An Accurate k*p Approximation of the Empirical Pseudopotential Hamiltonian for Confined States in Silicon Double Gate MOSFETs

Max Renner
Chair of Electromagnetic Theory
RWTH Aachen University
Aachen, Germany
0000-0002-5412-4936

Tobias Linn
Chair of Electromagnetic Theory
RWTH Aachen University
Aachen, Germany
0000-0002-7681-7725

Christoph Jungemann
Chair of Electromagnetic Theory
RWTH Aachen University
Aachen, Germany
0000-0002-3423-4046

Abstract—**The top part of the valence band structure is much more complex than the bottom part of the conduction bands and the valence bands can not be approximated by simple methods (e.g. effective mass approximation). Previously, it has been shown how to apply the local empirical pseudopotential method to confined states in MOS structures. A problem of this approach is the long CPU time owing to the large dimension of the Hamiltonian. With the k*p approximation the CPU time can be reduced by orders of magnitude with negligible error, in particular for the valence bands. Since the electrostatic potential can be included in the k*p approximation, the eigenvalue problem for the full pseudopotential Hamiltonian has to be solved only once for a given MOS structure regardless of the gate bias or temperature.**

Index Terms—**MOS devices, semiconductor device modeling, schrodinger equation**

I. Introduction

The valence band structure of silicon is much more complex near its maximum than the conduction bands near their minima and can not be approximated by an effective mass [1]. The local empirical pseudopotential method (EPM) yields a good approximation not only near the Γ-point but in the complete Brillouin zone (BZ) [2]. The EPM has been extended to the case of confined states in MOSFETs [3]. The CPU time for solving the corresponding eigenvalue problem is rather high due to the large size of the Hamiltonian. Here, we propose a k*p method for confined states by which the CPU time can be reduced by orders of magnitude.

II. The Pseudopotential Hamiltonian

In a perfect infinite crystal the lattice potential is periodic and can be approximated by a pseudopotential [4]. The periodic lattice potential $V_{\text{lat}}(\vec{r})$ is expanded into a Fourier sum

$$V_{\text{lat}}(\vec{r}) = \sum_l V_{\text{s}}\left(|\vec{G}_l|\right) \cos\left(\vec{G}_l \cdot \vec{\tau}\right) \exp\left(\text{i}\vec{G}_l \cdot \vec{r}\right) , \quad (1)$$

TABLE I
PSEUDOPOTENTIAL FORM FACTORS FOR $|\vec{G}_l|$ IN UNITS OF $2\pi/a_0$.

	$V_{\text{s},\sqrt{0}}$ [eV]	$V_{\text{s},\sqrt{3}}$ [eV]	$V_{\text{s},\sqrt{8}}$ [eV]	$V_{\text{s},\sqrt{11}}$ [eV]
Silicon	-10.2189	-3.0490	0.7497	0.9850
Oxide	-5.4189	-6.4030	1.5743	2.0686

where $V_{\text{s}}(|\vec{G}_l|)$ is a symmetric pseudopotential form factor, $\vec{\tau} = (\vec{e}_x + \vec{e}_y + \vec{e}_z)a_0/8$ (the positions of the two silicon atoms in the elementary cell are $\pm\vec{\tau}$), $a_0 = 0.54307\,\text{nm}$ the lattice constant of silicon, and $\vec{G}_l$ the lth vector of the reciprocal space of the diamond lattice

$$\vec{G}_l = \frac{2\pi}{a_0} \begin{pmatrix} -i_l + j_l + k_l \\ +i_l - j_l + k_l \\ +i_l + j_l - k_l \end{pmatrix} . \quad (2)$$

The indices i_l, j_l, k_l can take any integer value and l is a linear index for the index triple. In the Fourier sum only four nonzero form factors are taken into account which are given in Tab. I. The form factors $V_{\text{s},\sqrt{3}}$, $V_{\text{s},\sqrt{8}}$ and $V_{\text{s},\sqrt{11}}$ for silicon are taken from [5]. The form factor $V_{\text{s},\sqrt{0}}$ is introduced to shift the maximum of the valence bands to zero energy for bulk silicon.

