Comprehensive comparison and experimental validation of band-structure calculation methods in III-V semiconductor quantum wells

Author(s):
Zerveas, George; Caruso, Enrico; Baccarani, Giorgio; Czornomaz, Lukas; Daix, Nicolas; Esseni, David; Gnani, Elena; Gnudi, Antonio; Grassi, Roberto; Luisier, Mathieu; Markussen, Troels; Osgnach, Patrik; Palestri, Pierpaolo; Schenk, Andreas; Selmi, Luca; Sousa, Marilyne; Stokbro, Kurt; Visciarelli, Michele

Publication date:
2016-01

Permanent link:
https://doi.org/10.3929/ethz-b-000106406

Originally published in:
Solid-State Electronics 115(Part B), https://doi.org/10.1016/j.sse.2015.09.005

Solid-State Electronics 115 (2016) 92–102

Contents lists available at ScienceDirect

Solid-State Electronics

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

Comprehensive comparison and experimental validation of band-structure calculation methods in III–V semiconductor quantum wells

George Zerveas [a,*], Enrico Caruso [b], Giorgio Baccarani [c], Lukas Czornomaz [d], Nicolas Daix [d], David Esseni [b], Elena Gnani [c], Antonio Gnudi [c], Roberto Grassi [c], Mathieu Luisier [a], Troels Markussen [e], Patrik Osgnach [b], Pierpaolo Palestri [b], Andreas Schenk [a], Luca Selmi [b], Marilyne Sousa [d], Kurt Stokbro [e], Michele Visciarelli [c]

[a] Integrated Systems Laboratory, ETH Zürich, Gloriastrasse 35, 8092 Zürich, Switzerland
[b] DIEGM, Universitá degli studi di Udine, Via delle Scienze 206, 33100 Udine, Italy
[c] DEIS-ARCES, Universitá degli studi di Bologna, Via Risorgimento 2, 40136 Bologna, Italy
[d] IBM Zürich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland
[e] QuantumWise A/S, Fruebjergvej 3, Box 4, DK-2100 Copenhagen, Denmark

ARTICLE INFO

Article history:
Available online 2 October 2015

Keywords:
III-V semiconductors
Band-structure
DFT
Tight-binding
k · **p**
Non-parabolic effective mass models
Ultra-Thin Body MOSFET

ABSTRACT

We present and thoroughly compare band-structures computed with density functional theory, tight-binding, **k** · **p** and non-parabolic effective mass models. Parameter sets for the non-parabolic Γ, the L and X valleys and intervalley bandgaps are extracted for bulk InAs, GaAs and InGaAs. We then consider quantum-wells with thickness ranging from 3 nm to 10 nm and the bandgap dependence on film thickness is compared with experiments for $In_{0.53}Ga_{0.47}As$ quantum-wells. The impact of the band-structure on the drain current of nanoscale MOSFETs is simulated with ballistic transport models, the results provide a rigorous assessment of III–V semiconductor band structure calculation methods and calibrated band parameters for device simulations.

1. Introduction

The size- and bias-dependent band-structure of nanoscale Ultra Thin Body and Box Semiconductor on Insulator (UTBB ScOI) FETs, fin-FETs and quantum-well FETs is a key ingredient for accurate electrostatics and transport models. Models extending beyond the popular parabolic effective-mass-approximation (EMA) are desirable, especially for thin films III–V compound semiconductor channels, where strong carrier confinement coexists with a strong non-parabolicity (NP) of the conduction band minimum at the Γ point [1–4]. Several methods can be used to work out band-structure calculations in a 2D electron gas, each of them having its own parameters and limitations. While few of these are publicly available in research or commercial tools (e.g. [5–7]), unfortunately these methods have been rarely cross-checked. Thus, the relative accuracy is not rigorously assessed.

In this paper we compute the band-structure of InAs, $In_{0.53}Ga_{0.47}As$ and GaAs quantum wells at room temperature with thicknesses down to 3 nm by employing a hierarchy of modeling approaches ranging from atomistic density functional theory (DFT) and empirical tight binding (TB) to **k** · **p** and NP-EMA. These models are presented in Section 2. The analysis and parametrization of the in-plane dispersion relation ($E - \mathbf{k}$) for bulk materials and quantum well structures is reported in Section 3. Comparison with experimental band-gap vs. semiconductor thickness data is reported in Section 3.4. The manuscript notably extends the analysis presented in [8] and provides useful parameter sets for an accurate description of III–V MOS band-structures in nanoscale UTBB MOSFET architectures.

2. Model description

Four distinct models are taken into consideration. **DFT** calculations within the meta-generalized gradient approximation (meta-GGA) are performed using ATK-2014 [7]. The c-parameter of the Tran and Blaha [9] exchange–correlation functional is fitted to reproduce the experimental band gaps of the bulk materials (0.354 eV for InAs, 1.423 eV for GaAs and 0.742 eV for $In_{0.53}Ga_{0.47}As$). The Kohn–Sham wave functions are expanded in

* Corresponding author.
E-mail address: gzerveas@iis.ee.ethz.ch (G. Zerveas).

http://dx.doi.org/10.1016/j.sse.2015.09.005
0038-1101/

localized atomic orbitals (LCAO) with a double-zeta plus polarization basis set. The III–V surface is passivated with pseudo-hydrogen atoms containing the appropriate non-integer charges (fractional charge per atom = $(1/4)q$) [10]. The parameters of bulk InGaAs were obtained with the virtual crystal approximation (VCA), as implemented in ATK [11–13]. Here, one creates an average pseudo-potential describing an $In_xGa_{1-x}As$ virtual atom, i.e. one creates a new, virtual atom, which is a mixture of In and Ga with appropriate weights. The actual calculation of band-structures, is then performed on a primitive cell with only two atoms: the normal As atom and the virtual In-Ga atom. However, any effects due to random disorder are not included. The InGaAs quantum well structures were treated as follows: we define a unit cell, which is periodic in the in-plane directions. Within this unit cell we make a $In_{0.5}Ga_{0.5}As$ alloy, i.e. explicitly modeling both In and Ga atoms without relying on the VCA. Although random disorder in principle is included in the perpendicular direction (within the unit cell), the band-structure of the quantum well is still perfectly well defined, since the unit cell is repeated periodically in the in-plane directions. The VCA approach, widely used in literature, has been validated here for a 5 nm thin $In_{0.5}Ga_{0.5}As$ quantum well. If we explicitly model the In and Ga atoms we obtain a band gap of 0.97 eV, while VCA calculations yield a band gap of 0.94 eV.

Nearest neighbor sp3d5s* **empirical TB** calculations including spin–orbit coupling have been performed with OMEN [14]. The on-site orbital energies and the overlap integrals between atomic orbitals for different bond types are parameters optimized to fit experimental values for band gaps and effective masses at different points of the Brillouin zone, so as to reproduce the complete bulk band-structure [15]. Bowing parameters have been introduced to accurately capture the band gap and effective masses of $In_xGa_{1-x}As$ as a function of x in the virtual crystal approximation (VCA) [16]. For III–V compounds, 31 fitting parameters are used in total. Surface atoms are passivated by increasing the dangling bond energy [17]. For the purposes of this paper, the passivation energy has been set to 30 eV, which is equivalent to applying hard-wall boundary conditions. In contrast to the closed boundary conditions imposed in the $\mathbf{k}\cdot\mathbf{p}$ and NP-EMA models (see below), the wave-function is non-zero on the boundary atoms, but immediately vanishes outside the structure. This minor difference is not expected to have any observable effect on the reported results.

The $\mathbf{k}\cdot\mathbf{p}$ calculations have been performed using the 8-band model [18]. It is worth noting that the $\mathbf{k}\cdot\mathbf{p}$ expansion is expected to accurately describe the bands of direct bandgap semiconductors around the Γ point of the first Brillouin zone, and does not take into account satellite valleys which are instead embedded in DFT and TB calculations. For 2D subband calculations, the parameters from [19] have been assumed, with a modified energy parameter E_P according to [20] to eliminate spurious solutions. The energy parameter E_P is defined as $E_P = \left(\frac{2m_0}{\hbar^2}\right)P^2$, with m_0 the free electron mass and P the Kane parameter proportional to the momentum matrix element between the s-like conduction bands and the p-like valence bands. Closed boundary conditions are assumed at the semiconductor/oxide interface.

The **NP-EMA** model employed in the multi-subband Monte Carlo of [21], based on the NP corrections in [22], is compared with the models above. For ideal, infinite quantum-wells having a box energy profile with closed boundary conditions (zero wave-function at the edges) of width T_W, the in-plane energy dispersion for the n-th subband is given by:

$$E_n(k_x,k_y) = \frac{\sqrt{1+4\alpha_{3D}\left[\frac{\hbar^2}{2}\left(\frac{k_x^2}{m_{x,3D}}+\frac{k_y^2}{m_{y,3D}}\right)+\frac{\hbar^2(n\pi)^2}{2m_{z,3D}T_W^2}\right]}-1}{2\alpha_{3D}} \quad (1)$$

where z is the quantization direction, while x and y are the directions in the transport plane and $\mathbf{k}=(k_x,k_y)$ is the in-plane wave-vector. The sole inputs for such a NP-EMA model are the effective masses m_{3D} (along the various directions) and non-parabolicity coefficient α_{3D} of the bulk crystal. Throughout this paper those parameters are taken from the DFT calculations unless otherwise stated.

