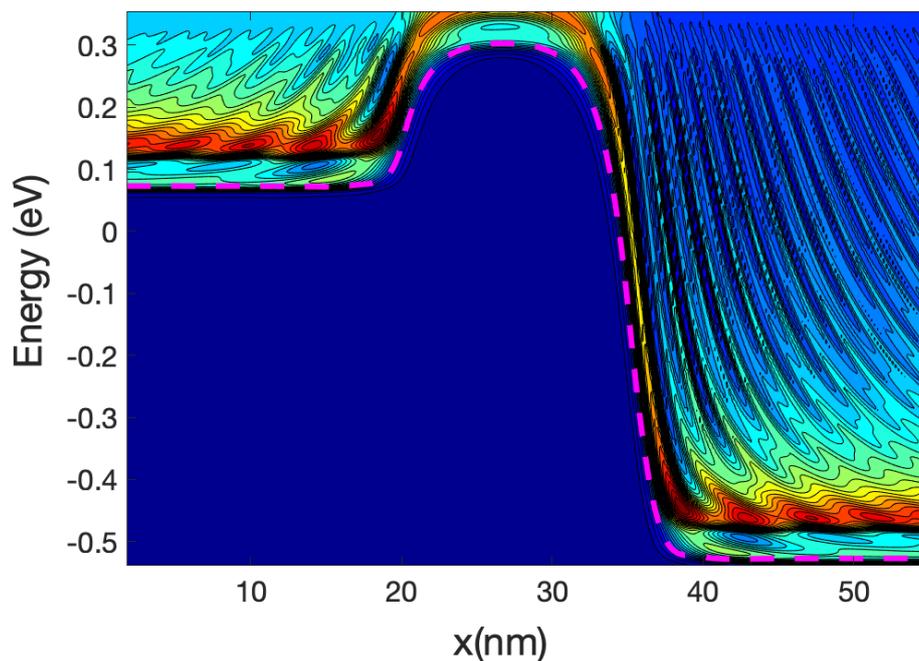


Figure 3. Local density of states (LDOS) along the nanowire axis. The dashed line depicts the first sub-band edge. Dark red indicates high LDOS and dark blue low LDOS. The drain bias is 0.6 V and the gate bias is 0.2 V.

3.2. Non Equilibrium Green Function Formalism: Density and Current Density

Equations (41) and (42), used for the calculation of the Green function, implicitly contain Pauli's exclusion principle [57,58,87]. There, $\Sigma^<$ is a sum of two terms, one is proportional to the electron-phonon scattering rate and the other representing the boundary self-energies that are different from zero only at the contacts. This latter term involves the product of the electron distribution of the contact and the density of states at the contact. This approximation in the contacts is a consequence of the fluctuation-dissipation theorem [88]. In one dimension, for each mode, the energy-resolved electron concentration and current density could be calculated from the mode-space Green functions by [83,84]:

$$n_l(x, \varepsilon) = -\left(\frac{i}{2\pi}\right)G_{ll}^<(x, x, \varepsilon) \quad (69)$$

$$J_{nl}(x, \varepsilon) = -\frac{e}{4\pi m^*} \lim_{x' \rightarrow x} \left(\frac{\partial}{\partial x} - \frac{\partial}{\partial x'}\right) G_{nl}^<(x, x', \varepsilon) \quad (70)$$

where x represents the spatial coordinate, e the electron charge and m^* the effective mass. The n and l in the RHS label the modes. Equations (69) and (70) must be modified by a factor of two for spin

degeneracy. The total 3D electron density is calculated from the above 1D electron density and the cross-section wave function by:

$$n(\mathbf{r}, \varepsilon) = \sum_l n_l(x, \varepsilon) \psi_x^{*l}(y, z) \psi_x^l(y, z) \quad (71)$$

Figure 4 shows the electron density at low and high gate bias, showing electrical neutrality, i.e., the potential is flat at the contact and the electron concentrations are similar in the source and drain contact. The electron density drops close to the oxide interface showing the volume inversion and the shape of the ground state transversal wave function. For narrow-body nanowires, this concentration of electrons along the wire axis is usually called the volume inversion or electron confinement. At very high gate voltage, the source-drain barrier disappears and, in order to fulfil electrical neutrality, the leads or ends of the source and drain region should be broadened (see Landauer [59]). At high gate bias, the electron potential energy along the wire direction resembles a step function or a linear drop with an energy difference proportional to the applied bias, so the right moving waves from the source reach the drain but the left moving waves from the drain are reflected by the energy drop. This means that the density of electrons in the drain will be always larger than in the source and therefore electrical neutrality would not be fulfilled. However, if simulations are carried out with broadened leads, electrical neutrality in the source/drain are independent of the gate potential and are controlled by the electrostatic of the leads.

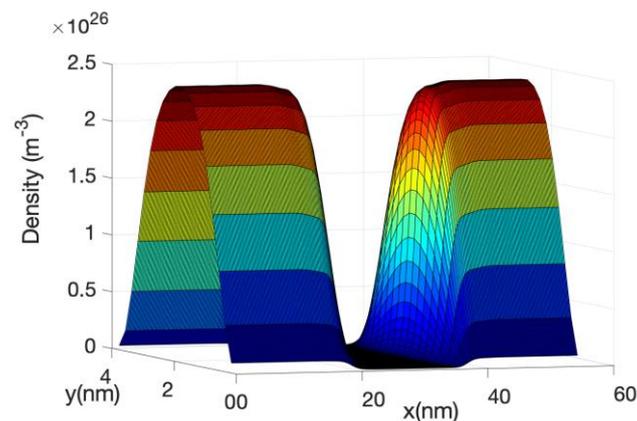


Figure 4. The electron density along the nanowire axis in a parallel xy plane at $V_g = 0.4$ V and $V_d = 0.6$ V. The figure shows the volume inversion and also the fulfilment of charge neutrality, see text.

Figure 5 shows the energy resolved current spectra calculated by Equations (69) and (70) along the nanowire transistor axis. The source contact is in the left end and the drain is at the right end. The first sub-band is shown as reference. The part of the current with energy lower than the top of the barrier sub-band energy represents tunneling current. As the device has a 15 nm channel length, the tunneling current is not negligible. From the figure it can be seen that electron at the source moving from the source towards the drain comes from two different energy regions:

- (i) energy close to the top of the barrier and
- (ii) energy close to the sub-band energy at the source contact.

Electrons entering the source region (i) will be able to go through the barrier, unless they emit phonons and lose energy. Those in region (ii) need to absorb phonons in order to cross the barrier and in doing so will cool the lattice locally. The temperature profile simulations presented in the next section confirm this statement. At the drain side, electrons injected from the source (at high energy) start to dissipate part of the energy and the current at the drain is more spread downwards in energy compared with

the current at the top of the barrier. However, the average energy of the electron at the drain end is approximately equal to 0.25 eV that is substantially larger than the sub-band energy (−0.55 eV).

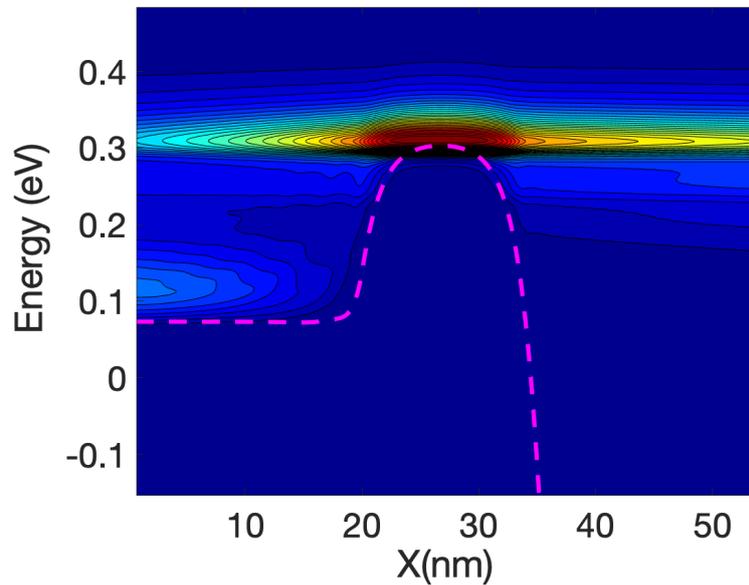


Figure 5. Energy resolved current spectra along the wire axis. The first sub-band is shown as a dashed line. The current at the source (in the left side of the figure) divides into two branches around energies 0.1 eV and 0.3 eV. The current in the low energy branch disappears when progressing from the source towards the right, vice-versa the current in the high energy branch increases. These means that the electrons from the lower energy branch are moving to the right branch. These electrons need to absorb phonons to increase their energies and therefore they tend to cool the lattice at the source.

3.3. Energy Balance and Power

The Equations (61)–(63) represent the energy balance between electrons and phonons and is discussed in several references [89–92].

These equations represent the total local power dissipated for the electron system into the phonon system. As indicated before, the last term on the RHS of Equation (63) is the local Joule power that integrated through the device produces the well-known Joule power $I \cdot V$. In Equation (63) the term $E(r)$ represents the local electric field along the nanowire axis. The first term on the RHS is the change in the “kinetic energy” current of the electron ensemble, if this kinetic energy current does not change it means that the electron system is adiabatically following the “bending” of the potential and therefore it is not increasing the electron mean energy relative to the conduction band minima, i.e., no hot electron phenomena occur. In this case, the electron system or more specifically the electrons taking part in the current are in equilibrium with the local potential. On the other hand, a large change in kinetic energy current implies the heating or cooling of the electron system.

Equation (63), for the local power, can be simplified for the case of local electron-phonon self-energies and the resulting equation obtained [56] is essentially a detailed balance equation (see Appendix E, F specifically Equation (A46)):

$$P(r) = \int d\varepsilon \varepsilon \{G^<(r, r, \varepsilon)\Sigma^>(r, r, \varepsilon) - G^>(r, r, \varepsilon)\Sigma^<(r, r, \varepsilon)\} \quad (72)$$

The electron-phonon self-energies used in this paper are also treated in a local approximation (diagonal approximation), so that Equation (72) is the local version of Equation (A46) in the Appendix F. The RHS of Equation (72) expresses the local power as a function of the Green functions and self-energies. It shows that the energy mismatch between the rate of ε -outgoing (leaving the state of energy ε) electrons (the first term in the RHS) and the rate of ε -outgoing holes (second term in the RHS) is

transferred to the phonon system. Note that for elastic scattering the expression in the curly brackets became zero [91] (Appendix E demonstrates this explicitly). As mentioned before, the first term inside the integral in the RHS represents the outgoing electrons from state ε and the second term is the ingoing electrons into state ε . For elastic scattering these rates are equal as the electron does not change its energy after scattering and $P(r)$ vanishes except in the contacts. Equation (72) represents the detailed energy balance between the electron and the phonon system.

Figure 6 shows the kinetic term and the Joule term of Equation (63). Their sum or the total power is also shown. Note that the kinetic term is plotted with reverse sign. The two terms are very similar but with opposite sign therefore in order to obtain an accurate value of the total power, the current needs to be calculated with great accuracy which implies a large number of Born iterations for the self-energies.

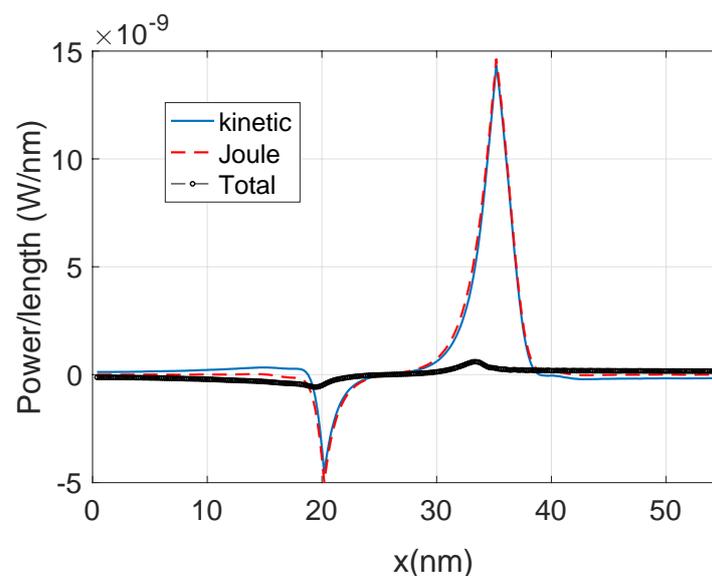


Figure 6. Local power dissipated along the device axis. $V_g = 0.4$ eV. The thin line and the dashed line represent the corresponding values of the kinetic term and the Joule term in Equation (48) of the text. The thick line represent the total power (the subtraction of the two graph). It should be noted that the actual power dissipated inside the device (signed area under the curve) is much smaller than the Joule power (Joule term).

