Impact-ionization theory consistent with a realistic band structure of silicon

Nobuyuki Sano and Akira Yoshii
NTT LSI Laboratories, 3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa 243-01, Japan
(Received 26 August 1991; revised manuscript received 22 October 1991)

We investigate impact-ionization processes in Si with use of a realistic band structure. The band structure and the corresponding wave functions, obtained with an empirical pseudopotential method, are used to evaluate the matrix elements for the ionization transitions. The matrix element includes the direct and the exchange terms with the umklapp terms associated with the periodic part of the Bloch function. It is shown that these ionization processes are *inherently* anisotropic and that it is crucial to take account of this anisotropy in analyzing the ionization processes. The anisotropy (wave-vector dependence) of the ionization probability is manifested through the strong restrictions imposed by energy and the momentum conservation during the transition under a realistic band structure.

I. INTRODUCTION

The interest in impact-ionization processes and high-field carrier transport in semiconductors becomes even greater as the size of semiconductor devices continues to shrink.[1] Such hot-carrier transport problems have, in general, been addressed using Monte Carlo methods, and there has been great success in explaining various *phonon-related* transport problems.[2,3] In spite of this progress, the impact-ionization process is still treated in a very primitive way in Monte Carlo simulations,[4–6] and thus the physics involved in the ionization process has not yet been clearly resolved. For example, it is not certain why a significant material dependence of the soft and hard ionization thresholds (P factor in the Keldysh formula[7]) exists in Si and GaAs.[6,8]

To clarify the mechanism of impact ionization in semiconductors, the most important quantity characterizing the ionization transitions is the ionization probability. The ionization probability is calculated formally from Fermi's golden rule in a straightforward way.[9,10] The problems that prevent the calculation of a realistic ionization probability and, in turn, the resolution of the problems mentioned above are mainly due to the fact that a realistic band structure in higher-energy regimes is very complicated in semiconductors. The evaluation of the matrix elements for the ionization transition and of the multidimensional integrals in wave-vector space requires both an accurate band structure and the wave functions, which results in a huge amount of numerical computations on a computer.

On the other hand, various investigators have extensively studied the ionization probability using analytical approaches[7,11] in which the investigators assumed analytical forms for the band structure, for example, a parabolic or nonparabolic band structure, for both the conduction and valence bands. These approaches may be partly justified for narrow- and direct-band-gap semiconductors in which the deviation of the band structure employed in the calculations from the realistic band structure would be insignificant in the energy range being investigated. However, it is clear that the above approach breaks down for wide- and/or indirect-band-gap semiconductors such as Si or GaAs.

Recently, the approximations employed in these calculations have been relaxed. A realistic valence-band structure distinctly different from the parabolic or nonparabolic band structure has been calculated by the $\mathbf{k}\cdot\mathbf{p}$ method,[12] and the corresponding wave functions have been used for evaluating the matrix elements.[13–15] This approach is not, however, accurate enough for wide-gap-band semiconductors in which the conduction band in higher-energy regimes also deviates from the analytical form of the band structure.

The present authors have recently proposed a generalized Keldysh formula with wave-vector-dependent ionization thresholds and have performed Monte Carlo simulations for Si.[16] We went on to point out the importance of the anisotropy imposed by the ionization processes. However, the formula employed there was somewhat phenomenological and no justification based on first principles was given. An alternative approach taking into account the wave-vector dependence of the ionization probability was described by Thoma *et al.*[17] They used an artificial band structure that reproduces a realistic Si density of states to model the entire band structure in the first Brillouin zone (BZ). From a technological (device application) point of view, this approach is advantageous insofar as it reduces the CPU time required for the computations. Unfortunately, it cannot answer the fundamental questions about the physics involved in impact ionization, because wave functions consistent with the band structure are not known.

In the present paper we perform a realistic evaluation of the matrix elements for the ionization transition and of the ionization probability consistent with a realistic band structure determined from an empirical pseudopotential method.[18] The symmetry property imposed by the crystal structure is fully exploited for efficient numerical calculations of the matrix elements and ionization probabilities. The overlap integrals, matrix elements, ionization threshold energies, and ionization probabilities are evalu-

ated and discussed with emphasis on the anisotropy associated with a realistic band structure. It is shown that the ionization process is *inherently* anisotropic. The wave-vector dependence (anisotropy) of the ionization probability is manifested in the strong restrictions of the energy and the momentum conservation under a realistic band structure, though the matrix elements also reveal strong wave-vector dependence.

