POLITECNICO DI TORINO
Repository ISTITUZIONALE

Quantum mechanical model of crossing and anti-crossing points in 3D full-band Monte Carlo simulations

Original
Quantum mechanical model of crossing and anti-crossing points in 3D full-band Monte Carlo simulations / Zhu, Mike; Bertazzi, Francesco; Matsubara, Masahiko; Bellotti, Enrico. - In: JOURNAL OF APPLIED PHYSICS. - ISSN 0021-8979. - STAMPA. - 135:6(2024), pp. 1-9. [10.1063/5.0194536]

Availability:
This version is available at: 11583/2990525 since: 2024-07-09T08:00:33Z

Publisher:
AIP Publishing

Published
DOI:10.1063/5.0194536

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

08 August 2024

Journal of Applied Physics

RESEARCH ARTICLE | FEBRUARY 09 2024

Quantum mechanical model of crossing and anti-crossing points in 3D full-band Monte Carlo simulations

Mike Zhu ; Francesco Bertazzi ; Masahiko Matsubara ; Enrico Bellotti

J. Appl. Phys. 135, 065702 (2024)
https://doi.org/10.1063/5.0194536

09 July 2024 07:50:08

Journal of Applied Physics

ARTICLE

pubs.aip.org/aip/jap

Quantum mechanical model of crossing and anti-crossing points in 3D full-band Monte Carlo simulations

Cite as: J. Appl. Phys. **135**, 065702 (2024); doi: 10.1063/5.0194536
Submitted: 27 December 2023 · Accepted: 17 January 2024 ·
Published Online: 9 February 2024

View Online

Export Citation

CrossMark

Mike Zhu,[1,a)] Francesco Bertazzi,[2,b)] Masahiko Matsubara,[3,c)] and Enrico Bellotti[1,3,d)]

AFFILIATIONS

[1]Department of Electrical and Computer Engineering, Boston University, 8 Saint Mary's Street, Boston, Massachusetts 02215, USA
[2]Department of Electronics and Telecommunications, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy
[3]Division of Materials Science and Engineering, Boston University, 8 Saint Mary's Street, Boston, Massachusetts 02215, USA

[a)]**Author to whom correspondence should be addressed:** mjzhu@bu.edu
[b)]**Electronic mail:** francesco.bertazzi@polito.it
[c)]**Electronic mail:** matsubar@bu.edu
[d)]**Electronic mail:** bellotti@bu.edu

ABSTRACT

This work presents a 3D quantum mechanics based model to address the physics at band structure crossing/anti-crossing points in full band Monte Carlo (FBMC) simulations. The model solves the Krieger and Iafrate (KI) equations in real time using pre-computed coefficients at *k*-points spatially sampled within the first Brillouin zone. Solving the KI equations in real time makes this model applicable for all electric fields, which enables its use in FBMC device simulations. In this work, a two-level refinement scheme is used to aggressively sample regions in proximity to band crossings for accurate solutions to the KI equations and coarsely sample everywhere else to limit the number of *k*-points used. The presented sampling method is demonstrated on the band structure of silicon but is effective for the band structure of any semiconductor material. Next, the adaptation of the fully quantum KI model into the semi-classical FBMC method is discussed. Finally, FBMC simulations of hole transport in 4H silicon carbide with and without the KI model are performed. Results along different crystallographic directions for a wide range of electric fields are compared to previously published simulation and experimental values.

09 July 2024 07:50:08

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/5.0194536

I. INTRODUCTION

Monte Carlo is a powerful and established method for simulating carrier transport in semiconductor materials. Its incorporation of the energy band structure makes it superior over other conventional methods, such as drift-diffusion, for simulating high-field effects in semiconductor materials and devices. For this reason, full-band Monte Carlo (FBMC) has been extensively used to study non-equilibrium carrier transport phenomena, such as velocity-field characteristics and impact ionization coefficients, of technologically significant and emerging semiconductor materials. Furthermore, FBMC also relies upon to simulate the performance of semiconductor devices operating in far from equilibrium conditions, such as power devices, high-speed transistors operating in the large signal regime, and avalanche photodiodes.

In the standard FBMC paradigm, carrier transport is treated semi-classically, meaning carriers are represented as wavepackets that are completely localized to an energy band. When accelerated by an applied field, the average wavevector of a carrier is altered, but the carrier remains localized to the same band. Thus, the standard semi-classical FBMC is only rigorously valid for simulating carrier transport within a singular band. Despite not including quantum mechanical effects, the standard FBMC transport model has still demonstrated its effectiveness by accurately simulating carrier transport properties for a wide variety of conventional semiconductors, especially those with cubic crystal structures.

However, for wide and ultra-wide bandgap semiconductors with non-cubic crystal structures, which have many band crossings and anti-crossings at high symmetry points, the standard semi-classical

© Author(s) 2024

FBMC model is insufficient for simulating high-field transport. Accurate simulations in these materials necessitate the integration of additional models to account for quantum mechanical effects associated with the crossing and anti-crossing points.[1] Examples of these crossing points in silicon and 4H silicon carbide (4H-SiC) are shown in Fig. 1, depicting a carrier approaching the band crossing point and tunneling into the neighboring bands. This process may result in the carrier completely transitioning into a different band or existing in multiple bands simultaneously in a mixed state, with neither effect included in the standard FBMC transport model. Prior Monte Carlo works on several important, non-cubic materials[2–5] have demonstrated the importance of including these quantum mechanical effects for high-field simulations to obtain results more closely aligned with the experiment.

Previous works that incorporate simulated physics at band crossings have used methods such as imposing continuous carrier velocity in free flight,[6] determining the amplitude probabilities through an overlap test[3,7] or solving the Krieger and Iafrate (KI) equation.[2,4,5,7–11] Since the KI equation produces the most physically correct results, it is the focus of this work. Prior FBMC works that include KI[4,5,7,10] were performed exclusively in 1D along a specific crystallographic direction or for a confined system such as a nanowire. However, for accurate 3D device simulations of non-cubic materials, a 3D method to solve the KI equations for any arbitrary electric field strength and direction is required. Furthermore, a 3D KI model would also simplify the process of extracting carrier transport properties along any arbitrary crystallographic direction. Unfortunately, adapting the currently available 1D methods to 3D would require an impractical amount of memory usage, which is explained in Secs II and III.

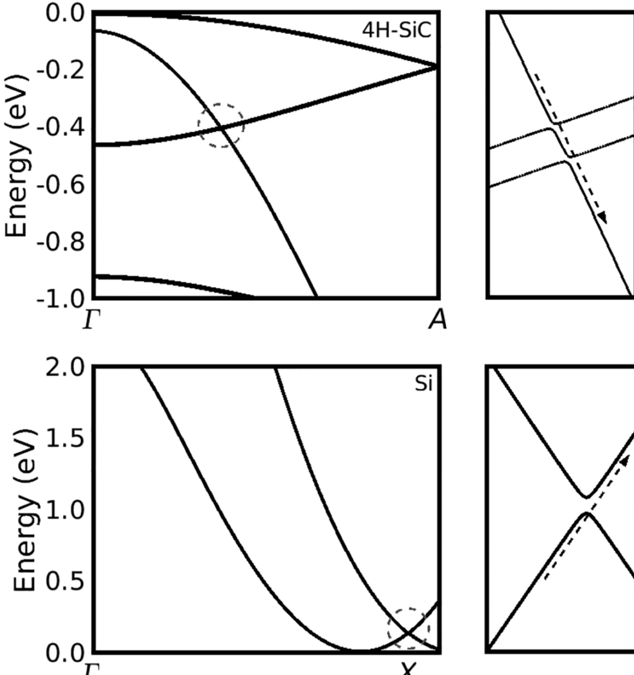


FIG. 1. Band crossings of the valence bands of 4H-SiC (top) and the conduction bands of Si (bottom) are plotted along the $\Gamma \rightarrow A$ and $\Gamma \rightarrow X$ axes, respectively. The close-ups of the circled crossings are shown in the right panels. The dotted arrows indicate a realistic transition for carriers through the crossing point.