The 3D Fourier heat equation used to calculate the local temperature profile is solved in the oxide and in the Si nanowire core. Neumann boundary conditions are used in the source and drain contacts and the Dirichlet boundary condition is used at the metal gate. The heat sources are calculated from the total power or using the Joule heat power, depending on which model we are using. Both sources include the volume inversion effects.

3.4. Results and Discussion

In this section, we report electrothermal simulations of a gate-all-around nanowire transistor (Figure 2). The cross-section of the silicon core is $2.2 \times 2.2 \text{ nm}^2$ and the source, channel, and drain have lengths of 20 nm, 15 nm, and 20 nm, respectively. The channel is undoped and the doping in the source and the drain are 10^{20} cm^{-3} . In all of this work, we assumed a drain bias of $V_d = 0.6$ V. We have compared the drain current vs gate voltage characteristic (transfer characteristic) for the case of standard scattering, i.e., assuming the phonon system or the lattice is at equilibrium at 300 K with cases in which the lattice is at higher temperature.

The local phonon temperatures are calculated from Fourier's law. Two cases have been considered: case (i): in which the electrons are in local equilibrium with the electrostatic potential, so the energy transfer from the electron system to the phonon system is given by the standard Joule law or the second

term in Equation (63); and case (ii), in which the electrons relax to equilibrium depending on the energy relaxation length (given by the inelastic electron-phonon scattering length) [92] and therefore the energy transfer are given by the LHS of Equation (63). This latter case considers that the electrons injected from the source into the drain leave the drain at a higher energy than the drain average (equilibrium) energy. These electrons are commonly called “hot electrons”. Consequently, these electrons dissipate less energy inside the device as compared with the electrons in local equilibrium and therefore produced less heating of the lattice in the active part of the device.

In a previous paper [92], we have simulated and examined different sizes of source and drain extensions for a nano-transistor with a similar cross section to the one studied here but self-heating was not considered. In that paper, we found that electrons needed a drain extension length larger than 100 nm in order to guarantee that the total power dissipated inside the device became equal to the Joule power.

However, even for larger source and drain extensions as long as the channel length is around 10 nm, inside the device, the potential energy of the electrons changes hundreds of millielectron volts in just a few nanometers. This length is too short compared with the energy relaxation length. As a consequence, the electron transport around the gate is essentially out of equilibrium or carried out by “hot electrons”. The energy relaxation depends on the density of states and on the electron-phonon scattering rates. However, previous papers [76] for small cross section devices confirm an increase in electron-phonon scattering rates due to an increase in scattering form factors. This fact favours a decrease in the energy relaxation length in small devices but this is not sufficient to allow full electron energy relaxation inside the device.

In general, devices with cross sections larger than $10 \times 10 \text{ nm}^2$ and with channel/source/drain of a few 100 of nanometers are expected to comply with the Joule law but as mentioned previously a more accurate assessment would look at the local spatial variation in the electrostatic potential or electron potential energy, which depends on the device structure and dimensions.

The current-voltage characteristic for these three scattering simulation cases: (i) no self-heating (NSH), (ii) hot electrons self-heating (HSH) and (iii) Joule self-heating (JSH) are shown in Figure 7.

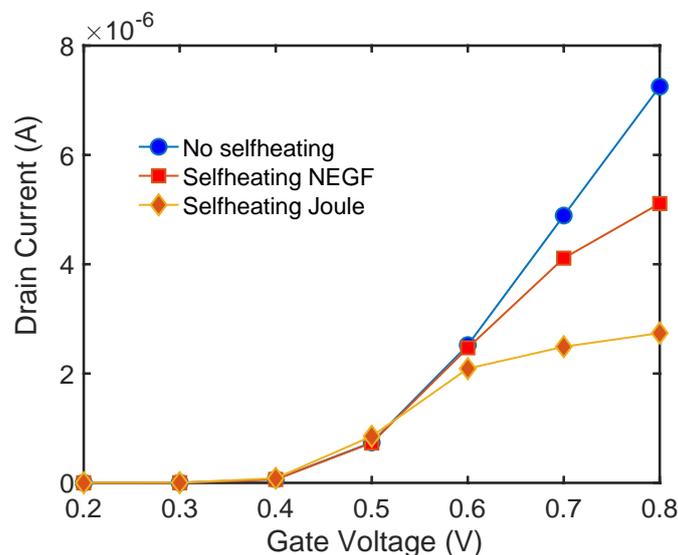


Figure 7. Transfer characteristic for cases (i) no self-heating (NSH), (ii) hot electrons self-heating (HSH) and (iii) Joule self-heating (JSH). In the no self-heating case the lattice is kept at 300 K. In the other cases the local lattice temperature depend on the net power dissipated by the electron system locally to the phonon system. For the case (iii), the electron system is considered at local equilibrium with the lattice. The drain bias is $V_d = 0.6 \text{ V}$. At high gate bias, the current for the of Joule selfheating is larger than the current for the hot electron one, this is a consequence of the higher temperature of the lattice for the joule case as compared to the hot electron case.

As it is clear from Figure 7 that under self-heating conditions, i.e., in which the lattice has a high temperature, the scattering rate is large and the current at $V_g = 0.8$ V is 30% smaller for the hot electrons case as compared with the non-self heating case. However, when the Joule self-heating is considered the effect is larger, and the drain current is reduced by 70%, compared to the NSH case. For low gate bias, as the current is small, the impact of self-heating and scattering in general is diminished. In what follows, we will discuss the results at low gate bias and large gate bias.

At low gate bias the current is low and consequently the power dissipated is low; however, the fact that we have a large source-drain bias barrier produce some interesting effects. Figures 8 and 9 show the current spectra at $V_g = 0.2$ V for the case of the hot electrons. Figure 9 shows an enlargement of the high energy region of Figure 8 where the electrons energy exchange occurs.

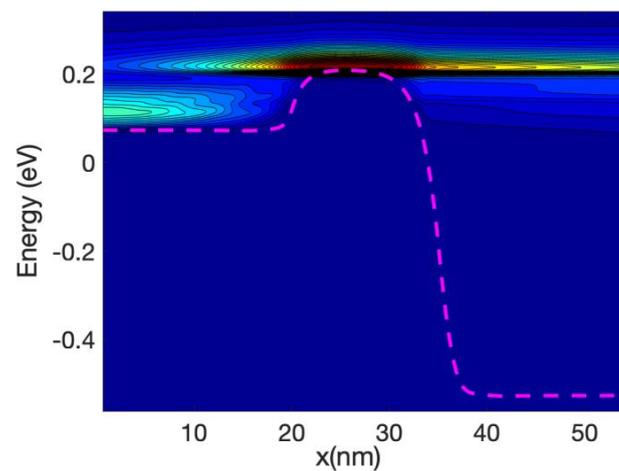


Figure 8. Energy-resolved current along the nanowire axis at $V_g = 0.2$ V. The first sub-band is shown by a dashed line. The electrons contributing to the current at the source (leftmost part in the figure) have lower energies than those at the middle of the device (or at the barrier top). Indicating large phonon absorption at the source (see the text).

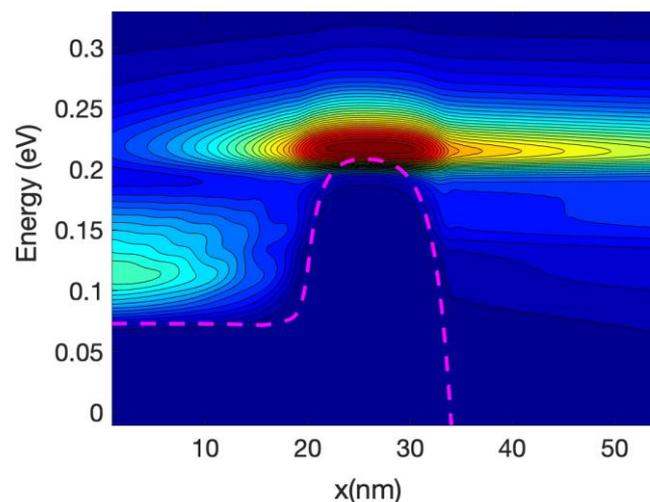


Figure 9. Close-up of Figure 8. The current spread in energy is very small at the middle of the wire or at the top of the source drain barrier. However, the spread in current increases as electrons travel towards the drain contact indicating energy relaxation. The average energy of the current is around 0.2 eV at the drain contact.

The figures show that the electrons at the bottom of the band entering in the source absorb phonons and therefore are able to cross the barrier, this phonon mediated effect tends to transfer energy

from the source to the drain of the device, and therefore it has a cooling effect on the source of the device and slight heating at the drain.

As the nanowire cross-section is small the LDOS is quasi-1D and therefore there is a relatively large DOS at the sub-band energy at the source end (and indeed at the drain end); consequently, there is an overwhelming number of electrons at that low energy. This induced a high probability of phonon absorption for the electrons in the source end. This is one of the reasons that the energy-resolved current at the source is split into two components for a large source-drain barrier.

Figure 10 shows the temperature profile across a plane of the device axis including the oxide at $V_g = 0.3$ V. The figure shows a slight cooling (from 300 K) of the source of the device. This effect is slightly enhanced at $V_g = 0.3$ V as the height of the barrier has decreased and therefore the electrons are more efficiently transferred lower energy in the source to the top of the barrier. The cooling of the source diminished at higher gate bias as a large fraction of electrons are able to tunnel or directly transit the barrier without the need of phonon assistance. This indicated an optimum length in which the cooling of the source is more efficient. At very low gate bias we have found that the total power dissipated by the electron current in the device is negative; this means that the electrons absorbing energy from the phonon system on the source are not able to release enough energy at the drain to counteract the cooling effect at the source. This net energy absorption of electrons in the source region has been observed in other papers [86,91] through the calculation of the power absorbed along the wire for the electrons using the Equation (9) as will be shown in the Figure 11, here. As mentioned before, it can also be inferred by the plot of the energy current resolved along the nanowire axis (see Figures 8 and 9). However, Figure 10 shows a decrease in temperature (cooling) of the lattice, which depends on the net interchange of heat of the lattice with the surrounding, as it is not sufficient that the electron system absorb energy from the lattice locally in order to cool the lattice. If the drain or other parts are hot enough, heat will be transfer to the source avoiding cooling. The fact that the electron going through the drain does not release enough energy there is one of the facts which allows for the cooling of the source. If we had assumed that the electrons are in equilibrium with the potential or, i.e., assuming Joule heating, the net power dissipation would be always positive, that is a power transfer from the electrons to the lattice. In addition, there would have been a large dissipation in the drain, heating the overall drain and surroundings including the source.

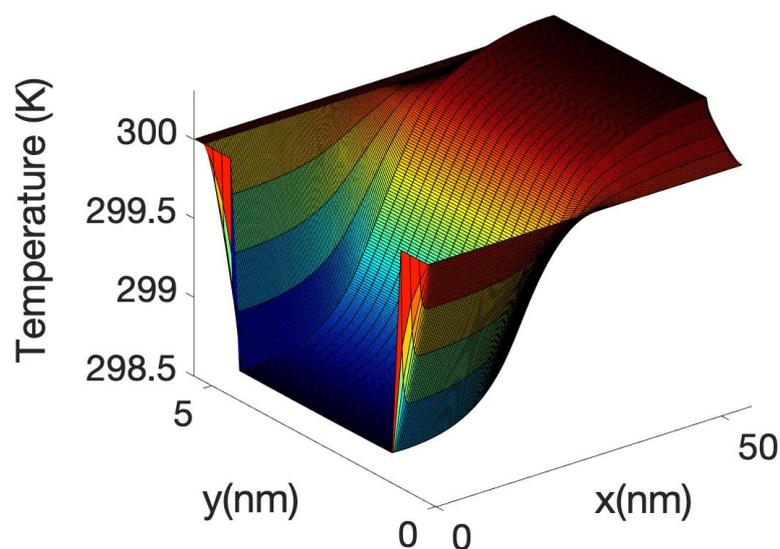


Figure 10. Temperature profile for (x,y) plane for the z coordinate located at the middle of wire cross section; for the coordinate system used and the device orientation see Figure 2. The profile corresponds to the hot electron self heating case (HSH) at a low bias of $V_g = 0.3$ V. The figure shows a slight cooling of the source of the transistor. The overall power dissipated inside the transistor is negative.

The local power density profile used as a source in the heat source is shown in Figure 10 along the nanowire axis for the case of hot electrons. It is shown that most of the power transfer occurs close to the source-drain barrier. The negative power density at the source end of the barrier indicates that the electron system is taking energy from the lattice or phonon system as described before.

