Solid-State Electronics 70 (2012) 81–91

Contents lists available at SciVerse ScienceDirect

Solid-State Electronics

journal homepage: www.elsevier.com/locate/sse

Confinement-induced carrier mobility increase in nanowires by quantization of warped bands

Neophytos Neophytou *, Hans Kosina

Institute for Microelectronics, TU Wien, Gußhausstraße 27-29/E360, A-1040 Vienna, Austria

ARTICLE INFO

Article history:
Available online 9 January 2012

Keywords:
Silicon nanowire
Ultra-thin-body
Mobility
Boltzmann
Atomistic
Bandstructure
$sp^3d^5s^*$ Tight binding model
Anisotropy
Warped bands

ABSTRACT

We calculate the transport characteristics of thin silicon nanowires (NWs) up to 12 nm in diameter. The $sp^3d^5s^*$-spin–orbit–coupled atomistic tight-binding (TB) model is used for the electronic structure calculation. Linearized Boltzmann transport theory is applied for transport, including carrier scattering by phonons and surface roughness (SRS). We show that for certain transport orientations, confinement in specific directions enhances the curvature of the subbands such that large improvements can be achieved in carrier velocity and mobility. This is observed for p-type [1 1 0] and [1 1 1] NWs for which confinement-induced bandstructure changes improve the phonon limited mobility by >2X compared to bulk values. Such improvement suggests very large carrier mean-free-paths for scattering and is large enough to compensate for the detrimental effect of SRS. This effect is observed at a smaller degree for n-type [1 1 0] NWs as well. The bandstructure engineering techniques we describe originate from the warped, anisotropic bulk bands, and can be also applied to 2D ultra-thin-body (UTB) layers as well as other channel materials. Our results may provide understanding of recent experimental observations, as well as guidance in the design of NW and UTB channel devices with improved transport properties.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

Silicon nanowire (NW) devices have recently attracted significant attention as candidates for a variety of applications. For field-effect-transistors, NWs offer the possibilities of enhanced electrostatic control and of quasi-ballistic transport [1–8]. For thermoelectric applications, recent experiments have shown that NWs can achieve high thermoelectric efficiency with potentially improved power factors and lower thermal conductivity [9–12]. Possible applications in optoelectronics [13,14] and biosensors [15] have also been suggested.

Ultra scaled NW devices of channel diameter down to $D = 3$ nm and lengths as short as 15 nm with excellent performance have already been demonstrated by various experimental groups [2–8]. Furthermore, large mobility and on-current densities in NWs compared to bulk have been observed experimentally [16,17]. Low-dimensional channels offer additional degrees of freedom in engineering their properties: (i) the length scale of the cross section, (ii) the transport orientation, (iii) the orientation of the confining surfaces [18,19]. The transport effective mass, carrier velocity, mean free path, and mobility can significantly vary with geometry and strongly influence the performance of devices.

To properly account for confinement and orientation effects, a well calibrated electronic structure model is needed. This is especially important for the valence band, where effective mass variations, valley splitting, and strong curvature variations are observed, and drastically affect the electronic properties [18,20,21]. In this work we employ the $sp^3d^5s^*$-spin–orbit–coupled (SO) tight-binding (TB) model which has been extensively used in electronic structure calculations for nanostructures [18,19,22–27], although other methods such as $k{\cdot}p$ [28,29] or pseudo-potentials can be equally applied.

For mobility calculation, we couple the atomistic model to linearized Boltzmann transport theory [30–32]. We include phonon and surface roughness (SRS) scattering. Similar studies that couple atomistic electronic structures to Boltzmann transport, usually consider NW diameters of only 1–3 nm due to computational reasons [11,23,33,34]. Here we describe a robust method to extend the atomistic computation to larger structures, up to 12 nm in diameter (~5500 atoms in the simulation domain), while employing a limited degree of parallelization.

The heavy-hole valence band in all semiconductors, but also the conduction band ellipsoids in silicon, are anisotropic and warped [18,19]. The effects of anisotropy and warping are more pronounced in nanostructures than in bulk channels. In this work, by taking advantage of this warping, we describe a method to achieve high performance by geometric confinement of the carriers in certain directions. We show that carrier velocity improvement

* Corresponding author.
E-mail addresses: neophytou@iue.tuwien.ac.at (N. Neophytou), kosina@iue.tuwien.ac.at (H. Kosina).

0038-1101/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.sse.2011.11.018

up to ∼2X can be achieved as the diameter is reduced, which translates to large low-filed mobility enhancements. Such strong effects can be observed for p-type [1 1 0] and [1 1 1] channels once the length scale of their cross section is scaled below ∼8 nm. This effect is also evident at a smaller degree in n-type [1 1 0] NWs.

We identify the confining directions responsible for this improvement, as well as directions that under confinement can actually cause performance degradation. Although our analysis is for Si NW channels, the bandstructure engineering techniques we describe originate from warped bulk bands, and can also be applied to 2D thin-body-layers as well as other channel materials. Our results potentially provide explanations for the large mobility enhancement in NWs [16,17] and for the strong anisotropic mobility trends in ultra-thin-layers [35] that have been observed experimentally, as well as useful insight into nanoscale device optimization strategies.

This paper is organized as follows: In Section 2 we describe the coupling of the TB electronic structures to linearized Boltzmann transport approach. We derive the scattering rates for phonons and surface roughness, and describe the approximations made to achieve computational efficiency. In Section 3 we present the bandstructure mechanisms responsible for the performance enhancement in warped bands and their effect on the electronic properties. In Section 4 we summarize and conclude the work.

2. Approach

2.1. Atomistic modeling

To obtain the bandstructure of the NWs both for electrons and holes for which spin–orbit coupling is important, a well calibrated atomistic model is used. The nearest neighbor $sp^3d^5s^*$-SO tight-binding model [18,36–39] captures all the necessary band features, and in addition, is robust enough to computationally handle larger NW cross sections as compared to *ab initio* methods. As an indication, the unit cells of the NWs considered in this study contain from ∼150 to ∼5500 atoms, and the computation time needed for the calculation of the mobility of a specific NW varies from a few to several hours for each case on a single CPU. Each atom in the NW unit cell is described by 20 orbitals including spin–orbit–coupling. The model itself and the parameterization used [36] have been extensively calibrated to various experimental data of various natures with excellent agreement without additional material parameter adjustments [40–43]. We consider silicon NWs in three different transport orientations [1 0 0], [1 1 0], and [1 1 1] and diameters varying from D = 12 nm down to D = 3 nm.

The electronic structure of ultra narrow NWs is sensitive to the diameter and orientation. For n-type NWs we previously showed that the effective masses change as the diameter is scaled below 7 nm by as much as ∼2X depending on the orientation [18]. More importantly, for p-type NWs we showed that the carrier velocities can change by a factor of ∼2X with confinement (∼4X variations in m^*), again depending on the orientation [19,22,44]. This behavior is attributed to the strong warping of the heavy-hole band. Indeed, the origin of most of these features can be traced back to the bandstructure of bulk Si.

2.2. Boltzmann theory

The linearized Boltzmann transport theory is used to extract the electronic transport properties for each wire using its dispersion relation. We consider electron–phonon and surface roughness scattering (SRS). The electrical low-field conductivity (σ) follows from the linearized Boltzmann equation as:

$$\sigma = q_0^2 \int_{E_0}^{\infty} dE \left(-\frac{\partial f(E)}{\partial E}\right) \Xi(E). \tag{1}$$

The energy E integration over the derivative of the Fermi distribution $f(E)$ is performed from the subband edge E_0 (conduction or valence band edges) though all energies. $\Xi(E)$ is the transport distribution function (TD) defined as [30,31]:

$$\begin{aligned}\Xi(E) &= \frac{1}{A}\sum_{k_x,n} v_n^2(k_x)\tau_n(k_x)\delta(E-E_n(k_x)) \\ &= \frac{1}{A}\sum_{n} v_n^2(E)\tau_n(E)g_{1D}^n(E).\end{aligned} \tag{2}$$

Here $v_n(E) = \frac{1}{\hbar}\frac{\partial E_n}{\partial k_x}$ is the bandstructure velocity of subband n with dispersion E_n, $\tau_n(k_x)$ is the momentum relaxation time for a carrier with wavenumber k_x in subband n,

$$g_{1D}^n(E) = \frac{1}{2\pi\hbar}\frac{1}{|v_n(E)|} \tag{3}$$

is the density of states for the 1D subbands (per spin), and A is the cross sectional area of the NW. The mobility (μ) is defined as:

$$\mu = \frac{\sigma}{q_0 \mathrm{n}}, \tag{4}$$

where n is the carrier concentration in the channel and q_0 is the electronic charge.

The electronic structure calculation is performed in k-space, where the energy $E_n(k_x)$, velocity $v_n(k_x)$, and density of states $g_n(k_x)$ are computed. The 1D k-space extends from $-\pi/a_0'$ to π/a_0' where a_0' is the length of the unit cell. We use 400 points for the k-space descritization, which is adequate to provide smooth and converged energy dependent functions for the quantities of interest. Due to symmetry only half of the k-space needs to be computed. Once the dispersion is computed, the states are sorted with respect to energy E in order to obtain $v_n(E), \tau_n(E)$, and $g_{1D}^n(E)$, the quantities used in the definition of the transport distribution function in Eq. (2).