In fact, Kane has already evaluated the ionization probability with a realistic Si band structure and has found a rather slow increase (soft threshold) of the ionization probability as the electron energy increases.[19] Very recently, Bude *et al.* have extended Kane's approach by including intracollisional field effects.[20] In these works, the ionization probability was averaged over the density of states, and the ionization probably was thus given as a function of the electron energy. Therefore, some interesting features, such as the inherent anisotropy of the ionization process (discussed in detail in the present paper), were overlooked. Furthermore, the average values of the matrix elements were used for the ionization transitions because of limitations in the numerical calculations. A similar approach has been also taken by Laks *et al.*[21] to obtain the lifetime due to nonradiative Auger recombinations. Laks *et al.* integrated the transition probability over initial states and, in doing so, missed the important feature mentioned above.

The present paper is organized as follows. The matrix elements of the impact-ionization transition and the ionization probability as a function of the wave vector are derived in Sec. II. The calculation procedures for efficient numerical evaluations of the matrix elements and the ionization probability are given in Sec. III. The results of the calculations of the overlap integrals, matrix elements, ionization threshold energies, and ionization probabilities are given and discussed in Sec. IV. Finally, we draw some conclusions from the present study in Sec. V.

II. IMPACT-IONIZATION PROBABILITY

Impact ionization involves electron-electron interactions taking place between the electrons in the conduction and valence bands in solids. Figure 1 schematically shows an impact-ionization process in Si. For wide- and indirect-gap band-semiconductors such as Si, the ionization transitions involve a substantial momentum and energy transfer during the transitions, as we shall explicitly show in Sec. IV. Therefore, a full knowledge of a realistic band structure and wave functions in the *entire* BZ is crucial in evaluating accurate matrix elements and transition probabilities.

In the present study the band structure and wave functions of Si are calculated with use of the empirical pseudopotential method, which produces a reliable description of the excitation spectrum.[18] In the pseudopotential method the periodic part of the Bloch wave function is expanded with a basis set of the reciprocal-lattice vectors (**G** vectors).[22] The wave function for the electron at position $\mathbf{r}_i$ with band index n_i and wave vector $\mathbf{k}_i$ is expressed by

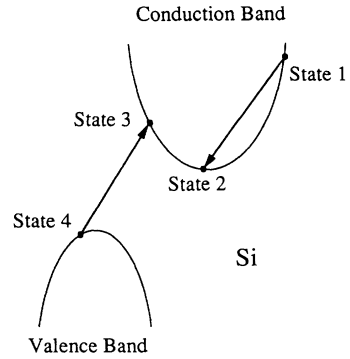

FIG. 1. Schematic drawing of an impact-ionization process in Si. The initial electrons in the conduction and valence bands before the transition are in states 1 and 4, respectively, and the final electrons in the conduction band after the transition are in states 2 and 3.

$$\varphi_{n_i\mathbf{k}_i}(\mathbf{r}_i)=\frac{1}{V^{1/2}}e^{i\mathbf{k}_i\cdot\mathbf{r}_i}\sum_{\mathbf{G}}c(n_i,\mathbf{k}_i;\mathbf{G})e^{i\mathbf{G}\cdot\mathbf{r}_i}\ , \tag{1}$$

where V is the volume of the crystal, $c(n_i,\mathbf{k}_i;\mathbf{G})$ the expansion coefficient, and i ($i=1,2,3,4$) denotes the electron state. States 1 and 4 are the states of the initial electrons in the conduction and valence bands before the transition, respectively, and states 2 and 3 those of the final electrons in the conduction band after the transition (see Fig. 1). In the summation over the **G** vectors, we have employed a set of 113 **G** vectors for the expansion. This is large enough to produce a good band structure and satisfactory convergence of the matrix elements. Figure 2 shows the band structure obtained using the pseudopotential method expanded with 113 **G** vectors. It should be noted, however, that the number of the **G** vectors in the present work cannot be chosen to be arbitrarily large. As we shall explain in Sec. III, we employ symmetric properties such that the basis set of 113 **G** vectors is invariant under the 48 symmetry transformations for the wave vectors in an irreducible wedge (IW) of the BZ. Employing this feature, most wave functions outside the IW can be constructed from the wave functions inside the

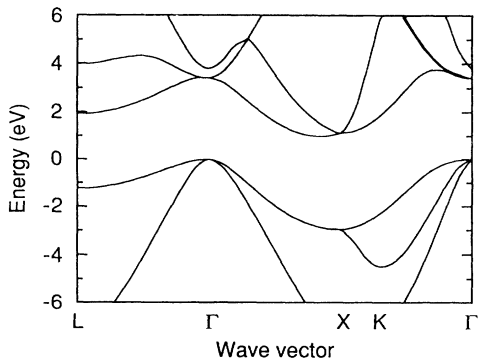

FIG. 2. Si band structure obtained from the empirical pseudopotential method. The lowest eight bands (four valence and four conduction bands) are used in the present study.

