

The 3D nanometer device project *nextnano*: Concepts, methods, results

Alex Trellakis · Tobias Zibold · Till Andlauer ·
Stefan Birner · R. Kent Smith · Richard Morschl ·
Peter Vogl

Published online: 9 December 2006
© Springer Science + Business Media, LLC 2007

Abstract *nextnano* is a simulation tool that aims at providing global insight into the basic physical properties of realistic three-dimensional mesoscopic semiconductor structures. It focuses on quantum mechanical properties such as the global electronic structure, optical properties, and the effects of electric and magnetic fields for virtually any geometry and combination of semiconducting materials. For the calculation of carrier dynamics, two models are currently implemented that provide results for the limiting cases of highly diffusive or purely ballistic quantum-mechanical transport. In this paper, we present an overview of *nextnano*'s present and future capabilities and discuss some key concepts in the areas of code structure, numerical techniques, and electronic structure principles.

Keywords Semiconductor devices · TCAD · Quantum · Electronic structure · Carrier transport

1 Introduction

nextnano is a large software package that aims at providing insight into a wide range of physical properties of realistic semiconductor nanostructures of virtually any shape and composition [1]. This package is not the first of its kind. For instance, Kumar, Laux, and Stern have developed a three-dimensional Schrödinger-Poisson solver called *Scraps* many years ago [2, 3] and many one-dimensional and some two-dimensional Schrödinger-Poisson solvers are available

for free or commercially on the internet [4, 5]. In addition, there are many commercial semiconductor simulation tools (TCAD) available. But their primary focus lies understandably on semiclassical carrier transport in silicon devices, and they often treat quantum mechanical effects as perturbations [6–8].

nextnano has the opposite goal by focusing on nanostructures where quantum mechanical effects such as electronic band structures, optical matrix elements, magnetic field effects, or tunneling effects play a vital role. Using an extensive materials database for Si/Ge and III-V materials including nitride semiconductors, their binary, ternary, and quaternary alloys, and both zincblende and wurtzite crystal structures, a wide variety of quantum structures with almost arbitrary device geometries can be modeled.

Since most nanostructures are strained materials, *nextnano* performs a global strain minimization based on macroscopic elasticity theory. Once the strain field has been calculated, the piezoelectric and (for wurtzite) pyroelectric charges are determined. Then the electronic structure is calculated within a multiband $k \cdot p$ envelope function approximation. Here, exchange-correlation effects are taken into account through the local density approximation and the Hartree electron-electron interaction by self-consistently solving the Poisson equation. In addition, we also include the spin-orbit interaction and the Zeeman term in order to deal with magnetic field and spin effects. Optical interband and intraband transition matrix elements between electron and hole wave functions can be calculated as well. Finally, we have also extended *nextnano* to include self-consistently semiconductor-electrolyte interactions in order to enable the simulation of ion sensitive field effect transistors (ISFETs).

All of the above features deal with equilibrium systems. Therefore, in order to enable the simulation of nanodevice properties under applied bias, *nextnano* also includes two

A. Trellakis · T. Zibold · T. Andlauer · S. Birner · R. K. Smith ·
R. Morschl · P. Vogl (✉)
Walter Schottky Institute, TU München, Am Coulombwall 3,
D-85748 Garching, Germany
e-mail: vogl@wsi.tum.de

models for the carrier transport. One is a WKB-type approach that we termed the quantum-drift-diffusion method, where we assume that the carriers are *locally* in equilibrium, characterized by a local Fermi level. The other approach, termed the contact block reduction method [9, 10], is based on an efficient method to compute the ballistic current through an arbitrarily shaped open system with any number of leads.

2 Program flow and code structure

The program flow of *nextnano* can be summarized as follows. The user specifies the input using a text file that defines the geometry and the materials of the nanostructure, the contact bias, and all other information needed to describe the physical system under consideration. The syntax of this input file is concise and allows even very complex structures to be characterized by a few lines. Then, in the initial phase, the program evaluates the bulk band structures of all constituent materials, performs a global strain calculation, and determines the new band edges and piezoelectric charges. Subsequently, the Poisson equation, current equation, and many-band Schrödinger equations are solved self-consistently. Finally, in a post-processing step optical matrix elements and ballistic currents can be computed.