In this work, we present a general method to solve the KI equations in 3D for any electric fields within our FBMC framework.[12] The 3D method incorporates an adaptive refinement k-point sampling technique that is used to produce accurate solutions of the KI equations while limiting the memory storage requirements. The effectiveness of the method is demonstrated on anti-crossing points in silicon, which have been the focus of previous studies.[9,10] The method is then integrated into our FBMC simulator, where the KI equations are solved along a carrier's drift path in reciprocal space, and also includes the development of mixed-state carrier drift and scatter models, while maintaining as much of the conventional FBMC paradigm as possible. The KI-integrated FBMC simulator is used to simulate the velocity-field characteristics and impact ionization coefficients for 4H-SiC along different crystallographic directions. The results are validated against available experimental data and also compared with previously computed FBMC results. Since the results from the previous FBMC simulations were primarily computed using a different and less physically accurate model for band crossings, with only a few datapoints computed using a KI model, the results of this work can also be considered an updated calculation of hole velocity-field curves and impact ionization coefficients of 4H-SiC.

All k-vector values given in this work are normalized by $2\pi/a$, where a is the lattice constant of the material.

09 July 2024 07:50:08

II. METHODOLOGY

A. Mathematical model

The most widely used equation for determining the time-dependent band amplitude probability coefficients, $C_n(t)$, of charge carriers traversing through band crossings was originally derived by Krieger and Iafrate (KI)[8] and is presented below:

$$i\hbar \frac{\partial C_n(t)}{\partial t} = E_n(\boldsymbol{k}(t))C_n(t) + q\boldsymbol{F} \cdot \sum_{m \neq n} \boldsymbol{X}_{m,n}(\boldsymbol{k}(t))C_m(t), \quad (1)$$

where E_n is the energy eigenvalue of the nth band, $\boldsymbol{k} = (k_x, k_y, k_z)$ is the reciprocal-space position, q is the electronic charge, and $\boldsymbol{F} = (F_x, F_y, F_z)$ is the electric field vector. The probability of the carrier being in band n is equal to $|C_n(t)|^2$. The off diagonal matrix coefficients $\boldsymbol{X}_{m,n}(\boldsymbol{k}) = (X_{m,n}(\boldsymbol{k}), Y_{m,n}(\boldsymbol{k}), Z_{m,n}(\boldsymbol{k}))$ are computed using the formula

$$\boldsymbol{X}_{m,n}(\boldsymbol{k}) = -i \int_{\Omega} u_n^*(\boldsymbol{k}, \boldsymbol{r}) \nabla_{\boldsymbol{k}} u_m(\boldsymbol{k}, \boldsymbol{r})\, d\boldsymbol{r}^3, \quad (2)$$

where the term $u_n(\boldsymbol{k}, \boldsymbol{r})$ is the periodic part of the Bloch wavefunction at $\boldsymbol{k}$ for band n, with $\boldsymbol{r} = (x, y, z)$ being the real-space position.

© Author(s) 2024

In matrix form, Eq. (1) becomes

$$i\hbar\begin{bmatrix}\dot{C}_1\\ \dot{C}_2\\ \cdots\\ \dot{C}_N\end{bmatrix}=\begin{bmatrix}E_1 & q\boldsymbol{F}\cdot\boldsymbol{X}_{1,2} & \cdots & q\boldsymbol{F}\cdot\boldsymbol{X}_{1,N}\\ q\boldsymbol{F}\cdot\boldsymbol{X}_{2,1} & E_2 & \cdots & q\boldsymbol{F}\cdot\boldsymbol{X}_{2,N}\\ \cdots & \cdots & \cdots & \cdots\\ q\boldsymbol{F}\cdot\boldsymbol{X}_{N,1} & q\boldsymbol{F}\cdot\boldsymbol{X}_{N,2} & \cdots & E_N\end{bmatrix}\begin{bmatrix}C_1\\ C_2\\ \cdots\\ C_N\end{bmatrix}, \tag{3}$$

where $\dot{C}_n$ is shorthand for $\partial C_n(t)/\partial t$ and N is the total number of bands.

The values of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ can be evaluated in two ways: with a finite difference approach[2,9,10] or with an approach originally derived from perturbation theory.[11] From previous works, the more widely used method has been the finite difference approach, which is more straightforward and requires computing the overlap integral between wavefunctions at adjacent k-points. For example, to calculate $Z_{m,n}(\boldsymbol{k})$, the term $\nabla_{\boldsymbol{k}}\, u_m(\boldsymbol{k},\boldsymbol{r})$ from Eq. (2) is replaced with $(u_m(\boldsymbol{k}+dk_z,\boldsymbol{r})-u_m(\boldsymbol{k},\boldsymbol{r}))/dk_z$, allowing Eq. (2) to be rewritten as

$$\begin{aligned}Z_{m,n}(\boldsymbol{k})&=-i\int_\Omega u_n^*(\boldsymbol{k},\boldsymbol{r})\,\frac{(u_m(\boldsymbol{k}+dk_z,\boldsymbol{r})-u_m(\boldsymbol{k},\boldsymbol{r}))}{dk_z}\\ &=-i\,\frac{O_{\boldsymbol{k},\boldsymbol{k}+d_z,m,n}-O_{\boldsymbol{k},\boldsymbol{k},m,n}}{dk_z}\\ &=-i\,\frac{O_{\boldsymbol{k},\boldsymbol{k}+d_z,m,n}}{dk_z},\end{aligned} \tag{4}$$

where $O_{\boldsymbol{k},\boldsymbol{k}',m,n}$ is the overlap integral $\langle u_n(\boldsymbol{k},\boldsymbol{r})|u_m(\boldsymbol{k}',\boldsymbol{r})\rangle$, whose magnitude has a maximum value of one. Although this method is both simple and effective for 1D implementations, it is non-local and is only accurate with a very small k-point spacing, Δk, as shown in Fig. 2. Previous works have noted that a Δk of 2×10^{-4} (Ref. 9) or 4×10^{-4} (Ref. 10), normalized by $2\pi/a$, is required to get consistent results. However, these Δk values are impractical for sampling the entire 3D first Brillouin zone (FBZ) or its irreducible wedge (IW) as it would result in approximately 10^{10} k-points.

For the general implementation of the KI equations into 3D Monte Carlo simulations, $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ must be evaluated using a method that is invariant to Δk to restrict to a more manageable number of required k-points. This can be achieved using the approach originally derived from the perturbation theory,[11] transforming Eq. (2) into

$$\boldsymbol{X}_{m,n}(\boldsymbol{k})=i\,\frac{\hbar}{m_0}\,\frac{\langle u_m(\boldsymbol{k},\boldsymbol{r})\,|\,\boldsymbol{p}\,|\,u_n(\boldsymbol{k},\boldsymbol{r})\rangle}{E_m(\boldsymbol{k})-E_n(\boldsymbol{k})}, \tag{5}$$

where $\boldsymbol{p}=(-i\hbar\frac{\partial}{\partial x},-i\hbar\frac{\partial}{\partial y},-i\hbar\frac{\partial}{\partial z})$ is the momentum operator and m_0 is the electron mass.