IW. As a result, the matrix elements of the ionization transition can be efficiently evaluated.

The wave vectors $\mathbf{k}_i$ are restricted to lie in the first BZ and satisfy the following momentum conservation:

$$\mathbf{k}_1+\mathbf{k}_4+\mathbf{G}_0=\mathbf{k}_2+\mathbf{k}_3 , \tag{2}$$

where $\mathbf{G}_0$ is the reciprocal-lattice vector that returns all wave vectors $\mathbf{k}_i$ to the first BZ (umklapp process). The electrostatic potential responsible for ionization transition between electrons is given by the screened Coulomb potential and is expressed by[21]

$$V(\mathbf{r}_1-\mathbf{r}_2)=\frac{e^2 e^{-\lambda|\mathbf{r}_1-\mathbf{r}_2|}}{\epsilon(q)|\mathbf{r}_1-\mathbf{r}_2|} , \tag{3}$$

where e is the elementary charge and λ the inverse of the screening length given by $[4\pi n_0 e^2/(k_B T)]^{1/2}$. Throughout the present study, the electron temperature T equals 300 K and the electron density n_0 equals 10^{16} cm^{-3}, which allows one to ignore the degeneracy. k_B is Boltzmann's constant and $\epsilon(q)$ is the dielectric function, the wave-vector dependence of which is taken into account via the form given by Nara and Morita.[23] As already pointed out by Laks *et al.*,[21] the inclusion of the wave-vector dependence of the dielectric function is important in calculating the matrix elements and, hence, also in the resulting transition probability because of the large momentum transfers involved.

The matrix element T_{ii} for the ionization transition consists of the direct and exchange processes and is given by

$$T_{ii}=\tfrac{1}{2}(|M_D|^2+|M_E|^2+|M_D-M_E|^2) , \tag{4}$$

where M_D and M_E represent the matrix elements of the direct and exchange processes, respectively. The matrix element M_D is calculated in terms of the pseudo-wave-functions given by Eq. (1) and is written as

$$M_D=\int d^3r_1 \int d^3r_2 \varphi^*_{n_1\mathbf{k}_1}(\mathbf{r}_1)\varphi^*_{n_4\mathbf{k}_4}(\mathbf{r}_2)V(\mathbf{r}_1-\mathbf{r}_2)\varphi_{n_2\mathbf{k}_2}(\mathbf{r}_1)\varphi_{n_3\mathbf{k}_3}(\mathbf{r}_2)$$
$$=\frac{4\pi e^2}{V}\sum_{\mathbf{G}_1,\mathbf{G}_2,\mathbf{G}_3,\mathbf{G}_4}\frac{1}{\epsilon(q)(\mathbf{q}^2+\lambda^2)}\delta_{\mathbf{k}_1+\mathbf{G}_1+\mathbf{k}_4+\mathbf{G}_4,\,\mathbf{k}_2+\mathbf{G}_2+\mathbf{k}_3+\mathbf{G}_3}c^*(n_1,\mathbf{k}_1;\mathbf{G}_1)c^*(n_4,\mathbf{k}_4;\mathbf{G}_4)c(n_2\mathbf{k}_2;\mathbf{G}_2)c(n_3,\mathbf{k}_3;\mathbf{G}_3) , \tag{5}$$

where $\delta_{\mathbf{k}_1+\mathbf{G}_1+\mathbf{k}_4+\mathbf{G}_4,\,\mathbf{k}_2+\mathbf{G}_2+\mathbf{k}_3+\mathbf{G}_3}$ is the Kronecker δ function and $\mathbf{q}$ $(=\mathbf{k}_1+\mathbf{G}_1-\mathbf{k}_2-\mathbf{G}_2)$ the momentum transfer. Using the momentum conservation given by Eq. (2) and defining $\mathbf{G}_U=\mathbf{G}_1-\mathbf{G}_2$, Eq. (5) can be simplified as

$$M_D=\frac{4\pi e^2}{V}\sum_{\mathbf{G}_U}\frac{I_{cc}(n_1,\mathbf{k}_1;n_2,\mathbf{k}_2;\mathbf{G}_U)I_{vc}(n_4;\mathbf{k}_4;n_3,\mathbf{k}_3;\mathbf{G}_U)}{\epsilon(q)(\mathbf{q}^2+\lambda^2)} , \tag{6}$$

where

$$I_{cc}(n_1,\mathbf{k}_1;n_2,\mathbf{k}_2;\mathbf{G}_U)=\sum_{\mathbf{G}_1}c^*(n_1,\mathbf{k}_1;\mathbf{G}_1)c(n_2,\mathbf{k}_2;\mathbf{G}_1-\mathbf{G}_U) , \tag{7}$$

$$I_{vc}(n_4,\mathbf{k}_4;n_3,\mathbf{k}_3;\mathbf{G}_U)=\sum_{\mathbf{G}_4}c^*(n_4,\mathbf{k}_4;\mathbf{G}_4)c(n_3,\mathbf{k}_3;\mathbf{G}_4+\mathbf{G}_U-\mathbf{G}_0) , \tag{8}$$

where $I_{vc}(n_i,\mathbf{k}_i;n_j,\mathbf{k}_j;\mathbf{G}_U)$ is the overlap integral between state i in the valence band and state j in the conduction band. Similar notation is used for I_{cc}. The exchange term M_E is given by simply exchanging the electron states of the two final electrons in the conduction band in Eq. (6); (2,3)→(3,2).