Originally, *nextnano* was written in Fortran 90. But as the code (termed *nextnano*³) grew organically instead of in a planned fashion, it eventually consisted of 250000 lines of very difficult to maintain code and a complete rewrite became necessary. Therefore, we are currently developing a new version called *nextnano++* that is written mostly in C++ except for core numerical libraries and that uses modern object-oriented programming methods as classes and generic programming in order to avoid code duplication. Furthermore, all input files use a powerful hierarchical syntax that can easily describe even complicated device geometries. These input files are then parsed using a *Bison* generated parser module and validated for errors using an approach similar to the one used for validating XML files.

3 Numerical techniques

The numerical methods used in *nextnano* are dictated by the simple fact that all equations we need to solve are partial differential equations (PDEs) in position space. Here we employ box integration finite differences for discretization. Compared to standard finite differences, this method has the advantage that discontinuities in a material parameter $M(\mathbf{x})$ are naturally taken into account for second order differential operators $\hat{\partial}_i M(\hat{\mathbf{x}}) \hat{\partial}_j$, since this method is in first order flux conservative due to the Gauss theorem. The grid itself is always assumed to be a nonuniform tensor product grid. Such

tensor grids have the advantage that no complicate meshing algorithm is required and the discretization can be easily and very efficiently implemented.

After discretization, all differential operators in the PDEs have now become banded sparse matrices which due to the underlying tensor grid have either diagonal or block diagonal sparsity structure. Here, every PDE that is a boundary value problem (such as the Poisson, the drift-diffusion, or the strain equations) becomes a large system of linear or nonlinear equations after discretization, while the k - p Schrödinger equation yields a large matrix eigenvalue problem with discretized wave functions.

The resulting linear systems are relatively easy to solve using iterative methods as the preconditioned conjugate gradient method, the preconditioned composite step conjugate gradient method, or (in case of non-Hermitian matrices) the preconditioned composite step biconjugate gradient method [11, 12]. As preconditioner we usually employ an incomplete Cholesky factorization, or, for Poisson operators, the very fast fill-in-free Dupont-Kendall-Rachford method [13]. If the original system is nonlinear in the solution vector, we use standard algorithms such as the Newton method with inexact line search to remap these problems into a sequence of linear solutions.

Solving the very large matrix eigenvalue systems resulting from the discretization of the k - p equations is computationally much more demanding. However, since for the $k \cdot p$ equations the occupation of electron and hole states falls off exponentially with the energy distance from the Fermi level, only relatively few quantum states near the Fermi level are needed to calculate the carrier densities. This allows us to use highly efficient iterative eigenvalue methods such as the Lanczos or Arnoldi iteration [12].

Since developing a fast and robust eigenvalue package is a very challenging and time consuming task, we usually rely on already available software libraries and customize these for our needs. For example, we find ARPACK [14] to be a very reliable and moderately fast software tool for finding the lowest energy eigenstates of sparse real and complex Hermitian matrices. However, ARPACK is still not fast enough for the huge system matrices that occur in three-dimensional quantum simulations and an additional preconditioner is needed.

Such a preconditioner can be easily obtained by utilizing the Chebyshev polynomials $T_n(E)$ for a spectral transformation. Here we use that if the Hermitian matrix \hat{H} has the spectrum $\{E_n\}$, the matrix $T(\hat{H})$ must have the spectrum $\{T(E_n)\}$. Since the polynomials $T_n(E)$ increase very rapidly for $|E| > 1$, we can use these $T_n(E)$ to make the desired eigenvalues at the lower end of the spectrum very large, which accelerates the convergence of the Arnoldi iteration in ARPACK by at least of factor of 10. For example, finding the lowest 30 eigenvalues of a test one-band problem with 10^5 grid nodes can now be achieved in less than 15 seconds

on a standard 3 GHz Pentium 4 PC. This result shows that memory exhaustion now surpasses excessive CPU time as the main concern for quantum simulations.