2.3. Scattering rate calculation

The transition rate $S_{n,m}(k_x, k_x')$ for a carrier in an initial state k_x in subband n to a final state k_x' in subband m is extracted from the numerically calculated dispersions and wave form overlaps using Fermi's Golden Rule as [32,44]:

$$S_{n,m}(k_x, k_x') = \frac{2\pi}{\hbar}\left|H_{k_x',k_x}^{m,n}\right|^2 \delta(E_m(k_x') - E_n(k_x) - \Delta E), \tag{5}$$

where $H_{k_x',k_x}^{m,n}$ is the matrix element for corresponding transition, $\delta(\cdot)$ denotes the Dirac-Delta function, and ΔE is the energy change for the specific process. The calculation of the relaxation times involves an integral equation for $\tau_n(k_x)$ [28,45–47]:

$$\frac{1}{\tau_n(k_x)} = \sum_{m,k_x'} S_{n,m}(k_x, k_x')\left(1 - \frac{v_m(k_x')\tau_m(k_x')f_m(k_x')}{v_n(k_x)\tau_n(k_x)f_n(k_x)}\right) \tag{6}$$

Self-consistent solutions of this are computationally very expensive, especially when involving atomistic calculations, and it is common practice to simplify the problem. For isotropic scattering (all phonon mechanisms) the term in the parenthesis reduces to unity upon integration over k-space due to symmetry considerations ($S_{n,m}$ is angle independent). For elastic intra-valley, intra-subband scattering (even if anisotropic, i.e. SRS, or impurity scattering), and under the parabolic band approximation, the term simplifies to:

$$1-\frac{v_m(k_x')\tau_m(k_x')f_m(k_x')}{v_n(k_x)\tau_n(k_x)f_n(k_x)}=1-\frac{v_m(k_x')}{v_n(k_x)}=1-\frac{\hbar v_m(k_x')/m_m^*}{\hbar v_n(k_x)/m_n^*}$$
$$=1-\frac{|p_x'|}{|p_x|}\cos\vartheta \quad (7)$$

which is the usual term used in the calculation of the momentum relaxation times [32,45]. Here, p_x denotes the momentum of the state k_x and ϑ the angle between the final and initial states (0 or π in 1D). Although this is strictly valid for intra-subband transitions only, it is often used for inter-subband transitions as well (assuming a weak k_x-dependence of $\tau(k_x)$ [28]). In general, however, one should solve the full integral equation, but often sufficiently accurate results are obtained using the above approximations, without the need to evaluate numerically demanding integral equations [28,46]. In this work, we calculate the relaxation times by:

$$\frac{1}{\tau_n(k_x)}=\sum_{m,k_x'}S_{n,m}(k_x,k_x')\left(1-\frac{v_m(k_x')}{v_n(k_x)}\right) \quad (8)$$

The Fermi functions in Eq. (6) cancel for elastic processes. For inelastic, isotropic processes (as all the inelastic processes we consider) the term in brackets reduces to unity anyway after integration over k_x. Although admittedly in 1D $\tau(k_x)$ in the numerator and denominator of Eq. (6) can be varying with k_x, we still drop it as it is commonly done in the literature [26,33,48,49]. This will only affect the SRS results at larger diameters where inter-subband transitions can be important, but at larger diameters the effect of SRS weakens anyway. We, therefore, believe that such approximation will not affect our results significantly.

The matrix element is computed using the scattering potential $U_S(\vec{r})$ as:

$$H_{k_x',k_x}^{m,n}=\int_{-\infty}^{\infty}\psi_{m,k_x'}^*(\vec{r})U_S(\vec{r})\psi_{n,k_x}(\vec{r})d^3r \quad (9)$$

where $\psi_{n,k_x}(\vec{r})$ is the wave function of a carrier in an initial state k_x in subband n, and $\psi_{m,k_x'}(\vec{r})$ the wave function of a final state k_x' in subband m.

The total wavefunction of a specific state can be decomposed into a plane wave in the x-direction, and a bound state $F_n(\vec{R})$ in the transverse plane $\vec{R}$ as:

$$\psi_n(\vec{r})=\frac{F_n(\vec{R})e^{ik_xx}}{\sqrt{\Omega}}, \quad (10)$$

where Ω is the normalization volume. The matrix element then becomes:

$$H_{k_x',k_x}^{m,n}=\frac{1}{\Omega}\int_{-\infty}^{\infty}F_m^*(\vec{R})e^{-ik_x'x}U_S(\vec{r})F_n(\vec{R})e^{ik_xx}d^2Rdx. \quad (11)$$

In the TB model we do not have access to the actual wavefunctions, but only the normalized expansion coefficients of the wavefunctions for each orbital on each atomic site. Numerically, therefore, Eqs. (9) and (11) are integrated over all the atomic sites, which form the actual descritization of our domain.

2.3.1. Phonon scattering

In the case of phonon scattering, we extend the approach described in [32] for bulk and 2D carriers, to 1D carriers. The perturbing potential is defined as:

$$U_S(\vec{r})=A_{\vec{q}}K_{\vec{q}}e^{\pm i(\vec{q}.\vec{r}-\omega t)} \quad (12)$$

where $A_{\vec{q}}$ is associated with the lattice vibration amplitude and $K_{\vec{q}}$ with the deformation potential. The scalar product is split as $\vec{q}\cdot\vec{r}=\vec{Q}\cdot\vec{R}+q_xx$ with the components $\vec{Q}\cdot\vec{R}$ in the transverse and q_xx in the transport direction. In this case, the matrix element is:

$$H_{k_x',k_x}^{m,n}=\int_{-\infty}^{\infty}\frac{F_m^*(\vec{R})}{\sqrt{A}}\frac{e^{-ik_x'x}}{\sqrt{L_x}}A_{\vec{q}}K_{\vec{q}}e^{\pm i\vec{Q}\cdot\vec{R}}e^{\pm iq_xx}\frac{F_n(\vec{R})}{\sqrt{A}}\frac{e^{ik_xx}}{\sqrt{L_x}}d^2Rdx, \quad (13)$$

where L_x is the length of the unit cell. The integral over the transport direction x becomes a Kronecker-delta expressing momentum conservation in the transport direction,

$$H_{k_x',k_x}^{m,n}=I_{k_x',k_x}^{m,n}(\vec{Q})A_{\vec{q}}K_{\vec{q}}\delta_{k_x',k_x\pm q_x} \quad (14)$$

with

$$I_{k_x',k_x}^{m,n}(\vec{Q})=\frac{1}{A}\int_R\rho_{k_x',k_x}^{m,n}(\vec{R})e^{\pm i\vec{Q}\cdot\vec{R}}d^2R, \quad (15a)$$

$$\rho_{k_x',k_x}^{m,n}(\vec{R})=F_{m,k_x'}(\vec{R})^*F_{n,k_x}(\vec{R}). \quad (15b)$$

When integrating and taking the square of the matrix element, the integral for the form factor is also evaluated as:

$$|I_{k_x',k_x}^{m,n}|^2=\frac{1}{A^2}\int_Rd^2R\int_{R'}d^2R'\rho_{k_x',k_x}^{m,n}(\vec{R})^*\rho_{k_x',k_x}^{m,n}(\vec{R})e^{\pm i\vec{Q}\cdot(\vec{R}-\vec{R}')} \quad (16)$$

The summation over the lateral momentum before substitution into Eq. (8) can be performed as:

$$\sum_{k_{\vec{R}}}\left|I_{k_x',k_x}^{m,n}\right|^2=\frac{A}{4\pi^2}\frac{1}{A^2}\int_Qd^2Qe^{\pm i\vec{Q}\cdot(\vec{R}-\vec{R}')}\int_Rd^2R\int_{R'}d^2R'\rho_{k_x',k_x}^{m,n}(\vec{R})^*\rho_{k_x',k_x}^{m,n}(\vec{R})$$
$$=A\int_R\left|\frac{\rho_{k_x',k_x}^{m,n}(\vec{R})}{A}\right|^2d^2R$$
$$=\frac{A}{A_{k_x',k_x}^{m,n}} \quad (17)$$

Here, $1/A_{k_x',k_x}^{m,n}$ has the units of m^{-2}. For computational efficiency, for the calculation of the form factor overlap, we add on each atom the components of the squares of each multi-orbital wavefunction, and afterwards perform the final/initial state overlap multiplication. In such way, we approximate the form factor components of a lattice atom at a specific location R_0 by:

$$\left|\rho_{k_x',k_x}^{m,n}(\vec{R})\right|^2=\left|F_{m,k_x'}(\vec{R})^*F_{n,k_x}(\vec{R})\right|^2$$
$$\approx\sum_\alpha\left|F_{n,k_x}^\alpha(\vec{R})\right|^2\sum_\alpha\left|F_{m,k_x'}^\alpha(\vec{R})\right|^2 \quad (18)$$
$$\equiv\|F_{n,k_x}\|^2\|F_{m,k_x'}\|^2$$

where α runs over the TB orbitals of a specific atom. This computes the overlaps using the probability density of each state (accounting for the contribution of all orbitals). This treatment is equivalent to a single orbital model (i.e. the effective mass approximation), although we still keep the k_x-dependence of the wavefunctions. It reduces the memory needed in the computation by 20X, allowing simulations of large NW cross sections with only slightly reduced accuracy. In the cases of sine/cosine wavefunctions for square NWs, it can be shown analytically that for 1D intra-band the form factor becomes $\frac{1}{A_{nm}}=\frac{9}{4A}$ [32,50]. For inter-band overlaps one obtains $\frac{1}{A_{nm}}=\frac{1}{A}$. Indeed, our numerical overlaps agree with these analytical expressions. To emphasize the importance of the approximation in Eq. (18), we note that even after this simplification the storage of the probability density functions for the larger diameter NWs requires several Giga bytes of memory. For smaller diameters the overlap integrals calculated using the actual wavefunctions, or their probability distributions are in very good agreement. It would have been computationally prohibitive, however, to calculate them using the actual wavefunctions for the larger NWs, at least on a single CPU. For consistency, therefore, we consider the probability density in all simulations, for NWs of both small and large diameters.