Note that I_{cc} and I_{vc} are both a function of the reciprocal-lattice vector $\mathbf{G}_U$. The terms with $\mathbf{G}_U\neq 0$ are often called the "umklapp" terms associated with the periodic part of the Bloch functions, and they have usually been ignored in past calculations.[24] The exclusion of the umklapp terms may be insignificant for direct- and narrow-band-gap semiconductors since the magnitude of the momentum transfer is usually small for such materials. However, for GaAs, which has a wide, direct band gap, this term is important—a fact already noted by Brand and Abram,[25] who found a 50% correction in the overlap integral as a result of the inclusion of the nonzero $\mathbf{G}_U$ terms. Therefore, it is expected that the inclusion of the terms with $\mathbf{G}_U\neq 0$ would be even more significant for Si.

Finally, the ionization probability per unit time for the initial electron with wave vector $\mathbf{k}_1$ and band index n_1 is calculated from Fermi's golden rule:[26]

$$w_{ii}(n_1,\mathbf{k}_1)=2\frac{2\pi}{\hbar}\frac{V^2}{(2\pi)^6}\sum_{n_2,n_3,n_4}\int d^3k_4\int d^3k_2 T_{ii}\delta(E_1+E_4-E_2-E_3) . \tag{9}$$

Here, the prefactor of 2 on the right-hand side of the equation represents spin degeneracy of the initial electron in the valence band (state 4 in Fig. 1) and $\hbar$ is Planck's constant divided by 2π. The summation over the band indices is performed over the lowest eight bands, i.e., the sum is taken over bands 5–8 for the conduction electron and over bands 1–4 for the valence electron. The momentum conservation given by Eq. (2) is implicitly imposed in Eq. (9).

III. CALCULATIONAL PROCEDURES

In order to carry out the numerical evaluations of the equations derived in the preceding section, the first BZ is discretized with the mesh spacing of 0.1 in units of $2\pi/a$, where a is the lattice constant of the crystal and is given by 5.43 Å for Si. The band structure is calculated for 152 $\mathbf{k}$ points in the IW of the first BZ. The region of the IW in wave-vector space is defined by

$$0 \le k_x + k_y + k_z \le 1.5(2\pi/a) ,$$
$$0 \le k_z \le k_y \le k_x \le 2\pi/a . \tag{10}$$

The band structure of the entire BZ (4481 points) can then be reproduced from the band structure in the IW by application of the 48 symmetry transformations U_k on the wave vectors inside the IW.[27] Table I lists the 24 symmetry transformations. An additional 24 symmetry transformations is generated by operating the inversion transformation $(-\mathbf{k}_1, -\mathbf{k}_2, -\mathbf{k}_3)$ on the 24 transformations listed in Table I. Here, the notation $(-\mathbf{k}_1, -\mathbf{k}_2, -\mathbf{k}_3)$ represents, for example, that $(\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3)$ is transformed into $(-\mathbf{k}_1, -\mathbf{k}_2, -\mathbf{k}_3)$.

It should be noted, however, that the wave functions at two *equivalent points* in the BZ, i.e., a point inside the IW and a corresponding point outside the IW obtained by operating one of the 48 symmetry transformations on the wave vector inside the IW, are *not*, in general, identical. Therefore, it is necessary, in general, to store the wave functions at every $\mathbf{k}$ point in the *entire* BZ and/or to explicitly carry out the band calculations whenever the wave functions are needed. This fact makes it very difficult to pursue the evaluations of the realistic ionization probability.