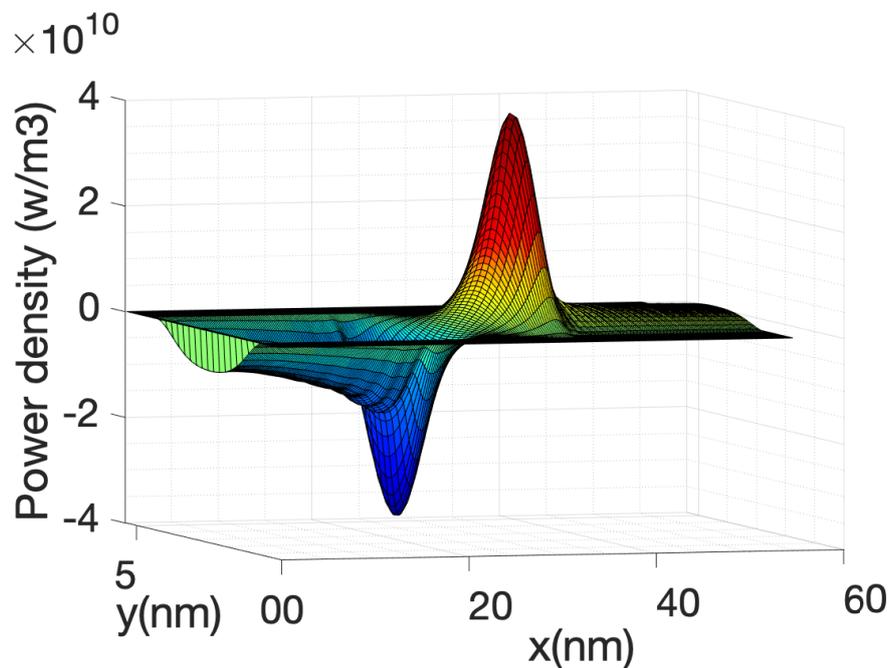


Figure 11. Power density profile used as heat source in the heat equation. The gate potential (V_g) and drain potential (V_d) are 0.3V and 0.6 V, respectively. This figure shows the 2D profile of the 3D analogue of the total power showed in Figure 6. Positive values indicate source of heat and negative values sink of heat or cooling.

At high gate voltage, the source-drain barrier is small or negligible. Figures 12 and 13 show the profiles of the heat source term for V_g equal to 0.6 V and 0.7 V, respectively. At $V_g = 0.6$ V, there is a small region in which the heat source has negative values: in this region the electrons are still absorbing energy from the lattice; however, there is a greater heat transfer from the electrons to the lattice at the drain region and as a consequence the overall temperature increases. When the barrier is sufficiently low, as is in the case of $V_g = 0.7$ V (Figure 13), there is no region of negative-valued heat source; in this case, the potential resembles a step potential and the electrons just release energy when transiting the device. The energy-resolved current density for $V_g = 0.6$ V and 0.7 V is shown in Figures 14 and 15, for the case when self-heating and hot electron effects are considered. The corresponding temperature profiles are shown in Figures 16 and 17. Note that although the current spectra look very similar, the change in maximum temperature is approximate 60 K. Figure 17 resembles the 1D temperature plots of Figure 7 in reference [93]. We have carried out a more thorough comparison in our previous paper [94]. However, it is worth mention that our nanowire has rectangular cross section and the nanowire in [93] has a circular cross section (not exactly matching the areas). We are using an effective mass approximation, and the reference [93] used full band. The scattering models are also different as they used the full phonon bands. The reference [93] simulated the phonons using a NEGF formalism instead of our simplified heat equation. However, the level of current obtained in our simulations are of similar order of magnitude to those in [93]. Considering all the differences between the models, our calculated temperature profiles qualitatively follow those found in [93]. The temperature profiles there range from 300–500 K and this also matches the range of our results.

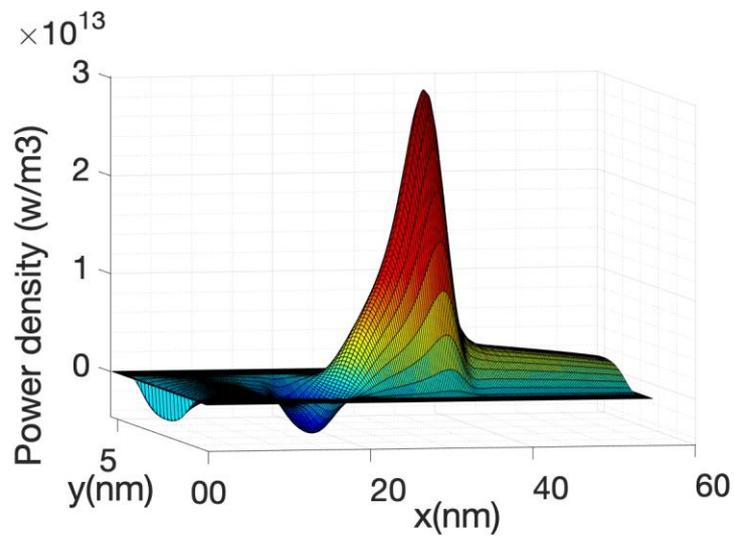


Figure 12. Heat source due to hot electrons at $V_g = 0.6$ V.

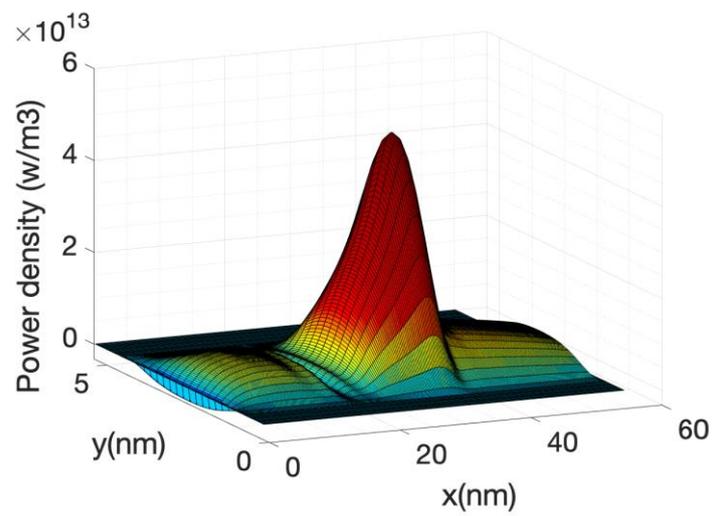


Figure 13. Heat source due to hot electrons system at $V_g = 0.7$ V.

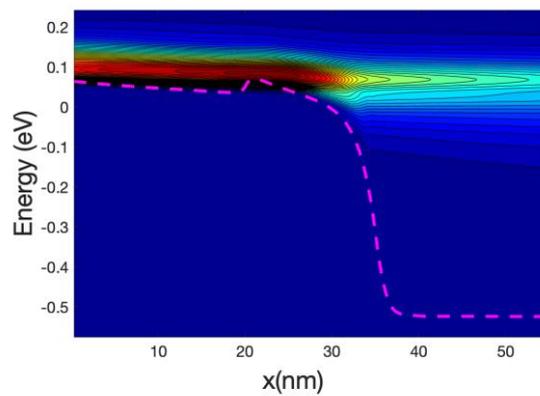


Figure 14. Energy resolved current spectra $V_g = 0.7$ V, hot electrons and selfheating.

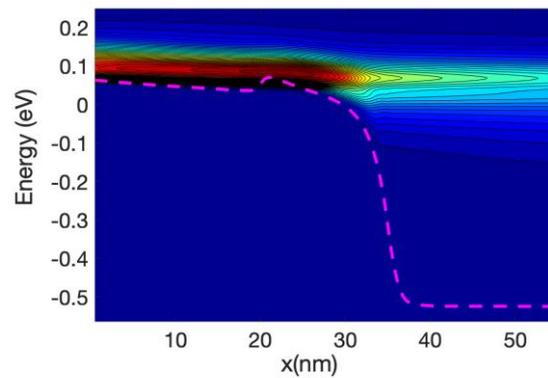


Figure 15. Energy resolved current spectra $V_g = 0.6$ V, hot electron source and self-heating.

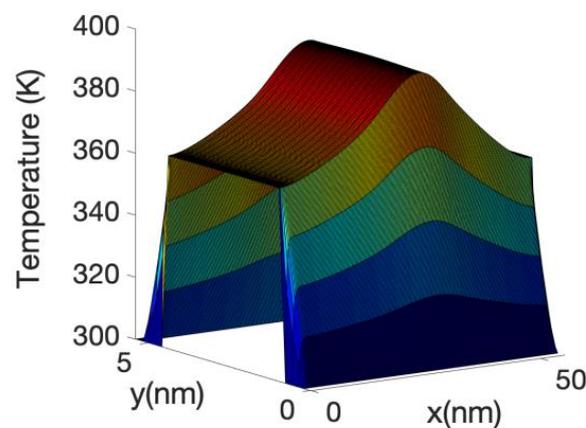


Figure 16. Lattice temperature profile at $V_g = 0.7$ V, hot electrons and self-heating.

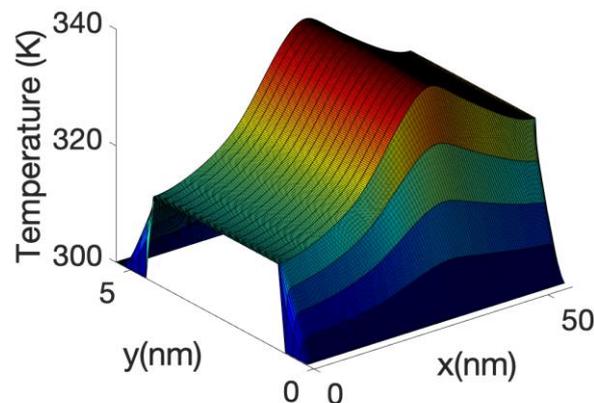


Figure 17. Lattice temperature profile at $V_g = 0.6$ V, hot electron source and self-heating.

The heating of the lattice due to the energy transfer from the electron system into the phonon system produces an increase in the electron-phonon scattering rate and therefore has the effect of reducing the current. This reduction is substantial (at $V_g = 0.7$ V), compared with the case in which the phonon system is considered at a temperature of 300 K or no self-heating. This is clearly shown in the current voltage characteristics depicted in Figure 7. However, the effect is even large if we neglected hot electron effects and assumed that the electron system is in local equilibrium with the potential, i.e., assuming the Joule law. The Figure 7 shows a very large reduction in current when Joule heating is considered. The corresponding temperature profile and heat source profile for this case are shown in Figures 18 and 19. The heat source in this case (Figure 20) is just proportional to the gradient of the

electrostatic potential. Note that the maximum temperature (over 600 K) is substantially larger than the previous case, indicating that the Joule heat grossly overestimates the power dissipation inside small nanowire transistors.

Figure 19 shows the temperature along the device corresponding to the case in which the electron system dissipates the Joule power into the lattice but where we fixed the lattice temperature of 300 K in computing the scattering of electrons. In this case as there is not a reduction in the electron energy current the lattice temperature increases substantially, the maximum peaking around 1000 K. This shows the danger of considering only the energy transfer from the electrons to the lattice, but not the impact of the heated lattice on the electron system for small nano-transistors.

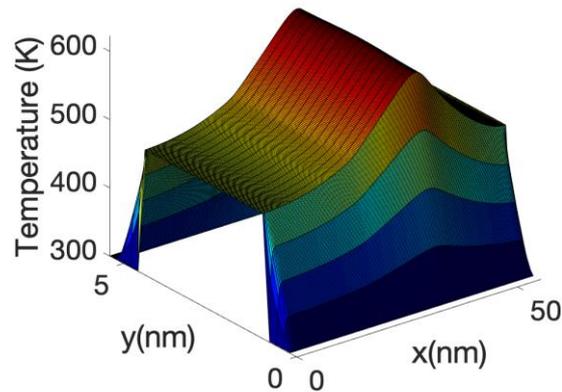


Figure 18. Lattice temperature profile for $V_g = 0.6$ V considering Joule self-heating. This figure is to be compared with figure 17. The peak temperature is almost twice the value of Figure 17. Indicating an overestimation of power dissipation.

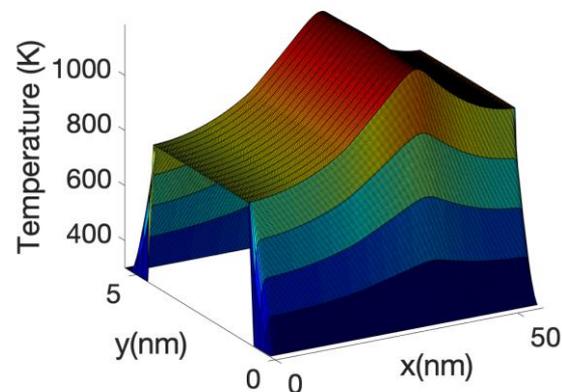


Figure 19. Lattice temperature profile for $V_g = 0.6$ V considering Joule heat source but no self-heating. The temperature peak is large than 1000 K. This is a substantial overestimation of lattice heating when compare with the more accurate hot electron heating of Figure 17.