The transition rate is given by:

$$S_{n,m}(k_x,k_x') = \frac{2\pi}{\hbar}\left|I^{m,n}_{k_x',k_x}\right|^2 \left|K_{\vec{q}}\right|^2 \left|A_{\vec{q}}\right|^2 \delta_{k_x',k_x\pm q_x}\delta(E_m(k_x') - E_n(k_x) \pm \hbar\omega_{ph}) \quad (19)$$

where

$$|A_{\vec{q}}|^2 = \frac{1}{\Omega}\frac{\hbar(N_\omega + \frac{1}{2} \mp \frac{1}{2})}{2\rho\omega_{ph}}, \quad (20)$$

ρ is the mass density, and N_ω is the number of phonons given by the Bose–Einstein distribution. The momentum relaxation rate for a carrier in a specific subband n as a function of energy is then given by:

$$\frac{1}{\tau^n_{ph}(E)} = \frac{\pi}{\hbar}\frac{(N_\omega + \frac{1}{2} \mp \frac{1}{2})}{\rho\hbar\omega_{ph}} \times \left(\frac{1}{L_x}\sum_{m,k_x'}\frac{|K_{\vec{q}}|^2}{A^{k_x k_{x'}}_{nm}}\delta_{k_x',k_x\pm q_x}\delta(E_m(k_x') - E_n(k_x) \pm \hbar\omega_{ph})\left(1 - \frac{\upsilon_m(k_x')}{\upsilon_n(k_x)}\right)\right), \quad (21)$$

where $\hbar\omega_{ph}$ is the phonon energy, and we have used $\Omega = AL_x$. For acoustic deformation potential scattering (ADP) or acoustic inter-valley scattering (IVS), it holds $|K_{\vec{q}}|^2 = q^2 D^2_{\text{ADP}}$, whereas for optical deformation potential scattering (ODP for holes, optical IVS processes for electrons) $|K_{\vec{q}}|^2 = D^2_O$, where D_{ADP} and D_0 are the scattering deformation potential amplitudes. For IVS in the conduction band we include all relevant g- and f-processes (acoustic and optical) [32]. On the other hand, specifically for elastic ADP, after applying the equipartition approximation, the relaxation rate becomes:

$$\frac{1}{\tau^n_{\text{ADP}}(E)} = \frac{2\pi}{\hbar} \times \frac{D^2_{\text{ADP}}k_BT}{\rho\upsilon_s^2}\left(\frac{1}{L_x}\sum_{m,k_x'}\frac{1}{A^{k_x k_{x'}}_{nm}}\delta_{k_x',k_x\pm q_x}\delta(E_m(k_x') - E_n(k_x))\left(1 - \frac{\upsilon_m(k_x')}{\upsilon_n(k_x)}\right)\right), \quad (22)$$

where υ_s is the sound velocity in Si.

The parameters we use are the same as in Ref. [32], with the exceptions of $D^{\text{holes}}_{\text{ODP}} = 13.24 \times 10^{10}$ eV/m, $D^{\text{holes}}_{\text{ADP}} = 5.34$ eV, and $D^{\text{electrons}}_{\text{ADP}} = 9.5$ eV from Refs. [26,33] which are more relevant for narrow NWs. These are higher than the bulk values [32]. Larger deformation potential values are commonly used to explain mobility trends in nanostructures, and different parameter sets appear in the literature [33,48,51,52]. Using different sets of parameters will only change our results quantitatively since qualitatively they only depend on the electronic structure.

Confinement of phonons, and the details of their full dispersion are neglected. Dispersionless bulk phonons provide an ease of modeling and still capture the qualitative dependence of the mobility on the electronic structure. Spatial confinement mostly affects acoustic phonons, whereas optical phonons are not affected significantly [53,54]. The degrading effect of confined phonons on the mobility for the thinnest NWs examined in this work can be of the order of 10–20%, and declines fast as the diameter increases [26,33,49,53]. As indicated by Donetti et al., this effect might be even weaker for holes than for electrons [55]. Bulk, dispersionless phonon treatment has been extensively utilized for treatment of scattering in bulk Si or Si ultra-thin-layers by the simulation community. Such methods have also been lately extended for treatment of scattering in thin NWs [48,50,56]. Such models could be improved for thin NWs by including a more sophisticated electron–phonon interaction model as used by Buin et al. [26,33], who, however, also showed qualitative similar results as the ones we present for the thin NWs. Our purpose, however, is to show that by proper confinement of warped bands, large mobility improvements can be achieved due to changes in the electronic structure alone. More sophisticated treatment of the scattering mechanisms could affect our results quantitatively, but this is not the point of this work. The effect we focus on is due to electronic structure, and we believe that qualitatively our results will still be valid.

A crucial approximation made here, is that an isotropic deformation potential $K_{\vec{q}}$ is employed for NWs in all orientations. This parameter could have its own anisotropy, which will add to the anisotropy due to bandstructure effects and might influence the results. The results presented in this work only provide insight into the effect of the anisotropy due to bandstructure alone, and should be confirmed by more sophisticated calculations that account for the anisotropy of the deformation potential as well [57].

2.3.2. Surface roughness scattering

For SRS, we assume a 1D exponential autocorrelation function [58] for the roughness given by:

$$\langle\delta(\rho)\delta(\rho' - \rho)\rangle = \Delta^2_{rms}e^{-\sqrt{2}|\rho'|/L_C}, \quad (23)$$

where Δ_{rms} is the roughness amplitude and L_C the roughness correlation length. Surface roughness is assumed to cause a band edge shift. The scattering strength is derived from the shift in the band edge with quantization $\frac{\Delta E_{C,V}}{\Delta D}$ [59,60], and the transition rate here is derived as:

$$S^{\text{SRS}}_{n,m}(k_x,k_x') = \frac{2\pi}{\hbar}\left(\frac{q_0\Delta E_{C,V}}{\Delta D}\right)^2\left(\frac{2\sqrt{2}\Delta^2_{rms}L_C}{2 + q_x^2L_C^2}\right)\delta(E_m(k_x') - E_n(k_x)), \quad (24)$$

where $q_x = k_x - k_x'$.

This SRS model is more simplified than the ones described in Refs. [48,61–63] that account for the wavefunction deformation/shift at the interface, as well as additional Coulomb effects [64]. The band edge variation, however, is the dominant SRS mechanism in ultra scaled channels as described by various authors, both theoretically and experimentally [48,59,60,65,66]. It has been shown in Ref. [48,60] that the SRS limited low-field mobility in ultra thin nanostructures follows a D^6 behavior, where D is the confinement length scale which is in this case the diameter of the channel. This is an indication that the dominant SRS contribution originates from the subband edge shift due to quantization. Additional Coulomb effects as described by Jin et al. in Ref. [64] can have contributions to SRS, but they mostly cause quantitative rather than qualitative changes in our results.

2.4. General scattering considerations

Elastic and inelastic scattering processes are considered. We assume bulk phonons and follow the *bulk* Si scattering processes. Following the bulk Si scattering "selection rules", we allow only intra-valley transition for elastic processes (elastic phonons, SRS), whereas we allow only inter-valley transitions for inelastic processes (inelastic phonons), where the final scattering state is selected according to whether the process is of f- or g-type. All six relevant phonon modes in Si are included [32,67]. We note that elastic inter-valley transitions although weak can indeed be possible, but for consistency to the bulk model we ignore them. For p-type NWs we include ADP and ODP processes allowing for all intra- and inter-band transitions. Quantum confinement induces mixing of the heavy-hole, light-hole and split-off bands. Inter-valley transitions are also allowed, following the bulk scattering process.

We note that these are approximations for the sake of consistency with the treatment of bulk Si scattering mechanisms, rather than strict rules. In reality, these "selection rules" rise automatically from the symmetry of the wavefunctions and scattering potential. Since we only store the probability density and not the

actual wavefunctions, the phase information is lost and these “selection rules” are imposed “by hand” in the simulation. This is also common practice when using simplified effective mass models and under the bulk dispersionless phonon approximation, where the waveform overlaps are separately computed using the probability densities of the states, still with good accuracy in the results [32,48,50,56]. However, this, as well as the rest of the approximations we make such as bulk phonons, or the simplified SRS scattering model, will affect our results only quantitatively. The qualitative behavior in terms of geometry and orientation effects we describe originates mostly from electronic structure and will not be strongly affected by these approximations.

3. Results and discussion

3.1. Confinement of the heavy-hole band

Fig. 1a shows the heavy-hole (1 0 0) energy surface (for $k_z = 0$) of bulk Si. The 45° degree lines indicate a discrete wave vector component for a quantization in [1 1 0] direction. The subbands formed correspond to “cut”-planes through the bulk bandstructure and will be the relevant energy subbands/surfaces that will form the dispersion of NW or UTB channels under $(1\bar{1}0)$ surface quantization. The stronger the confinement, the farther away from the center the lines shift. Because of the warped anisotropic shape, the relevant bands acquire larger curvatures, and therefore higher hole velocities. Indeed, the actual quantization involves significant band mixing, however, this analysis provides an indication as to where the main dispersion features originate from [19].

This is illustrated in Fig. 1b and c, which show the energy surface contours of a UTB film with $(1\bar{1}0)$ surface confinement and thicknesses of $W = 15$ nm and $W = 3$ nm, respectively. In this case three observations can be made regarding strong contour variations: (i) For the $W = 15$ nm UTB film there is strong anisotropy between the [1 1 0] direction (arrow, x-axis) and the [0 0 1] (y-axis); (ii) When the thickness reduces (Fig. 1c), the contour in the [1 1 0] direction changes, acquiring a larger curvature (the contours now point towards the center); (iii) The change in the energy contours is anisotropic, i.e. no significant changes are observed to the contours along the [0 0 1] direction (y-axis). This behavior is a consequence of the warped shape of the heavy-hole band. Detailed explanations on the effect of confinement on the band shapes are provided in our previous works [18,19,32,68].