For this work, wavefunctions obtained from band-structure calculations through methods such as empirical pseudopotential (EPM) or Density Functional Theory (DFT) are represented as a sum of plane waves, as shown below:

$$u_n(\boldsymbol{k},\boldsymbol{r})=e^{i\boldsymbol{k}\cdot\boldsymbol{r}}\sum\nolimits_h A_{n,\boldsymbol{k}}(\boldsymbol{G}_h)\,e^{i\boldsymbol{G}_h\cdot\boldsymbol{r}}, \tag{6}$$

where $A_{n,\boldsymbol{k}}(\boldsymbol{G}_h)$ is the complex basis of the plane wave component of the wavefunction at $\boldsymbol{k}$ and in band, n, for a corresponding G-vector, $\boldsymbol{G}_h=(G_{h,x},G_{h,y},G_{h,z})$. The value of $Z_{m,n}(\boldsymbol{k})$ using the z component of the momentum operator can be calculated as

$$\left\langle u_m(\boldsymbol{k},\boldsymbol{r})\right|-i\hbar\frac{\partial}{\partial z}\left|u_n(\boldsymbol{k},\boldsymbol{r})\right\rangle=\sum\nolimits_h A^*_{m,\boldsymbol{k}}(\boldsymbol{G}_h)\,G_{h,z}\,A_{n,\boldsymbol{k}}(\boldsymbol{G}_h). \tag{7}$$

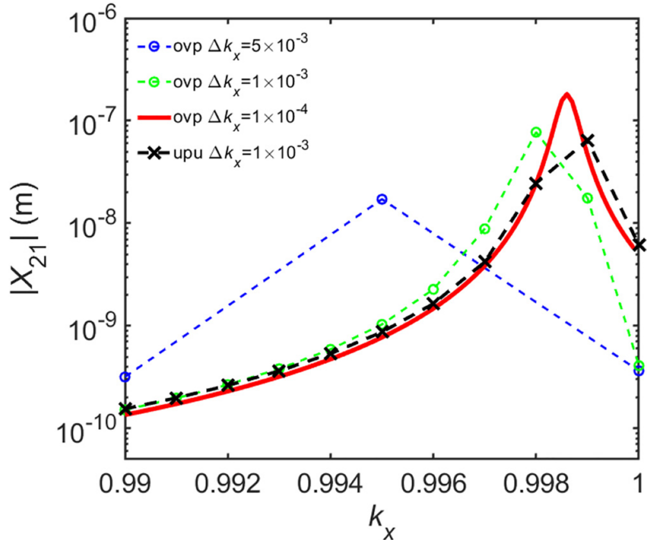


FIG. 2. Comparison of the calculated magnitudes $|X_{2,1}(\boldsymbol{k})|$ between the finite difference overlap integral method (ovp) and the perturbation theory method (upu) with different Δk. The band structure of silicon was used, with the slice performed along the k_x axis from $0.99\le k_x\le 1.00$, normalized by $2\pi/a$. (k_y,k_z) is at (0.005, 0.005).

The values of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$ near the X-point in silicon calculated using the perturbation theory method [Eq. (5)] are compared to those calculated using the finite difference method [Eq. (4)] in Fig. 2. The peak value of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$ occurs where the energy separation between the first and second conduction bands of silicon is smallest. This comparison demonstrates that obtaining accurate values of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$ using the finite difference method requires sampling with a small Δk but can be achieved with a more relaxed Δk using the perturbation theory method.

Previous works have shown that the arbitrary phase of the wavefunctions obtained by band structure solvers, such as EPM and DFT, creates a discontinuities in the phase of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$[11] and results in erroneous solutions of the KI equation [Eq. (1)]. The method outlined in Ref. 11, which has been used in prior FBMC works,[4,5] aligns all wavefunction phases to a reference wavefunction in order to enforce phase continuity of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$, but has only been used in 1D. In this work, we find that this approach is inconsistent when used for the full 3D IW, with differing solutions to the KI equation depending on the location of the reference k-point. Instead, we find that simply using purely real values of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ and preserving the skew symmetry of the matrix in Eq. (3), which can be accomplished by replacing $\langle u_m(\boldsymbol{k},\boldsymbol{r})\mid\boldsymbol{p}\mid u_n(\boldsymbol{k},\boldsymbol{r})\rangle$ from Eq. (5)

09 July 2024 07:50:08

© Author(s) 2024

with $-i \mid \langle u_m(\boldsymbol{k}, \boldsymbol{r}) \mid \boldsymbol{p} \mid u_n(\boldsymbol{k}, \boldsymbol{r})\rangle \mid$, results in consistent and well-behaved solutions to the KI equations.

Finally, with the values $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ computed, the probability coefficients $C_n(t)$ are obtained by solving Eq. (1) with a fourth order Runge–Kutta (RK4) scheme.

B. Sampling in the 3D Brillouin zone

The general 3D treatment for band crossings requires solving the KI equation [Eq. (1)] along an arbitrary k-vector with an arbitrary electric field. Regions with strong band interaction correspond to the highest values of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$, with the peak of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$ located where the energy separation between interacting bands is the smallest. In these regions, $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$ must be finely sampled to obtain accurate solutions to the KI equation. In this section, the first and second conduction bands of silicon are used to demonstrate the sampling methodology.

Figure 3 exemplifies how the Δk affects the RK4 solution of the KI equations along $0.99 < k_x < 1.00$ near the X-point of the silicon band structure for the interaction between the first and second conduction bands. In the first case, a fixed Δk of 1.67×10^{-3} is used for the entire k range, with values of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$ in between sampled k-points exponentially interpolated, resulting in inaccurate values of C_n. However, by increasing the sampling only around the peak of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$, with $\Delta k = 5.56 \times 10^{-4}$ for $0.9983 < k_x < 1.00$, as shown in the second case, the resulting accuracy of C_n can be significantly improved. Also important is the behavior for $0.99 < k_x < 0.995$, where the small values of $|\boldsymbol{X}_{m,n}(\boldsymbol{k})|$ does not significantly affect the amplitude probabilities. This means that the Δk far away band crossing points can be further relaxed, reducing the total number of k-points while retaining accuracy.

To minimize the number of k-points used, a two-level adaptive refinement meshing technique is used so that the smallest Δk is used in regions with the strongest band-to-band interaction. Using the IW reduces the required k-points by 48 times for zincblende structures, and 24 times for wurtzite structures, but requires additional rotation operations in its implementation in Monte Carlo. The initial sampling uses a uniform mesh to detect the regions with strong band interactions. The first level (L1) of mesh refinement is then applied only in regions with band crossings to increase the sampling in those regions. A second level (L2) mesh refinement is then subsequently applied to the L1 refinements to further enhance the sampling. The L2 refinement is implemented so that a Δk on the order of 10^{-4} can be achieved with a much fewer number of k-points than using only a single level of refinement. With this methodology, the normalized Δk within the same mesh can range from 10^{-4} to 10^{-2}, which results in a number of k-points on the order of 10^6. Although this many k-points is relatively computationally expensive for standard FBMC simulations, it is well within the capabilities of modern-day computers, and is drastically fewer than the unattainable 10^{10} k-points required for a fixed Δk of 2×10^{-4} or 4×10^{-4} used in the prior 1D methods.