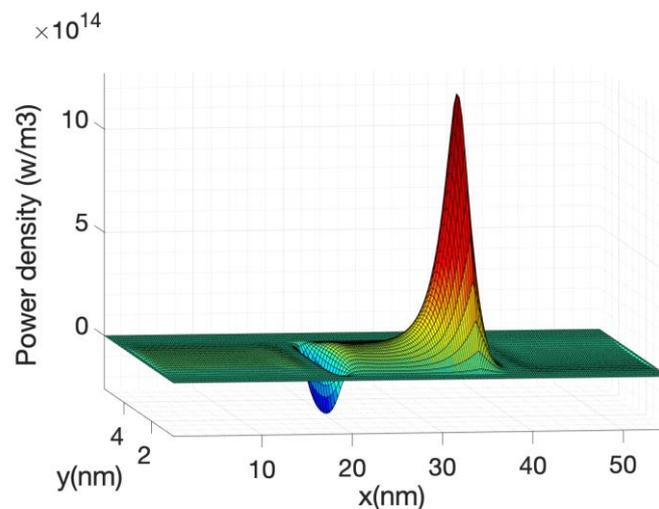


Figure 20. Power density profile used as heat source (Joule heating) in the heat equation. $V_g = 0.7$ V.

3.5. Perspectives and Challenges

In this section, we describe certain open issues and limitations of the simulation of carrier transport in nano-transistors relevant to our work. Even though substantial progress in nanowire transistor simulation has been achieved, reliable and unified models do not exist yet. The majority of simulation methodology has a simplified description of the contacts. Usually, the contacts have the same cross section as the active part of the device. This is very convenient when using the NEGF formalism as it requires a region of uniformity in order to properly inject and absorb carriers through self-energies. A more realistic description of the contacts would require [59] a broadening of the contacts relative to the device active region. This broadening, of course, would depend on the realistic geometry of the device and needs a large simulation domain as a large part of the extension regions would need to be added. Calculations using different source/drain geometries have been carried out in a ballistic context [95]. In the self-consistent NEGF/Poisson formalism used to simulate devices, the Fermi levels are fixed in the boundary of the source and drain and the potential is allowed to float to achieve charge neutrality. This fixing of the Fermi level would be more accurate as the electrical neutrality in the contact should not be affected by the device active region, as the contacts will ultimately represent reservoirs in equilibrium. However, these reservoirs could be made of or connect 1D, 2D and 3D electron states. In addition, the broadening of the contacts will create new challenges from a computational point of view as scattering needs to be considered in those contacts in order to inject electrons in the narrow regions with the proper equilibrium distribution, the proper quantum states and energies. The transition region between the simulated contacts and the narrow device region will induce unwanted reflections if the transition is abrupt, therefore a smooth change in cross section, from wide to narrow channel is required in conjunction with decoherence and randomization due to electron-phonon scattering. A proper smoothness of the transition will require the use of the finite element method [96] to treat the change of geometry in an intrinsic way. However, this will stretch the use of the recursive algorithm [97], which is the pinnacle of success in NEGF computational versatility. Electron-electron scattering in the interfaces and contacts could play a role but it would be challenging to introduce into reliable, computationally manageable and even theoretical models [51,52]. In addition, electrons in the channel experience remote interactions from interfaces and contacts such as soft optical interface phonon scattering [53,54] and remote Coulomb scattering. The latter references indicate that these mechanisms could have substantial impact on device performance. The modelling and description of these mechanisms are not unique and represent a challenge due to the reduced and changing dimensionality of the nanostructure.

The present study has not considered the influence of atomistic effects on electro-thermal modelling, such as stray discrete impurities, trapped charge at the interfaces or discrete effects of the atomistic interfaces, although we have made extensive studies of these effects in purely electron transport modelling (e.g., [11,79]). However, the high surface to volume ratio for nanowires raises the issue of electronic interface states and their effect on electro-thermal properties which deserves future scrutiny. In addition, the role of localised phonon modes at the interfaces could lead to important thermal channels and local non-equilibrium phonon effects.

However, there is one aspect of simplification that needs caution: it has been usual in most studies to neglect the real part of the self-energies. Unfortunately, this is not justified in general. For example, the retarded self-energy has to satisfy causality requirements and these lead to Kramers-Kronig relations (Hilbert transform) that relate the real part of the self-energy to the imaginary part. As we have shown, elsewhere (see Appendix G and references herein), neglect of the real part may lead to serious violations of the spectral density sum rule and possible violations of charge conservation. Significant differences may be demonstrated in device modelling [77]. In general, the real part of the retarded self-energy leads to energy-dependent level shifts and may lead to bifurcation of the spectral density of states. The drawback is that the procedure for the full self-energy (within the self-consistent Born approximation) is very much more compute-intensive (further details in references Appendix G).

The use of a heat equation in this work and others as an alternative to more complex methodologies and moving towards more computational efficient methods is promising. The similarity in results with other methodologies demonstrated in ref [94] is encouraging in spite of the strong dissimilarities between methodologies. Their use of NEGF description for phonon transport and the full phonon dispersion law, instead of the simple heat equation, made the methodology extremely computationally demanding and therefore limiting its scope. We use a very simplified version of the phonon scattering mechanisms [85,98] that has been proven to produce similar values and trends for low field mobilities [76] and which has been widely used in the Monte Carlo/Greenwood-Kubo methodologies and adapted for nanowire [98]. The calculation of phonon modes in confined nanostructures is also a challenge, as they strongly depend on the boundary conditions used [99]. In our study, the acoustic deformation potential has been parameterised to partially take care of the decrease in mobility for narrow nanowire following [86,98]. Using a heat equation has the advantage of extending thermal simulations in larger regions surrounding the device and interconnects. In addition, interfaces and boundaries between materials of different thermal conductivities can be handled. This opens up the possibility of using more realistic thermal environments [100,101] and compact models as thermal resistance. Thermal conductivities which vary with the temperature could also be incorporated. Simple phonon balance equations can be used instead of the heat law if necessary, alternatively, different heat equations for different phonon types (for example separately for polar and acoustic) could also be used and this could partially address heat transport through interfaces and phonon-phonon interaction. Finally, thermal transport processes in low-dimensional and non-trivial spatial topology nanostructures (so-called “holey” materials) are becoming important [102] for applications in thermo-electrics, photonics and batteries. Reference [102] provides a detailed review of this challenging field and showed that thermal conductivity measurements at room temperature are considerably lower than the Casimir limit for structures smaller than a few hundreds of nanometers. The traditional continuum models of thermal diffusion and conduction no longer apply. The power of the spatially resolved NEGF methodology may prove important in that context.

4. Conclusions

In this paper, we revised and reviewed some of the quasi-classical and quantum methodologies used for the electrothermal simulation of nanoelectronics devices. Details and insights about the physics behind the formalism were also provided and these were complemented with the relevant literature and clarifying appendixes. The rest of this work focused on the application of a blended electrothermal methodology to a gate all around nanowire transistor.

Specifically, we have carried out electrothermal simulations of a narrow gate all-around nanowire transistor using a quantum transport formalism. The non-equilibrium Green function formalism has been used to describe the electron transport and the electron-phonon scattering is described within the self-consistent Born approximation

We have modelled the lattice heating by a Fourier law, and used as a heat source the true loss or gain of energy of the electrons (hot electrons) to the phonons.

These simulations have been compared with the use of the Joule law as a heat source. The drain on-current is reduced around 30% when the hot phonon source is used and 70%, compared to when the Joule law alone is used.

We have found that at low gate bias a slightly cooling of the source region by a few degrees occurs. The effect is large at an intermediate gate bias that depends on the electron-phonon scattering strength and the length of the source. This effect might be used for cooling the device if the source material and the dimension of the device are properly engineered.

At high gate bias, the injected electrons go straight to the drain region without the need of phonon absorption as the source drain barrier height is negligible. The dissipation of energy from the high energy electrons injected from the source induces an overall heating in the active region of the device. This process is more pronounced in the drain side as has been confirmed by the simulations. This effect becomes larger as the current increases. However, as the lattice heats, the average phonon energy or the equivalent temperature of the phonon system increases, and as a consequence, the electron-phonon scattering rate increases, the electron current is decreased from its value when the lattice is at room temperature. This results in less heating of the lattice. Our simulation produces a maximum device temperature of 400 K at large gate bias.

We have also investigated the case that the electrons are in local equilibrium with the lattice and used the Joule power as a source of the heat equation, this approximation is only valid for large devices. This produces a large heating of the lattice and consequently a large decrease in the drain current. The maximum temperature is approximate 600 K.

The methodology presented in this paper is computationally efficient and allows the accurate prediction of the effect of self-heating in small devices and produces very similar results to more sophisticated but computationally intensive methodologies. As indicated, some simple models can reproduce the trends of more sophisticated ones in a more computational efficient way; however, these models require verification or at least justification of the principle used. These models usually provide a simpler and concise explanation of the behaviour of the devices as they used few physical parameters. This work is a small step in this direction of the search for more simplified and phenomenologically-based models allowing the incorporation of more phenomena and larger parts of the device environment in the simulation domain. The solution of the 3D heat equation has a similar computational complexity to the 3D Poisson equations, which are solved quite efficiently by iterative methods. Therefore, quantum transport can be naturally connected with self-heating and thermal transport in a large domain. This provides more realistic boundary conditions, i.e., boundaries at room temperature should be far away from the active region of the device. For a device, designer/engineer accuracy and speed are equally important as substantial numbers of devices and conditions would need to be considered before an optimum decision is taken.

It is anticipated that our methodology may find wide application in the design of power efficient nanodevices.

Author Contributions: Conceptualization, A.M. and J.R.B.; methodology, A.M. and J.R.B.; software, A.M.; validation, J.R.B. and A.M.; formal analysis, J.R.B.; investigation, J.R.B. and A.M.; writing of the original draft preparation, J.R.B. and A.M.; writing of review and editing, J.R.B. and A.M.; visualization, J.R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A. Local Density of States for Simple Systems

Here, we compare classical and simple quantum expressions for the LDOS of a simple model of the potential energy profile for a silicon NWFET of total length 55 nm; the channel is between 20 nm and 35 nm.

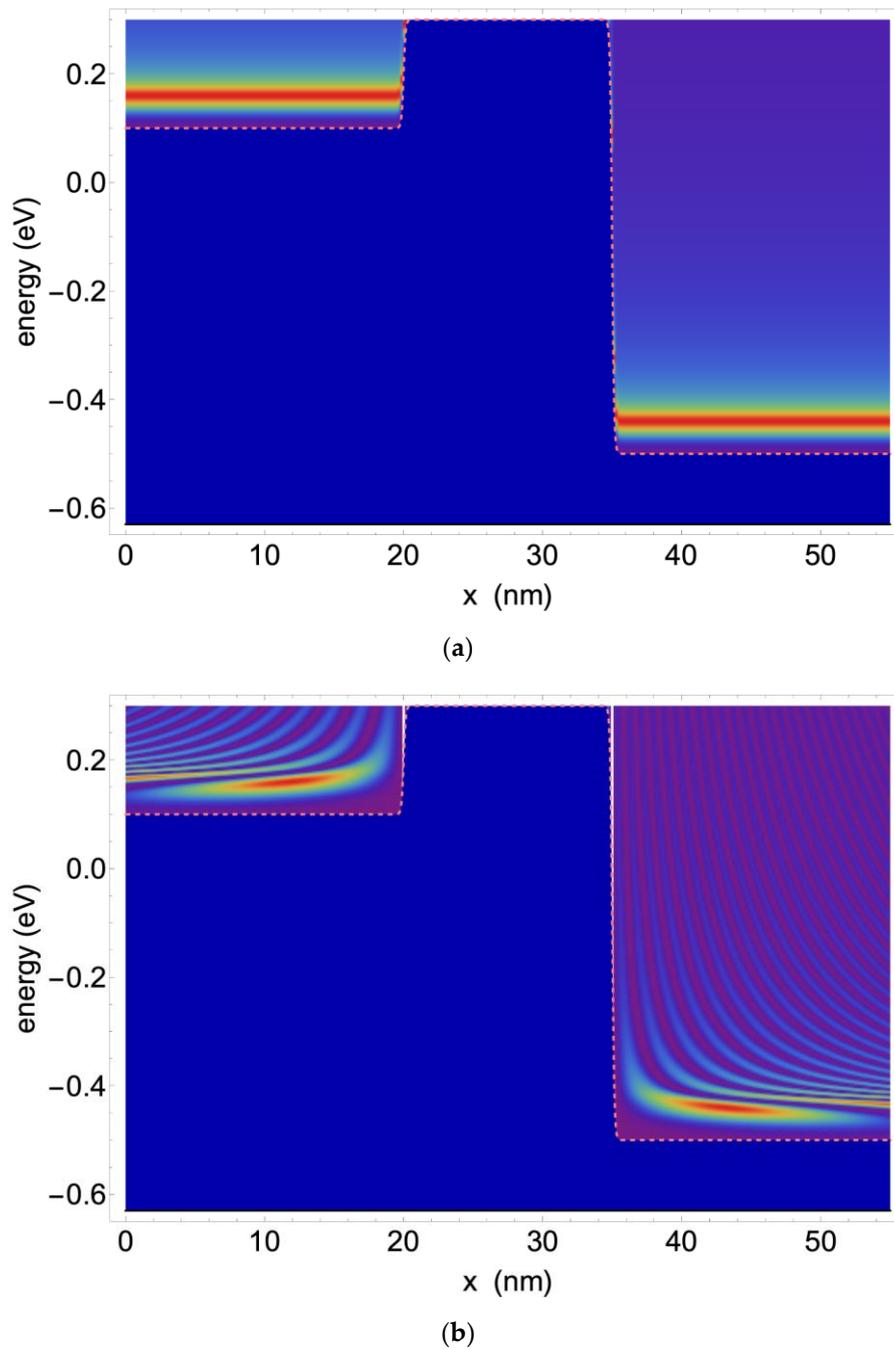


Figure A1. LDOS for a silicon nanowire-FET (NWFET) simple device potential profile (dashed pink line) (a) classical result; (b) quantum result showing influence of strong reflections on either side of the barrier.