For the MOS structure a pseudopotential for the silicon dioxide is required. This pseudo oxide is based on the same lattice as silicon and its pseudopotential form factors are chosen in such a way that the conduction and valence band barriers between silicon and oxide match the experimental values [3]. In Fig. 1 the double gate PMOS structure is shown, which is assumed to be periodic and the (001) surface orientation is chosen. The lattice potential for this structure is assumed to change abruptly when the material changes [3]. The position of the interface w.r.t. the atoms has a strong impact on the 2D subband structure and its symmetry. The left interface is located at $z = 2a_0$ and thus falls in between the two atoms of the elementary cell centered at $(0, 0, 2a_0)$. This configuration has only a
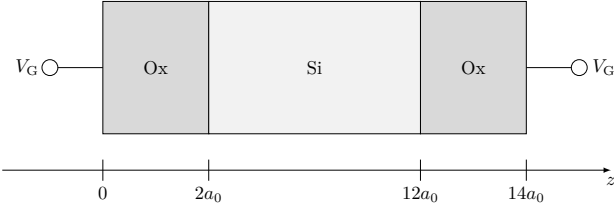
Fig. 1. Double gate PMOS structure.

fourfold symmetry in the 2D BZ. In order to obtain the expected eightfold symmetry the lattice is shifted by $\vec{\tau}$ and one atom of the elementary cell falls now onto the interface between silicon and oxide. The corresponding lattice potential for the PMOS structure is given by

$$V_{\text{lat}}^{\text{MOS}}(\vec{r}) = \begin{cases} V_{\text{lat}}^{\text{si}}(\vec{r} + \vec{\tau}) & \text{for } 2a_0 \leq z \leq 12a_0 \\ V_{\text{lat}}^{\text{ox}}(\vec{r} + \vec{\tau}) & \text{for } z < 2a_0 \text{ or } z > 12a_0 \end{cases} \quad (3)$$

with a total length in z-direction of $L_z = 14a_0$ (Fig. 1).

The Schrödinger equation (SE) is given by [3]

$$\underbrace{\left[-\frac{\hbar^2}{2m_0} \nabla^2 + V_{\text{lat}}^{\text{MOS}}(\vec{r}) - e\varphi(z) \right]}_{=H(\vec{r})} \Psi_{\eta,\vec{k}}(\vec{r}) = \varepsilon_{\eta,\vec{k}} \Psi_{\eta,\vec{k}}(\vec{r}) ,$$

$$(4)$$

where $\hbar$ is the reduced Planck constant, m_0 the free electron rest mass, e the elementary charge, $\Psi_{\eta,\vec{k}}(\vec{r})$ the wave function of the ηth eigen state, $\varepsilon_{\eta,\vec{k}}$ the corresponding eigen energy for the wavevector $\vec{k}$, and $\varphi(z)$ the electrostatic potential.

The lattice periodic part of the Bloch states is expanded in a Fourier series resulting in a wave function of the form

$$\Psi_{\eta,\vec{k}}(\vec{r}) = \frac{1}{\sqrt{a_0^2 L_z}} \sum_\lambda a_{\eta,\vec{k};\lambda} \exp\left(\mathrm{i} \left(\vec{k} + \vec{\Gamma}_\lambda \right) \cdot \vec{r} \right) \quad (5)$$

with

$$\vec{\Gamma}_\lambda = \left[(i_\lambda + j_\lambda) \vec{e}_x + (i_\lambda - j_\lambda) \vec{e}_y \right] \frac{2\pi}{a_0} + \frac{\kappa_\lambda 2\pi}{L_z} \vec{e}_z . \quad (6)$$

The indices i_λ, j_λ and κ_λ can take any integer value and λ is again a single index. The prefactor is set to fulfill the normalization condition on the device volume $a_0 \times a_0 \times L_z$. The x- and y-components of $\vec{\Gamma}_\lambda$ form the lattice of the 2D reciprocal space [3]. Since the confined subbands are essentially flat w.r.t. k_z due to the thick oxide, we always use $k_z = 0$. In Fig. 2 the 2D BZ and the irreducible wedge are shown. The eight symmetry operations of the energy are four reflections and one permutation: $\varepsilon(\pm k_x, \pm k_y) = \varepsilon(\pm k_y, \pm k_x)$.