3. Results

3.1. $E-\mathbf{k}$ relation of bulk materials

Tables 1–3 report the band-structure parameters for InAs, GaAs and InGaAs as obtained with the models in Section 2. The masses are extracted as the second derivative of the energy dispersion at the conduction band minimum, whereas α_{3D} for the Γ valley is obtained by fitting the $E-\mathbf{k}$ over a 0.5 eV energy interval. The c-parameter of the DFT model have been adjusted to reproduce the gap (valence band to Γ minimum of the conduction band) as in [9], but this results in a slightly larger effective mass of the Γ valley compared to other approaches for the GaAs and InGaAs cases. Also the masses of the satellite valleys and their energy gap with respect to the Γ valley are slightly different.

The corresponding energy dispersion for these bulk materials are shown in Fig. 1 along the main symmetry directions in the 3D Brillouin zone. Due to the differences among the internal energy references, the $E-\mathbf{k}$ relations have been rigidly shifted to zero the energy of the conduction band minimum at the Γ point. The same energy shift is then applied in the next subSection 3.2 when comparing the $E-\mathbf{k}$ relation of quantum wells.

Overall, Fig. 1 shows a good mutual agreement among the models for the Γ valley and to a lesser extent for X and L valleys. The large $m_l^*(X)$ of InAs in the DFT is the result of a very flat band structure in this point, which is not well-described by a parabolic model, but instead by a k^n relation with $n>2$. This means that the effective mass, extracted as the second derivative of the $E-\mathbf{k}$ curves, tends to infinity. Note that the numerically calculated second order derivative of the bands (inversely proportional to the effective mass) depends critically on the finite difference step size, Δk. We have explicitly checked that the exceptionally high DFT value for InAs is not due to a miscalculation but is simply the result for the given pseudopotential and basis set. Using a different pseudopotential of the Open-MX type, we indeed obtain a smaller effective mass ($1.565\cdot m_0$).

The satellite valleys are not included in the $\mathbf{k}\cdot\mathbf{p}$ calculation, whereas the differences between TB and DFT results are due to the parameters of the latter being adjusted to reproduce the VB-Γ energy gap.

3.2. $E-\mathbf{k}$ relation in quantum wells

Comparisons among the $E-\mathbf{k}$ dispersion relations for quantum wells of 7 nm $In_{0.53}Ga_{0.47}As$, 5 nm InAs and 3 nm GaAs are reported in Figs. 2a, 3a and 4a, respectively. Preliminary results for a few other thicknesses were examined in [8]. In these plots the energy reference ($E=0$) for TB, $\mathbf{k}\cdot\mathbf{p}$ and NP-EMA is the minimum of the Γ valley in the bulk crystal, as already mentioned in Section 3.1. As regards DFT calculation, the method provides the system's ground state but the energy reference is not immediately obvious. In order to compare DFT with the other methods and overcome the above difficulty we first calculated the Fermi energy (E_{f0}) and the conduction band minimum (CBM0) of the bulk crystal. Then the $E-\mathbf{k}$ of each quantum well is computed assuming charge neutrality and then is rigidly shifted along the vertical axis to align the Fermi energy to E_{f0}. Finally CBM0 is subtracted to the shifted $E-\mathbf{k}$ of the quantum well.

Table 1
Band-structure parameters for InAs bulk material.

Method	$m^*(\Gamma)$ $[m_0]$	$\alpha(\Gamma)$ $\left[\text{eV}^{-1}\right]$	$m_t^*(L)$ $[m_0]$	$m_l^*(L)$ $[m_0]$	L − Γ [eV]	$m_t^*(X)$ $[m_0]$	$m_l^*(X)$ $[m_0]$	X − Γ [eV]
InAs								
DFT	0.0231	3.12	0.122	1.801	0.83	0.251	423.386	1.54
TB	0.0236	2.5	0.094	1.54	1.16	0.176	1.123	1.91
k · p	0.026	3.15	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
[19]	0.026	–	0.05	0.64	0.716	0.16	1.13	1.01

Table 2
Band-structure parameters for GaAs bulk material.

Method	$m^*(\Gamma)$ $[m_0]$	$\alpha(\Gamma)$ $\left[\text{eV}^{-1}\right]$	$m_t^*(L)$ $[m_0]$	$m_l^*(L)$ $[m_0]$	L − Γ [eV]	$m_t^*(X)$ $[m_0]$	$m_l^*(X)$ $[m_0]$	X − Γ [eV]
GaAs								
DFT	0.0829	0.75	0.124	1.622	0.12	0.225	0.866	0.52
TB	0.067	0.73	0.100	1.787	0.288	0.174	1.261	0.49
k · p	0.067	0.65	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
[19]	0.067	–	0.0754	1.9	0.28	0.23	1.3	0.48

Table 3
band-structure parameters for $In_{0.53}Ga_{0.47}As$ bulk material.

Method	$m^*(\Gamma)$ $[m_0]$	$\alpha(\Gamma)$ $\left[\text{eV}^{-1}\right]$	$m_t^*(L)$ $[m_0]$	$m_l^*(L)$ $[m_0]$	L − Γ [eV]	$m_t^*(X)$ $[m_0]$	$m_l^*(X)$ $[m_0]$	X − Γ [eV]
$In_{0.53}Ga_{0.47}As$								
DFT	0.0507	1.55	0.132	3.343	0.99	0.367	0.979	1.90
TB	0.043	1.25	0.096	1.588	0.67	0.198	2.824	1.09
k · p	0.043	1.35	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
[19]	0.043	–	0.0619	1.2322	0.5498	0.1929	1.2099	0.5498

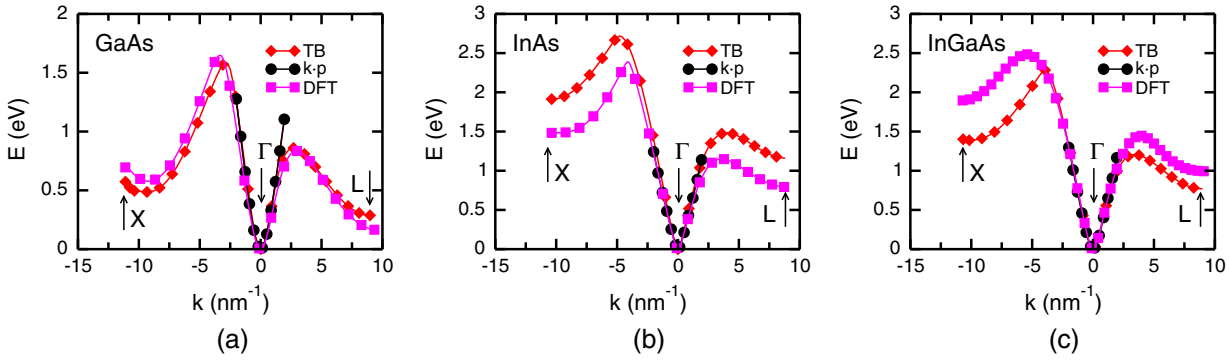

Fig. 1. Bulk band-structures of (a) GaAs, (b) InAs, (c) $In_{0.53}Ga_{0.47}As$ along the [100–110] directions (X–Γ–L).

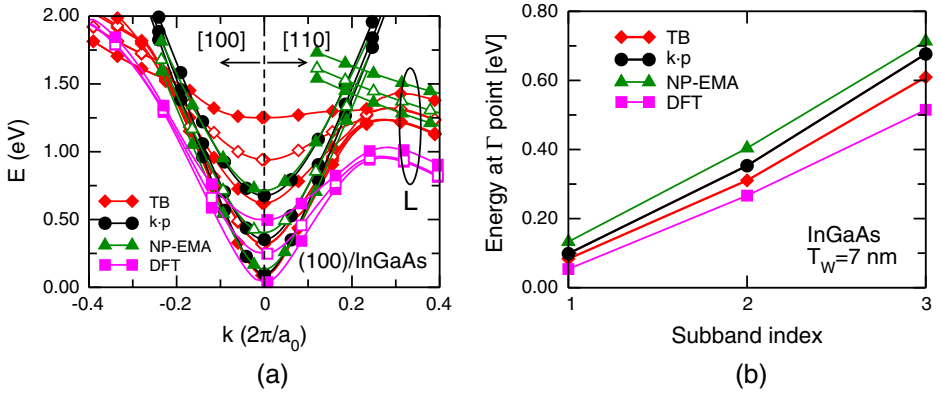

Fig. 2. (a): In-plane $E - \mathbf{k}$ for a 7 nm $In_{0.53}Ga_{0.47}As$ quantum well. Quantization is along the [001] direction and the in-plane plotting directions are [100] and [110]. The three subbands of the NP-EMA located on the right side of the chart, whereas the ones in the center are the subbands originating from the Γ valley. (b): Energy at the Γ point for the lowest three sub-bands.