Finally, in order to solve the coupled Poisson-Schrödinger system, we employ an approximate quantum charge density inside of Poisson’s equation in order to estimate the dependence of the density on the potential through Schrödinger’s equation. Using this estimator the coupling between both equations is much decreased and rapid convergence is achieved [15]. Recently, we have refined this technique further by projecting the Hamiltonian into the subspace spanned by all already known eigenvectors, and then diagonalizing this small subspace matrix. With this method yet again only half as much work is needed to obtain the correct solution.

4 Electronic structure principles and techniques

nextnano relies on the $k \cdot p$ envelope function approximation to compute the global electronic structure of the nanostructure. The basic idea of this method is to patch up the bulk $k \cdot p$ Hamiltonian of each constituent material such that the global Hamiltonian remains Hermitian. While this method is inferior to microscopic electronic structure methods such as pseudopotential or empirical tight-binding schemes on an atomic scale, it is the method of choice for structures that extend over many tens or hundreds of nanometers and reflects accurately the extended electronic states in a mesoscopic device.

Even though the envelope function approach is well established, it is less widely known that this method is plagued by several ambiguities and instabilities that can lead to unphysical ghost states [16, 17], artifactual oscillatory wave functions, and incorrect bound states at interfaces [18]. We have been able to eliminate these problems by a careful treatment of so-called far-band contributions to the bulk $k \cdot p$ Hamiltonians, by defining an operator ordering that leads to manifestly self-adjoint Hamiltonians, and by employing an upwinding scheme for discretizing derivatives. Since a detailed discussion of all of these effects goes far beyond this short review, we will only briefly sketch the rationale behind the upwinding scheme as a typical example.

In the envelope function approach, every bulk $k \cdot p$ Hamiltonian $\hat{H}_{\text{bulk}}(\mathbf{k})$ is Fourier transformed into real space and replaced by a global Hamiltonian $\hat{H}_{\text{global}}(-i\nabla)$ for the envelope functions. However, when \hat{H}_{global} is mapped onto a discrete real-space lattice, the discretization of the first order space derivatives is not unique. For example, one could use centered differences

$$\frac{d}{dx}F(x) \rightarrow \frac{F(x+h) - F(x-h)}{2h}, \tag{1}$$

which unfortunately lead to serious problems as unphysical solutions $F(x)$ that rapidly oscillate between nearest grid lines. And plain forward or backward differencing results into a non-Hermitian Hamilton matrix. In order to eliminate this problem, we have employed a combination of forward and backward differencing that guarantees the hermiticity of the Hamiltonian, i.e.

$$\frac{d}{dx}F(x) \rightarrow \pm \frac{F(x \pm h) - F(x)}{h}, \tag{2}$$

$$H_{\text{global}}(-i\nabla) = \begin{bmatrix} & \text{forward} \\ \text{backward} & \end{bmatrix}. \tag{3}$$

With this discretization a second-order derivative term is added to each first order derivative that vanishes in the limit of zero grid spacing $h \rightarrow 0$. This second order term now suppresses the high-frequency components which constitute the spurious solutions, while the physically accurate low-frequency components remain unchanged. This approach is analogous to upwinding schemes that are used for hyperbolic differential equations in hydrodynamics [19].

5 Inclusion of magnetic fields

An important feature of *nextnano* is its ability to solve the relativistic multi-band Schrödinger equation in an arbitrarily oriented magnetic field. Somewhat unexpectedly, this is a very demanding and complex problem. It turns out that the minimal coupling Hamiltonian

$$\hat{H} = \frac{1}{2m}[\hat{\mathbf{p}} + e\hat{\mathbf{A}}(\mathbf{x})]^2 + V(\hat{\mathbf{x}}) \tag{4}$$

is unsuitable for spatial discretization [20]. This is a consequence of the fact that, even for a spatially constant magnetic field, the vector potential $\mathbf{A}(\mathbf{x})$ must increase at least linearly with $|\mathbf{x}|$. But since any chosen $\mathbf{A}(\mathbf{x})$ will be large on some boundaries and small on others, this leads inevitably to a pronounced gauge dependence and therefore large errors of the resulting eigenvalues and eigenfunctions.