What described above, will affect NWs even stronger since they are subject to 2D confinement. In Fig. 2a and b we show the electronic structure of a cylindrical [1 1 0] NW with $D = 12$ nm and $D = 3$ nm, respectively. As the diameter is reduced to 3 nm, the bands acquire large curvatures, which indicate smaller effective mass and larger carrier velocities. The relevant confinement surfaces in this case are (1 1 0) and (1 0 0). This large increase in the curvature originates mostly from the (1 1 0) confinement as we will explain. Fig. 2c and d shows the effect of confinement for the [1 1 1] NWs. The electronic structure also acquires large curvature as the diameter is reduced. The relevant confinement surfaces in this case are (1 1 0) and (1 1 2). Both confinement directions are responsible almost equally for this curvature increase as we will explain. On the other hand, this effect does not appear in all NW orientations. In Fig. 3 we show the same features as in Fig. 2, but this time for the [1 0 0] NWs. In this case, the curvature of the bands does not change with confinement, and in addition, an oscillating behavior appears in the dispersion, which can limit the carrier velocity due to the creation of several zero slope regions.

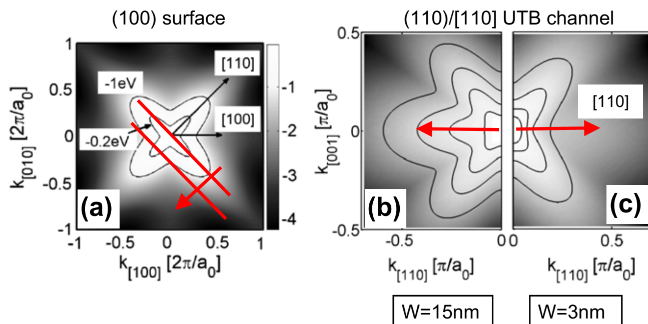

Fig. 1. The energy surfaces of: (a) Heavy-hole band of bulk Si. Contour lines are 0.2 eV and 1 eV below the valence band edge. (b) Width $W = 15$ nm UTB layer with $(1\bar{1}0)$ surface orientation in [1 1 0] transport. (c) Width $W = 3$ nm UTB layer with $(1\bar{1}0)$ surface orientation in [1 1 0] transport. The contour lines (from the center outwards) represent energy contours at 0.02 eV, 0.05 eV, 0.1 eV and 0.2 eV below the top of the valence band edge.

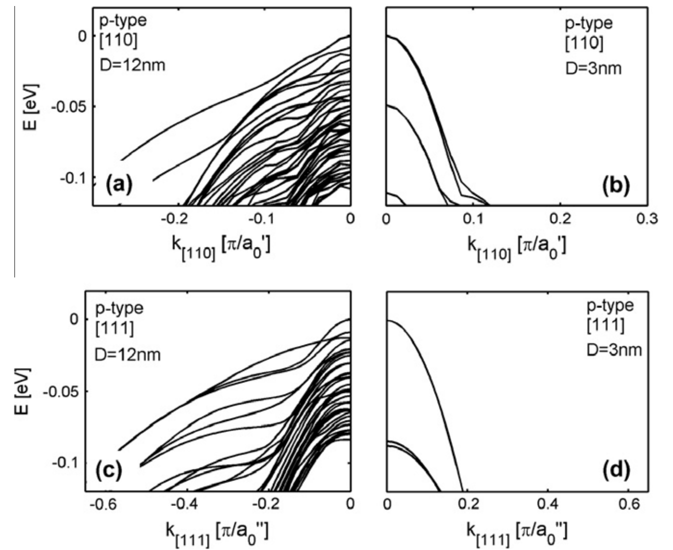

Fig. 2. Dispersions of p-type NWs for various orientations and diameters (D). The left column is for $D = 12$ nm, and the right column for $D = 3$ nm. (a) [1 1 0], $D = 12$ nm. (b) [1 1 0], $D = 3$ nm. (c) [1 1 1], $D = 12$ nm. (d) [1 1 1], $D = 3$ nm. a_0' and a_0'' are the unit cell lengths for the wires in the [1 1 0] and [1 1 1] orientations, respectively.

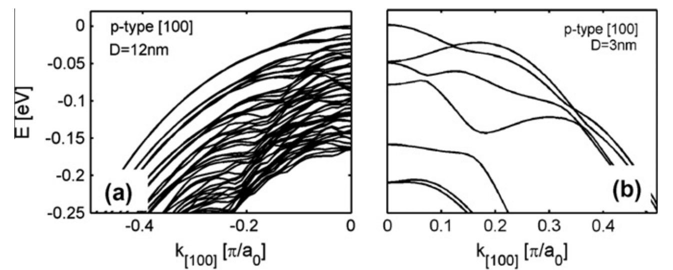

Fig. 3. Dispersions of p-type [1 0 0] NWs. (a) Diameter $D = 12$ nm. (b) $D = 3$ nm. a_0 is the unit cell length.

3.2. Carrier velocities upon confinement and orientation

In order to identify which confining surface is responsible for the large curvature increase in each NW case, using a simple semiclassical ballistic model [18,32,69] we calculate the hole injection velocity in rectangular NWs of different transport and surface orientations, assuming a source-to-drain bias of $V_{DS} = 0.5$ V and non-degenerate conditions. The injection velocity provides a measure

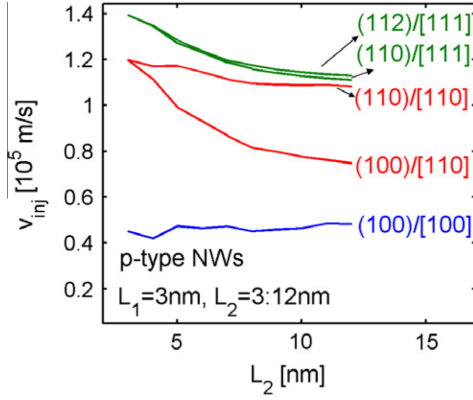

Fig. 4. Low bias (non-degenerate limit) hole velocities for rectangular NWs in the [1 0 0], [1 1 0] and [1 1 1] transport orientations as one of the side length scales from L_2 = 12 nm down to L_2 = 3 nm. The other side length is kept constant at L_1 = 3 nm. The surface indicated (·), represents the L_1 = 3 nm side on which the strongest confinement is applied. For example, the (1 1 0)/[1 1 0] NW case has constant width scaled to $W = L_1$ = 3 nm while the height changes from $H = L_2$ = 12 nm down to $H = L_2$ = 3 nm. The (1 0 0)/[1 1 0] NW case has constant height $H = L_1$ = 3 nm, while the width changes from $W = L_2$ = 12 nm down to $W = L_2$ = 3 nm.

of the velocities of all the subbands of the particular channel. We identify how the velocity changes as the length of each confining side independently varies (see Fig. 4). The hole velocities of the NWs in [1 0 0], [1 1 0] and [1 1 1] transport orientations are shown. In each case, the length of one of the sides is kept at L_1 = 3 nm, whereas the length of the other is varied from L_2 = 3 nm to L_2 = 12 nm. The surface notation (in parenthesis) in each case indicates the surface inducing the strongest confinement (perpendicular to the L_1 side direction, whose length is kept at L_1 = 3 nm). Strong anisotropic behavior is observed. [1 1 1] transport-oriented NWs (green lines) have the highest hole velocities, followed by the [1 1 0] NWs (red lines), whereas the [1 0 0] ones (blue line) have the lowest hole velocities. In the case of the [1 1 1] NWs, scaling of either of the two different side surfaces (1 1 0) or (1 1 2) increases the hole velocities in a similar way from 1.1×10^5 m/s to 1.4×10^5 m/s. For the [1 1 0] cases, (1 1 0) surface confinement provides higher hole velocities compared to (1 0 0) surface confinement (due to subbands with larger curvature as shown in Fig. 1). Indeed, in the (1 1 0)/[1 1 0] case, further scaling of the [0 0 1] side length from L_2 = 12 nm to L_2 = 3 nm does not provide any advantage in the hole velocities. On the other hand, scaling the $[1\bar{1}0]$ side length from L_2 = 12 nm to L_2 = 3 nm in the (1 0 0)/[1 1 0] case, increases the hole velocities. Finally, for (1 0 0)/[1 0 0] NWs, the velocities are low, and scaling causes only minor variations. This is due to the oscillating behavior of the subbands in this orientation [19] which keeps the velocities low.

Unlike thin layers, cylindrical NWs are subject to have a continuously varying confining orientated surface, which makes the analytical band construction more complicated as the diameter is reduced. The basic guidelines to achieve large band curvature that we described above, however, are still valid in this case as well: (i) For [1 0 0] p-type NWs neither (1 1 0) nor (1 0 0) confinement can provide bands with large curvature, (ii) For [1 1 0] NWs (1 1 0) confinement provides bands with large curvature, whereas (1 0 0) confinement does not, and (iii) For [1 1 1] NWs both (1 1 0) and (1 1 2) confining surfaces can provide bands with large curvature.