Figs. 2b, 3b and 4b directly compare the subband energy at the Γ point for all models. For InAs and InGaAs all these subbands originate from the CB minimum in Gamma, whereas in GaAs also X valleys play a role, which explains the difference between $\mathbf{k}\cdot\mathbf{p}$ and the rest of the models.

Judging by the leftmost graphs of Figs. 2–4, the shape of the lowest subbands at Γ is in overall good mutual agreement up to a few hundreds meV along the [100] direction. Discrepancies are larger along the $\langle 110\rangle$ direction in most cases and especially for the highest subbands. L and X valleys are not included in the $\mathbf{k}\cdot\mathbf{p}$, and are added as separate subbands in the NP-EMA (see for instance the green lines departing from $k=-0.4(2\pi/a_0)$ and for graphical purposes terminated at $k\approx -0.2(2\pi/a_0)$ in Figs. 3a and 4a). Thus, they lose the smooth transition observed in TB and DFT calculations. The small energy gap between X and Γ valleys in GaAs slabs farther complicates the picture. Differences between DFT and TB reflect the different energy and effective masses of the L valleys in the corresponding bulk materials shown in Fig. 1 and Tables 1–3. However these discrepancies have a small effect for most InAs and InGaAs quantum-wells because the satellite valleys play a marginal role in the electronic transport. They are more relevant in GaAs wells with small $T_W < 5$ nm. The latter, however, are of less technological interest for future III–V MOSFETs.

Similarly to Fig. 2, Fig. 5 shows the in-plane $E-\mathbf{k}$ dispersion relationship for $In_{0.53}Ga_{0.47}As$ quantum wells of varying thickness. Expectedly, as the thickness decreases and the confinement becomes stronger, the conduction band minimum is shifted towards higher energies, while at the same time the energy difference between subbands increases. The agreement between different models proves to be more challenging for structures enforcing stronger confinement.

An important band parameter for the study of electron transport along the MOSFET channel is the in-plane effective mass m_{2D}. The m_{2D} is extracted from the second derivative of the energy dispersion of the lowest subband at the Γ point. Fig. 6 shows m_{2D} as a function of the well thickness. As expected, and consistently with other works [23,24], the in-plane effective mass increases when the well thickness is decreased. The masses extracted from the NP-EMA track the DFT results, since NP-EMA calculations were based exactly on the m_{3D} and α_{3D} parameters of the bulk crystal extracted from the DFT results. Since the bulk parameters of the DFT differ from the TB and $\mathbf{k}\cdot\mathbf{p}$ parameters, we see that the curvature (hence, the m_{2D}) of the quantized $E-\mathbf{k}$ relation is different. Good match between the in-plane effective mass given by the TB and $\mathbf{k}\cdot\mathbf{p}$ methods is observed.

The NP coefficient α_{2D} indicates how much the in-plane $E-\mathbf{k}$ in the quantized electron gas deviates from a parabola. Fig. 7 plots the NP coefficient as extracted from the $E-\mathbf{k}$ along the $\langle 100\rangle$ direction, i.e. by fitting the $E-\mathbf{k}$ of the lowest subband with the expression:

$$E(k_x,k_y)\left(1+\alpha_{2D}E(k_x,k_y)\right)=\frac{\hbar^2\left(k_x^2+k_y^2\right)}{2m_{2D}} \tag{2}$$

The α_{2D} values are different from the α_{3D} ones; moreover they decrease when the well thickness is reduced, meaning that the in-plane $E-\mathbf{k}$ becomes more parabolic. This behavior is consistent with the limit of Eq. (1) as T_W approaches zero. The DFT, NP-EMA and $\mathbf{k}\cdot\mathbf{p}$ models provide quite different α_{2D} values although with similar trends, whereas the TB results show a weaker dependence on the well thickness.

Fig. 8 plots the energy minimum for the lowest subband as a function of the well thickness. As discussed previously, the zero energy is at the bottom of the Γ valley of the bulk crystal. The $E-\mathbf{k}$ dispersion relations have been rigidly shifted to make the conduction band minima coincide in the bulk case and the same shift has been applied to the $E-\mathbf{k}$ of the quantum wells.

The group velocity υ_g in the transport direction is an important ingredient of any semi-classical transport model where particles are associated to wavepackets. We thus investigated υ_g in the same quantum wells of Figs. 2–4 by computing the derivative of the in-plane dispersion relationship along the [100] direction as $\upsilon_g(k_{100})=\frac{1}{\hbar}\frac{dE}{dk_{100}}$.

Fig. 9 reports υ_g for the same cases of Figs. 2–4 over the $\pm 0.2(2\pi/a_0)$, interval, which corresponds to an energy range of a few hundred meV. The large effective mass of the Γ valley predicted by DFT calculation yields a slightly lower group velocity compared to the other methods. Since the NP-EMA results are based on the bulk parameters of DFT, they well agree with the DFT υ_g for quantum wells. Further good mutual agreement between TB and $\mathbf{k}\cdot\mathbf{p}$ is observed up to the point where satellite valleys come into play.

Fig. 10 depicts the group velocity along the [100] direction for $In_{0.53}Ga_{0.47}As$ quantum wells of varying thickness. Increasing the well thickness corresponds only to a mild increase in group velocity (about 10% increase of the maximum value for a thickness change from 3 to 10 nm). However, the increase is more prominent for low k, resp. for low energies, because the effective mass decreases with increasing thickness (as observed in Fig. 6) and therefore the initial slope of the group velocity graphs is steeper

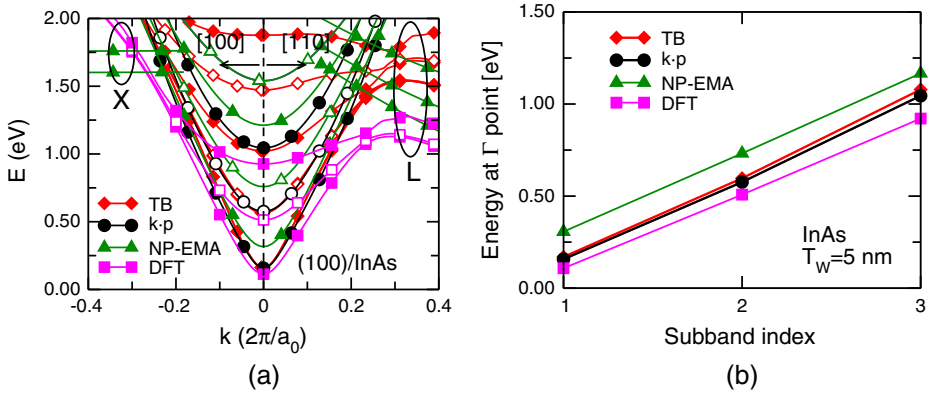

Fig. 3. (a): In-plane $E-\mathbf{k}$ for a 5 nm InAs quantum well. Quantization is along the [001] direction, in-plane directions are [100] and [110]. The three subbands of the NP-EMA located on the right and left edges of the graph are associated to L and X valleys respectively, whereas the ones in the center are the subbands originating from the Γ valley. (b): Energy at the Γ point for the lowest three subbands.

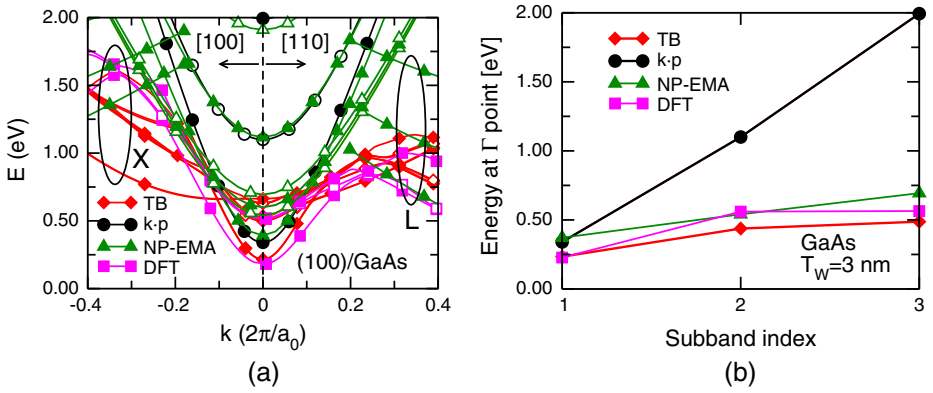

Fig. 4. (a): In-plane $E - \mathbf{k}$ for a 3 nm GaAs quantum well. Quantization is along the [0 0 1] direction, in-plane directions are [1 0 0] and [1 1 0]. The subbands of the NP-EMA that are appearing on the right and left edges of the graph are associated to L and X valleys, respectively, whereas the ones in the center are the subbands originating from the Γ valley. (b): Energy at the Γ point for the lowest three subbands.