We have found that this difficulty can be overcome by employing the symmetry properties implied by the crystal structure, as we shall explain below. Recall that the basis set of the 113 $\mathbf{G}$ vectors transforms into itself under the 48 symmetry transformations U_k.[28] Another transformation matrix U_F, represented by a symmetric 113×113 matrix in the present case, operating on the wave function at a point inside the IW, can then be constructed for each 48 symmetry transformation U_k. However, the transformation matrix U_F so obtained alone does not produce the correct phase factor of each element of the wave function because of the fact that the diamond structure contains two atoms per primitive cell. We have found that the correct phase factor of each element is uniquely determined by the diagonal matrix S, whose matrix element S_{ii} $(i=1-113)$ is defined by

$$S_{ii} = \exp[i|(\mathbf{G}_i \mp U_k \mathbf{G}_i)\cdot\mathbf{t}|] \tag{11}$$

for each of the 48 symmetry transformations U_k. Here, $\mathbf{t}=(a/8)(1,1,1)$, and the upper sign is taken when U_k is one of the 24 symmetry transformations *without* inversion (listed in Table I) and the lower sign is taken for the transformations with inversion. $\mathbf{G}_i$ denotes the ith basis vector in the expansion of Eq. (1). Apart from a *global* phase factor of the wave function that has no physical consequences in the present calculations, the correct wave function at a point *outside* the IW can be reproduced by operating the transformation matrix U_F on the wave function at *the equivalent point inside* the IW first and then by operating the diagonal matrix S on it.

Unfortunately, this method does not work for some points (71 in number) in the BZ. For those points, the correct wave functions must be stored before the evaluations of the matrix elements. Still, the correct wave functions at most $\mathbf{k}$ points (over 4200 points) outside the IW for the lowest eight conduction bands are reproducible by the present method, i.e., *the wave functions in the entire BZ are generated from the wave functions at only the 223 points.* The memory in currently available mainframe computers is entirely sufficient for storing the wave functions. Also, this method is applicable to any covalent semiconductors having a diamond structure, and would therefore reduce a large amount of computation time and/or memory size.

With a full knowledge of both the band structure and the wave functions in the entire BZ, the calculation of the ionization probability is, *in principle,* straightforward. Equation (9) is discretized and the multidimensional integrations over the wave vectors are replaced by the summations. Assuming that the matrix elements inside the cubic mesh are constant, Eq. (9) becomes

TABLE I. Symmetry transformations from the irreducible wedge.

$(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)$	$(-\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_3)$	$(-\mathbf{k}_1,\mathbf{k}_2,-\mathbf{k}_3)$	$(\mathbf{k}_1,-\mathbf{k}_2,-\mathbf{k}_3)$
$(\mathbf{k}_1,\mathbf{k}_3,\mathbf{k}_2)$	$(-\mathbf{k}_1,-\mathbf{k}_3,\mathbf{k}_2)$	$(-\mathbf{k}_1,\mathbf{k}_3,-\mathbf{k}_2)$	$(\mathbf{k}_1,-\mathbf{k}_3,-\mathbf{k}_2)$
$(\mathbf{k}_2,\mathbf{k}_3,\mathbf{k}_1)$	$(-\mathbf{k}_2,-\mathbf{k}_3,\mathbf{k}_1)$	$(-\mathbf{k}_2,\mathbf{k}_3-\mathbf{k}_1)$	$(\mathbf{k}_2,-\mathbf{k}_3,-\mathbf{k}_1)$
$(\mathbf{k}_2,\mathbf{k}_1,\mathbf{k}_3)$	$(-\mathbf{k}_2,-\mathbf{k}_1,\mathbf{k}_3)$	$(-\mathbf{k}_2,\mathbf{k}_1,-\mathbf{k}_3)$	$(\mathbf{k}_2,-\mathbf{k}_1,-\mathbf{k}_3)$
$(\mathbf{k}_3,\mathbf{k}_1,\mathbf{k}_2)$	$(-\mathbf{k}_3,-\mathbf{k}_1,\mathbf{k}_2)$	$(-\mathbf{k}_3,\mathbf{k}_1,-\mathbf{k}_2)$	$(\mathbf{k}_3,-\mathbf{k}_1,-\mathbf{k}_2)$
$(\mathbf{k}_3,\mathbf{k}_2,\mathbf{k}_1)$	$(-\mathbf{k}_3,-\mathbf{k}_2,\mathbf{k}_1)$	$(-\mathbf{k}_3,\mathbf{k}_2-\mathbf{k}_1)$	$(\mathbf{k}_3,-\mathbf{k}_2,-\mathbf{k}_1)$

$$w_{ii}(n_1,\mathbf{k}_1)=\frac{4}{\hbar\pi}\left[\frac{e^2}{a}\right]^2\frac{(\Delta k)^6}{\delta E}\times\sum_{n_2,n_3,n_4}\sum_{\Delta\mathbf{k}_4,\Delta\mathbf{k}_2}T_{ii}\delta_{\mathbf{k}_1+\mathbf{k}_4+\mathbf{G}_0,\mathbf{k}_2+\mathbf{k}_3}=\frac{4}{\hbar\pi}\left[\frac{e^2}{a}\right]^2\frac{(\Delta k)^6}{\delta E}g_{ii}(n_1,\mathbf{k}_1)\ , \tag{12}$$