In Fig. 5 we calculate the hole injection velocities for cylindrical [1 1 0] and [1 1 1] NWs. The hole velocity increases by a factor ~2X as the diameter reduces in both cases as indicated. The values at D = 3 nm are very similar to the [1 1 0] and [1 1 1] rectangular NWs with L_1 = 3 nm, L_2 = 3 nm (most left in Fig. 4). Comparing the velocity values of the largest cylindrical NWs with D = 12 nm

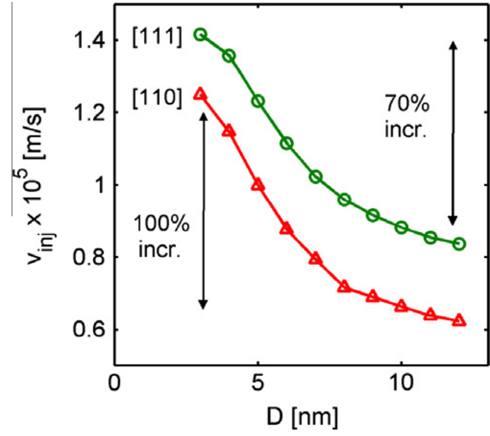

Fig. 5. The carrier velocities of cylindrical p-type NWs in [1 1 0] and [1 1 1] transport orientations versus diameter under non-degenerate conditions. The data are same as in [44].

in Fig. 5 (most right), to the ones confined in one direction only (most right in Fig. 4), we can make two observations: (i) For the [1 1 0] NWs, (1 1 0) confinement increases the hole velocity from 0.61×10^5 m/s (Fig. 5, most right) to 1.08×10^5 m/s (Fig. 4, most right), an 77% increase. The benefit from (1 0 0) confinement is smaller, increasing the velocity to only $\sim 0.75 \times 10^5$ m/s (23%); (ii) For the [1 1 1] NWs, both (1 1 0) or (1 1 2) confinement increase the hole velocity from $\sim 0.82 \times 10^5$ m/s (Fig. 5, most right) to $\sim 1.12 \times 10^5$ m/s (37%) each (Fig. 4, most right).

3.3. Carrier mobility and mean-free-path (MFP)

The two-fold increase in the carrier velocity with diameter scaling results in a large increase in the carrier mobility. The dependence of the low-field mobility on the carrier velocity can be shown in a simplified way be as follows: The 3D carrier concentration is given by

$$n_{3D} = M \int_E \frac{g_{1D}^n(E)}{A} f(E) dE, \tag{25}$$

where M is the number of subbands and A is the normalization cross section area. We assume that $\tau_n(E)$ follows a simple relation (at least for elastic isotropic processes):

$$\tau_n(E) \propto \frac{A}{M g_{1D}^n(E)} \tag{26}$$

which just means that the scattering rates are proportional to the density of states that a carrier can scatter into. Now we substitute these relations in the mobility equation Eq. (4) (and after performing the summation over the subbands in Eq. (2)):

$$\begin{aligned} \mu = \frac{\sigma}{q_0 n} &\propto \frac{\int_E v_n^2(E) \frac{A}{M g_{1D}^n(E)} \frac{M g_{1D}^n(E)}{A} \left(-\frac{\partial f(E)}{\partial E}\right) dE}{M \int_E \frac{g_{1D}^n(E)}{A} f(E) dE} \\ &= \left(\frac{A}{M}\right) \frac{\int_E v_n^2(E) \left(-\frac{\partial f(E)}{\partial E}\right) dE}{\int_E g_{1D}^n(E) f(E) dE} \end{aligned} \tag{27}$$

Using the relations $g_{1D}^n(E) \propto \sqrt{m_{eff}/\tilde{E}}$ and $v_n(E) \propto \sqrt{\tilde{E}/m_{eff}}$ (valid for 1D and under the parabolic band approximation), where $\tilde{E} = E - E_0$, with E_0 being the energy of the subband edge, and m_{eff} the effective mass of the subband, then:

$$\begin{aligned} \mu &\propto \left(\frac{A}{M}\right) \frac{\int_{E_0}^{\infty} \tilde{E}/m_{eff} \left(-\frac{\partial f(E)}{\partial E}\right) dE}{\int_{E_0}^{\infty} \sqrt{\frac{m_{eff}}{\tilde{E}}} f(E) dE} \\ &= \frac{1}{m_{eff}^{3/2}} \left(\frac{A}{M}\right) \tilde{F}(\eta_F) \end{aligned} \tag{28}$$

where $\tilde{F}(\eta_F)$ is some function of $\eta_F = E_F - E_0$, where E_F is the Fermi level. In the non-degenerate limit, this function is independent of bandstructure. The mobility, therefore, has an effective mass dependence $m_{eff}^{-3/2}$ as also mentioned in Ref. [26] and consequently, a third power dependence on the injection velocity:

$$\mu \propto m_{eff}^{-3/2}\left(\frac{A}{M}\right) \propto v_{inj.}^3\left(\frac{A}{M}\right). \quad (29)$$

The mobility is therefore proportional to the third power of the injection velocity. Improvements in the carrier velocity will result in much larger improvements in the mobility. This is clearly indicated in Fig. 6, which shows the phonon-limited mobility of the D = 3 nm (solid) and D = 12 nm (dashed) NWs versus the carrier concentration. Both [1 1 0] (red) and [1 1 1] (green) NW orientations are shown. The D = 12 nm mobility values are ~100 cm²/Vs and ~250 cm²/Vs for the [1 1 0] and [1 1 1] NWs, respectively, whereas the corresponding values for the D = 3 nm NWs jump to ~1300 cm²/Vs and ~2100 cm²/Vs, respectively. We note here that the values for the larger NW diameters are lower than bulk values because we used larger deformation potential amplitudes. Additionally, such a larger relative increase in the mobility will only be observed if the deformation potentials were constant under confinement, which is not the case in reality.

The large increase in the mobility with diameter scaling has also been described by Buin et al. [26,33] theoretically using a similar approach and two different sets of deformation parameters (square points in Fig. 6), but also experimentally by Trivedi et al. [17] for the [1 1 0] NWs. In that work, [1 1 0] NWs with smooth surfaces and different side lengths were fabricated. The mobility of the NWs with cross section side lengths down to 3–5 nm was measured to peak at μ_p = 1235 cm²/Vs (diamond). It drops to μ_p ~ 400 cm²/Vs (triangle) as the (1 1 0) surface side thickness increases to 169 nm while the height of the (1 0 0) surface remains at 4.3 nm (forming a nano-belt structure). Although high mobility measurements in NWs is a controversial issue because surface charges make it difficult to extract the correct carrier density, the results of Ref. [17] could be an experimental confirmation of the effect we describe in this work. It also supports our claim that the mobility improvement in [1 1 0] NWs originates from (1 1 0) confinement, rather than (1 0 0) confinement. Other experimental works [35,70,71] have also reported higher mobilities on (1 1 0) surfaces compared to other surfaces, which are improved, or at least not degraded, as the channel thickness is reduced down to 3 nm [70].

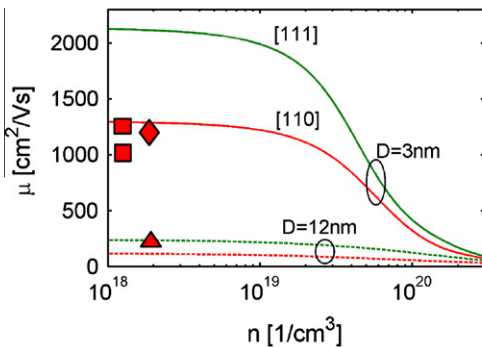

Fig. 6. The low-field phonon-limited mobility for p-type NWs in [1 1 0] (red) and [1 1 1] (green) transport orientations versus carrier concentration. Diameters D = 3 nm (solid) and D = 12 nm (dashed) are shown. Square points: simulation results for the D = 3 nm [1 1 0] NW from Ref. [33] with two different deformation parameter sets. Diamond: Experimental mobility measurement for a rectangular [1 1 0] NW with side lengths 4.3 nm × 3.4 nm in the [1 0 0] and [1 1 0] confinement directions respectively. Triangle: experimental mobility measurement for a [1 1 0] nano-belt with side lengths 4.3 nm × 169 nm in the [1 0 0] and [1 1 0] confinement directions respectively. The experimental results are from Ref. [17]. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

A large orientation dependence is also observed, with the mobility of the [1 1 1] NWs being almost twice as that of the [1 1 0] NWs. This anisotropy, however, is reduced as the carrier concentration increases. At concentrations above 5×10^{19}/cm³ the differences significantly decrease. The mobility differences between NWs of different diameters decline as well. At concentrations above 10^{20}/cm³, any diameter or orientation differences diminish. The large drop in mobility for the smaller diameter NWs, as well as the reduction in anisotropy at higher concentrations originates from the fact that at higher energies the electronic structures of the NWs become more similar to each other. The bands of all NWs acquire similar bandstructure consisting of an increased number of heavier subbands, and the light subband advantage of the thinner NWs is lost. We note here that we only consider phonon-limited mobility in this calculation and assume that the high concentrations will be achieved by electrostatic gating (although we do not calculate the full self-consistent electrostatic solution on the 2D cross section of the NW channel as in Refs. [18,19]). In the case of doped channels, impurity scattering will dominate the mobility, and large mobility degradation will be observed at much lower concentrations, beginning from $p_0 = 10^{16}$/cm³.