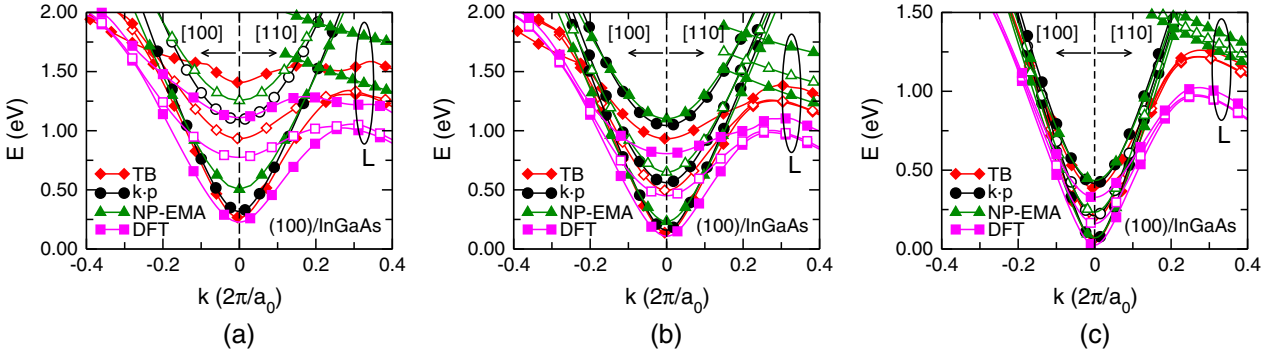

Fig. 5. In-plane $E - \mathbf{k}$ for a 3 nm (a), 5 nm (b) and 10 nm (c) $In_{0.53}Ga_{0.47}As$ quantum well. Quantization is along the [001] direction, in-plane directions are [1 0 0] and [1 1 0].

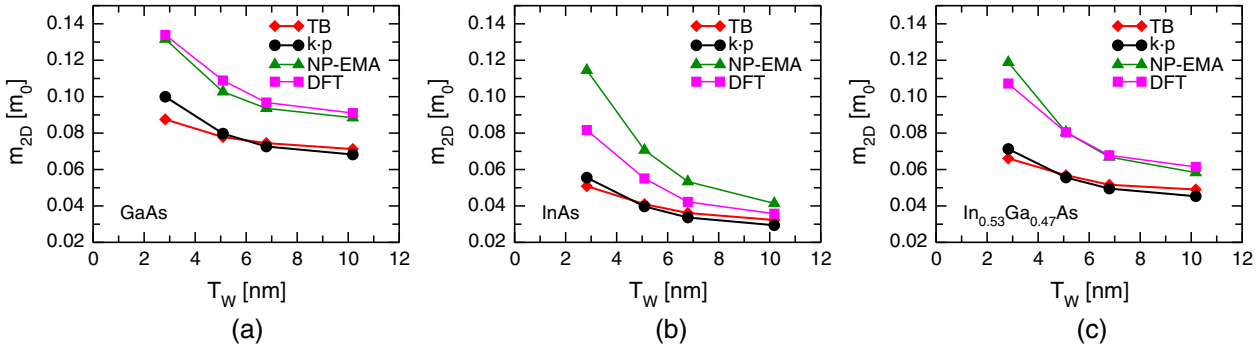

Fig. 6. Effective mass (extracted from the second derivative of the energy dispersion of the lowest subband at the Γ point) vs. well thickness for GaAs (a), InAs (b) and InGaAs (c).

for larger thicknesses. One may also observe that the atomistic models predict a saturation of the group velocity for high energies, as opposed to the $\mathbf{k} \cdot \mathbf{p}$ and NP-EMA models which predict a group velocity that is strictly monotonically increasing with k, resp. energy.

3.3. Simulation time for band-structure calculation

In order to effectively facilitate the transistor design process, numerical models need to take into account computational needs and efficiency. Therefore, an important consideration when comparing different models is the simulation time. A direct comparison is not trivial, as on the one hand the models examined here describe semiconductors at notably different levels of abstraction, providing a different degree of physical detail, and on the other hand, software implementations reflect the different priorities of developers, favouring completeness and accuracy for academic simulators, as opposed to robustness and ease of use for commercial programs. Nevertheless, a meaningful comparison examining execution times and how these scale with the computational problem size would assist in identifying the role that each model could play in the industrial design process of semiconductor devices.

In the present section we compare execution times for quantum well band-structure calculations. Two of the parameters that determine the size of the computational problem are firstly the well thickness, which sets the number of atoms in the DFT and TB codes and the size of the real space grid in $\mathbf{k} \cdot \mathbf{p}$ codes, and secondly the number of k-space discretization points. Models based

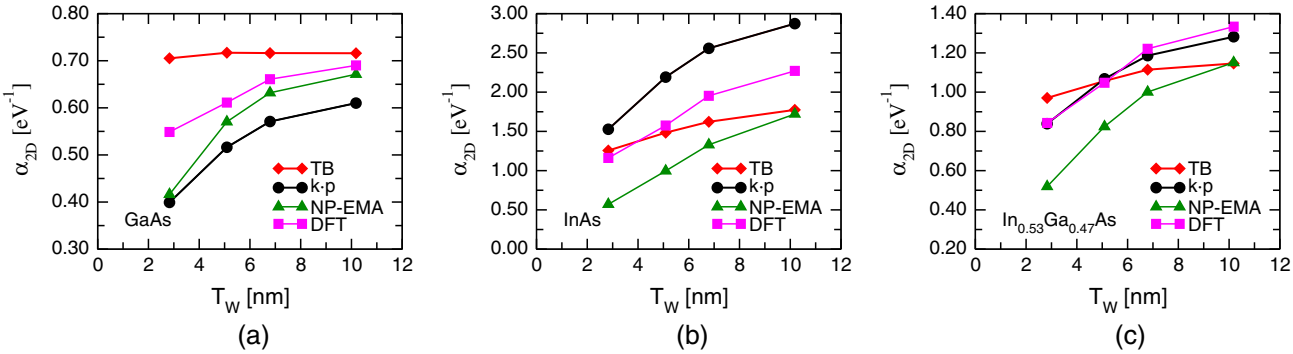

Fig. 7. NP coefficient of the lowest subband α_{2D} of GaAs (a), InAs (b) and InGaAs (c) as extracted from the 2D energy dispersion over an energy range of 0.5 eV from the bottom. For a 7 nm quantum well of $In_{0.53}Ga_{0.47}As$, the NP coefficient is 1.00 eV^{-1} for the first subband, 0.65 eV^{-1} for the second one.

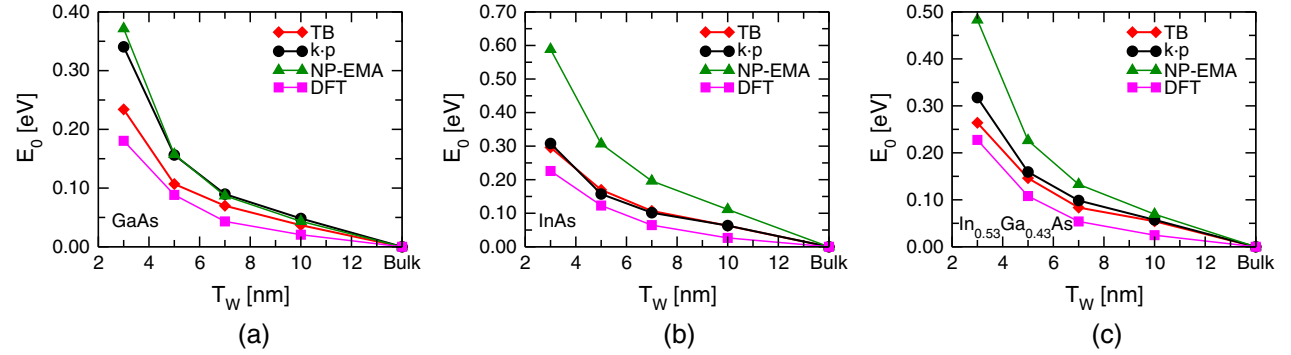

Fig. 8. Energy of the first subband as a function of the well thickness for GaAs (a), InAs (b) and InGaAs (c) slabs. The energy reference ($E = 0$) is the minimum of the Γ valley in the bulk crystal. This means that we have shifted the $E - \mathbf{k}$ obtained with the various models by the same amount used to align them in the bulk case.