where the matrix element T_{ii} and the wave vectors are expressed in units of $[(ea)^2/(\pi V)]^2$ and $2\pi/a$, respectively. The momentum conservation [Eq. (2)] is explicitly included in Eq. (12). The energy-conserving δ function has been approximated by the square of unit area of height $1/\delta E$ and of width δE,

$$\delta(E-E_0)=\begin{cases}1/\delta E & \text{for } |E-E_0|\le\delta E/2\ ,\\ 0 & \text{otherwise}\ ,\end{cases} \tag{13}$$

where the energy interval $\delta E=0.2$ eV and the mesh spacing $\Delta k=0.1$ are used throughout the study.[29]

It should be noted here that the present calculation is exact within the uncertainty δE associated with the energy-conserving δ function and with the finite size, Δk, of mesh spacing in the BZ. As the mesh spacing and energy uncertainty reduce, the present results, of course, approach the exact solutions.

IV. RESULTS AND DISCUSSION

In this section we present the concrete results of our calculations for the overlap integrals, matrix elements, ionization threshold energies, and ionization probabilities, focusing our attention on the wave-vector dependence implied by the realistic band structure of Si.

A. Overlap integral and matrix element

We first examine the overlap integrals, which are of great importance in the analysis not only of the impact-ionization processes, but also of phonon scattering. The overlap integrals for direct-band-gap semiconductors under a realistic band structure have been investigated by several people,[30–32] and the importance of the anisotropy imposed by the realistic band structure have been stressed. To our knowledge, concrete results of the overlap integrals and the matrix elements of the ionization transitions under a realistic Si band structure have not been presented to date.

We investigate the overlap integrals for two particular cases: the overlap integrals, I_{cc}, between the conduction bands and the overlap integrals, I_{vc}, between the conduction and the valence band. Since it is impossible to show all possible values occurring in the impact-ionization processes, we show only some specific cases. The final state ($\mathbf{k}$ and n) of the electron expressed by the second argument of I_{cc} and I_{vc} in Eqs. (7) and (8) is fixed at $\mathbf{k}=(-0.85,0,0)$ with $n=5$, where the minimum of the lowest conduction band is located. Hereafter, the wave vectors and reciprocal-lattice vectors are expressed in units of $2\pi/a$. In addition, the reciprocal-lattice vectors ($\mathbf{G}_U$ and $\mathbf{G}_0$) included in Eqs. (7) and (8) are assumed to be zero. The initial state expressed by the first argument of I_{cc} and I_{vc} is varied along three different crystallographic directions: [100], [110], and [111]. Figure 3 shows the squares ($|I_{cc}|^2$ and $|I_{vc}|^2$) of the overlap integrals as a function of the wave vector along these directions. $|I_{cc}|^2$ is plotted for $n=5$ for both the initial and final states, and $|I_{vc}|^2$ is between $n=3$ and 5 for the valence and the conduction band, respectively.[33] The wave vector is normalized by the value of the wave vector at the intersection with the zone edge for each direction. It is clear from Fig. 3 that the overlap integrals indeed depend on the wave vector of the electron. In particular, the overlap integrals at some regimes become very close to zero. This is because the cases plotted here are somewhat *special* in the sense that both the initial and final states of the electron are on the symmetry lines.

It should be pointed out, however, that most impact-ionization transitions take place between the states away from the symmetry lines, which means the values of the overlap integrals are not always close to zero. This point is very important in calculating the ionization probability and should be stressed in another way. As we shall show below, the number of impact-ionization events increases steeply as the electron energy becomes larger (see Table III). Since most of these events are the transitions between states away from the symmetry lines, the wave-

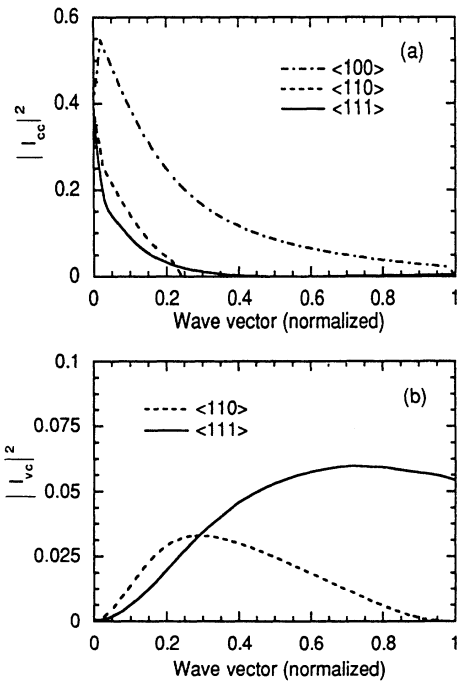

FIG. 3. (a) Square, $|I_{cc}|^2$, of the overlap integral between the conduction bands ($n=5$) as a function of the wave vector of the initial electron along the three crystallographic directions [100], [110], and [111]. (b) Square, $|I_{vc}|^2$, of the overlap integral between the valence ($n=3$) and conduction bands ($n=5$) as a function of the wave vector of the valence electron along the [110] and [111] directions. The state of the final electron in (a) and (b) is fixed at $\mathbf{k}=(-0.85, 0, 0)$. The wave vector is normalized with the magnitude of the wave vector at the intersection with the zone edge for each direction.