We have developed a discretization of the multiband envelope function Hamiltonian that reestablishes gauge invariance by replacing the finite difference equation resulting from (2) by the discrete version of the covariant derivative D_i used in lattice gauge theory [21]. This derivative is defined by

$$D_i \psi(\mathbf{x}) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} [U(\mathbf{x} + \epsilon \hat{\mathbf{e}}_i, \mathbf{x}) \psi(\mathbf{x} + \epsilon \hat{\mathbf{e}}_i) - \psi(\mathbf{x})]$$

$$U(\mathbf{x} + \epsilon \hat{\mathbf{e}}_i, \mathbf{x}) = \exp \left[i \frac{e}{\hbar} \int_{\mathbf{x}}^{\mathbf{x} + \epsilon \hat{\mathbf{e}}_i} \mathbf{A}(\mathbf{x}') d\mathbf{x}' \right], \tag{5}$$

and the corresponding Hamiltonian now reads as

$$\hat{H} = \frac{\hat{\mathbf{D}}^2}{2m} + V(\hat{\mathbf{x}}). \quad (6)$$

It is easy to verify that this Hamiltonian is equivalent to (4) in the limit $\varepsilon \rightarrow 0$. Since now the vector potential only appears in a phase factor U , the matrix elements cannot grow indefinitely anymore for large \mathbf{x} and gauge invariance can be manifestly ensured.

6 Carrier transport

We now turn to the calculation of carrier transport in *nextnano*. As already mentioned, there are currently two models available for studying the electronic structure and carrier transport under an applied bias, namely the quantum-drift-diffusion method [22, 23] and the contact block reduction method [9, 10, 24]. The latter is based on the solution of the Schrödinger equation for an open quantum mechanical system and does not include any scattering or decoherence.

We will focus here on the first method which uses the first moment of the Boltzmann equation to determine the current and quantum mechanics to calculate the carrier density. In many nanostructures, the current $\mathbf{j}(\mathbf{x})$ is limited by several heterobarriers and therefore very small. In such a situation, it is reasonable to assume that the carriers are close to thermodynamic equilibrium. For a unipolar device and in the absence of recombination or generation, the basic equations read then

$$n(\mathbf{x}) = \sum_i |\psi_i(\mathbf{x})|^2 f\left(\frac{E_F(\mathbf{x}) - E_i}{k_B T}\right)$$

$$\mathbf{j}(\mathbf{x}) = \mu n(\mathbf{x}) \nabla E_F(\mathbf{x}), \quad \nabla \cdot \mathbf{j}(\mathbf{x}) = 0, \quad (7)$$

where $n(\mathbf{x})$ is the carrier density, ψ_i and E_i the eigenstates from solving the Schrödinger equation, $E_F(\mathbf{x})$ the quasi-Fermi level, and μ the mobility. Note that the quantum states E_i get occupied nonuniformly in this model which is only meaningful if the potential varies slowly and the total current through the device is small.

Recently, we have performed a detailed assessment of this model by comparing it with fully self-consistent nonequilibrium Green's function calculations (NEGF) [25]. In Fig. 1, we compare *nextnano* and NEGF calculations of the conduction band profile for a 10 nm InGaAs quantum well structure for 0 V and 0.08 V bias, respectively. This quantum well supports two bound states that become resonances under sufficiently high bias. For small bias, we find that the NEGF results agree excellently with *nextnano* even though the balance between