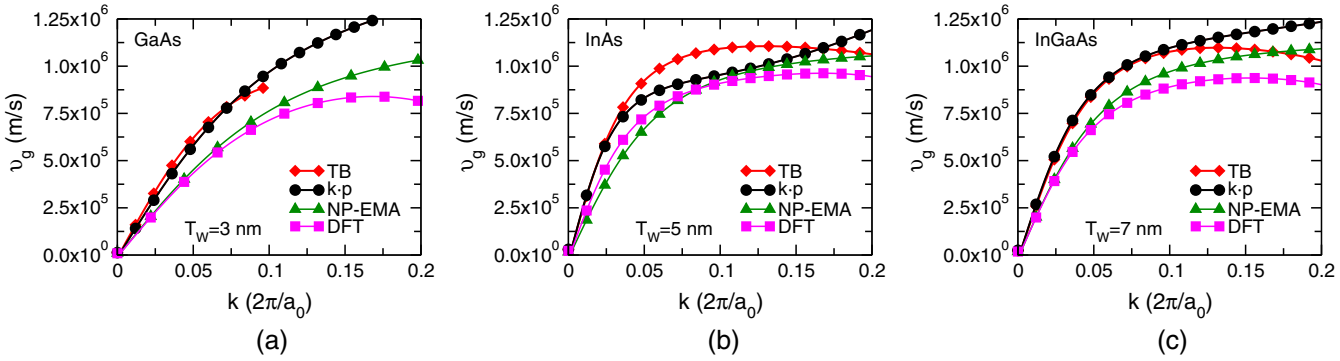

Fig. 9. Group velocity of the first subband for GaAs (a), InAs (b) and InGaAs (c) QW of various thickness.

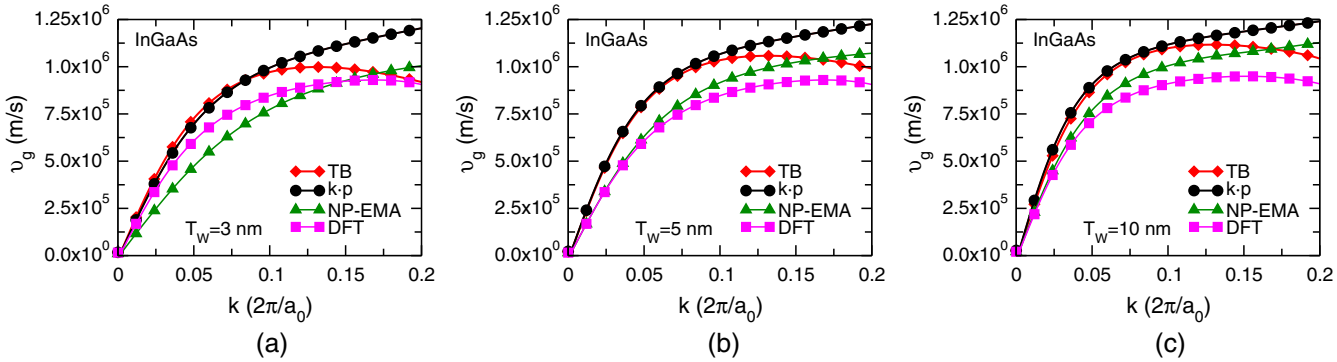

Fig. 10. Group velocity of the first subband for a 3 nm (a), 5 nm (b) and 10 nm (c) $In_{0.53}Ga_{0.47}As$ quantum well.

on an effective mass approximation are not considered in this comparison, as they rely on analytical band-structures and have a negligible computational burden compared to the rest of the models. Instead, in addition to the 8-band $\mathbf{k}\cdot\mathbf{p}$ method presented thus far, in this section we report data for the simulation times of a lower order, 4-band implementation of the $\mathbf{k}\cdot\mathbf{p}$ method. The execution times reported below have been rescaled to a reference processing power of 40 GFlops in order to account for the differences in computational power of the hardware used. Also, they refer to computations which include 64 subbands. Because there is no way of arbitrarily setting the number of subbands in the DFT model and instead this number depends on the well thickness, DFT raw execution times have been linearly rescaled down to 64 subbands as a rough approximation which allows comparison with the rest of the models.

The left graph of Fig. 11 shows how execution time changes with an increasing quantum well thickness. The number of k-space discretization points is in all cases 1001 points. As expected, the execution time increases with the well thickness, with atomistic methods gaining a competitive advantage as the well thickness decreases (i.e. the number of atoms is reduced). Overall, at a given thickness, the ratio of execution times between the slowest and the fastest algorithm is at most an order of magnitude or less. When comparing the 8-bands $\mathbf{k}\cdot\mathbf{p}$ model to the 4-band model, we observe an increase in execution time by approximately a factor of four, as the 8-band Hamiltonian matrix has four times the number of elements of the 4-band matrix. However, an interesting result is that although increasing the order of a $\mathbf{k}\cdot\mathbf{p}$ model is expected to improve the accuracy of band-structure calculations, this comes at an additional computational cost that might render the model slower than the 10-orbital TB model, which is in general expected to offer even higher accuracy [25]. The dependence of simulation time on the number of k-space discretization points is depicted in the right graph of Fig. 11 for well thicknesses of 5 and 9 nm. One observes that the execution time scales linearly with this number for the TB code. Despite its weak linear dependence, the execution time for DFT is almost constant with respect to the number of k-points, since only a proportionally small part of the simulation time is spent on solving the eigenvalue problem and as such depends on this parameter. The same argument holds for the $\mathbf{k}\cdot\mathbf{p}$ solver, where most of the execution time is dedicated to computing the Hamiltonian matrix and its transformation to the reduced equivalent matrix used in the mode space approach. For a small number of k-space points the best performing method is TB, while the 4-band $\mathbf{k}\cdot\mathbf{p}$ solver becomes the fastest as the number of k-space points increases. However, this superior performance of $\mathbf{k}\cdot\mathbf{p}$ vanishes when the order of the model is increased to 8 bands.

Lastly, it is important to note that the performance results presented thus far depend on the number of CPU cores used for executing the simulations. As recent developments in hardware technology have established a trend for systems with an ever-increasing number of cores, the parallelizability of a code, i.e. better scaling of its performance with the number of CPU cores, would render it increasingly advantageous. Table 4 shows the speed-up factor of each method for every additional core employed and corresponds to the slope of a graph of speed-up as a function of CPU cores. An ideal scaling behavior would correspond to a value of 1.0, meaning that an increase in the number of cores would lead to exactly the same proportional increase in performance, e.g. using ten cores instead of one would yield a ten times faster simulation. As can be seen from the table, although a single-threaded execution yielded an approx. 4.5 times faster simulation for the 8-band $\mathbf{k}\cdot\mathbf{p}$ code compared to the TB code, the significantly stronger ability for parallelization exhibited by the TB code compared to the 8-band $\mathbf{k}\cdot\mathbf{p}$ code can account for its superior performance reported in Fig. 11, for which 20 CPU cores have been used. Similarly, although the DFT code yields very long execution times for a small number of cores, its high degree of parallelizability could render it more competitive on many-core systems, assuming that the same scaling trend persists for even higher numbers of cores than those examined here. It should nevertheless be noted that, despite the importance of parallelizability, there exist also other factors of computational efficiency which can drastically affect performance and are not examined in this paper, such as the use of optimized numerical algorithms, optimized data structures and more.

The analysis above indicates that for a device scale below 10 nm, as far as band-structure calculations are concerned, an implementation of the empirical TB method can be computationally efficient enough to be competitive with generally more lightweight methods such as $\mathbf{k}\cdot\mathbf{p}$, additionally offering the benefit of higher accuracy. However, being an atomistic method, above the limit of approx. 10 nm examined here it suffers an increasing penalty, while the band-structure converges to that of a bulk crystal, which in turn can be calculated with simpler algorithms on simpler physical systems. DFT, being an ab initio method, can be valuable for calibrating all other models. Nevertheless, simulation times still remain significantly longer than the other methods, especially when considering the absolute times of execution instead of implicitly assuming that the number of subbands can be reduced in order to linearly expedite simulations, as done in Fig. 11 for the sake of comparison. These results outline complementary roles for atomistic and non-atomistic models, based on their distinct advantages.

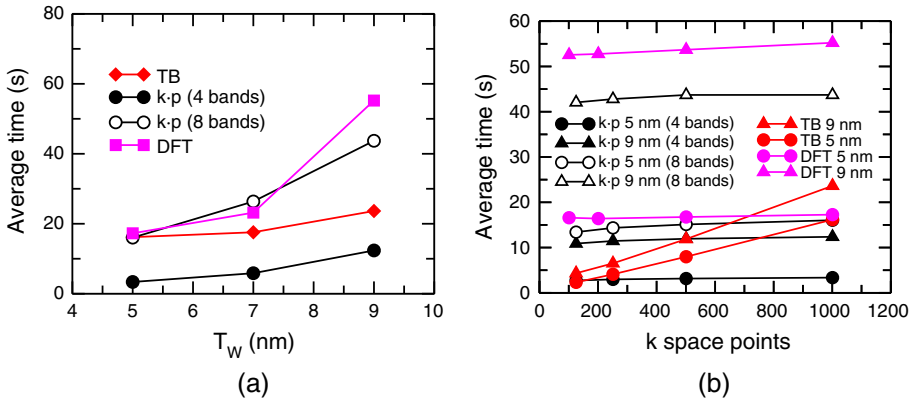

Fig. 11. (a): Execution time as a function of the quantum well thickness. (b): Dependence of simulation time on the number of k-space discretization points.