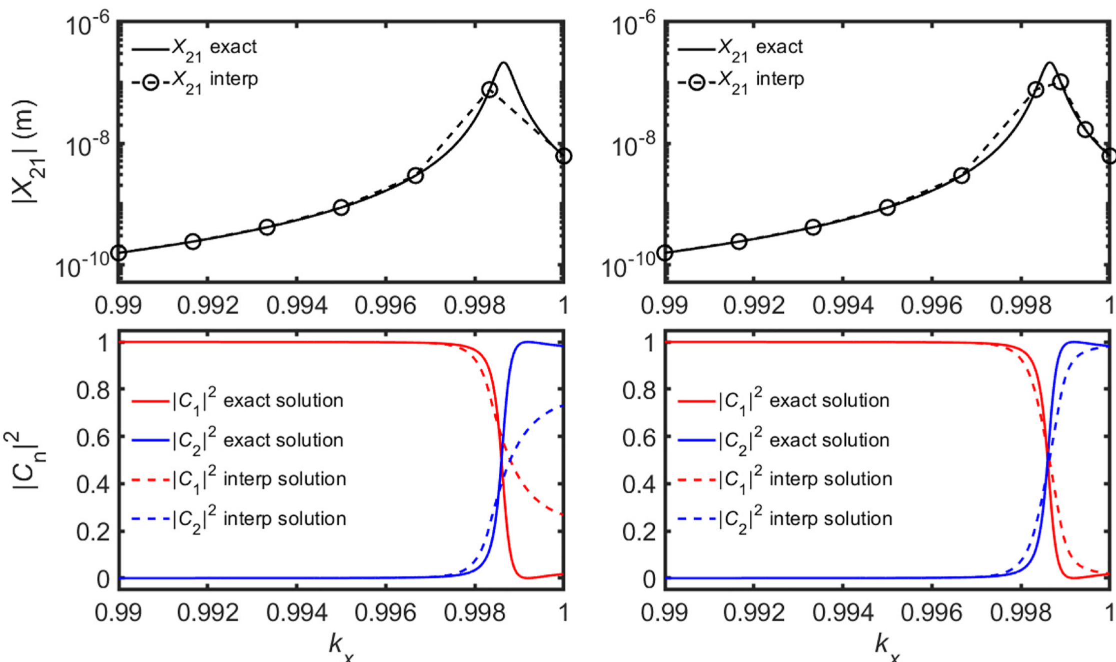


FIG. 3. Comparison of the exact and interpolated (interp) solutions of the KI equations with different sampling of $X_{m,n}(\boldsymbol{k})$. The top left figure shows $|X_{mn}|$ sampled with a fixed Δk spacing of 1.67×10^{-3} for $0.99 < k_x < 1$. The top right figure shows a fixed Δk spacing of 1.67×10^{-3} for $0.99 < k_x < 0.9983$, and a fixed spacing of 5.56×10^{-4} for for $0.9983 < k_x < 1$. $|X_{mn}|$ in between sampled k-points are exponentially interpolated. The resulting solutions of the KI equations are correspondingly shown in the bottom row. All k values are normalized by $2\pi/a$.

09 July 2024 07:50:08

© Author(s) 2024

TABLE I. Adaptive refinement methodology used for silicon. Δ*k*-partition refers to how each side of the cubic element is subdivided for refinement, resulting in "num *k*-cubes" number of additional cubic elements, and "num *k*-points" additional number of *k*-points.

$O_{\max}$	Δ*k*-partition	Num cubes	Num *k*-points
>0.999, ≤ 1.00	1/6	216	343
>0.90, ≤ 0.999	1/5	125	216
>0.75, ≤ 0.90	1/4	64	125
>0.60, ≤ 0.75	1/3	27	64
>0.30, ≤ 0.60	1/2	8	27

To demonstrate the adaptive refinement method on silicon, the IW of the zincblende reciprocal space is uniformly sampled with a Δ*k* of 1×10^{-2} requiring 96 876 *k*-points and 87 125 cubic elements. Within each cubic element, potential band crossings are detected by computing the overlap integrals $O_{m,n,\boldsymbol{k}_f,\boldsymbol{k}_i} = \langle u_n(\boldsymbol{k}_i, \boldsymbol{r}) | u_m(\boldsymbol{k}_j, \boldsymbol{r}) \rangle$ for all $n \neq m$ and $i \neq j$, where *n* and *m* are the band indices, and $\boldsymbol{k}_i$ and $\boldsymbol{k}_j$ are the *k*-point nodes of each cubic element. The max value of this calculation, $O_{\max}$, correlates with the probability of carriers experiencing interband transitions somewhere within the cubic element. Cubic elements with high $O_{\max}$ are subsequently refined and subdivided into smaller cubic elements to form the L1 refinements, following the refinement rules presented in Table I. For the first two conduction bands of silicon, the L1 refinements require an additional 151 928 *k*-points and 79 836 cubic elements.

The next step is to create the L2 refinements, following the same set of rules as before, but applied onto the L1 refinements. For this step, only the cubic elements with $O_{\max} > 0.75$ are further refined to limit the number of additional *k*-points. For the example case, this step adds 545 875 *k*-points and 277 599 cubic elements. In total, the mesh used to map the values of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ for the interaction between first and second conduction bands of silicon required 794 679 *k*-points and 444 560 cubic elements, with a normalized Δ*k* ranging from 3.33×10^{-4} to 1×10^{-2}. The calculated values of $O_{\max}$ on the $k_z = 0$ plane of the IW for silicon and the resulting refined cubic mesh with both L1 and L2 refinements are shown in Fig. 4.

Finally, the values of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ are computed at all sampled *k*-points, which includes the uniform mesh, L1 refinements, and L2 refinements, with the results shown in Fig. 5. In total, the required memory to store the $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ values is calculated to be $N_b \times N_b \times N_k \times 3 \times 8$ bytes, where N_b is the number of bands included and N_k is the number of *k*-points. The factor of 3 accounts for the storage of $X_{m,n}(\boldsymbol{k})$, $Y_{m,n}(\boldsymbol{k})$, and $Z_{m,n}(\boldsymbol{k})$ values, and the factor of 8 accounts for the byte size of double-precision floating-point data.

C. Integration into full-band Monte Carlo

This section discusses how the 3D KI model is integrated into FBMC simulations. Prior implementations, which were primarily used for calculation of bulk transport properties, used a lookup table approach, requiring a different lookup table for each different electric field strength and direction.[4,5,7] In this work, the KI equations are solved in real time with RK4 during the FBMC simulations using a 3D mesh of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ values as discussed in Sec. II B.