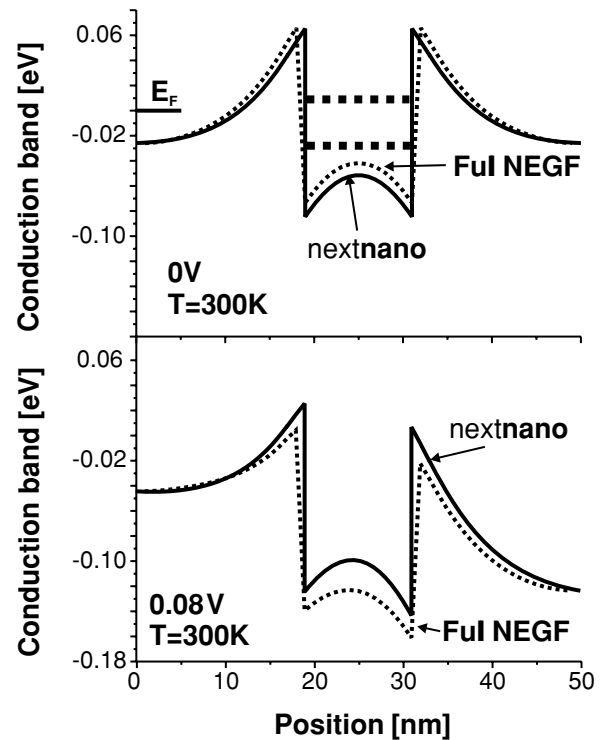


Fig. 1 Assessment of the quantum-drift-diffusion method by self-consistent non-equilibrium Green's function calculations

the carrier concentration outside and inside the quantum well is far from trivial.

7 Results and applications

Many groups around the world are using *nextnano* to model devices and quantum structures in the Si-SiGe and III-V material systems, as far as we can judge from the feedback of our users and the 10 to 15 downloads per day from our web site. Examples for such applications are for instance semiconductor based Mach-Zehnder interferometers [26], double gate and wrap-gate MOSFETs, three-dimensional III-V nanostructures fabricated by cleaved edge overgrowth, excitonic energies in quantum dot molecules and arrays under bias, and GaN based ISFETs for biosensing applications.

As an illustration, we show here two concrete applications. As first example, we have used the relativistic 8-band $k \cdot p$ model in *nextnano* to calculate the effective g -tensor of InAs/InP nanowire quantum dots as a function of the dot length (see Fig. 2). A comparison of our results with experimentally determined values [27] finds an excellent agreement without any fitting parameters; we find that the g -factors are close to the free electron values for the smallest dots and tend towards the bulk InAs values with increasing length of the nanowire dots.

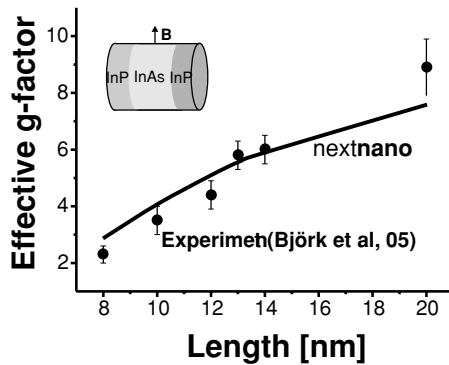


Fig. 2 Effective g -factor of InAs/InP nanowire-based quantum dots with different lengths for magnetic field perpendicular to the wire axis. The full line indicates *nextnano* results, the data points are experimental values from Ref. [27]. The diameter of the wire ranges from 50 to 70 nm

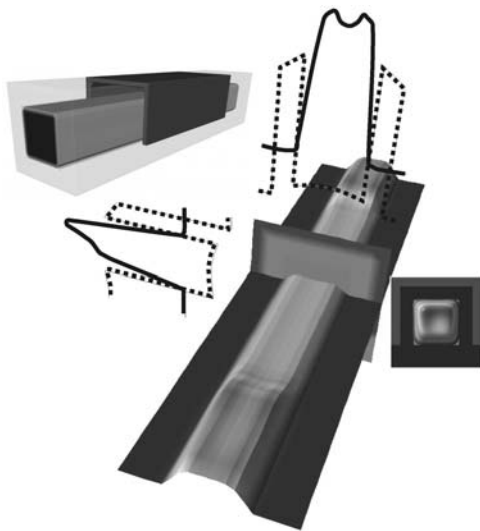


Fig. 3 Carrier density in a triple-gate MOSFET with a 5×5 nm channel and 25 nm gate length

Secondly, we show a simulation of a triple-gate MOSFET in Fig. 3. The Si wire forming the channel has a 5 by 5 nm cross section. The total length of the Si wire is 45 nm, the gate length is 25 nm, and the oxide thickness is 1.5 nm. The figure shows the charge density in the open channel at a gate voltage of 0.5 V but no applied source-drain voltage. For the center of the gate region, cross sections of the conduction band profile and the carrier density are shown as well. Note the asymmetry in the charge density that is caused by the gate extending over three sides of the Si wire.