The Hamiltonian of the SE is projected onto the plane-wave basis

$$\left[\hat{H}^{\text{EPM}} \right]_{\lambda,\lambda'} = \frac{1}{a_0^2 L_z} \int_0^{L_z} \int_0^{a_0} \int_0^{a_0} \exp\left(-\mathrm{i} \left(\vec{k} + \vec{\Gamma}_\lambda \right) \cdot \vec{r} \right)$$
$$H(\vec{r}) \exp\left(\mathrm{i} \left(\vec{k} + \vec{\Gamma}_{\lambda'} \right) \cdot \vec{r} \right) \mathrm{d}x \mathrm{d}y \mathrm{d}z . \quad (7)$$
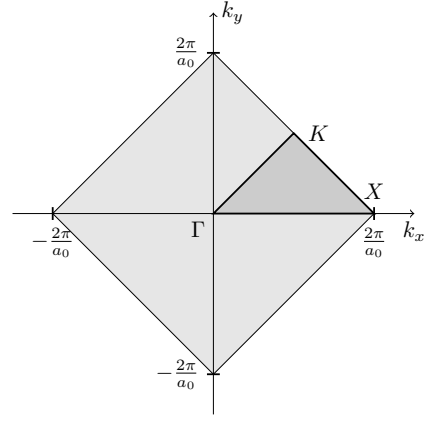

Fig. 2. 2D Brillouin zone for a (001) surface orientation. The dark grey area is the irreducible wedge.

The expansion is limited to a finite number of plane waves by the condition for the index λ

$$\left| \vec{\Gamma}_\lambda \right| < 5.7 \frac{2\pi}{a_0} , \quad (8)$$

which cuts off the expansion at very high energies similar to the EPM for the bulk case [2]. The cutoff sphere is not shifted with the wavevector to avoid discontinuities in the bands. Instead, a rather large value $(5.7 \times 2\pi/a_0)$ is used for the cutoff radius and k_x, k_y are restricted to the first BZ.

The space charge density in the silicon region due to holes and acceptors with a concentration $N_{\text{A}}(z) = 10^{18} \, \text{cm}^{-3}$ is averaged in the x- and y-directions

$$\rho(z) = \frac{e}{2\pi^2} \sum_\eta \iint_{\text{BZ}} \frac{1}{1 + \exp\left(\frac{\mu_{\text{F}} - \varepsilon_\eta(k_x, k_y)}{k_{\text{B}} T_0} \right)}$$
$$\int_0^{a_0} \int_0^{a_0} \left| \Psi_{\eta,k_x,k_y}(x,y,z) \right|^2 \mathrm{d}x \mathrm{d}y \, \mathrm{d}k_x \mathrm{d}k_y - eN_{\text{A}}(z) , \quad (9)$$

where μ_{F} is the Fermi energy and $k_{\text{B}} T_0$ the thermal energy at the lattice temperature. The sum runs over all valence subbands.

The electrostatic potential $\varphi(z)$ is the solution of the 1D Poisson equation (PE) [6]

$$\frac{\mathrm{d}}{\mathrm{d}z} \left(\epsilon(z) \frac{\mathrm{d}\varphi}{\mathrm{d}z} \right) = -\rho(z) . \quad (10)$$

The permittivity $\epsilon(z)$ changes abruptly when moving from silicon into oxide resulting in an interface condition that requires the continuity of the electric flux density in z-direction and of the potential. On the gates Dirichlet boundary conditions are applied: $\varphi(0) = \varphi(L_z) = V_{\text{G}}$. The PE is discretized on an equidistant grid with the finite volume method. The SE and PE are solved self-consistently for the double gate PMOS structure by a Gummel loop [6].

III. k*p Approximation

We project the Hamiltonian (7) onto its eigenvector basis at the Γ-point ($k_x = k_y = 0$) and $\varphi(z) = 0$

$$\hat{H}^\Gamma = \hat{a}_0^\dagger \hat{H}_0^{\text{EPM}} \hat{a}_0 \ , \tag{11}$$

where $\hat{a}_0$ is the matrix of which the column vectors are the eigenvectors at the Γ-point and zero potential. Nonzero values of the wave numbers k_x, k_y and potential are treated as perturbations resulting in a perturbation Hamiltonian

$$\hat{P} = \hat{P}_{k_x} k_x + \hat{P}_{k_y} k_y + \hat{P}_\varphi \ . \tag{12}$$