vector dependence of the overlap integrals are quickly washed out when the electron energy exceeds the ionization threshold energy. As a result, the average value of the overlap integrals does not vary significantly with respect to the wave vectors. In other words, the wave-vector dependence of the overlap integrals shown in Fig. 3 is insignificant, except for regions very close to the ionization threshold energies, as far as the impact-ionization probability is concerned.

The matrix elements for the ionization transition are calculated by Eq. (6) and include the summation over the umklapp terms with nonzero $\mathbf{G}_U$ vectors and the q-dependent dielectric function $\epsilon(q)$.

In Fig. 4 we show the average magnitudes of the momentum transfers, $|\mathbf{q}_D|$ and $|\mathbf{q}_E|$ for the direct and exchange processes, respectively, for a typical ionization transition. The ionization transition considered here is that for the initial states, $\mathbf{k}_1=(1.0, 0.2, 0.0)$ and $\mathbf{k}_2=(0.0, 0.0, -0.9)$, and that for the final states, $\mathbf{k}_4=(0.0, -0.1, 0.1)$ and $\mathbf{k}_3=(0.0, -0.9, 0.0)$, with $\mathbf{G}_0=(-1,-1,-1)$. Note that the average magnitude of the momentum transfer is indeed large (on the order of $2\pi/a$).[34] This also implies that the dielectric function $\epsilon(q)$ becomes an order of magnitude smaller than that of the static value (~ 11.9) (Ref. 35) usually used. As already asserted by Laks *et al.*,[21] this is mainly due to the fact that Si is an indirect- *and* wide-band-gap material, so that the transitions take place between two states greatly separated from each other in wave-vector space.

Since the magnitude of the momentum transfer is large, the summation over the umklapp terms ($\mathbf{G}_U$ vectors) in Eq. (6) is also important. Figure 5 shows the squares of the matrix elements, $|M_D|^2$ and $|M_E|^2$, and T_{ii} for that particular ionization transition as a function of the number of the umklapp terms in Eq. (6). Good convergence is obtained when several tens of the umklapp terms are included in the summation to get a reasonable result. Therefore, the inclusion of the umklapp terms, usually ignored in other calculations, is essential for Si, as clearly pointed out by Laks *et al.*[21]

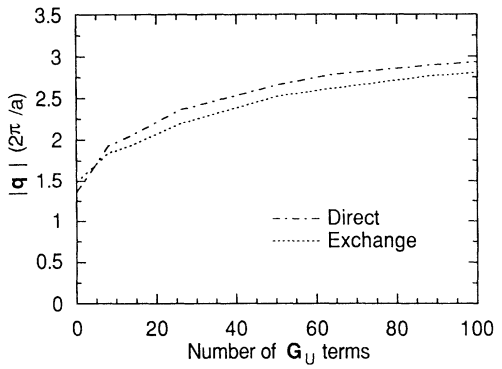

FIG. 4. Average magnitudes of momentum transfer for the direct process, $|\mathbf{q}_D|$, and for the exchange process, $|\mathbf{q}_E|$, as a function of the number of the umklapp (nonzero $\mathbf{G}_U$ vector) terms. The momentum transfer is expressed in units of $2\pi/a$.

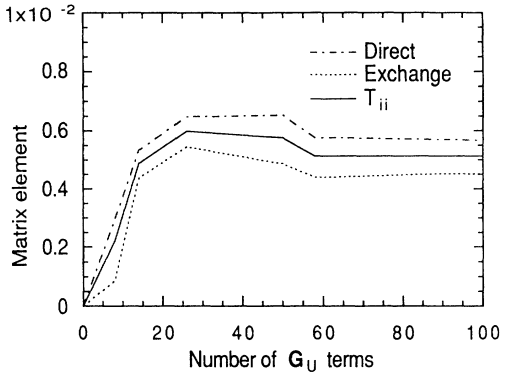

FIG. 5. Matrix elements for the direct process, $|M_D|^2$, and for the exchange process, $|M_E|^2$, and T_{ii} as a function of the number of the umklapp terms. The matrix elements are expressed in units of $[(ea)^2/(\pi V)]^2$.