Acknowledgment Financial support by the Deutsche Forschungsgemeinschaft (SFB 631), the Österreichische FWF (Projekt IR-ON F025),

and the Office of Naval Research under Contract No. N00014-01-1-0242 is acknowledged. Helpful comments and encouragement by S. Selberherr, H. Kosina, and F. Stern in the early stages of the project are gratefully acknowledged.

References

1. See *nextnano* website <http://www.wsi.tum.de/nextnano3> for executables and documentation
2. Kumar, A., Laux, S.E., Stern, F.: Phys. Rev. B **42**, 5166 (1990)
3. Laux, S.E.: In: Miller, J.J.H. (ed.) Proc. 5th Int. conf. on Numerical Analysis of Semiconductor Devices and Integrated Circuits (NASECODE V), p. 270. Boole, Dun Laoghaire, Ireland (1987)
4. nanoHUB portal for nanodevice simulation software <http://www.nanohub.org>
5. 1D Poisson by G. Snider available at <http://www.nd.edu/~gsnider>
6. SIMBA by W. Klix, R. Stenzel available at <http://www.htw-dresden.de/~klix/simba/welcome.html>
7. Atlas Device Simulation Framework by Silvaco Int., see <http://www.silvaco.com>
8. Palankovski, V., Selberherr, S.: J. Telecomm. Info. Tech. **1**, 15 (2004)
9. Mamaluy, D., Vasileska, D., Sabathil, M., Zibold, T., Vogl, P.: Phys. Rev. B **71**, 245321 (2005)
10. Mamaluy, D., Sabathil, M., Vogl, P.: J. Appl. Phys. **93**, 4628 (2003)
11. Bank, R.E., Chan, T.F.: Numerische Mathematik **66**, 295 (1993)
12. Bai, Z., Demmel, J., Dongarra, J., Ruhe, A., van der Vorst, H.: (eds.): Templates for the Solution of Algebraic Eigenvalue Problems: A Practical Guide. SIAM, Philadelphia (2000)
13. Dupont, T., Kendall, R.P., Rachford, H.H.: SIAM J. Numer. Anal. **5**, 559 (1968)
14. See <http://www.caam.rice.edu/software/ARPACK/> for obtaining the ARPACK libraries and related publications.
15. Trellakis, A., Galick, A.T., Pacelli, A., Ravaioli, U.: J. Appl. Phys. **81**, 7880 (1997)
16. Foreman, B.A.: Phys. Rev. B **48**, 4964 (1993)
17. Foreman, B.A.: Phys. Rev. B. **56**, R12748 (1997).
18. Rodina, A.V., Yu. Alekseev, A., Efros, A.L., Rosen, M., Meyer, B.K.: Phys. Rev. B **65**, 125302 (2002)
19. Bank, R.E., Bürgler, J.F., Fichtner, W., Smith, R.K.: Numer. Math. **58**, 185 (1990)
20. Governale, M., Ungarelli, C.: Phys. Rev. B **58**, 12 (1998)
21. Wilson, K.: Phys. Rev. D **10**, 2445 (1974)
22. Hackenbuchner, S.: Elektronische Struktur von Halbleiter-Nanobaulementen im thermodynamischen Nichtgleichgewicht. Ph.D. Thesis, TU München, München (2002)
23. Sabathil, M., Hackenbuchner, S., Majewski, J.A., Zandler, G., Vogl, P.: J. Comp. Electro. **1**, 81 (2002)
24. Sabathil, M.: Opto-electronic and quantum transport properties of semiconductor nanostructures. Ph.D. Thesis, TU München, München (2005)
25. See article by T. Kubis and P. Vogl in this volume.
26. Sabathil, M., Mamaluy, D., Vogl, P.: Semicond. Sci. Tech. **19**, S137 (2004)
27. Björk, M.T., Fuhrer, A., Hansen, A.E., Larsson, M.W., Fröberg, L.E., Samuelson, L.: Phys. Rev. B **72**, 201307(R) (2005)