For example, the perturbation Hamiltonian $\hat{P}_{k_x}$ is obtained by

$$\hat{P}_{k_x} = \hat{a}_0^\dagger \frac{\partial \hat{H}^{\text{EPM}}}{\partial k_x}\Big|_{k_x = k_y = 0, \varphi = 0} \hat{a}_0 \ . \tag{13}$$

With the Löwdin perturbation theory we can derive the k*p approximation for the Hamiltonian [7]. To this end we split the eigen states into two sets. The set $\mathcal{A}$ contains all states with an eigen energy of less than 20 eV and the set $\mathcal{B}$ all other states. The k*p Hamiltonian is given with $\eta, \eta' \in \mathcal{A}$ by

$$\left[\hat{H}^{\text{k*p}}\right]_{\eta,\eta'} = \left(\varepsilon_{\eta,0} + \frac{\hbar^2}{2m_0}(k_x^2 + k_y^2)\right)\delta_{\eta,\eta'}$$

$$+ \left[\hat{P}\right]_{\eta,\eta'} + \sum_{\nu \in \mathcal{B}} \frac{\left[\hat{P}\right]_{\eta,\nu}\left[\hat{P}\right]_{\nu,\eta'}}{\varepsilon_{\text{ref}} - \varepsilon_{\nu,0}} \ , \tag{14}$$

where $\varepsilon_{\eta,0}$ are the eigen energies at the Γ-point and zero potential. The reference energy ε_{ref} is set to the value of the top valence subband at the Γ-point [1].

IV. Simulation Results

The dimension of the EPM Hamiltonian for the PMOS structure in Fig. 1 is 5311 and it takes 80 CPU seconds to solve the corresponding eigenvalue problem for a single k-vector on a single CPU core. The dimension of the k*p Hamiltonian is only 431 and the CPU time 63 milliseconds corresponding to a reduction by more than three orders of magnitude. In addition, the smaller size of the eigenvalue problem results in a more cache-efficient parallelization. This does not mean that the solution of the SE with the EPM Hamiltonian dominates the total CPU time, because it has to be solved for a given MOS structure only once regardless of the applied bias or temperature. If desired, the size of the k*p Hamiltonian can be even further reduced by shrinking the set $\mathcal{A}$, still obtaining a good approximation near the Γ-point.

The subband energies along the k_x-axis for a flat electrostatic potential are shown in Fig. 3 evaluated with the full EPM Hamiltonian and k*p. The calculation yields the usual minima for the conduction subbands: the central valleys and the satellite ones at $0.85 \cdot 2\pi/a_0$. Close to the X-point the deviation of the k*p approximation from EPM
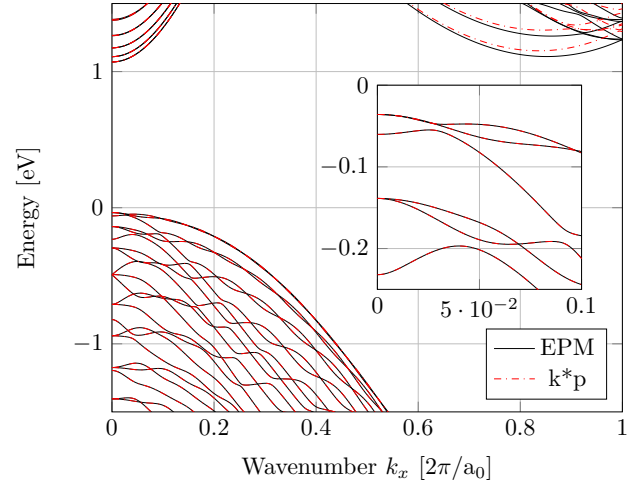

Fig. 3. Subband structure of the double gate PMOS structure for a flat electrostatic potential from the Γ- to the X-point based on the EPM and k*p Hamiltonians. In the inset the top six valence subbands are shown.

becomes larger, as expected for a Γ centric expansion. This does not matter in the case of holes, because the top of the valence subbands is centered at the Γ-point, where the error is much smaller. In the inset the valence subbands are shown close to the Γ-point and the threefold degeneracy of the local EPM at the Γ-point for bulk silicon is reduced to a twofold one because one subband splits off due to the reduced symmetry of the PMOS structure.