The large increase in the carrier velocity for the smaller diameter NWs also results in long scattering mean-free-paths (MFPs) in these channels. The MFP is defined as $\lambda_n(E) = 2\pi\hbar\Xi(E)/M$, where M is the number of modes [72]. This is shown in Fig. 7 for the D = 3 nm NWs. Here, the energy reference is shifted to the Fermi level ($E_F = 0eV$), which is calculated for a hole concentration of $p = 10^{19}$/cm³. Although we consider p-type NWs, all quantities are presented in positive energies for presentation purposes. The MFP(E) is almost a linear function of energy in the low energy region where only the first subband contributes to transport. The slope of the MFP in the linear region depends on the effective mass of the subbands. It results from the fact that:

$$\begin{aligned}\lambda_n(E) &= 2\pi\hbar\Xi(E)/M \\ &\propto M v_n^2(E)\tau_n(E)\frac{g_{1D}^n(E)}{A}/M \\ &\propto v_n^2(E)\frac{A}{g_{1D}^n(E)}\frac{g_{1D}^n(E)}{A} \\ &\propto \frac{2E}{m^*}\end{aligned} \quad (30)$$

Slight deviations from a straight line originate from the non-parabolicity of the bands. As shown, the phonon-limited MFP peaks around ~100 nm for both the [1 1 1] (green) and [1 1 0] (red) NWs almost only within k_BT into the valence band edge. At $2k_BT$ into the valence band the MFP for these two NWs reaches ~200 nm and ~150 nm respectively. These are long MFPs compared to bulk Si values, both for electrons or for holes, and explain

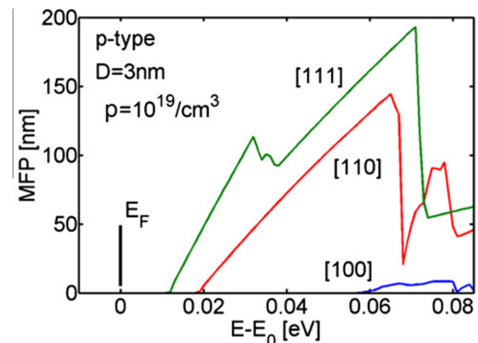

Fig. 7. The phonon-limited mean-free-path (MFP) for the holes in D = 3 nm NWs of different transport orientations. Carrier concentration $p = 10^{19}$/cm³ is assumed.

the large increase in the mobility as calculated in Fig. 6. On the other hand, such long MFPs for phase coherent processes for the cases of the [1 1 0] and [1 1 1] NWs can make these NWs more sensitive to phase coherent carrier localization originating from surface roughness for example, or other elastic, coherent scattering processes. In comparison, the MFP of the [1 0 0] NW for which there is no increase in the band curvature with decreasing diameter is much lower, reaching only a few nanometers, indicating low carrier mobility.

The almost 2X larger mobility for the [1 1 1] NWs compared to the [1 1 0] NWs can be understood from the larger MFP values, but also from the fact that valence band is closer to the Fermi level in the [1 1 1] case at the same carrier concentration, which makes the MFP(E) curve start at smaller values of $E - E_F$. Averaged over the Fermi distribution, in the [1 1 1] case the carriers will utilize their large MFP more effectively. The positioning of valence band compared to E_F is determined by the mass of the subbands and their degeneracy. The smaller these quantities are, the closer to the Fermi level the band edge will reside. In the D = 3 nm p-type NW the first band is single-degenerate (excluding spin degeneracy) but the mass of the [1 1 1] NW is smaller (as indicated from the larger carrier velocities in Fig. 5), which at the same carrier concentration places the valence band of this NW closer to E_F and further improves hole mobility.

3.4. The effect of surface roughness scattering (SRS)

Although the hole mobility of the [1 1 0] and [1 1 1] NWs largely increases as the diameter is reduced, in reality channels with thickness below 6 nm suffer from enhanced SRS [60] that degrades the mobility. Fig. 8 demonstrates the effect of SRS on the mobility of the p-type [1 1 1] NW. The solid lines are the phonon plus SRS-limited low-field mobility results for the D = 3 nm NW for cases with different roughness parameters: (i) Δ_{rms} = 0 (phonon-limited), (ii) Δ_{rms} = 0.14 nm, (iii) Δ_{rms} = 0.48 nm, (iv) Δ_{rms} = 0.8 nm. In all cases we keep L_C = 1.3 nm [48]. The dashed line is the phonon-limited result for the larger diameter D = 12 nm NW. As the roughness height progressively increases the mobility decreases. For low carrier concentrations the mobility can decrease by ~3X for the Δ_{rms} = 0.8 nm case. For higher carrier concentrations the deviation from the phonon limited result is less because faster carriers are affected less from this elastic anisotropic scattering mechanism. The magnitude of this degradation is in qualitative agreement with other works on the mobility degradation due to SRS in nanodevices [48], although it is subject to the parameters used in Eq. (23). Nevertheless, the large mobility improvement as the diameter scales in these NWs, can partially compensate for SRS and still provide mobility benefits compared to the D = 12 nm NW, even for fairly large roughness height values. The same can also be observed for the [1 1 0] NW. As an example, we mention that an effective way to design high efficiency nanostructured thermoelectric devices is to scale the cross section feature sizes in order to reduce the lattice thermal conductivity. In that scope, efficient thermoelectric devices based on silicon NWs have already been demonstrated [11,12]. High electronic conductivity though is still needed. The [1 1 0] and [1 1 1] p-type channels, in which the mobility (and conductivity) increases with feature size reduction, might be suitable for such applications.

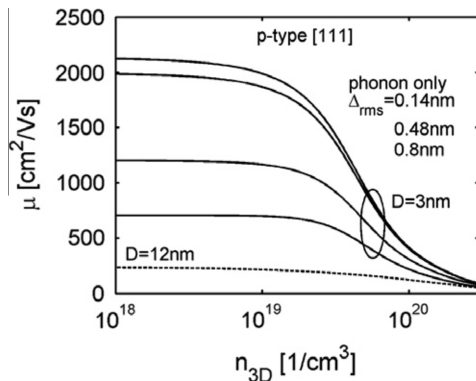

Fig. 8. The low-field phonon plus SRS-limited mobility for p-type [1 1 1] NWs versus carrier concentration. Diameters D = 3 nm (solid) and D = 12 nm (dashed) are shown. For the D = 3 nm, the following roughness height Δ_{rms} values are used: (i) Δ_{rms} = 0 (phonon-limited), (ii) Δ_{rms} = 0.14 nm, (iii) Δ_{rms} = 0.48 nm, (iv) Δ_{rms} = 0.8 nm. For the D = 12 nm NW, Δ_{rms} = 0.

3.5. Tolerance to SRS

Proper choice of confinement and transport orientations in channels with highly warped bands can, therefore, offer opportunities for performance improvement as the feature sizes of the devices are scaled to nanometer sizes. This is observed for the [1 1 0] transport direction under strong (1 1 0) surface confinement, and for the [1 1 1] transport orientation under both, (1 1 0) or (1 1 2) surface confinement. Here we show that not only such choice of confinement is beneficial for the carrier velocities and mobility, but also it might make the NWs less susceptible to SRS as well. As described in Section 2.3, the strongest contribution to SRS at such low diameters originates from the shift in the band edges with quantization $dE_{C,V}/dD$. The strength of this shift is determined by the confinement effective mass. The heavier the confinement mass, the weaker the shift and the more tolerant the channel to SRS is. Fig. 9a shows the band edges versus the NW diameter as it is scaled from D = 12 nm down to D = 3 nm for the [1 0 0], [1 1 0] and [1 1 1] p-type NWs. The band edge variation is the larg-

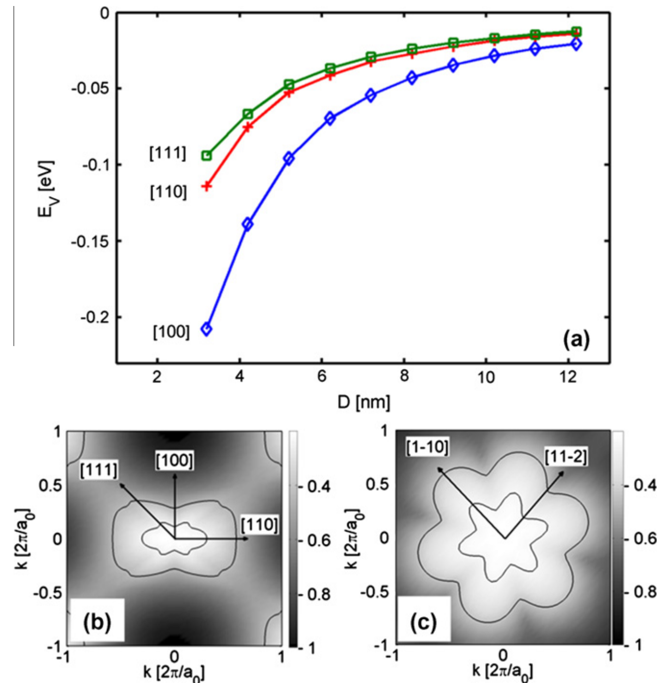

Fig. 9. (a)The valence band edge for p-type NWs in the [1 0 0] (diamond-blue), [1 1 0] (cross-red), and [1 1 1] (square-green) orientations versus the NW diameter. (b) The (1 1 0) heavy-hole energy surface of bulk Si valence band. This is the corresponding confinement surface for the [1 1 0] NW. (c) The (1 1 1) heavy-hole energy surface of bulk Si valence band. This is the corresponding confinement surface for the [1 1 1] NW. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

est for the [1 0 0], lesser for the [1 1 0] and the smallest for the [1 1 1] NW. The [1 1 1] and the [1 1 0] NWs that have the largest improvement in mobility are also more tolerant to SRS, which makes them even more attractive as device channel materials.