The gate bias is $V_g = 0.2$ V, with a drain bias of 0.6 V. We consider only a single sub-band for transport along the [100] direction. The quantum expression used is derived from (25) for a piece-wise constant potential as:

$$\rho_{LDOS}(x, \varepsilon) = |\psi_{k(\varepsilon)}(x)|^2 \left(\frac{2m}{\pi \hbar^2} \right) / \sqrt{\frac{2m(\varepsilon - V(x) - \varepsilon_c)}{\hbar^2}} \quad (\text{A1})$$

In the classical case, the term $|\psi_{k(\varepsilon)}(x)|^2$ becomes unity. The inverse square root gives rise to very high values for the density of states at the sub-band edge. The strong fringing modulation in the quantum case derives from the strong reflections of electrons at source side and the drain side of the gate barrier.

For realistic semiconductor models, the orientation of the 1D axis is crucial (usually along the [100] direction) and multiple valleys must be considered. This leads to a superposition of 1D LDOS terms associated with different valleys. The full range of Δ valleys considered is shown in the iso-energy surfaces in k-space in Figure A2, showing the principal valleys for transport in orange and the other valleys in silver.

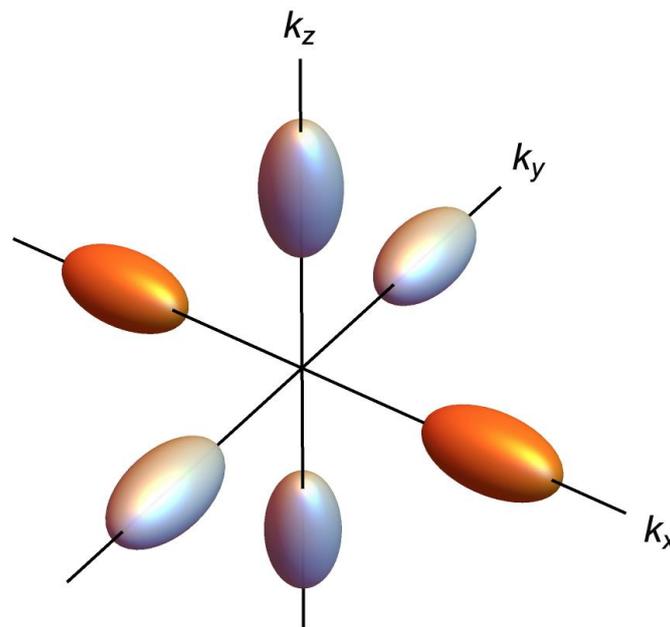


Figure A2. Constant energy surfaces for the Δ valleys in the k-space used to represent the effective mass Hamiltonian for a silicon nanowire device. The principal valleys for transport are shown in orange. In our work the k_x aligns with the channel of the wire.

Appendix B. Use of Self-Energies to Project Coupling of Contacts into Finite Device Domain

Consider the Green function operator $G_{total}(\varepsilon)$ for the open system comprising a finite nano-device connected through essentially open source (S) and drain contacts (D), which have well-defined temperatures and chemical potentials at the interface to the device (W). $G_{total}(\varepsilon)$ satisfies the eigenvalue equation (I is a unit matrix):

$$(\varepsilon I - H_{total}) G_{total}(\varepsilon) = I \quad (\text{A2})$$

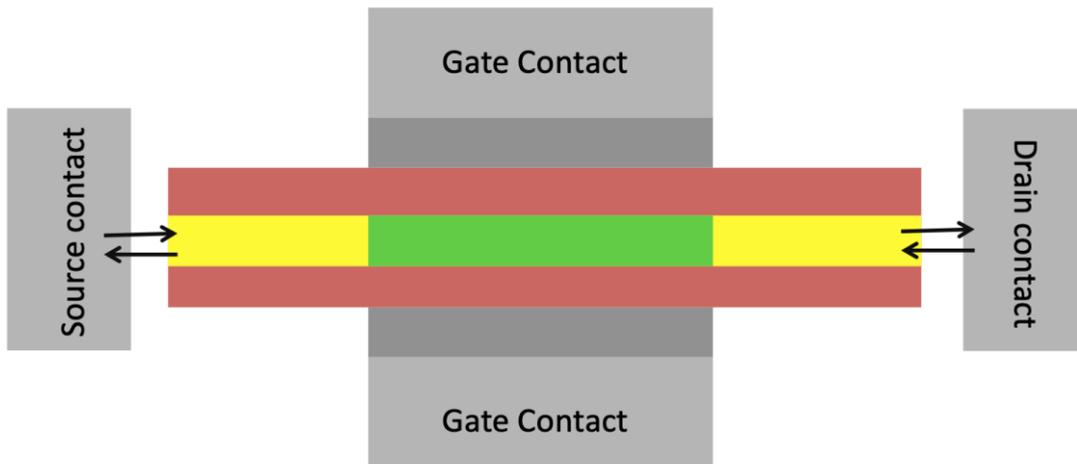


Figure A3. Simplified plan of simulation domain (in yellow-source extension, green-channel, brown-oxide, dark grey-wrap-round gate) and macroscopic environment /reservoirs (light grey-macroscopic contact domains). No channel: gate current.

In any representation (Equation (A3)), is a huge intractable matrix equation. In matrix operator form we may formally partition this Hamiltonian into a matrix of sub-Hamiltonians following reference [103], representing the finite device in interaction with the open source/drain contact domains. Consider the Green function G_{WW} for the finite device (W) where electrons are injected or extracted at the surface of the Drain (D) and Source (S) contacts. The total Green function operator is partitioned into the matrix

$$G_{total} = \begin{bmatrix} G_{SS} & G_{SW} & G_{SD} \\ G_{SW} & G_{WW} & G_{WD} \\ G_{DS} & G_{DW} & G_{DD} \end{bmatrix} \quad (A3)$$

and we similarly partition the total Hamiltonian

$$H_{total} = \begin{bmatrix} H_{SS} & H_{SW} & 0 \\ H_{SW}^\dagger & H_{WW} & H_{DW}^\dagger \\ 0 & H_{DW} & H_{DD} \end{bmatrix} \quad (A4)$$

The source/drain contact regimes are represented by Green functions G_{SS}/G_{DD} and Hamiltonians H_{SS}/H_{DD} . The finite nanowire device (simulation region) is modelled by a Hamiltonian H_{WW} . The transfer Hamiltonians are non-hermitian forms, H_{SW}, H_{DW} , which inject or extract carriers at the source contact: source extension boundary surface and at the drain contact: drain extension boundary surface. Introducing a matrix of the sub-domain Green functions (in an obvious notations) Equation (A3) becomes the matrix equation:

$$\begin{bmatrix} \varepsilon - H_{SS} & -H_{SW} & 0 \\ -H_{SW}^\dagger & \varepsilon - H_{WW} & -H_{DW}^\dagger \\ 0 & -H_{DW} & \varepsilon - H_{DD} \end{bmatrix} \begin{bmatrix} G_{SS} & G_{SW} & G_{SD} \\ G_{SW} & G_{WW} & G_{WD} \\ G_{DS} & G_{DW} & G_{DD} \end{bmatrix} = \begin{bmatrix} I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \quad (A5)$$

Solving the equations corresponding to the second column of the identity matrix in Equation (A5), it is simple to derive the Green functions for source and drain, G_{SW} and G_{DW} :

$$G_{SW} = g_{SS} H_{SW} G_{WW} \quad (A6)$$

$$G_{DW} = g_{DD} H_{DW} G_{WW} \quad (A7)$$

$$G_{WW} = (\varepsilon - H_{WW} - \Sigma_{SW} - \Sigma_{DW})^{-1} \quad (A8)$$

where

$$g_{SS} = (\varepsilon - H_{SS})^{-1}; g_{DD} = (\varepsilon - H_{DD})^{-1} \quad (\text{A9})$$

and the contact-coupling self-energies are:

$$\Sigma_{SW} = H_{SW}^\dagger g_{SS} H_{SW}; \Sigma_{DW} = H_{DW}^\dagger g_{DD} H_{DW} \quad (\text{A10})$$

Equation (A8) illustrates how to project out a finite domain Green function from the infinite open system by introducing contact self-energies

$$\Sigma_{contacts} = \Sigma_{SW} + \Sigma_{DW} \quad (\text{A11})$$

at the boundaries. The retarded and advanced Green functions are retrieved by adding $\pm i\eta$ ($\eta > 0$) to the energy ε . A similar scheme applies to the lesser and greater Green functions. The details depend on the system.

Appendix C. The Keldysh NEGF Equations

The Keldysh formalism describes a number of different NEGF, only 3 of which are independent. Here, it is assumed that three independent Green functions are G^R , G^A , $G^<$. In matrix operator form, the steady-state Green functions satisfy:

$$G^R = (\varepsilon + i\eta - H_0 - H_{ext} - \Sigma^R)^{-1} = G_0^R + G_0^R \Sigma^R G^R \quad (\text{A12})$$

$$G_0^{R-1} = (\varepsilon + i\eta - H_0 - H_{ext}) \quad (\text{A13})$$

$$G^A = G^{R\dagger} \quad (\text{A14})$$

$$G^< = G^r \Sigma^< G^A \quad (\text{A15})$$

$$G^> = G^r \Sigma^> G^A \quad (\text{A16})$$

Here, for convenience, we include any confinement potential energy $H_c(\mathbf{r})$ into the term H_{ext} ,

$$H_{ext} \rightarrow H_c(\mathbf{r}) + e\phi(\mathbf{r}) \quad (\text{A17})$$

which makes explicit the applied electrostatic potential.

Equation (A12) is the first Keldysh equation for the retarded Green function G^R ; it describes the dynamical states of the system (current carrying states, applied field corrections, confinement potential renormalisation, bound states, resonances and lifetime broadening of states). Equation (A15) is the second Keldysh equation describing the statistical distribution of states due to scattering (analogous to the Boltzmann equation). $H_{ext}(\mathbf{r})$ is the total potential energy comprising, confinement potential including possible interface roughness, plus the electrostatic potential energy $e\phi(\mathbf{r})$ determined self-consistently from Poisson's equation.

A connection to the Boltzmann transport equation may be made by defining the relative and centre coordinates r_{rel} and R , respectively:

$$r_{rel} = \mathbf{r} - \mathbf{r}'; R = (\mathbf{r} + \mathbf{r}')/2 \quad (\text{A18})$$

Assuming the Hamiltonian for the electron kinetic energy is a function of the momentum operator \mathbf{p} , and by applying the inverse Fourier transform of the relative space variable into momentum space, after some manipulation, Equations (A12)–(A19) may be used to give:

$$\begin{aligned} & \left\{ \mathbf{v}(\mathbf{p}) \cdot \frac{\partial}{\partial \mathbf{r}} - \frac{\partial}{\partial \mathbf{r}} \{ H_c(\mathbf{r}) + e\phi(\mathbf{r}) \} \cdot \frac{\partial}{\partial \mathbf{p}} \right\} G^<(\mathbf{p}, \mathbf{R}, \varepsilon) \\ & = \Sigma^>(\mathbf{p}, \mathbf{R}, \varepsilon) G^<(\mathbf{p}, \mathbf{R}, \varepsilon) - \Sigma^<(\mathbf{p}, \mathbf{R}, \varepsilon) G^>(\mathbf{p}, \mathbf{R}, \varepsilon) \end{aligned} \quad (\text{A19})$$

Equation (A19) has close similarity to the Boltzmann Equations (4) and (6), but $G^<$ and $G^>$ are generalised Wigner distributions similar to $f(p,R)$ and $(1-f(p,r))$ in semi-classical theory. The scattering integral on the RHS of Equation (A19) acquires the Boltzmann like exclusion principle structure $f(1-f)$ when we deconstruct the self-energies $\Sigma^< \propto G^<$, $\Sigma^> \propto G^>$ as shown in Appendix D.

Equation (A19) is difficult to solve compared to the standard NEGF spatial matrix element energy formulation. Discretising the space dimensions and taking the simple difference form of H_0 yields a set of coupled matrix equations that may be solved recursively. For nanowire devices, the problem is sufficiently tractable to be solved quickly on a laptop. 2D or full 3D is much more computationally expensive.