After 21 Gummel-type iterations of the SE and PE the change in the potential is less than $1 \cdot 10^{-7}$ V at $V_G = -1$ V, $\mu_F = 0$ and room temperature. In Fig. 4 (upper graph) the energy of the first valence subband is shown in the irreducible wedge of the first BZ. The black line indicates an energy of -0.3 eV below which the Fermi-Dirac distribution for holes can be neglected at room temperature and $\mu_F = 0$. In the lower graph the difference between the k*p and EPM calculations is shown. The line indicates an error of 1 meV and the error in energy for the top 300 meV of the first valence subband is much smaller than 1 meV. Thus, the k*p approximation is sufficiently accurate to calculate the valence subband structure.

The corresponding hole density averaged over the two lateral dimensions and the potential are shown in Fig. 5. The hole density oscillates on the scale of the elementary cell of silicon, but these oscillations have a minor impact on the potential due to the low-pass filter characteristics of the PE. As expected, the hole density is symmetric w.r.t. $z = 7a_0$.

The valence subbands are strongly nonparabolic and the mass is not constant. In Fig. 6 the averaged inverse mass
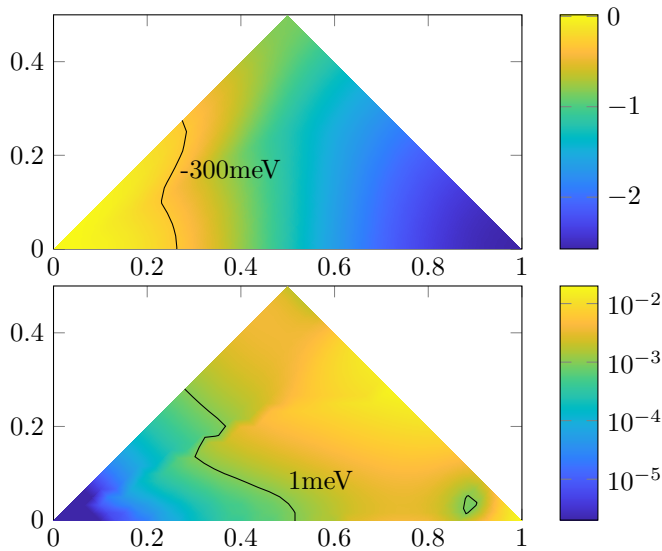
Fig. 4. Energy of the first valence subband in the irreducible wedge of the BZ at $V_G = -1\,\text{V}$ and room temperature (upper graph, linear scale). The axes are in units of $[2\pi/a_0]$. The absolute value of the difference of EPM and k*p (lower graph, log scale). The lines indicate an energy of $-300\,\text{meV}$ in the upper graph and an error of $1\,\text{meV}$ in the lower graph.
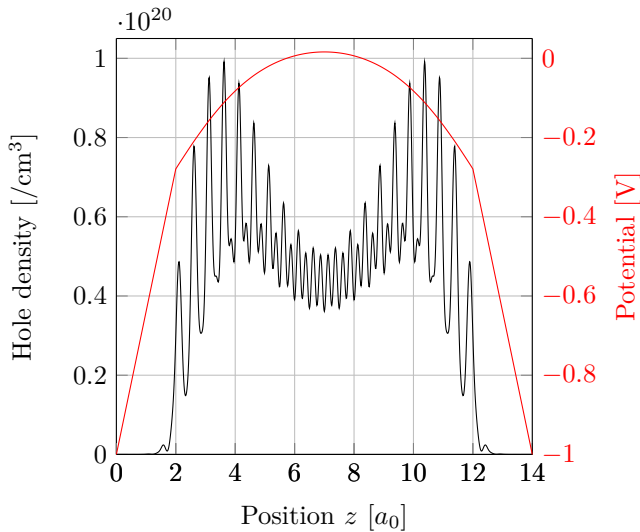

Fig. 5. Hole density and potential along a vertical cut of the double gate PMOSFET for $V_G = -1\,\text{V}$ at room temperature and $\mu_F = 0$.