This behavior is explained by examining the confining surfaces of the NWs in each orientation. The NWs are cylindrically confined from the different surfaces perpendicular to the transport direction. From Fig. 1a, however, it can be observed that the [100] NW is mainly confined by (1 0 0) and (1 1 0) equivalent confinement surfaces each of them spanning half of the *k*-space. The (1 0 0) confinement is light, whereas the (1 1 0) one is heavy. On the other hand, the [1 1 0] NW as shown in Fig. 9b, is mostly subject to (1 0 0), (1 1 0), and (1 1 1) equivalent confinement surfaces, with the (1 1 1) confinement having a stronger influence because four equivalent [1 1 1] orientations exist on the confinement plane; whereas there are only two [1 0 0] and two [1 1 0] orientations. The (1 1 1) surface provides heavy confinement as well. Therefore, the [1 1 0] NW is overall more heavily confined than the [1 0 0] NW. The [1 1 1] NW is mostly subject to (1 1 1) and (1 1 2) surface confinement, as shown in Fig. 9c. These are both surfaces of heavy confinement masses, which make this NW the one most heavily confined one.

3.6. Anisotropy in n-type channels

Strongly warped energy contours that result in the large improvement of the carrier velocities under confinement along their long axis and transport perpendicular to that, appear not only in the valence band of Si (as well as other p-type semiconductors), but also in the conduction band of Si [18,27]. Fig. 10a shows the six equivalent ellipsoids in the conduction band of Si. By "cutting" an ellipsoid in the center as shown in Fig. 10b, and plotting the (k_x, k_y) energy surface (for $k_z = 0$), a warped shape is also observed (Fig. 10c), similar to the one shown in Fig. 1a for the heavy-hole band of the valence band. The warping, however, is weaker in this case and as a result the confinement induced performance enhancement is also smaller. For the n-type [1 1 0] NW, as the diameter scales down to D = 3 nm, the effective mass of the lowest band reduces from 0.19m_0 (bulk) to 0.16m_0 [Neo08], which results in a velocity improvement of ~30% [21].

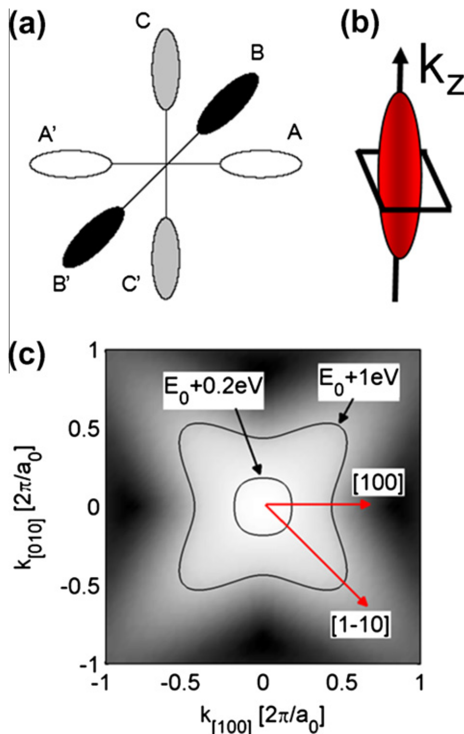

Fig. 10. Energy surfaces of the conduction band of Si. (a) The six equivalent ellipsoids. (b) A "cut" through the middle of one of the ellipsoids. (c) The 2D $E(k_x, k_y)$ (1 0 0) energy surface through the middle of the ellipsoid.

Fig. 11 shows the phonon-limited mobility of n-type NWs as a function of the electron concentration. In this case, for comparison we show three different orientations [1 0 0] (blue), [1 1 0] (red), and [1 1 1] (green), for diameters D = 3 nm (solid) and D = 12 nm (dashed). The behavior is different from what was observed for p-type NWs in Fig. 6. In this case, the mobility for the larger diameter D = 12 nm is higher than that for D = 3 nm for all cases. The mobility reduces because the waveform overlaps increase and increase the scattering rate. Alternatively, the number of subbands M do not decrease linearly with diameter, at some point only one subband is reached and M stays contact at M = 1, such that the ratio A/M in Eq. (29) decreases with diameter and decreases the mobility. Once one subband remains, then at the same carrier concentration, further reducing the diameter will shift the band edges further from the Fermi level to keep the concentration constant (simply stated, less states need to be filled to keep the concentration constant is a smaller area). This reduces the conductivity exponentially and reduces the mobility.

The mobility of the D = 12 nm NWs in Fig. 11 shows anisotropic behavior, with the [1 0 0] NW having the highest mobility, followed by the [1 1 1] and finally by the [1 1 0] NW. Interestingly for the D = 3 nm NWs, however, although this order is retained for the [1 0 0] and [1 1 1] NWs, the mobility of the [1 1 0] NWs is the highest of the three. The reason is the carrier velocity increase in this NW because it takes advantage of its warped bands shown in Fig. 10. In this case, the warping is weaker than what is observed for the p-type NWs to cause mobility increase. However, it minimizes the degradation that the other n-type NWs experience with diameter scaling. The reduction in the mobility of the n-type [1 1 0] NW as the diameter is reduced from D = 12 nm down to D = 3 nm is small. As in the case of p-type NWs in Fig. 6, the anisotropic behavior in terms of both diameter and orientation is retained for carrier concentrations up to n = 5 × 10^{19}/cm^3, whereas for higher concentrations it is minimized since a lot more subbands come into play and scattering increases for all cases.

Fig. 12 shows the MFP of the first subband of the D = 3 nm n-type NWs for the three orientations, similar to Fig. 7 for the p-type NWs. The MFP of the [1 1 0] NW is longer than the other two orientations, attributed to the lower effective mass it acquires upon confinement. We mention here that the masses of the other two NWs increase with cross section scaling as shown in Ref. [18]. Although for the [1 1 0] NW, the warped anisotropic band shape increases the

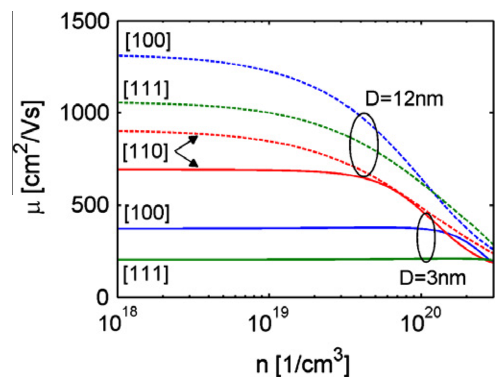

Fig. 11. The low-field phonon-limited mobility for n-type NWs in [1 0 0] (blue), [1 1 0] (red) and [1 1 1] (green) transport orientations versus carrier concentration. Diameters D = 3 nm (solid) and D = 12 nm (dashed) are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

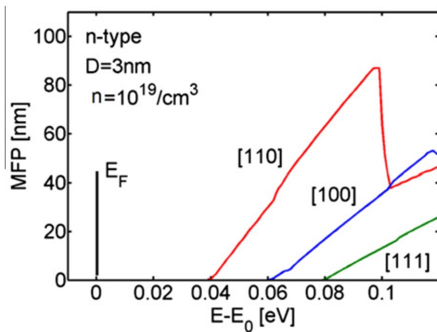

Fig. 12. The phonon-limited mean-free-path (MFP) for the electrons in $D = 3$ nm NWs of different transport orientations. Carrier concentration $n = 10^{19}/cm^3$ is assumed.

carrier mobility upon confinement, for the [1 0 0] NWs it has the opposite effect, degrading curvature, carrier velocities, and mobility. For the [1 0 0] NW the effective mass can reach up to $0.27m_0$ at $D = 3$ nm [18]. The lower mass and degeneracy (two-fold) of the [1 1 0] NW puts its band edge closer to E_F at the same carrier concentration, which results in an overall longer MFP and mobility compared to the other two NWs. A comparison of the MFPs at k_BT above the conduction band for each case shows that for the [1 1 0] NW it reaches around 40 nm, for the [1 0 0] around 20 nm, and for the [1 1 1] NW around 10 nm. These are much smaller MFP values than what is shown for holes in the [1 1 0] and [1 1 1] NWs in Fig. 7, but larger than the [1 0 0] p-type NW.

3.7. Summary of the approximations made

We would like to note that similar transport studies, combining atomistic tight-binding and Boltzmann transport are also found in the literature for either NWs of smaller diameters, fewer scattering processes [23,33,34], or different channel structures and materials [73,74]. In this work, we include all basic scattering mechanisms, and much larger diameters up to $D = 12$ nm (~5500 atoms in the unit cell). To be able to couple TB electronic structures and Boltzmann transport calculations for such large structures, several approximations are made: (i) The assumption of deformation potential theory and dispersionless bulk phonons instead of confined phonons, (ii) the use of the probability density instead of the actual wavefunctions, (iii) the use of a simplified SRS treatment, (iv) flat (zero) electrostatic potential in the 2D cross section of the NW channel, and (v) unrelaxed, unstrained structures instead of reconstructed ones were considered. These are approximations that mostly affect our results quantitatively. The quantitative behavior in terms of geometry and orientation we describe originates mostly from warped electronic structures, and is only weakly affected by these approximations. Approximation (v), the use of unrelaxed surfaces, could have some effect on the effective masses of the $D = 3$ nm NWs as shown in Ref. [75], especially for the p-type [1 0 0] NW. The effect, however, does not seem to be strong enough to alter the anisotropy trends we describe. Such effects, however, depend on the passivation agents, or the type of oxides the NW could be surrounded by, and more elaborate studies will be needed to fully comprehend the effects of structure relaxation on the electronic structure. Although more sophisticated scattering and structure geometry models could be potentially employed, this is beyond the scope of this work.