Appendix D. Phonon Green Functions and Electron-Phonon Self-Energies

To fully specify the basic Equations (A12)–(A16), we must construct the self-energies for the contacts $\Sigma_{contacts}^<$, $\Sigma_{contacts}^R$ and construct the self-energies for electron-phonon scattering using the phonon Green functions D^R and $D^<$. In the conventional decoupling approximation, we may assign the phonon Green functions in their local equilibrium forms. The decomposition equations are in symbolic operator form [74]:

$$\Sigma^R = G^R D^R + G^R D^< + G^< D^R \quad (A20)$$

$$\Sigma^< = G^< D^< \quad (A21)$$

where the matrix products involve an implicit energy convolution integral of the form

$$\Sigma(\varepsilon) \sim \int d\varepsilon' G(\varepsilon - \varepsilon') D(\varepsilon') \quad (A22)$$

Several authors have neglected the third term on the RHS of Equation (A20), which can be fatal as it controls the Pauli exclusion, as pointed out in detail by Kubis and Vogel [104,105].

Here, we give explicit expressions for the lowest order self-consistent expression for the electron-phonon “lesser” self-energy written as a sum over the phonon branches b with energy $\hbar\omega_{bq}$, where q denotes the phonon wave vector.

It is assumed, for generality, that the phonon distributions are at individual lattice temperatures T_b . A 3D description is given for generality.

The free phonon Hamiltonian is:

$$H_{ph} = \sum_b \int d^3q \left(N_{bq} + \frac{1}{2} \right) \hbar\omega_{bq} \quad (A23)$$

where $N_{bq} = a_{bq}^\dagger a_{bq}$ is the phonon number operator in terms of the usual Bosonic annihilation and creation operators a_{bq} , a_{bq}^\dagger , respectively. The condition that the phonons comprise a thermal equilibrium heat bath (possibly locally) requires us to define the phonon Green functions in terms of the following transformed operators (see for example [71])

$$A_{bq} \equiv (a_{bq} + a_{b-q}^\dagger); A_{bq}^\dagger \equiv (a_{b-q} + a_{bq}^\dagger) \quad (A24)$$

in the Heisenberg picture

$$A_{bq}(t) = A_{b-q}^\dagger(t) = \exp[-i\omega_{bq}t] a_{bq} + \exp[i\omega_{bq}t] a_{b-q}^\dagger \quad (A25)$$

The basic double-time phonon Green function is time-ordered on the Keldysh contour (using the Boson time-ordering rules)

$$D_b = \langle T_{Keldysh} A_{bq}(t) A_{b-q}^\dagger(t) \rangle_{0eq} \quad (A26)$$

The relevant piece-wise time-ordered phonon Green functions are

$$D_b^<(q, \varepsilon) = D_b^>(-q, \varepsilon) = -2\pi i \{n_{bq} \delta(\varepsilon - \hbar\omega_{bq}) + (n_{bq} + 1) \delta(\varepsilon + \hbar\omega_{bq})\} \tag{A27}$$

$$D^R(q, \varepsilon) = \{(\varepsilon + i\eta - \hbar\omega_{bq})^{-1} - (\varepsilon + i\eta + \hbar\omega_{bq})^{-1}\} \tag{A28}$$

The basic electron-phonon interaction [71] is

$$H_{int}(r, t) = \psi^\dagger(r, t) \sum_b \int \frac{d^3q}{(2\pi)^3} U_b(q) e^{iq \cdot r} \{a_{bq}(t) + a_{b-q}^\dagger(t)\} \psi(r, t) \tag{A29}$$

$U_b(q)$ is the matrix element of the coupling potential for the b th mode phonon in the lattice wave-vector representation. In many-body perturbation theory the lowest self-consistent approximation for the total electron-phonon self-energies are obtained in the self-consistent Born approximation (SCBA), using Equations (A21) and (A27):

$$\begin{aligned} \Sigma^< &= G^< D^< \rightarrow \Sigma^<(r_1, r_2, \varepsilon) \\ &= \sum_b \int \frac{d^3q}{(2\pi)^3} |U_b(q)|^2 e^{iq \cdot (r_1 - r_2)} \int \frac{d\varepsilon'}{2\pi} D^<(q, \varepsilon') G^<(r_1, r_2, \varepsilon - \varepsilon') \\ &= \sum_b \int \frac{d^3q}{(2\pi)^3} |U_b(q)|^2 e^{iq \cdot (r_1 - r_2)} \times \{n_{bq} G^<(r_1, r_2, \varepsilon - \hbar\omega_{bq}) + \\ &\quad (n_{bq} + 1) G^<(r_1, r_2, \varepsilon + \hbar\omega_{bq})\} \alpha \end{aligned} \tag{A30}$$

This may be compared with the semi-classical Equation (7). Here, the phonon distributions are assumed to be Bose-Einstein at temperature T_b :

$$n_{bq} = \left(\exp\left[\frac{\hbar\omega_{bq}}{k_B T_b}\right] - 1 \right)^{-1} \tag{A31}$$

Here, the lattice temperature is potentially allowed to be different from the ambient temperature due to heat dissipation in the non-equilibrium device.

The SCBA also yields Σ^R which assumes a much more complicated form, again carrying out the implicit convolution integral over energy, using Equations (A20) and (A28):

$$\begin{aligned} \Sigma^R &= G^R D^R + G^R D^< + G^< D^R \rightarrow \Sigma^R(r_1, r_2, \varepsilon) \\ &= \sum_b \int \frac{d^3q}{(2\pi)^3} |U_b(q)|^2 e^{iq \cdot (r_1 - r_2)} \times \{n_{bq} G^R(r_1, r_2, \varepsilon + \hbar\omega_{bq}) \\ &\quad + (n_{bq} + 1) G^R(r_1, r_2, \varepsilon - \hbar\omega_{bq}) + \left(\frac{1}{2}\right) (G^<(r_1, r_2, \varepsilon - \hbar\omega_{bq}) - G^<(r_1, r_2, \varepsilon + \hbar\omega_{bq})) \\ &\quad - \left(\frac{i}{2\pi}\right) \int d\varepsilon' \left\{ P \frac{G^<(r_1, r_2, \varepsilon - \varepsilon')}{\varepsilon' + \hbar\omega_{bq}} - P \frac{G^<(r_1, r_2, \varepsilon - \varepsilon')}{\varepsilon' - \hbar\omega_{bq}} \right\} \} \end{aligned} \tag{A32}$$

Note the coupling to $G^<$ and principal-part (P), integral over the intermediate energy, that derives from Equation (A28) for D^R . The above equations use the identity

$$\lim_{\eta \rightarrow 0^+} \frac{1}{\varepsilon - \varepsilon' + i\eta} = \frac{P}{\varepsilon - \varepsilon'} - i\pi \delta(\varepsilon - \varepsilon') \tag{A33}$$

The principal-part integrals cancel out for *elastic* scattering: $\omega_{bq} \rightarrow 0$. It should be noted that if the interaction potential is independent of q and if the phonon energies are dispersionless, then the self-energy is diagonal.

It is tempting to replace the Green functions $G^<$ in Equation (A22) by the unperturbed forms $G_0^<$, which gives essentially the semi-classical Equation (7); however, the standard Born approximation fails the continuity equation which for the steady state assumes the form:

$$\nabla \cdot \mathbf{J}(\mathbf{r}) = \int d\varepsilon \left(\frac{e}{\hbar} \right) \langle \mathbf{r} | \{ G^R \Sigma^< - \Sigma^< G^A + G^< \Sigma^A - \Sigma^R G^< \} | \mathbf{r} \rangle = \int d\varepsilon \langle \mathbf{r} | I_{current} | \mathbf{r} \rangle \quad (\text{A34})$$

where we recognise the current operator $I_{current}$ [57]

$$I_{current}(\varepsilon) = \left(\frac{e}{\hbar} \right) \{ G^R \Sigma^< - \Sigma^< G^A + G^< \Sigma^A - \Sigma^R G^< \} \quad (\text{A35})$$

Appendix E. Elastic Phonon Scattering Gives no Heat Dissipation

Using the specific forms for the self-energies (Equations (A30) and (A32)) for the elastic scattering limit of acoustic phonon scattering (high temperature model), we may demonstrate that Equation (A30) vanishes within the device volume. The coupling interaction is:

$$|U(\mathbf{q})|^2 = \frac{\hbar \Xi_d^2}{2\rho_d \omega_q} q^2 \quad (\text{A36})$$

where the acoustic phonon dispersion relation is assumed to be the linear form

$$\omega_q = v_s q \quad (\text{A37})$$

where v_s is the speed of sound in the channel. At high temperatures, the Bose-Einstein functions are given in the equipartition limit as

$$n_q \sim (n_q + 1) \sim \frac{k_B T}{\hbar \omega_q} \quad (\text{A38})$$

Using Equation (A38) and the dispersion relation (A37) the lesser self-energy becomes diagonal:

$$\Sigma^<(\mathbf{r}_1, \mathbf{r}_2, \varepsilon) = \int \frac{d^3 q}{(2\pi)^3} |C| e^{i\mathbf{q} \cdot (\mathbf{r}_1 - \mathbf{r}_2)} G^<(\mathbf{r}_1, \mathbf{r}_2, \varepsilon) = |C| G^<(\mathbf{r}_1, \mathbf{r}_1, \varepsilon) \delta(\mathbf{r}_1 - \mathbf{r}_2) \quad (\text{A39})$$

$$|C| = \frac{\Xi_d^2 k_B T}{\rho_d v_s^2} \quad (\text{A40})$$

In a similar fashion, using Equation (A22), it is found that Σ^R decouples from $G^<$ to give a diagonal form:

$$\Sigma^R(\mathbf{r}_1, \mathbf{r}_2, \varepsilon) = |C| G^<(\mathbf{r}_1, \mathbf{r}_1, \varepsilon) \delta(\mathbf{r}_1 - \mathbf{r}_2) \quad (\text{A41})$$

Inserting the diagonal Equations (A39) and (A46) into Equation (A34) gives

$$\begin{aligned} \nabla \cdot \mathbf{J}(\mathbf{r}) &= \int d\varepsilon \left(\frac{e}{\hbar} \right) \langle \mathbf{r} | (G^R - G^A) \Sigma^< + (\Sigma^A - \Sigma^R) G^< | \mathbf{r} \rangle \\ &= \int d\varepsilon \left(\frac{e}{\hbar} \right) \langle \mathbf{r} | (G^R - G^A) |C| G^< + |C| (G^A - G^R) G^< | \mathbf{r} \rangle = 0 \end{aligned} \quad (\text{A42})$$

Similarly, the divergence of the total energy current density (Equation (67)) may be shown to vanish for elastic acoustic phonon scattering:

$$\begin{aligned} \nabla \cdot \mathbf{J}_{Energy}(\mathbf{r}) &= \int d\varepsilon \varepsilon \langle \mathbf{r} | (G^R - G^A) \Sigma^< + (\Sigma^A - \Sigma^R) G^< | \mathbf{r} \rangle \\ &= \int d\varepsilon \varepsilon \langle \mathbf{r} | (G^R - G^A) |C| G^< + |C| (G^A - G^R) G^< | \mathbf{r} \rangle = 0 \end{aligned} \quad (\text{A43})$$

Appendix F. Detailed Balance

Integrating Equation (A34) over all space in the device we obtain

$$\int d^3r \nabla \cdot \mathbf{J}(\mathbf{r}) = \int d\varepsilon \text{Tr}[\mathbf{I}_{current}(\varepsilon)] = 0 \quad (\text{A44})$$

The expression

$$\text{Tr}[\mathbf{I}_{current}(\varepsilon)] = \text{Tr}[G^<\Sigma^> - G^>\Sigma^<] \quad (\text{A45})$$

is basically a trace over the RHS of Equation (A19) and is similar to the principle of detailed balance in semi-classical theory, but here the trace does not vanish in general to satisfy Equation (A20). It is the energy integral over Equation (A45) that vanishes. We note there is a wealth of physics hidden in Equation (A45), especially if the contact self-energies are made explicit.

For inelastic scattering, the power dissipation to the lattice is

$$P(\mathbf{r}) = \int d\varepsilon \varepsilon \sum_{\mathbf{r}'} \{G^<(\mathbf{r}, \mathbf{r}'\varepsilon)\Sigma^>(\mathbf{r}, \mathbf{r}', \varepsilon) - G^>(\mathbf{r}, \mathbf{r}'\varepsilon)\Sigma^<(\mathbf{r}, \mathbf{r}', \varepsilon)\} \quad (\text{A46})$$

Appendix G. Kramers-Kronig Relations

There have been many instances of device modelling studies that have neglected the real part of the retarded self-energy often on the grounds that is a “small” correction. In fact this cannot be the case because the self-energy has a definite analyticity structure arising from causality and this leads to a Kramers-Kramers-Kronig relation (Hilbert transform) between the real part Δ and the imaginary part Γ of the self-energy:

$$\Delta(\varepsilon) = -\frac{P}{\pi} \int_{-\infty}^{\infty} d\varepsilon' \frac{\Gamma(\varepsilon')}{\varepsilon' - \varepsilon} \quad (\text{A47})$$

As can be seen from Equation (47), the calculation of the real part requires an extra energy integration loop for each energy resulting in an additional computational load. Our studies have shown that this leads to significant differences in device modelling predictions from those neglecting the real part. In particular, the local density of states may be seriously mis-calculated and the spectral density sum rule [71] is violated. There are then significant errors caused in electro-thermal predictions. Detailed discussion may be found in the references [77,106–108]. In addition, it has been demonstrated that the neglect of the real part of the self-energy, for simulation of low dimensional nanostructures induces a violation of charge conservation of 5% [108].