Values of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ are determined by first reducing an arbitrary *k*-vector in the FBZ to its equivalent position in the IW with the appropriate rotation matrix, $\boldsymbol{R}_{FBZ \rightarrow IW}$ and localizing the cubic element with the highest level of refinement. Interpolation is performed using an exponential first-order serendipity interpolation scheme[15] using the values of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ computed at the eight corner

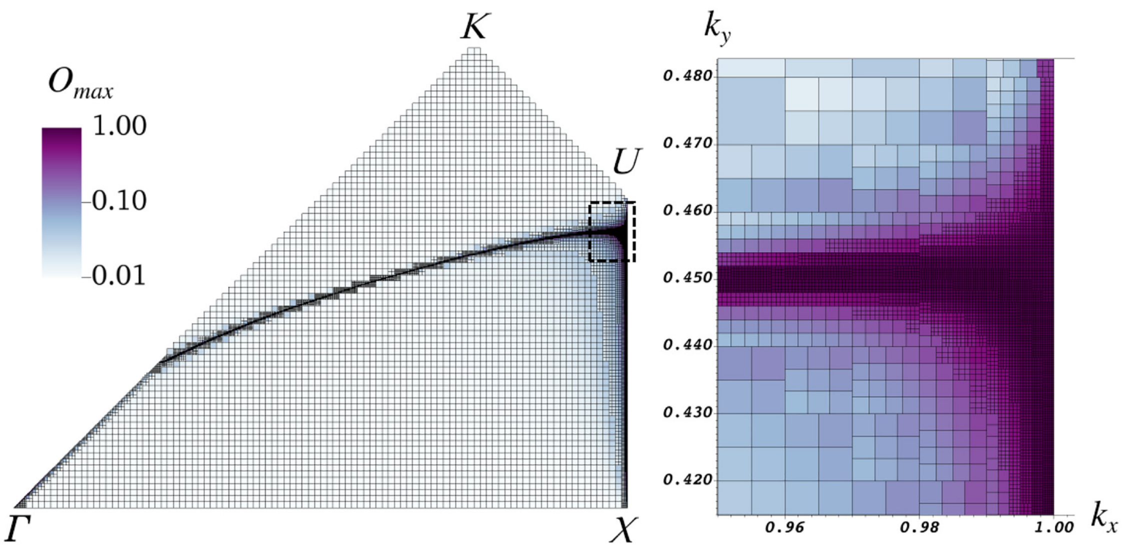


FIG. 4. The computed values of $O_{\max}$ for the first conduction band of silicon along the $z = 0$ plane of the IW (left) and the closeup of the mesh near the *U*-point (right) are shown. High values of $O_{\max}$ correspond to where the first conduction band interacts with the second conduction band. L1 and L2 mesh refinements are only placed in regions with high $O_{\max}$ values. Images were generated using VisIt.[13]

© Author(s) 2024

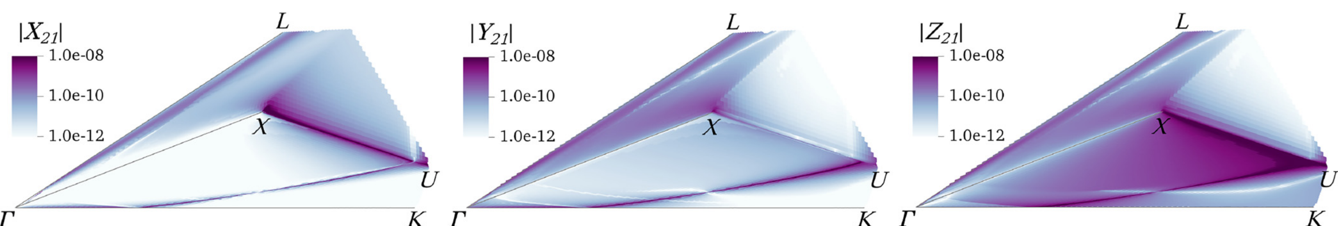


FIG. 5. Calculated values of $|X_{2,1}|$, $|Y_{2,1}|$, and $|Z_{2,1}|$ in the IW of silicon. The high values seen in $X_{2,1}$ and $Z_{2,1}$ along $\Gamma \rightarrow U$ indicate band crossings for carriers traversing in the k_x and k_z directions, respectively. The value of $|Y_{2,1}|$ along $\Gamma \rightarrow U$ is low since the bands are degenerate in the k_y direction. Images were generated using VisIt.[13]

nodes of each cubic element. For a carrier with k-vector rotated into the IW with $\boldsymbol{R}_{FBZ \rightarrow IW}$, the electric field acting on that carrier must also be rotated using the same rotation matrix.

FBMC simulations are discretized into drift and scatter cycles, where a carrier ballistically drifts for a finite drift time, Δt_{dr}, then scatters depending on the scattering rate. The KI equations are solved during the drift cycle of an FBMC simulation. For a carrier in an electric field, $\boldsymbol{F}$, its displacement vector in normalized k-space is computed as $\Delta \boldsymbol{k}_{dr} = \pm q/\hbar \cdot \boldsymbol{F} \cdot \Delta t_{dr} \cdot 2\pi/a$. The KI equations are solved using RK4 along this vector, with an initial k-vector, $\boldsymbol{k}_0$, and final k-vector, $\boldsymbol{k}_f = \boldsymbol{k}_0 + \Delta \boldsymbol{k}_{dr}$.

For accurate solutions using RK4, the selection of the time step, Δt_h, is crucial. In this work, the value of Δt_h used is 2×10^{-17} s, which we observe provides a good trade-off between the accuracy of the solution and its computational cost. For high electric fields, it is important to also consider limiting the value of $|\Delta \boldsymbol{k}_h|$, defined as $|\Delta \boldsymbol{k}_h| = q/\hbar \cdot |\boldsymbol{F}| \cdot \Delta t_h \cdot 2\pi/a$, in order to maintain accuracy. In this work, the max value of $|\Delta \boldsymbol{k}_h|$ is imposed as half the edge length of the localized cubic element, with Δt_h reduced accordingly where necessary. With the adaptively refined mesh, limiting $|\Delta \boldsymbol{k}_h|$ to half the cubic element edge length naturally results in a smaller $|\Delta \boldsymbol{k}_h|$ in regions with the highest band-to-band interaction. Finally, the RK4 is solved in $N_h = \Delta t_{dr}/\Delta t_h$ number of steps, with each RK4 $\Delta \boldsymbol{k}_h$ set as $\Delta \boldsymbol{k}_h = \Delta \boldsymbol{k}_{dr}/N_h$. Examples comparing the exact and the FBMC-implemented solutions of the KI equations using this method for varying strength of the band interactions using exact and FBMC-interpolated values of $|\boldsymbol{X}_{2,1}(\boldsymbol{k})|$, respectively, are provided in Fig. 6. The same methodology is applied to the simulations of 4H-SiC in Sec. III.

After solving the KI equations, the carriers' amplitude probability coefficients, C_n, are used to determine its energy and velocity. At any given time, the probability of a carrier being in band n is determined by $|C_n|^2$. For carriers in a mixed state, its instantaneous energy and velocity are computed by randomly selecting the band of the carrier using a weighted probability according to $|C_n|^2$. For a carrier determined to be in band m, its energy and velocity are $E_m(\boldsymbol{k})$ and $1/\hbar\ (\partial E_m/\partial \boldsymbol{k})$, respectively. Finally, the values of C_n are reset only after a successful scattering event. For a carrier determined to be in band m after scattering, C_m is set to 1 and C_n is set to 0 for $m \neq n$.

III. RESULTS AND DISCUSSIONS

The methodology presented in this work is applied to perform calculations of the velocity, energy, and impact ionization coefficients of holes in 4H-SiC. Using the 12 highest valence bands, $\boldsymbol{X}_{mn}(\boldsymbol{k})$ are computed at the nodes for the adaptively refined mesh and are used to solve the KI equations for all the simulations at different electric fields and crystallographic directions presented in this section. The results are compared with prior FBMC and experimental data.