of the holes

$$\frac{1}{m_{xx}} = -\frac{\sum_\eta \iint_{\text{BZ}} \frac{\frac{1}{\hbar^2}\frac{\partial^2 \varepsilon_\eta}{\partial k_x^2}\,\text{d}k_x\,\text{d}k_y}{1+\exp\left(\frac{\mu_F - \varepsilon_\eta(k_x - k_{x,\text{dis}}, k_y)}{k_B T_0}\right)}}{\sum_\eta \iint_{\text{BZ}} \frac{\text{d}k_x\,\text{d}k_y}{1+\exp\left(\frac{\mu_F - \varepsilon_\eta(k_x, k_y)}{k_B T_0}\right)}} \quad (15)$$

is shown together with the drift velocity in x-direction for a Fermi-Dirac distribution, which is displaced by a wavevector $k_{x,\text{dis}}$ in the x-direction. The signs of the mass and velocity are chosen according to the hole picture. The rapid drop of the inverse mass by a factor of 2 when the
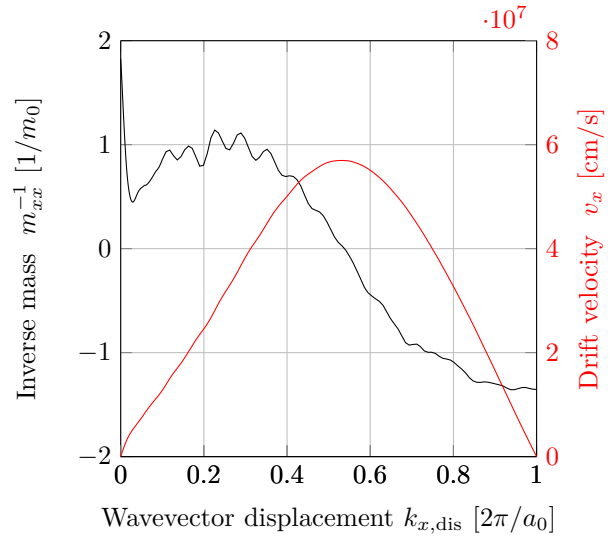

Fig. 6. Inverse mass and drift velocity of the holes evaluated with a displaced Fermi-Dirac distribution for the double gate PMOSFET at $V_G = -1\,\text{V}$, room temperature and $\mu_F = 0$.

displacement changes from 0 to $0.05 \times 2\pi/a_0$ demonstrates the strong nonparabolicity of the valence bands, which has a strong impact on the hole mobility, Hall factor etc. [8].

V. Conclusions

In conclusion, a k*p approximation for the EPM Hamiltonian of confined states in a MOS structure has been demonstrated by which the CPU time can be reduced by orders of magnitude without sacrificing accuracy. The approach is not limited to silicon and can be easily applied to other semiconductors by changing the EPM parameters.

References

[1] J. M. Luttinger and W. Kohn, "Motion of electrons and holes in perturbed periodic fields," *Phys. Rev.*, vol. 97, pp. 869–883, Feb 1955. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.97.869

[2] M. L. Cohen and T. K. Bergstresser, "Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures," *Phys. Rev.*, vol. 141, pp. 789–796, Jan 1966. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.141.789

[3] M. G. Pala and D. Esseni, "Quantum transport models based on NEGF and empirical pseudopotentials for accurate modeling of nanoscale electron devices," *Journal of Applied Physics*, vol. 126, no. 5, p. 055703, 2019. [Online]. Available: https://doi.org/10.1063/1.5109187

[4] M. L. Cohen and J. R. Chelikowsky, *Electronic Structure and Optical Properties of Semiconductors*, 2nd ed. New York: Springer, 1989.

[5] M. M. Rieger and P. Vogl, "Electronic-band parameters in strained Si$_{1-x}$Ge$_x$ alloys on Si$_{1-y}$Ge$_y$ substrates," *Phys. Rev. B*, vol. 48, pp. 14 276–14 287, 1993.

[6] S. Selberherr, *Analysis and Simulation of Semiconductor Devices*. Wien: Springer, 1984.

[7] P. O. Löwdin, "A note on the quantum-mechanical perturbation theory," *J. Chem. Phys.*, vol. 19, pp. 1396–1401, 1951.

[8] W. Brauer and H. W. Streitwolf, *Theoretische Grundlagen der Halbleiterphysik*, 2nd ed. Braunschweig: Vieweg, 1977.