References

1. Pop, E. Energy dissipation and transport in nanoscale devices. *Nano Res.* **2010**, *3*, 147–169. [[CrossRef](#)]
2. Farrah, H.; Steinberg, R. Analysis of double-gate thin-film transistor. *IEEE Trans. Electron Devices* **1967**, *14*, 69–74. [[CrossRef](#)]
3. Colinge, J.P. *FinFETs and Other Multi-Gate Transistors*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2008.
4. Martinez, J.; Martinez, R.V.; Garcia, R. Silicon nanowire transistors with a channel width of 4 nm fabricated by atomic force microscope nano-lithography. *Nano Lett.* **2008**, *8*, 3636–3639. [[CrossRef](#)] [[PubMed](#)]
5. Wang, D.; Wang, Q.; Javey, A.; Tu, R.; Dai, H.; Kim, H.; McIntyre, P.C.; Krishnamohan, T.; Saraswat, K.C. Germanium nanowire field-effect transistors with SiO₂ and high-k HfO₂ gate dielectrics. *Appl. Phys. Lett.* **2003**, *483*, 2432–2434. [[CrossRef](#)]
6. Vashaee, D.; Shakouri, A.; Goldberger, J.E.; Kuykendall, T.; Pauzauskie, P.; Yang, P. Electrostatics of nanowire transistors with triangular cross sections. *J. Appl. Phys.* **2006**, *99*, 054310. [[CrossRef](#)]
7. Suk, S.D.; Li, M.; Yeoh, Y.Y.; Yeo, K.H.; Cho, K.H.; Ku, I.K.; Cho, H.; Jang, W.; Kim, N.-W.; Park, D.; et al. Investigation of nanowire size dependency on TSNWFET. In Proceedings of the 2007 IEEE International Electron Devices Meeting, Washington, DC, USA, 5–7 December 2011.

8. Jeong, M.; Doris, B.; Kedzierski, J.; Rim, K.; Yang, M. Silicon Device Scaling to the Sub-10-nm Regime. *Science* **2004**, *306*, 2057–2060. [[CrossRef](#)]
9. Majima, H.; Ishikuro, H.; Hiramoto, T. Experimental evidence for quantum mechanical narrow channel effect in ultra-narrow MOSFET's. *IEEE Elect. Dev. Lett.* **2000**, *21*, 396–398. [[CrossRef](#)]
10. Barker, J.R.; Martinez, A. Image charge models for accurate construction of the electrostatic self-energy of 3D layered nanostructure devices. *J. Phys. Condens. Matter* **2018**, *30*, 134002. [[CrossRef](#)]
11. Martinez, A.; Aldegunde, M.; Seoane, N.; Brown, A.; Barker, J.R.; Asenov, A. Quantum-Transport Study on the Impact of Channel Length and Cross Sections on Variability Induced by Random Discrete Dopants in Narrow Gate-All-Around Silicon Nanowire Transistors. *IEEE Trans. Electron Devices* **2011**, *58*, 2209–2217. [[CrossRef](#)]
12. Pop, E.; Sinha, S.; Goodson, K. Heat Generation and Transport in Nanometer-Scale Transistors. *Proc. IEEE* **2006**, *94*, 1587–1601. [[CrossRef](#)]
13. Feser, J.P.; Sadhu, J.S.; Azeredo, B.P.; Hsu, K.H.; Ma, J.; Kim, J.; Seong, M.; Fang, N.X.; Li, X.; Ferreira, P.M.; et al. Thermal conductivity of silicon nanowire arrays with controlled roughness. *J. App. Phys.* **2012**, *112*, 114306. [[CrossRef](#)]
14. Karg, S.; Mensch, P.; Gotsmann, B.; Schmid, H.; Das Kanungo, P.; Ghoneim, H.; Schmidt, V.; Björk, M.T.; Troncale, V.; Riel, H. Measurement of Thermoelectric Properties of Single Semiconductor Nanowires. *J. Electron. Mater.* **2013**, *42*, 2409–2414. [[CrossRef](#)]
15. Menges, F.; Riel, H.; Stemmer, A.; Gotsmann, B. Quantitative Thermometry of Nanoscale Hot Spots. *Nano Lett.* **2012**, *12*, 596–601. [[CrossRef](#)] [[PubMed](#)]
16. Menges, F.; Mensch, P.; Schmid, H.; Riel, H.; Stemmer, A.; Gotsmann, B. Measurement of Thermoelectric Properties of Single Semiconductor Nanowires. *Nat. Commun.* **2016**, *7*, 10874. [[CrossRef](#)] [[PubMed](#)]
17. Goodson, K.; Flik, M. Effect of microscale thermal conduction on the packing limit of silicon-on-insulator electronic devices. *IEEE Trans. Compon. Hybrids Manuf. Technol.* **1992**, *15*, 715–722. [[CrossRef](#)]
18. Shakouri, A.; Lee, E.Y.; Smith, D.L.; Narayanamurti, V.; Bowers, J.E. Thermoelectric Effects In Submicron Heterostructure Barriers. *Microscale Thermophys. Eng.* **1998**, *2*, 37–47. [[CrossRef](#)]
19. Chen, G.; Shakouri, A. Heat Transfer in Nanostructures for Solid-State Energy Conversion. *J. Heat Transf.* **2001**, *124*, 242–252. [[CrossRef](#)]
20. Shakouri, A.; Zebarjadi, M. Nanoengineered Materials for Thermoelectric Energy Conversion. In *Thermal Nanosystems and Nanomaterials; Topics in Applied, Physics*, Volz, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2009; Volume 118, pp. 225–299.
21. Borkar, S. Design challenges of technology scaling. *IEEE Micro* **1999**, *19*, 23–29. [[CrossRef](#)]
22. Whitney, R.S.; Sánchez, R.; Splettstoesser, J. Quantum Thermodynamics of Nanoscale Thermoelectrics and Electronic Devices. In *Thermodynamics in the Quantum Regime. Fundamental Theories of Physics*; Binder, F., Correa, L., Gogolin, C., Anders, J., Adesso, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2018; Volume 195.
23. Van Roosbroeck, W. Theory of the Flow of Electrons and Holes in Germanium and Other Semiconductors. *Bell Syst. Tech. J.* **1950**, *29*, 560–607. [[CrossRef](#)]
24. Shockley, W. *Electrons and Holes in Semiconductors*; D. Van Nostrand Company, Inc.: Princeton, NJ, USA, 1950.
25. Sze, S.M.; Mattis, D.C. *Physics of Semiconductor Devices*; Wiley-Interscience: New York, NY, USA, 1981.
26. Rohr, P.; Lindholm, F.A.; Allen, K.R. Questionability of drift-diffusion transport in the analysis of small semiconductor devices. *Solid State Electron.* **1974**, *17*, 729–734. [[CrossRef](#)]
27. Barker, J.; Ferry, D. On the physics and modeling of small semiconductor devices—I. *Solid State Electron.* **1980**, *23*, 519–530. [[CrossRef](#)]
28. Jacoboni, C.; Lugli, P. *The Monte Carlo Method for Semiconductor Device Simulation*; Springer Science: Berlin/Heidelberg, Germany, 2012.
29. Fischetti, M.V. Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. I. Homogeneous transport. *IEEE Trans. Electron Devices* **1991**, *38*, 634–649. [[CrossRef](#)]
30. Frank, D.J.; Laux, S.E.; Fischetti, M.V. Monte Carlo simulation of a 30 nm dual-gate MOSFET: How short can Si go? In *IEDM Tech. Dig.* **1992**, 553–556. [[CrossRef](#)]
31. Sadi, T.; Kelsall, R.W.; Pilgrim, N.J. Electrothermal Monte Carlo simulation of submicrometer Si/SiGe MO DFETs. *IEEE Trans. Electron Devices* **2007**, *54*, 332–339. [[CrossRef](#)]
32. Chen, Y.; Li, D.; Lukes, J.R.; Majumdar, A. Monte Carlo Simulation of Silicon Nanowire Thermal Conductivity. *ASME J. Heat Transfer* **2005**, *127*, 1129–1137. [[CrossRef](#)]

33. Mazumder, S.; Majumdar, A. Monte Carlo Study of Phonon Transport in Solid Thin Films including Dispersion and Polarization. *ASME J. Heat Transfer* **2001**, *123*, 749–759.
34. Sinha, S.; Goodson, K.E. Review: Multiscale Thermal Modeling in Nanoelectronics. *Int. J. Multiscale Comput. Eng.* **2005**, *3*, 107–133. [[CrossRef](#)]
35. Raleva, K.; Vasileska, D.; Goodnick, S.M.; Nedjalkov, M. Modeling Thermal Effects in Nanodevices. *IEEE Trans. Electron Devices* **2008**, *55*, 1306–1316. [[CrossRef](#)]
36. Vasileska, D.; Goodnick, S.M.; Klimeck, G. *Computational Electronics: Semiclassical and Quantum Transport Modeling*; CRC Press: Boca Raton, FL, USA, 2010.
37. Hossein, A.; Raleva, K.; Vasileska, D.; Goodnick, S.M. Self-Heating effects in nanowire transistors. *NST Nanotech.* **2010**, *2*, 45–48.
38. Callen, H.B. *Thermodynamics and an Introduction to Thermostatistics*, 2nd ed.; John Wiley and Sons: Chichester, UK, 1985.
39. Lindefelt, U. Heat generation in semiconductor devices. *J. Appl. Phys.* **1994**, *75*, 942–957. [[CrossRef](#)]
40. Langreth, D.C. Linear and Non Linear Response Theory with Applications. In *Linear and Non-Linear Electron Transport in Solids*; Devreese, J.T., van Boren, E., Eds.; Plenum: New York, NY, USA, 1976; pp. 3–32.
41. Khan, F.S.; Davies, J.H.; Wilkins, J.W. Quantum transport equations for high electric fields. *Phys. Rev. B* **1987**, *36*, 2578–2597. [[CrossRef](#)] [[PubMed](#)]
42. Schwinger, J. Brownian Motion of a Quantum Oscillator. *J. Math. Phys.* **1961**, *2*, 407. [[CrossRef](#)]
43. Keldysh, L.V. Diagram technique for non-equilibrium processes. *Sov. Phys. JETP* **1965**, *20*, 1018.
44. Kadanoff, L.P.; Baym, G. *Quantum Statistical Mechanics*; CRC Press: Boca Raton, FL, USA, 2018.
45. Barker, J.R. Quantum theory of high field transport in semiconductors. *J. Phys. C* **1973**, *6*, 2663–2685. [[CrossRef](#)]
46. Barker, J.R.; Ferry, D.K. Self-Scattering Path-Variable Formulation of High-Field, Time-Dependent, Quantum Kinetic Equations for Semiconductor Transport in the Finite-Collision-Duration Regime. *Phys. Rev. Lett.* **1979**, *42*, 1779–1781. [[CrossRef](#)]
47. Barker, J.R. Fundamental aspects of quantum transport. In *Handbook of Semiconductors*; Paul, W., Ed.; North Holland: Amsterdam, The Netherlands, 1982; Volume 1, Chapter 13; pp. 617–659.
48. Pop, E. Heat Generation and Transport in SOI and GOI Devices. *ECS Meet. Abstr.* **2007**, *6*, 151–157. [[CrossRef](#)]
49. Pala, M.; Cresti, A. Increase of self-heating effects in nanodevices induced by surface roughness: A full-quantum study. *J. Appl. Phys.* **2015**, *117*, 084313. [[CrossRef](#)]
50. Rhyner, R.; Luisier, M. Influence of anharmonic phonon decay on self-heating in Si nanowire transistors. *Appl. Phys. Lett.* **2014**, *105*, 062113. [[CrossRef](#)]
51. Fischetti, M.V.; Laux, S.E. Performance degradation of small silicon devices caused by long-range Coulomb interactions. *Appl. Phys. Lett.* **2000**, *76*, 2277–2279. [[CrossRef](#)]
52. Fischetti, M.V. Long-Range Coulomb interactions in small Si devices. Part II. Effective electron mobility in thin-oxide structures. *J. Appl. Phys.* **2001**, *89*, 1232–1250. [[CrossRef](#)]
53. Fischetti, M.V.; Neumayer, D.A.; Cartier, E.A. Effective electron mobility in Si inversion layers in metal oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering. *J. Appl. Phys.* **2001**, *90*, 4587–4608. [[CrossRef](#)]
54. Barker, J.R.; Martinez, A. Remote soft-optical phonon scattering in Si nanowire FETs. In Proceedings of the 2014 International Workshop on Computational Electronics (IWCE 2014), Institute of Electrical and Electronics Engineers (IEEE), Paris, France, 3–6 June 2014; pp. 143–145.
55. Srivastava, G.P.; Kresin, V. The Physics of Phonons. *Phys. Today* **1991**, *44*, 75. [[CrossRef](#)]
56. Broido, D.; Malorny, M.; Birner, G.; Mingo, N.; Stewart, D.A. Intrinsic lattice thermal conductivity of semiconductors from first principles. *Appl. Phys. Lett.* **2007**, *91*, 231922. [[CrossRef](#)]
57. Datta, S. *Electronic Transport in Mesoscopic Systems*; Cambridge University Press: Cambridge, UK, 1997.
58. Datta, S. *Quantum Transport: Atom to Transistor*; Cambridge University Press: Cambridge, UK, 2005.
59. Landauer, R. Spatial Variation of Currents and Fields Due to Localized Scatterers in Metallic Conduction. *IBM J. Res. Dev.* **1957**, *1*, 223–231. [[CrossRef](#)]
60. Wachutka, G. Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **1990**, *9*, 1141–1149. [[CrossRef](#)]
61. Sadi, T.; Thobel, J.-L.; Dessenne, F. Self-Consistent electrothermal Monte Carlo simulation of single InAs nanowire channel metal-insulator field-effect transistors. *J. Appl. Phys.* **2010**, *108*, 084506. [[CrossRef](#)]