Table 4
Execution times on a single core and scaling behavior of different methods with an increasing number of cores (measured for up to 20 cores).

Method	Exec. time on 1 core [s]	Speed-up per core
DFT	653.10	0.59
TB	242.65	0.54
$\mathbf{k}\cdot\mathbf{p}$ (8-band)	52.98	0.06

3.4. Comparison with bandgap energy measurements

For the sake of a thorough model calibration we compare the models described in Section 2 with data for the energy gap of thin InGaAs layers measured on test device fabricated ad-hoc.

The fabrication of UTBB InGaAs on Si wafers [26] (described in Fig. 12) starts with 2″ semi-insulating (1 0 0)-oriented InP donor wafers loaded in a MOVPE system. An $In_{0.53}Ga_{0.47}As/In_{0.52}Al_{0.48}As$ etch-stop heterostructure is then grown at 550 °C followed by the growth of the active layer, namely: an $In_{0.53}Ga_{0.47}As$ channel of varying thickness. Subsequently, the wafers are loaded in an Atomic Layer Deposition tool (ALD) where the Al_2O_3 buried oxide (BOX) is deposited at 250 °C. The target 4″ (1 0 0)-oriented p-type ($N_A = 1\times10^{17}$ cm^{-3}) Si wafers are wet cleaned for organic contaminants and their native oxide is stripped in diluted hydrofluoric acid (DHF). A thin and high quality native oxide is then chemically regrown by exposing the wafers to ozone-rich deionized water (DIO_3), making their surfaces hydrophilic. The donor and target wafer surfaces are brought into intimate contact at room temperature and ambient atmosphere to initiate the bonding. The wafers are then annealed at 300 °C for 2 h in order to raise the bonding energy. The InP donor wafers are etched in concentrated hydrochloric acid until reaching the InGaAs/InAlAs etch-stop heterostructure. This etch-stop structure is finally etched in diluted acids to ensure a soft-landing on the active layer.

The resulting structure is thus $In_{0.53}Ga_{0.47}As$ on Al_2O3, that is relaxed for $x = 0.53$.

The $In_{0.53}Ga_{0.47}As/Al_2O_3/Si$ layers have been characterized by spectroscopic ellipsometry using a variable angle equipment (VASE®) from J.A. Woollam Co. The measurements have been carried out between 0.496 eV and 2.48 eV at 3 different angles of incidence (65°, 70° and 75°) and the $In_{0.53}Ga_{0.47}As$ energy gap has been determined using a parametric semiconductor model [27] for the data fitting.

Fig. 13 compares the simulated energy gap versus film thickness spectroscopic ellipsometry measurements for relaxed layers.

All models demonstrate a remarkable mutual consistency and overall good mutual agreement with the experiments. Agreement with experiments is good around 10 nm. As T_W increases, all simulations methods predict a more rapid convergence to the bulk material energy gap value compared to experiments. To shed more light on this behavior $\mathbf{k}\cdot\mathbf{p}$ results are shown for both closed and open boundary conditions. In the latter case the Al_2O_3 substrate of the experimental samples is described with the same $\mathbf{k}\cdot\mathbf{p}$ parameters as the $In_{0.53}Ga_{0.47}As$ layer but with appropriate affinity (χ_{InGaAs} = 4.5 eV and $\chi_{Al_2O_3}$ = 1.0 eV). For very thin wells, the simulated band gap is larger than the experimental one. The discrepancy which is only partially reconciled adopting open boundary conditions.

The $\mathbf{k}\cdot\mathbf{p}$ results with open and closed boundary conditions are very similar, pointing out that in these structure the treatment of the boundaries is not critical.

4. Comparison among III–V-MOSFET ballistic currents

4.1. Ballistic current calculated using multiple transport models

To assess the impact of band-structure on electron transport in intrinsic device designs relevant for ultimate technology nodes, we have simulated a double-gate InAs MOSFET with $L_G = 10$ nm, 1 nm of SiO_2 gate dielectric, film thickness $T_W = 2.35$ nm, S/D doping $N_D = 5\cdot10^{19}$ cm^{-3} and channel doping of $N_A = 1\cdot10^{15}$ cm^{-3}, as sketched in the left plot of Fig. 14.

The right plot of Fig. 14 shows the ballistic currents computed by Non-Equilibrium Green Function (NEGF) solvers based on TB [14], DFT [7] and $\mathbf{k}\cdot\mathbf{p}$ bands [28] and by semiclassical Multi-Subband Monte Carlo (MSMC [21]) solver of the Boltzmann transport equation for the NP-EMA model. The parameters for the NP-EMA model are the bulk masses, non-parabolic coefficient and energy offsets from the DFT.

The curves have been shifted horizontally to match the off-current with the ITRS specification for high performance devices (100 nA/μm). The sub-threshold swing of the MSMC simulation is the smallest because semi-classical transport neglects source-to-drain-tunneling. This discrepancy is easily reconciled by using the subband smoothing method proposed in [29]. A gaussian

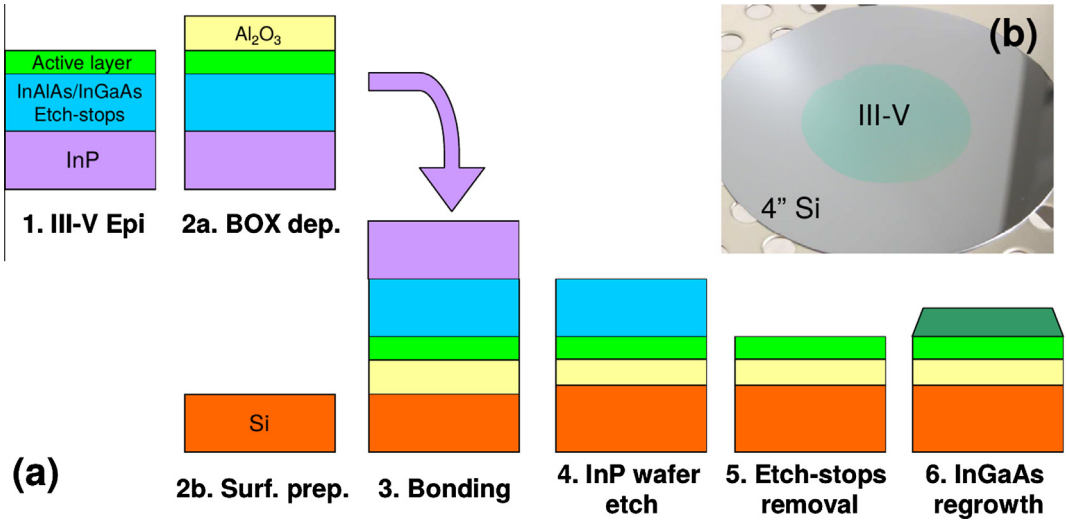

Fig. 12. (a) Schematic process flow of the fabrication of UTBB InGaAs-on-insulator substrates (1 to 5) and InGaAs regrowth module for S/D regions (6). (b) Picture of an InGaAs-on-insulator substrate.

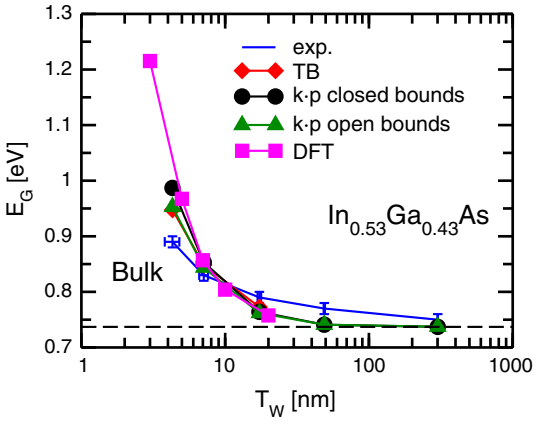

Fig. 13. Simulated and experimental energy gap for unstrained $In_{0.53}Ga_{0.47}As$ quantum well on Al_2O_3.

wavepacket with standard deviation of $\sigma = 5$ nm (open triangles), yields a good agreement with the NEGF full-quantum transport results. This value is larger than the one used in [30] for Si ($\sigma = 1.1$ nm), which is a reasonable finding since the small effective mass of InAs should result in broader wavepackets than in silicon.