To account for the band crossings, the prior works used overlap tests (ovp) between adjacent k-points to determine the final amplitude probabilities, with a normalized Δk of 0.02 in the k_x, k_y, and k_z directions.[3,7] For example, given a simulated carrier starting in band n near $\boldsymbol{k}_i$, its probability of being in band m near $\boldsymbol{k}_f$ after the band crossing is $O_{m,n,\boldsymbol{k}_f,\boldsymbol{k}_i}$ for all bands, m. However, computing $O_{m,n,\boldsymbol{k}_f,\boldsymbol{k}_i}$ using a coarse Δk leads in inaccurate results, as discussed in Sec. II and may lead to different results depending on the value

09 July 2024 07:50:08

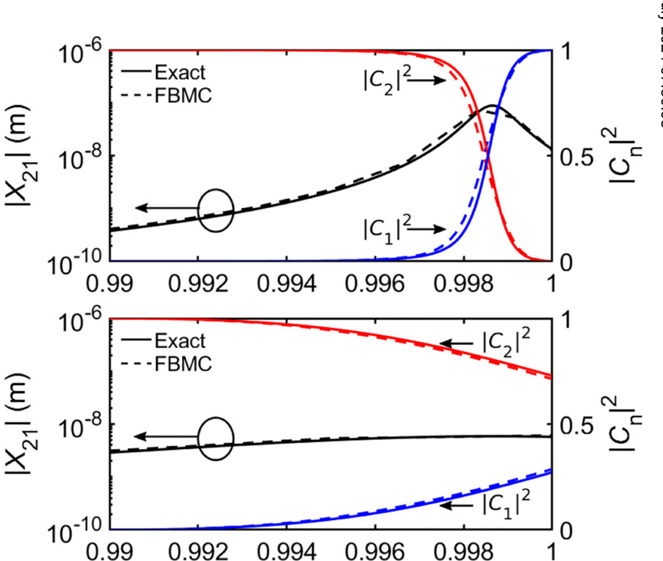


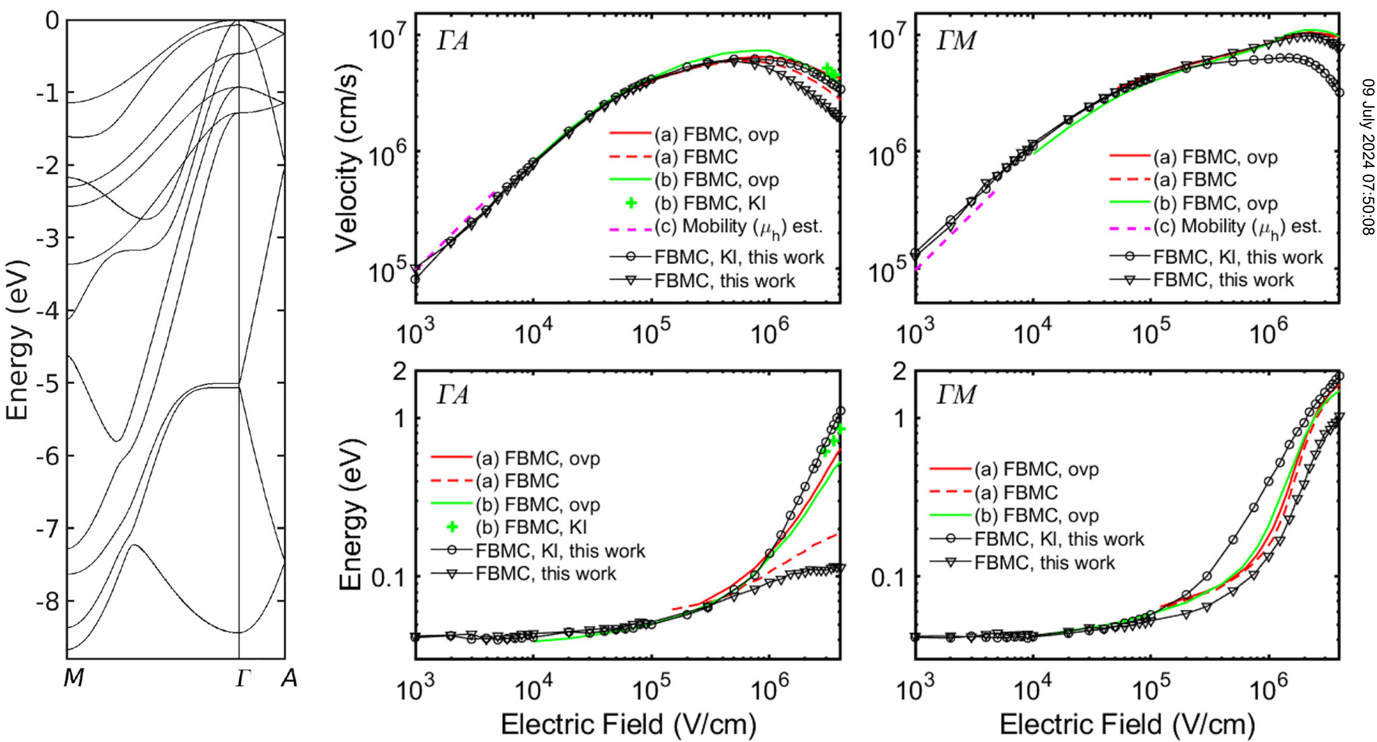
FIG. 6. Comparing the exact and FBMC solutions to the KI equations for the band crossing between the first and second conduction bands of silicon near the X point, with $0.99 < k_x < 1.00$. The (k_y, k_z) positions are (0.005, 0.005) (top), and (0.025, 0.015) (bottom), each having a different strength of their band-to-band interaction. Values computed for $|X_{2,1}|$ use an exponential serendipity interpolation scheme using the corner nodes of the cubic elements of the mesh.

© Author(s) 2024

of Δk used. Regardless, their results showed that band crossings must be accounted for in order to obtain accurate simulations of hole transport at high fields, especially in the $\Gamma \rightarrow A$ direction. In addition to the overlap test, carriers exiting the FBZ in the k_z direction which are in the first and second valence bands are automatically transitioned to the third and fourth valence bands, respectively, to account for the band crossing near the A point. These automatic transitions were also implemented for higher indexed bands as well, where appropriate. Reference 7 also provides limited number of simulated data points for hole velocities, energies, and impact ionization coefficients using band amplitude probabilities obtained by solving the KI equations using a lookup table approach. Their results using the KI model showed significant differences to their results using the *ovp* method, with the KI model achieving impact ionization coefficients more closely aligned with the available experimental data. However, their method required a different lookup table for each electric field strength and direction, which limits its applicability for device simulations.

In this work, the simulations for 4H-SiC are performed at 300 K using EPM-generated energies and wavefunctions mapped in the entire FBZ. Full band scattering rates for acoustic and optical phonon modes are computed using previously established methods.[16] Following a standard Monte Carlo approach, three acoustic modes are represented by a single *effective* acoustic mode, with a dispersion of the longitudinal acoustic (LA) mode. The 21 optical modes are represented one *effective* non-polar transverse optical (TO) mode, and one *effective* longitudinal optical (LO) mode, both using a constant energy dispersion approximation, with the energies given in Ref. 3. The calculated k-dependent rates are converted to energy-dependent rates for use in the FBMC simulations. The LA and TO *effective* modes are scaled by a deformation potential to approach the values of *ab initio* DFT computed rates, with the same approach used in Ref. 17. Finally, the 12 highest valence bands are used in the simulation, with the mesh of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$ constructed for each band-to-band interaction using the previously described methods.