62. Lai, J.; Majumdar, A. Concurrent thermal and electrical modeling of sub-micrometer silicon devices. *J. Appl. Phys.* **1996**, *79*, 7353–7361. [[CrossRef](#)]
63. Majumdar, A.; Fushinobu, K.; Hijikata, K. Effect of gate voltage on hot-electron and hot phonon interaction and transport in a submicrometer transistor. *J. Appl. Phys.* **1995**, *77*, 6686–6694. [[CrossRef](#)]
64. Goodnick, S.; Raleva, K.; Vasileska, D. Self-Consistent thermal electron-phonon simulator for SOI devices. In Proceedings of the 11th Annual NSTI Nanotechnology Conference and Trade Show, Boston, MA, USA, 1–5 June 2008; pp. 537–540.
65. Goodnick, S.M.; Raleva, K.; Vasileska, D. Heating Effects in Dual-Gate Devices. In Proceedings of the 2008 8th IEEE Conference on Nanotechnology, Arlington, TX, USA, 18–21 August 2008; pp. 10–13.
66. Goodnick, S.M.; Honsberg, C. Ultrafast carrier relaxation and nonequilibrium phonons in hot carrier solar cells. In Proceedings of the 2011 37th IEEE Photovoltaic Specialists Conference, Seattle, WA, USA, , 19–24 June 2011; pp. 002066–002070.
67. Raleva, K.; Bury, E.; Kaczer, B.; Vasileska, D. Uncovering the temperature of the hotspot in nanoscale devices. In Proceedings of the 2014 International Workshop on Computational Electronics, Paris, France, 3–6 June 2014; pp. 104–106.
68. Mohamed, M.; Aksamija, Z.; Vitale, W.; Hassan, F.; Park, K.H.; Ravaoli, U. A Conjoined Electron and Thermal Transport Study of Thermal Degradation Induced During Normal Operation of Multigate Transistors. *IEEE Trans. Electron Devices* **2014**, *61*, 976–983. [[CrossRef](#)]
69. Mohamed, M.; Aksamija, Z.; Godoy, A.; Martin, P.; Hahm, H.-S.; Lee, W.; Lee, K.; Ravaoli, U. Size effects and performance assessment in nanoscale multigate MOS FET structures. *J. Comput. Theoret. Nanosci.* **2009**, *6*, 1927–1936. [[CrossRef](#)]
70. Lifshitz, E.M.; Pitaevski, L.P. *Lehrbuch der Theoretischen Physik*; Band X Physikalische Kinetik; Akademie-Verlag: Berlin, Germany, 1983.
71. Bruus, H.; Flensberg, K. *Many-Body Quantum Theory in Condensed Matter Physics*; Oxford University Press: New York, NY, USA, 2004.
72. Rammer, J.; Smith, H. Quantum field-theoretical methods in transport theory of metals. *Rev. Mod. Phys.* **1986**, *58*, 323–359. [[CrossRef](#)]
73. Danielewicz, P. Quantum Theory of Non-Linear Processes. *Ann. Phys.* **1984**, *152*, 239. [[CrossRef](#)]
74. Haug, H.; Jauho, A. -P. *Quantum Kinetics in Transport and Optics of Semiconductors*; Springer: Berlin/Heidelberg, Germany, 1996.
75. Mathieu, L.; Rhyner, R. Atomistic simulation of electron and phonon transport in nano-devices. In Proceedings of the European Solid-State Device Research Conference (ESSDERC), Bucharest, Romania, 16–20 September 2020; pp. 308–313.
76. Aldegunde, M.; Martinez, A.; Barker, J.R. Study of individual phonon scattering mechanisms and the validity of Matthiessen's rule in a gate-all-around silicon nanowire transistor. *J. Appl. Phys.* **2013**, *113*, 14501. [[CrossRef](#)]
77. Valin, R.; Aldegunde, M.; Martinez, A.; Barker, J.R. Quantum Transport of a Nanowire FET with Complex Phonon Self-Energy. *J. Appl. Phys.* **2014**, *116*, 084507. [[CrossRef](#)]
78. Luttinger, J.M.; Kohn, W. Motion of Electrons and Holes in Perturbed Periodic Fields. *Phys. Rev.* **1955**, *97*, 869–883. [[CrossRef](#)]
79. Martínez, A.; Bescond, M.; Barker, J.R.; Svizhenko, A.; Anantram, M.; Millar, C.; Asenov, A. A Self-Consistent Full 3-D Real-Space NEGF Simulator for Studying Nonperturbative Effects in Nano-MOSFETs. *IEEE Trans. Electron Devices* **2007**, *54*, 2213–2222. [[CrossRef](#)]
80. Nehari, K.; Cavassilas, N.; Autran, J.; Bescond, M.; Munteanu, D.; Lannoo, M. Influence of band structure on electron ballistic transport in silicon nanowire MOSFET's: An atomistic study. *Solid State Electron.* **2006**, *50*, 716–721. [[CrossRef](#)]
81. Barker, J.R. Theory of quantum transport in lateral nanostructures. In *Physics and Fabrication of Nanostructures*; Read, M., Kirk, W.P., Eds.; Academic Press: Cambridge, MA, USA, 1989; pp. 253–262.
82. Barker, J.; Pepin, J.; Finch, M.; Laughton, M. Theory of non-linear transport in quantum waveguides. *Solid State Electron.* **1989**, *32*, 1155–1159. [[CrossRef](#)]
83. Venugopal, R.; Ren, Z.; Datta, S.; Lundstrom, M.S.; Jovanović, D. Simulating quantum transport in nanoscale transistors: Real versus mode-space approaches. *J. Appl. Phys.* **2002**, *92*, 3730–3739. [[CrossRef](#)]
84. Luisier, M.; Schenk, A.; Fichtner, W. Quantum transport in two- and three-dimensional nanoscale transistors: Coupled mode effects in the nonequilibrium Green's function formalism. *J. Appl. Phys.* **2006**, *100*, 43713. [[CrossRef](#)]

85. Lundstrom, M. *Fundamentals of Carrier Transport*; Cambridge University Press: Cambridge, UK, 2000.
86. Martinez, A.; Price, A.; Valin, R.; Aldegunde, M.; Barker, J. Impact of phonon scattering in Si/GaAs/InGaAs nanowires and FinFets: A NEGF perspective. *J. Comput. Electron.* **2016**, *15*, 1130–1147. [[CrossRef](#)]
87. Supriyo, D. Nanoscale device modeling: the Green's function method. *Superlattices Microstruct.* **2000**, *28*, 253–278.
88. Kubo, R. The fluctuation-dissipation theorem. *Rep. Prog. Phys.* **1966**, *29*, 255–284. [[CrossRef](#)]
89. Smith, A.C.; Janak, J.F.; Adler, R.B. *Electronic Conduction in Solids*; McGraw Hill: New York, NY, USA, 1967.
90. Lake, R.; Datta, S. Nonequilibrium Green's-Function method applied to double-barrier resonant-tunneling diodes. *Phys. Rev. B* **1992**, *45*, 6670–6685. [[CrossRef](#)]
91. Lake, R.; Datta, S. Energy balance and heat exchange in mesoscopic systems. *Phys. Rev. B* **1992**, *46*, 4757–4763. [[CrossRef](#)]
92. Martinez, A.; Barker, J.R.; Aldegunde, M.; Valin, R. Study of Local Power Dissipation in Ultrascaled Silicon Nanowire FETs. *IEEE Electron Device Lett.* **2014**, *36*, 2–4. [[CrossRef](#)]
93. Rhyner, R.; Luisier, M. Atomistic modeling of coupled electron-phonon transport in nanowire transistors. *Phys. Rev. B* **2014**, *89*, 235311. [[CrossRef](#)]
94. Price, A.; Martinez, A. Electrothermal simulations of Si and III-V nanowire field effect transistors: A non-equilibrium Green's function study. *J. Appl. Phys.* **2017**, *122*, 074502. [[CrossRef](#)]
95. Laux, S.; Kumar, A.; Fischetti, M. QDAME simulation of 7.5 nm double-gate Si nFETs with differing access geometries. In Proceedings of the Digest International Electron Devices Meeting, San Francisco, CA, USA, 8–11 December 2002; pp. 715–718. [[CrossRef](#)]
96. Kurniawan, O.; Bai, P.; Li, E. Ballistic calculation of nonequilibrium Green's function in nanoscale devices using finite element method. *J. Phys. D Appl. Phys.* **2009**, *42*, 105109. [[CrossRef](#)]
97. Svizhenko, A.; Anantram, M.P.; Govindan, T.R.; Biegel, B.; Venugopal, R. Two-Dimensional quantum mechanical modeling of nano-transistors. *J. Appl. Phys.* **2002**, *91*, 2343–2354. [[CrossRef](#)]
98. Jin, S.; Fischetti, M.V.; Tang, T.-W. Modeling of electron mobility in gated silicon nanowires at room temperature: Surface roughness scattering, dielectric screening, and band nonparabolicity. *J. Appl. Phys.* **2007**, *102*, 83715. [[CrossRef](#)]
99. Yu, S.; Iafrate, G.J.; Kim, K.W.; Stroschio, M.A. Electron-Acoustic-phonon scattering rates in cylindrical quantum wires. *Phys. Rev. B* **1995**, *51*, 4695–4698. [[CrossRef](#)] [[PubMed](#)]
100. Tanaka, H.; Suda, J.; Kimoto, T. Insight into phonon scattering in Si nanowires through high-field hole transport: Impacts of boundary condition and comparison with bulk phonon approximation. *J. Phys. Conf. Ser.* **2017**, *864*, 012046. [[CrossRef](#)]
101. Ajami, A.; Banerjee, K.; Pedram, M. Modeling and analysis of nonuniform substrate temperature effects on global ULSI interconnects. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **2005**, *24*, 849–861. [[CrossRef](#)]
102. Ma, J.; Sadhu, J.S.; Ganta, D.; Tian, H.; Sinha, S. Thermal transport in 2 and 3 dimensional periodic holey nanostructures. *AIP Adv.* **2014**, *4*, 124502. [[CrossRef](#)]
103. Caroli, C.; Combescot, R.; Lederer, D.; Nozieres, P.; Saint-James, D. Direct calculation of the tunneling current. *J. Phys. C Solid State Phys.* **1971**, *4*, 916. [[CrossRef](#)]
104. Kubis, T.; Vogl, P. Self-Consistent quantum transport theory: Applications and assessment of approximate models. *J. Comput. Electron.* **2006**, *6*, 183–186. [[CrossRef](#)]
105. Kubis, T.; Vogl, P. Assessment of approximations in non-equilibrium Green's function theory. *Phys. Rev. B* **2011**, *83*, 195304. [[CrossRef](#)]
106. Barker, J.R.; Martinez, A. Self-Energy Models for Scattering in Semiconductor Nanoscale Devices: Causality Considerations and the Spectral Sum Rule. *Mat. Res. Soc. Proc.* **2013**, *1551*, 17–22. [[CrossRef](#)]
107. Barker, J.; Martinez, A.; Aldegunde, M.; Valin, R. Causal self-energies for NEGF modelling of quantum nanowires. *J. Phys. Conf. Ser.* **2014**, *526*, 012001. [[CrossRef](#)]
108. Svizhenko, A.; Anantram, M.P. Effect of scattering and contacts on current and electrostatics in carbon nanotubes. *Phys. Rev. B* **2005**, *72*. [[CrossRef](#)]