As regards the on-current curves, we note that the solver used in combination with NP-EMA, TB and $\mathbf{k}\cdot\mathbf{p}$ models enforce von Neumann boundary conditions for the potential at the source/drain contacts when solving the Poisson equation, whereas DFT has Dirichlet boundary conditions. This means that TB, NP-EMA and $\mathbf{k}\cdot\mathbf{p}$ based simulations lower the conduction band edge of the source to maintain charge neutrality at high V_{GS} and thus yield a higher on-current than the DFT based solution. Another source of mismatch in the on-current stems from the decision to match the off-current; in fact, in this case small differences in terms of subthreshold slope translate in an appreciable change of the threshold voltage and thus of I_{ON}.

4.2. Ballistic current calculated with MSMC simulator and different band-structure

The results in Section 4.1 cannot separate the effects of a different transport model from those of the band structure. To isolate the effect of the band-structure alone, we decided to run ballistic MSMC simulation of the same devices as in Fig. 14 employing the same transport model with all the examined band models. To this purpose the parameters for bulk InAs from Table 1 have been used. The ballistic current simulations are shown in the top-left plot of Fig. 15 and demonstrate a very good mutual agreement, meaning that the differences among curves in Fig. 14 are mainly due to the transport models rather than the band-structure parameters.

We have repeated the analysis for a InGaAs double-gate transistor with $T_W = 5$ nm, sketched in the bottom of Fig. 16. The gate length is $L_G = 14$ nm, and the gate dielectric is HfO_2 ($\kappa = 22$), with an equivalent oxide thickness of 0.7 nm. The ballistic current is reported in Fig. 16 (left plot) and it is almost independent of the bulk band-structure. In fact, the $E-\mathbf{k}$ for a 5 nm well as obtained from Eq. (1) with parameters sets extracted from the DFT, TB or $\mathbf{k}\cdot\mathbf{p}$ models (right-top plot in Fig. 16) is only weakly affected by the band-structure, because the changes in α_{3D} and m_{3D} tend to compensate each other. This is in line with the $E-\mathbf{k}$ for a 5 nm InGaAs well as shown in Fig. 1 of reference [8], where good agreement between the DFT, TB and $\mathbf{k}\cdot\mathbf{p}$ simulations was found.

We thus see that for the technology-relevant case of a few nm thick $In_{0.53}Ga_{0.47}As$ well, the models predict very consistent results. This is because at low drain voltage electron transport is dominated by the lowest subband, and satellite valleys provide a modest contribution to the total inversion charge.

One should bear in mind, however, that current simulations are ballistic and thus neglect scattering and its effects on the electrostatics [31]. Electron–phonon scattering can easily be included in semi-classical Monte Carlo transport models. Its inclusion in full-quantum atomistic and $\mathbf{k}\cdot\mathbf{p}$ models, however, is less trivial and results in time-consuming simulations. We have used the MSMC to assess the influence of phonons (polar and non-polar) on the drain current of the device in Fig. 14: the current at $V_{GS} = 1.0$ V is reduced by approximately 16%. Surface roughness is expected to have an additional detrimental effect. At this stage an estimate by MSMC is not possible, since there is no experimental data for the surface roughness spectrum of thin InAs layers. Surface roughness can also be described via full-quantum NEGF models [32], although this requires generating sample devices with rough interfaces and averaging the results over a family of such sample structures.

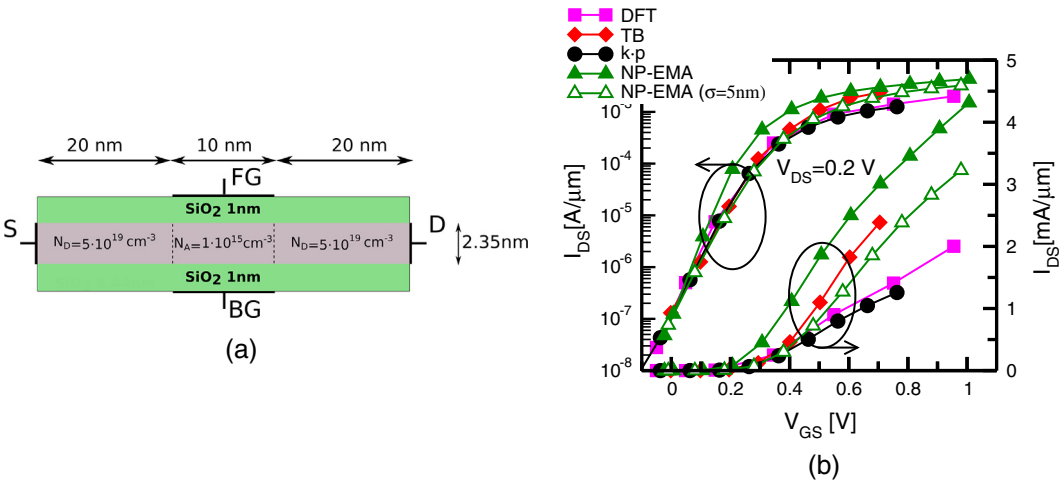

Fig. 14. (a): Sketch of the simulated device. (b): Ballistic current vs. gate voltage at a source-drain voltage of 0.2 V. The work functions of the metal gate are adjusted in order to have a 100 nA/μm off-current, according to the ITRS high performance specification.

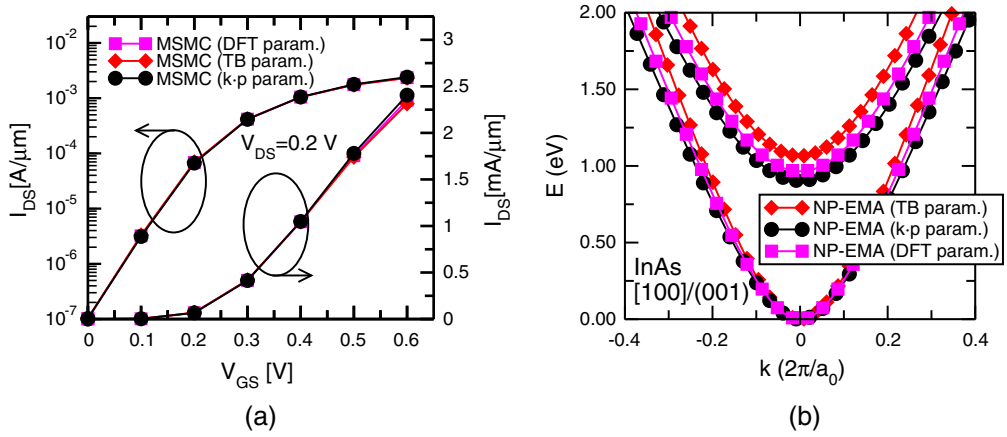

Fig. 15. (a): Current vs. gate voltage for $V_{DS} = 0.2$ V. (b): In-plane $E - \mathbf{k}$ for a 2.35 nm InAs quantum well, calculated with NP-EMA model using the bulk parameters of DFT, TB and $\mathbf{k} \cdot \mathbf{p}$ reported in Table 1.

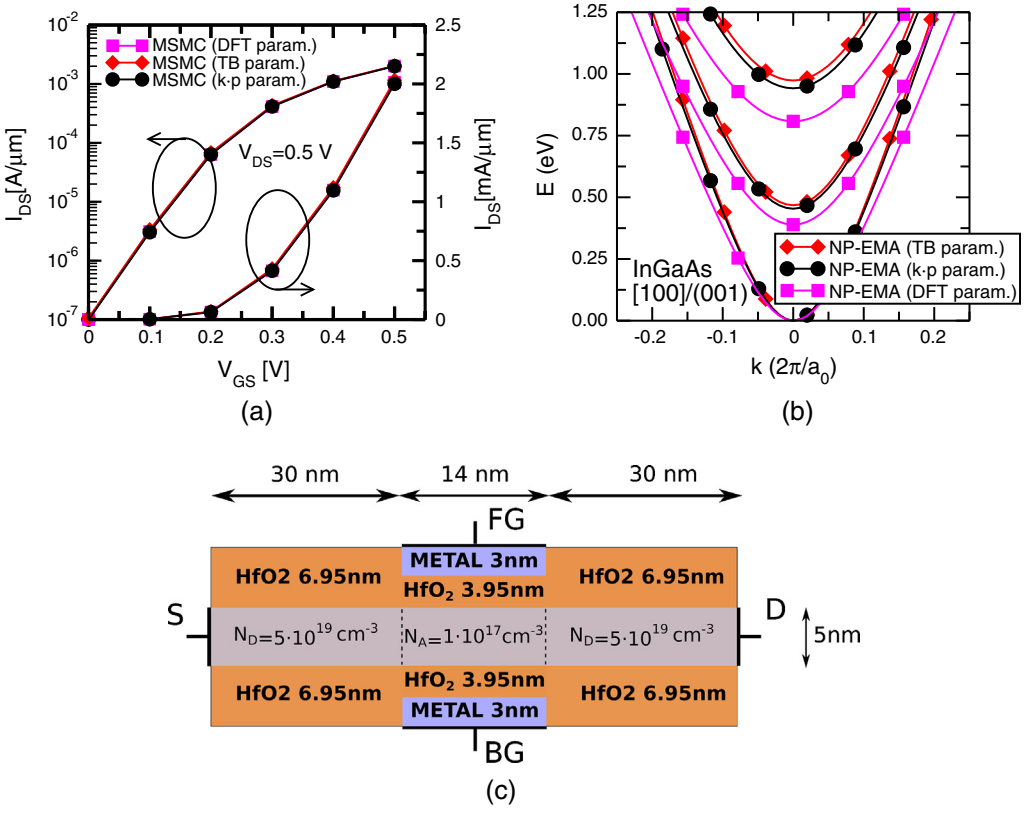

Fig. 16. (a): Current vs. gate voltage for $V_{DS} = 0.5$ V. (b): In-plane $E - \mathbf{k}$ for a 5 nm $In_{0.53}Ga_{0.47}As$ quantum well, calculated with the NP-EMA model using the bulk parameters from DFT, TB and $\mathbf{k} \cdot \mathbf{p}$ methods reported in Table 3. (c): Sketch of the simulated device.