The simulated velocities and energies in the $\Gamma \rightarrow A$ and $\Gamma \rightarrow M$ crystallographic directions for electric fields ranging from 1×10^3 to $4 \times 10^6\ \mathrm{V\,cm^{-1}}$ are shown in Fig. 7, with all band crossings treated using the same mesh of $\boldsymbol{X}_{m,n}(\boldsymbol{k})$. Furthermore, all band-to-band interactions are treated the same, with no special considerations given to band crossings at zone edges as seen in the prior works. First, it is important to note that carrier energies and velocities in both crystallographic directions below a field of $100\ \mathrm{kV\,cm^{-1}}$ do not see differences with and without the KI model, suggesting that as expected, carriers do not transition to other bands at low



09 July 2024 07:50:08

FIG. 7. The 12 highest energy valence bands of 4H-SiC in the $\Gamma \rightarrow M$ ($\langle 1\bar{1}00 \rangle$) and $\Gamma \rightarrow A$ ($\langle 0001 \rangle$) directions are shown in the far left panel. In the right panels, carrier velocities (top row) and carrier energies (bottom row) vs electric field for electric fields in the $\Gamma \rightarrow A$ (left column) and $\Gamma \rightarrow M$ (right column) directions. The simulations are performed with and without the inclusion of KI. Prior FBMC simulations, (a) from Ref. 3 and (b) from Ref. 7, used an overlap test (*ovp*) to account for band crossings. Mobility estimate, $\mu_h \cdot |F|$ use hole mobility values, (c) from Ref. 14. Black lines and symbols depict the results from this work.

© Author(s) 2024

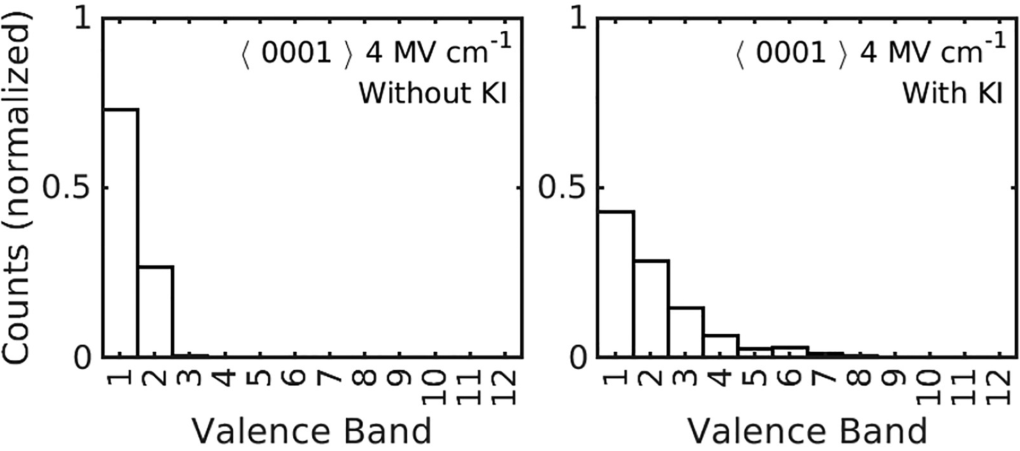


FIG. 8. Valence band occupation of holes at 4 MV cm^{-1} for simulations in the $\Gamma \rightarrow A$ (⟨0001⟩) crystallographic direction without KI (left) and with KI (right). KI allows carriers to tunnel into higher indexed bands, which change their saturation velocities and energies. Valence band 1 refers to the top valence band.

fields. The effects of KI model is most prominent in the $\Gamma \rightarrow A$ direction, with the average hole energy increasing from 0.1 to 1 eV at 4 MV cm^{-1} when KI is included, and showing agreement with the prior FBMC simulation using KI. Figure 8 presents the simulated hole band occupation in the $\Gamma \rightarrow A$ direction, showing that the differences in energy and velocity at high fields are directly attributed to carriers being able to access higher indexed bands with KI. Without KI, carriers in the first and second valence bands are rarely scattered to the higher indexed bands up to 4 MV cm^{-1}.

In the $\Gamma \rightarrow M$ direction, the effects of the KI model are less pronounced but still significant, with the carrier energy at 4 MV cm^{-1} increasing from 1 to 1.8 eV with KI. A significant effect on the simulated velocities is also observed, with the velocities decreasing from 7×10^6 to 3×10^6 cm s^{-1} with KI. The lower simulated velocity with KI is in contrast to the simulated velocities in the prior FBMC works, which saw negligible changes in their simulated velocities with and without using the *ovp* method. This discrepancy can be explained by the *ovp* method not allowing holes to access higher bands, which can be deduced from the energies in the prior simulations with and without *ovp*. However, since a crossing between the second and third valence bands exists at 0.06 eV below the top of the valence band in the $\Gamma \rightarrow M$ direction, as seen in Fig. 7, the carrier energy should still increase significantly when band crossing physics is included. Thus, the differences between the velocities between the KI and *ovp* model are likely attributed to a more refined treatment of band crossings using the KI model.

Finally, the calculated impact ionization coefficients in the $\Gamma \rightarrow A$ direction, shown in Fig. 9, are also compared to previously calculated results and experimental data. The energy-dependent ionization rates, $R_{ii}(E)$, following the Keldysh formulation,[18] are the same as the ones previously used[7] and is given by

$$R_{ii}(E) = \begin{cases} 6\times10^{14} \cdot \left(\frac{E-3.23}{3.23}\right)^4 \text{ s}^{-1} & E > 3.23 \text{ eV}, \\ 0 & \text{Otherwise.} \end{cases} \tag{8}$$

Since ionizations in 4H-SiC can only occur for carriers above its bandgap energy of 3.23 eV, only carriers in the upper tail of the energy distribution can contribute to the ionization coefficients for fields up to 4 MV cm^{-1}. The coefficients calculated in this work show agreement with prior experimental results,[19–22] suggesting that the KI model also properly accounts for the highest energy carriers. Furthermore, the KI model in this work shows improvement over

09 July 2024 07:50:08

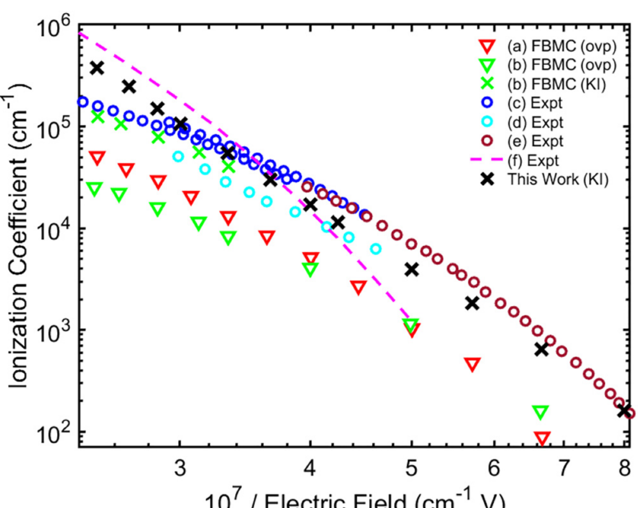


FIG. 9. Ionization coefficients in the $\Gamma \rightarrow A$ (⟨0001⟩) direction. The values computed using the KI model of this work are compared with previously performed simulations and experimental data. The experimental data are obtained from (c) Ref. 19, (d) Ref. 20, (e) Ref. 21, and (f) Ref. 22. The prior FBMC ionization coefficient values are from (a) Ref. 3 and (b) Ref. 7. The calculated ionization coefficients of this work fall in between experimental values. Simulations without KI produced no ionizations.