4. Conclusion

We couple the $sp^3d^5s^*$-spin–orbit–coupled TB model to linearized Boltzmann transport formalism to calculate the low field mobility in ultra narrow Si NWs. We show that the electronic properties of NWs have strong orientation and diameter dependence. Proper choice of confinement and transport orientations in ultra narrow NWs of sub-ten nanometers in diameter can lead to large improvement in carrier velocity and mobility compared to the bulk values as the diameter is scaled below 8 nm. This can be achieved when confining channels with strongly warped anisotropic bands, i.e. the heavy-hole valence band, or even the conduction band in Si. [1 1 0] and [1 1 1] p-type channels, as well as [1 1 0] n-type Si channels can benefit from this effect. In the case of p-type NWs, the phonon-limited mobility largely increases as the diameter scales from $D = 12$ nm down to $D = 3$ nm. This leads to large scattering MFPs in the order of ~100 nm or more. The effect is strong enough to compensate for the detrimental effect of SRS as the channel thickness is drastically scaled. Our results agree with other computational studies [11,33], and could potentially provide explanations for the large mobility enhancement in NWs that has been observed experimentally [16,17] as well as insight into the strong mobility orientation effects in ultra scaled MOSFET device measurements [35,70,71].

Acknowledgment

This work was supported by the Austrian Climate and Energy Fund, Contract No. 825467.

References

[1] ITRS Public Home Page. <http://www.itrs.net/reports.html>.
[2] Singh N, Lim FY, Fang WW, Rustagi SC, Bera LK, Agarwal A, et al. In: Int electron dev meeting, IEDM; 2006.
[3] Cho KH, Jung YC, Hong BH, Hwang SW, Oh JH, Ahn D, et al. In: Int electron dev meeting, IEDM; 2006.
[4] Cho KH, Yeo KH, Y Yeoh Y, Suk SD, Li M, Lee JM, et al. Appl Phys Lett 2008;92:052102.
[5] Kobayashi M, Hiramoto T. J Appl Phys 2008;103:053709.
[6] Xiang J, Lu W, Hu Y, Wu Y, Yan H, Lieber Charles M. Nature 2006;441(25).
[7] Yeo KH, Suk SD, Li M, Yeoh Y-Y, Cho KH, Hong K-H, et al. In: Int electron dev meeting, IEDM; 2006.
[8] Bangsaruntip S, Cohen GM, Majumdar A, Zhang Y, Engelmann SU, Fuller NC, et al. In: IEEE Int electron dev meeting, IEDM, vol. 297; 2009.
[9] Hicks LD, Dresselhaus MS. Phys Rev B 1993;47(24):16631.
[10] Neophytou N, Wagner M, Kosina H, Selberherr S. J Electron Mater 2010;39(9):1902–8.
[11] Boukai AI, Bunimovich Y, Tahir-Kheli J, Yu J-K, Goddard III WA, Heath JR. Nature 2008;451:168–71.
[12] Hochbaum AI, Chen R, Delgado RD, Liang W, Garnett EC, Najarian M, et al. Nature 2008;451:163–8.
[13] Winkelmann CB, Ionica I, Chevalier X, Royal G, Bucher C, Bouchiat V. Nano Lett 2007;7(6):1454–8.
[14] Law M, Greene LE, Johnson JC, Saykally R, Yang P. Nature Mat 2005;4:455–9.
[15] Huang S, Chen Y. Nano Lett 2008;8(9):2829–33.
[16] Cui Y, Zhong Z, Wang D, Wang WU, Lieber CM. Nano Lett 2003;3(2):149–52.
[17] Trivedi K, Yuk H, Floresca HC, Kim MJ, Hu W. Nano Lett 2011;11:1412.
[18] Neophytou N, Paul A, Lundstrom M, Klimeck G. IEEE Trans Electron Dev 2008;55(6):1286–97.
[19] Neophytou N, Paul A, Klimeck G. IEEE Trans Nanotechnol 2008;7(6):710–9.
[20] Neophytou N, Paul A, Lundstrom M, Klimeck G. J Comput Electron 2008;7(3).
[21] Neophytou N, Kim SG, Klimeck G, Kosina H. J Appl Phys 2010;107:113701.
[22] Neophytou N, Klimeck G. Nano Lett 2009;9(2):623–30.
[23] Persson MP, Lherbier A, Niquet Y-M, Triozon F, Roche S. Nano Lett 2008;8:4146.
[24] Luisier M, Klimeck G. In: Proc of simulation of semiconductor processes and devices, SISPAD; 2008.
[25] Liang GC, Xiang J, Kharche N, Klimeck G, Lieber CM, Lundstrom M. Nano Lett 2007;7:642–6.
[26] Buin AK, Verma A, Svizhenko A, Anantram MP. Nano Lett 2008;8(2):760–5.
[27] Liu Y, Neophytou N, Low T, Klimeck G, Lundstrom MS. IEEE Trans Electron Dev 2008;55(3):866–71.
[28] Fischetti MV, Ren Z, Solomon PM, Yang M, Rim K. J Appl Phys 2003;94:1079.
[29] Wang EX, Matagne P, Shifren L, Obradovic B, Kotlyar R, Cea S, et al. IEEE Trans Electron Dev 2006;53(8).
[30] Mahan GD, Sofo JO. Proc Natl Acad Sci USA 1996;93:7436–9.
[31] Scheidemantel TJ, -Draxl CA, Thonhauser T, Badding JV, Sofo JO. Phys Rev B 2003;68:125210.
[32] Lundstrom M. Fundamentals of carrier transport. Cambridge University Press; 2000.

[33] Buin AK, Verma A, Anantram MP. J Appl Phys 2008;104:053716.
[34] Vo TTM, Williamson AJ, Lordi V, Galli G. Nano Lett 2008;8(4):1111–4.
[35] Yang M, Chan VWC, Chan KK, Shi L, Fried DM, Stathis JH, et al. IEEE Trans Electron Dev 2006;53(5):965–78.
[36] Boykin TB, Klimeck G, Oyafuso F. Phys Rev B 2004;69:115201–10.
[37] Klimeck G, Ahmed S, Bae H, Kharche N, Clark S, Haley B, et al. IEEE Trans Electron Dev 2007;54(9):2079–89.
[38] Slater JC, Koster GF. Phys Rev 1954;94(6).
[39] Lee S, Oyafuso F, Von Allmen P, Klimeck G. Phys Rev B 2004;69:045316–23.
[40] Bowen RC, Klimeck G, Lake R, Frensley WR, Moise T. J Appl Phys 1997;81:3207.
[41] Wang J, PhD thesis. Purdue University; 2005.
[42] Kharche N, Prada M, Boykin TB, Klimeck G. Appl Phys Lett 2007;90(9):092109.
[43] Rahman R, Wellard CJ, Bradbury FR, Prada M, Cole JH, Klimeck G, et al. Phys Rev Lett 2007;99(3):036403.
[44] Neophytou N, Kosina H. Nano Lett 2010;10(12):4913–9.
[45] Ferry DK, Goodnick SM. Transport in nanostructures. Cambridge University Press; 1997.
[46] Fischetti MV. J Appl Phys 2001;89:1232.
[47] Esseni D, Abramo A. IEEE Trans Electron Dev 2003;50(7):1665.
[48] Jin S, Fischetti MV, Tang T-W. J Appl Phys 2007;102:083715.
[49] Ramayya EB, Vasileska D, Goodnick SM, Knezevic I. J Appl Phys 2008;104:063711.
[50] Kim R, Lundstrom M. IEEE Trans Electron Dev 2009;56(1):32–139.
[51] Fischetti MV, Laux SE. J Appl Phys 1996;80:2234.
[52] Yamada T, Ferry DK. Solid-State Electron 1995;38:881.
[53] Donetti L, Gamiz F, Roldan JB, Godoy A. J Appl Phys 2006;100:013701.
[54] Balandin A, Wang KL. Phys Rev B 1998;58:1544–9.
[55] Donetti L, Gamiz F, Rodriquez N, Godoy A. IEEE Electron Dev Lett 2009;30(12):1338.
[56] Kotlyar R, Obradovic B, Matagne P, Stettler M, Giles MD. Appl Phys Lett 2004;84(25):5270.
[57] -Armando FM, Fagas G, Green JC. Nano Lett 2010;10:869.
[58] Goodnick SM, Ferry DK, Wilmsen CW, Liliental Z, Fathy D, Krivanek OL. Phys Rev B 1985;32:8171.
[59] Sakaki H, Noda T, Hirakawa K, Tanaka M, Matsusue T. Appl Phys Lett 1987;51(23):1934.
[60] Uchida K, Takagi S. Appl Phys Lett 2003;82(17):2916–8.
[61] Prange RE, Nee T-W. Phys Rev 1968;168:779.
[62] Ando T, Fowler A, Stern F. Rev Mod Phys 1982;54:437.
[63] Esseni D. IEEE Trans Electron Dev 2004;51(3):394.
[64] Jin S, Fischetti M, Tang T-W. IEEE Trans Electron Dev 2007;54:2191.
[65] Fang T, Konar A, Xing H, Jena D. Phys Rev B 2008;78:205403.
[66] Wang J, Polizzi E, Ghosh A, Datta S, Lundstrom M. Appl Phys Lett 2005;87:043101.
[67] Jacoboni C, Reggiani L. Rev Mod Phys 1983;55:645.
[68] Neophytou N, Klimeck G, Kosina H. J Appl Phys 2011;109:053721.
[69] Rahman A, Guo J, Datta S, Lundstrom M. IEEE Trans Electron Dev 2003;50(9):1853–64.
[70] Tsutsui G, Saitoh M, Hiramoto T. IEEE Electron Dev Lett 2005;26:836.
[71] Chen J, Saraya T, Hiramoto T. IEEE Electron Dev Lett 2010;31:1181.
[72] Jeong C, Kim R, Luisier M, Datta S, Lundstrom M. J Appl Phys 2010;107:023707.
[73] Fomin VM, Kratzer P. Phys Rev B 2010;82:045318.
[74] Fomin VM, Kratzer P. Physica E 2010;42:906.
[75] Vo T, Williamson AJ, Galli G. Phys Rev B 2006;74:045116.