5. Conclusions

The careful and detailed comparison and calibration of four band-structure calculation methods pointed out a number of critical issues for the accurate description of III–V compound semiconductor films with nanoscale thickness. Beside the importance of setting a well defined energy reference for all methods, the calculations highlighted a strong non-parabolicity of the Γ valley and the multi-valley nature of materials such as GaAs. The NP-EMA only partially captures non-parabolicity effects, and since the conduction bands are treated as separate valleys, it fails to reproduce the smooth transitions between valleys along some critical directions. The 8 bands $\mathbf{k} \cdot \mathbf{p}$ method is limited to the Γ valley, which is adequate for high indium-content semiconductors. Empirical TB and DFT provide a detailed atomistic description of the device, but can become demanding in terms of absolute computation time, especially for increasing device dimensions. As transistor dimensions decrease, atomistic approaches gain a competitive edge. Furthermore, they inherently capture complex multi-valley phenomena. When applied to technologically relevant InAs and InGaAs wells in the 5 nm thickness range, all the models provide fairly consistent results, aligned with experimental data on energy bangaps. Consistent parameters sets for effective mass band structure models have been extracted which will be useful

for device simulations of advanced MOSFET architecture in III–V semiconductor channels. Ballistic simulations of template decananometric MOSFETs show that the different approaches provide different drain current mainly due to different transport models, whereas the impact of the band structure parameters is limited.

Acknowledgements

The research leading to these results has received funding from the European Commission's Seventh Framework Programme (FP7/2007–2013) under Grant Agreement III–V-MOS Project No. 619326 via the IUNET Consortium.

We would like to acknowledge the reviewers of the EU project III–V-MOS for suggesting us to compare the different models in terms of calculation times.

References

[1] Bhattacharya S, Mahapatra S. Influence of band non-parabolicity on few ballistic properties of III–V quantum wire field effect transistors under strong inversion. J Comput Theor Nanosci 2009;6(7):1605–16.
[2] Liu Y, Neophytou N, Klimeck G, Lundstrom M. Band-structure effects on the performance of III–V ultrathin-body SOI MOSFETs. IEEE Trans Electron Dev 2008;55(5):1116–22.
[3] Wang L, Asbeck PM, Taur Y. Self-consistent 1-D Schrödinger-Poisson solver for III–V heterostructures accounting for conduction band non-parabolicity. Solid State Electron 2010;54(11):1257–62.
[4] Adachi S. Properties of Group-IV, III–V and II-VI semiconductors. Wiley Ser Mater Electron Optoelectron Appl. Wiley; 2005.
[5] Klimeck G, McLennan M, Brophy SP, Adams III GB, Lundstrom MS. nanohub.org: Advancing education and research in nanotechnology. Comput Sci Eng 2008;10(5):17–23.
[6] Sentaurus Device, Synopsys inc. <http://www.synopsys.com>.
[7] Atomistix ToolKit. QuantumWise A/S. <http://www.quantumwise.com>.
[8] Caruso E, Zerveas G, Baccarani G, Czornomaz L, Daix N, Esseni D, et al. Modeling approaches for band-structure calculation in III–V FET quantum wells. In: Proc. EuroSOI-ULIS; 2015. p. 101–04.
[9] Tran F, Blaha P. Accurate band gaps of semiconductors and insulators with a semilocal exchange-correlation potential. Phys Rev Lett 2009;102:226401.
[10] Li J, Wang L-W. Band-structure-corrected local density approximation study of semiconductor quantum dots and wires. Phys Rev B 2005;72:125325.
[11] Atomistix ToolKit Manual. QuantumWise A/S. <http://www.quantumwise.com/documents/manuals/latest/ReferenceManual/index.html/>.
[12] Brandbyge M, Mozos J-L, Ordejón P, Taylor J, Stokbro K. Density-functional method for nonequilibrium electron transport. Phys Rev B 2002;65:165401.
[13] Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, et al. The SIESTA method for ab initio order-N materials simulation. J Phys: Condens Matter 2002;14(11):2745.
[14] Luisier M, Schenk A, Fichtner W, Klimeck G. Atomistic simulation of nanowires in the $sp^3d^5s^*$ tight-binding formalism: from boundary conditions to strain calculations. Phys Rev B 2006;74:205323.
[15] Boykin TB, Klimeck G, Bowen RC, Oyafuso F. Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory. Phys Rev B 2002;66:125207.
[16] Luisier M, Klimeck G. Investigation of $In_xGa_{1-x}As$ Ultra-thin-body tunneling FETs using a full-band and atomistic approach. In: Proc SISPAD; 2009. p. 1–4.
[17] Lee S, Oyafuso F, von Allmen P, Klimeck G. Boundary conditions for the electronic structure of finite-extent embedded semiconductor nanostructures. Phys Rev B 2004;69:045316.
[18] Bahder TB. Eight-band $\mathbf{k}\cdot\mathbf{p}$ model of strained zinc-blende crystals. Phys Rev B 1990;41:11992–2001.
[19] Vurgaftman I, Meyer JR, Ram-Mohan LR. Band parameters for III–V compound semiconductors and their alloys. J Appl Phys 2001;89(11):5815–75.
[20] Foreman BA. Elimination of spurious solutions from eight-band $\mathbf{k}\cdot\mathbf{p}$ theory. Phys Rev B 1997;56:R12748–51.
[21] Lizzit D, Esseni D, Palestri P, Osgnach P, Selmi L. Performance benchmarking and effective channel length for nanoscale InAs, $In_{0.53}Ga_{0.47}As$, and sSi n-MOSFETs. IEEE Trans Electron Dev 2014;61(6):2027–34.
[22] Jin S, Fischetti MV, 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(8):83715.
[23] Pal H, Low T, Lundstrom M. NEGF analysis of InGaAs Schottky barrier double gate MOSFETs. In: IEEE IEDM technical digest; 2008. p. 1–4.
[24] van der Steen J-L, Esseni D, Palestri P, Selmi L, Hueting R. Validity of the parabolic effective mass approximation in silicon and Germanium n-MOSFETs with different crystal orientations. IEEE Trans Electron Dev 2007;54(8):1843–51.
[25] Tamir T, Griffel G, Bertoni H. Guided-wave optoelectronics. Springer; 2013.
[26] Czornomaz L, Daix N, Kerber P, Lister K, Caimi D, Rossel C, et al., Scalability of ultra-thin-body and BOX InGaAs MOSFETs on silicon. In: Proc European solid state device res conf; 2013. p. 143–6.
[27] Herzinger C, Johs B. Dielectric function parametric model, and method of use, US Patent 5796983 A; 1998.
[28] Baravelli E, Gnani E, Grassi R, Gnudi A, Reggiani S, Baccarani G. Optimization of n- and p-type TFETs integrated on the same $InAs/Al_xGa_{1-x}Sb$ technology platform. IEEE Trans Electron Dev 2014;61(1):178–85.
[29] Ferry D, Akis R, Vasileska D. Quantum effects in mosfets: use of an effective potential in 3d monte carlo simulation of ultra-short channel devices. In: IEEE IEDM technical digest; 2000. p. 287–90.
[30] Palestri P, Lucci L, Tos SD, Esseni D, Selmi L. An improved empirical approach to introduce quantization effects in the transport direction in multi-subband Monte Carlo simulations. Semicond Sci Technol 2010;25(5):055011.
[31] Palestri P, Esseni D, Eminente S, Fiegna C, Sangiorgi E, Selmi L. Understanding quasi-ballistic transport in nano-mosfets: Part I – scattering in the channel and in the drain. IEEE Trans Electron Dev 2005;52(12):2727–35.
[32] Nier O, Rideau D, Niquet Y, Monsieur F, Nguyen V, Triozon F, et al. Multi-scale strategy for high-k/metal-gate UTBB-FDSOI devices modeling with emphasis on back bias impact on mobility. J Comput Electron 2013;12(4):675–84.