© Author(s) 2024

previously calculated results using the *ovp* method, and an even more dramatic difference vs simulations in this work that do not account for band crossings, which produced no ionizations at all.

IV. CONCLUSIONS

We have presented a method for a fully quantum mechanics based treatment of band crossings in 3D within the FBMC framework. The method uses a local perturbation theory approach to calculate the necessary coefficients, $\boldsymbol{X}_{mn}(\boldsymbol{k})$, which are used to solve the KI equations. In order to reduce the memory and storage used in the simulations while maintaining the accuracy of the solutions to the KI equations, an adaptive meshing technique is used for the sampling of $\boldsymbol{X}_{mn}(\boldsymbol{k})$. First, band crossing and anti-crossing points are automatically detected within the 3D Brillouin zone using overlap tests between adjacent k-points. Next, enhanced sampling is placed in proximity to the detected band crossing and anti-crossing points using a two-level refinement scheme. The resulting mesh consists of a coarse Δk of 1×10^{-2} away from band crossings, and a fine Δk of 3×10^{-4} in proximity to band crossings, both normalized by $2\pi/a$. With the KI coefficients $|X_{mn}|$, $|Y_{mn}|$, and $|Z_{mn}|$ calculated at every k-point, the KI equation, Eq. (1) can be solved for any electric field, $\boldsymbol{F}$.

This method is tested by computing the hole velocities in 4H-SiC along different crystallographic directions for a wide range of electric field strengths. Hole transport in the $\Gamma \rightarrow A$ direction is particularly impacted by band crossings. The calculated hole velocities and energies using the 3D KI model show agreement with previously computed and experimental values for all simulated electric fields. The calculated impact ionization coefficients using the KI model also show agreement with experimental values, while simulations without the KI model produced no ionizations at all. The differences in the computed parameters are directly attributed to the carrier band occupation, with carriers being able to reach higher indexed bands when the KI model is included in the simulations.

With the ability to account for band crossings and anti-crossings in any arbitrary k direction and for arbitrary electric field strengths, the capabilities of our FBMC simulator are expanded to include accurate device simulations of semiconductors with many band crossings and anti-crossing points.

ACKNOWLEDGMENTS

The development of the computational tools was supported by the US Army Research Laboratory through the Center for Semiconductor Materials and Device Modeling (Grant No. W911NF-18-2-0027) managed by Dr. M. Reed and Dr. J. Schuster. The authors would like to thank the support of ARPA-E's OPEN+ Kilovolt Devices Cohort directed by Dr. Isik Kizilyalli. The computational resources were provided by the DoD HPC Systems and the 2019 Army Research Office DURIP Award (Grant No. W911NF-19-1-0161).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Mike Zhu: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Investigation (lead); Methodology (lead); Software (lead); Supervision (equal); Validation (lead); Visualization (lead); Writing – original draft (lead); Writing – review & editing (equal). **Francesco Bertazzi:** Methodology (supporting); Software (supporting); Writing – review & editing (equal). **Masahiko Matsubara:** Methodology (supporting); Software (supporting); Writing – review & editing (supporting). **Enrico Bellotti:** Conceptualization (equal); Data curation (equal); Funding acquisition (lead); Investigation (equal); Methodology (supporting); Project administration (lead); Resources (lead); Software (supporting); Supervision (equal); Validation (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]K. Brennan, E. Bellotti, M. Farahmand, H.-E. Nilsson, P. Ruden, and Y. Zhang, IEEE Trans. Electron Devices **47**, 1882 (2000).
[2]H.-E. Nilsson, A. Martinez, E. Ghillino, U. Sannemo, E. Bellotti, and M. Goano, J. Appl. Phys. **90**, 2847 (2001).
[3]E. Bellotti, H.-E. Nilsson, K. F. Brennan, P. P. Ruden, and R. Trew, J. Appl. Phys. **87**, 3864 (2000).
[4]E. Bellotti and F. Bertazzi, J. Appl. Phys. **111**, 103711 (2012).
[5]F. Bertazzi, M. Moresco, and E. Bellotti, J. Appl. Phys. **106**, 063718 (2009).
[6]M. V. Fischetti and S. E. Laux, Phys. Rev. B **38**, 9721 (1988).
[7]M. Hjelm, H.-E. Nilsson, A. Martinez, K. Brennan, and E. Bellotti, J. Appl. Phys. **93**, 1099 (2003).
[8]J. Krieger and G. Iafrate, Phys. Rev. B **33**, 5494 (1986).
[9]D. Esseni and P. Palestri, J. Appl. Phys. **105**, 053702 (2009).
[10]R. Hathwar, M. Saraniti, and S. M. Goodnick, J. Appl. Phys. **120**, 044307 (2016).
[11]U. Lindefelt, H.-E. Nilsson, and M. Hjelm, Semicond. Sci. Technol. **19**, 1061 (2004).
[12]I. Prigozhin, S. Dominici, and E. Bellotti, IEEE Trans. Electron Devices **68**, 279 (2020).
[13]H. Childs, E. Brugger, B. Whitlock, J. Meredith, S. Ahern, D. Pugmire, K. Biagas, M. Miller, C. Harrison, G. H. Weber, H. Krishnan, T. Fogal, A. Sanderson, C. Garth, E. W. Bethel, D. Camp, O. Rübel, M. Durant, J. M. Favre, and P. Navrátil, in *High Performance Visualization–Enabling Extreme-Scale Scientific Insight* (CRC Press, 2012), pp. 357–372.
[14]H. Tanaka, S. Asada, T. Kimoto, and J. Suda, J. Appl. Phys. **123**, 245704 (2018).
[15]O. C. Zienkiewicz, R. L. Taylor, and J. Z. Zhu, *The Finite Element Method: Its Basis and Fundamentals* (Elsevier, 2005).
[16]C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer Science & Business Media, 2012).
[17]M. Zhu, M. Matsubara, and E. Bellotti, Phys. Rev. Appl. **20**, 034055 (2023).
[18]L. Keldysh, Sov. Phys. JETP **21**, 1135 (1965).
[19]A. Konstantinov, Q. Wahab, N. Nordell, and U. Lindefelt, Appl. Phys. Lett. **71**, 90 (1997).
[20]W. Loh, B. Ng, J. Ng, S. I. Soloviev, H.-Y. Cha, P. M. Sandvik, C. M. Johnson, and J. P. David, IEEE Trans. Electron Devices **55**, 1984 (2008).
[21]Y. Zhao, H. Niwa, and T. Kimoto, Jpn. J. Appl. Phys. **58**, 018001 (2018).
[22]T. Hatakeyama, T. Watanabe, T. Shinohe, K. Kojima, K. Arai, and N. Sano, Appl. Phys. Lett. **85**, 1380 (2004).

09 July 2024 07:50:08

© Author(s) 2024