B. Threshold energy and ionization probability

Another important quantity characterizing the ionization transition is the ionization threshold energy. The ionization threshold energy is defined as the minimum energy required for the electron in the conduction band to excite an electron in the valence band. The ionization threshold energy is determined from the rules of energy and momentum conservation under a *realistic* band structure. If both the conduction and valence bands are assumed to be parabolic, the well-known $\frac{3}{2}$ rule holds true: the ionization threshold energy $E_{\rm th}$ is given by $3E_g/2$, where E_g is the band gap.[10] However, this is not true, especially for wide-band-gap materials. The conduction bands in high-energy regimes are greatly deformed and different from the analytical band structures assumed by the parabolic or nonparabolic approximations. In addition, the valence bands are usually deformed. Therefore, it is essential to incorporate a realistic band structure in order to evaluate the ionization threshold energies.

Since energy and momentum conservation under a realistic band structure is explicitly taken into account in the present work, the ionization threshold energies found here are exact *within the energy interval* δE. The ionization threshold energies are then simply given by the minimum energies for the electron to induce impact ionization. As we stressed in our previous paper, the ionization threshold energies are dependent on the directions of the wave vector of the initial electron in state 1.[16] For example, although the minimum threshold energy is given by about 1.1 eV along the [100] direction, the electron

TABLE II. Ionization threshold energies along the three crystallographic directions concerned.

Direction	This work	AC method
[100]	1.0 eV	1.1 eV
[110]	1.4 eV	2.0 eV
[111]	1.8 eV	3.0 eV

with energy above 1.1 eV does not necessarily impact-ionize if the electron is located away from the [100] direction. This is seen in Table II, where the ionization threshold energies found from the present method along the [100], [110], and [111] directions are listed. We also determined the threshold energies using the Anderson-Crowell (AC) method,[36] and they are listed for comparison. The same band structure was used to calculate the threshold energies in both approaches.

A discrepancy between the threshold energies found in this work and those from the AC method is seen in the [110] and [111] directions. This difference is beyond the energy uncertainty δE in the present calculations and is due to the breakdown of the approximations *implicitly* assumed in the AC method: the ionization transitions *at the ionization threshold energies* are assumed to take place in parallel or antiparallel. This is correct only if the band structures are *locally* isotropic. If the materials have a direct-band gap, this approximation might be sufficient. However, the approximation becomes questionable for materials having an indirect and wide band gap, such as Si. This is probably the reason for the discrepancy between the present results and those obtained by use of the AC method. It should be emphasized, however, that it is still true that the ionization threshold energy depends on the direction of the wave vector of the initial electron. As we shall explain below in more detail, taking this fact into account is crucial in analyzing the ionization processes.

Table III lists the number, C_{ii}, of possible ionization transition events and the average matrix element, $\langle T_{ii} \rangle$, along the [100], [110], and [111] directions. Here, we have considered only the lowest conduction band ($n_1=5$) for the initial electron (state 1). Note that C_{ii} is directly related to the energy and momentum conservation under a realistic band structure and corresponds to $g_{ii}(n_1,\mathbf{k}_1)$ in Eq. (10) *with constant matrix elements T_{ii} (=1) for each ionization event.* The most important feature seen here is that C_{ii} is strongly dependent on the wave vector of the initial electron, as opposed to the wave-vector dependence of $\langle T_{ii} \rangle$. As the electron energy increases, the magnitude of C_{ii} increases several orders. This suggests that, when the electron energy is above the threshold energies, the number of possible ionization transitions becomes so large that the wave-vector dependence of the matrix elements would be averaged out, as we mentioned before. Figure 6 shows the average matrix elements $\langle |M_D|^2 \rangle$, $\langle |M_E|^2 \rangle$, and $\langle T_{ii} \rangle$ as a function of the number of ionization events for the initial electron with $\mathbf{k}_1=(0,0,0)$. The average matrix elements do not vary significantly as the number of the ionization events increases and, therefore, the wave-vector dependence of the matrix element is indeed averaged out. As a consequence, the wave-vector dependence of the ionization probability calculated from Eq. (12) is dominated by C_{ii}. This implies that the wave-vector dependence of the ionization probability is manifested by the strong restrictions imposed by energy and momentum conservation under a *realistic* band structure. This fact may be used to develop a new model of the ionization probability consistent with a realistic band structure, in which the matrix elements are replaced by the *wave-vector-independent* constant and treated as a fitting parameter.[37]

Figure 7 shows the ionization probabilities $w_{ii}(n_1,\mathbf{k}_1)$ along the [100], [110], and [111] directions as a function of the wave vector for the lowest conduction band ($n_1=5$). Here, the wave vector is again normalized by the value of the wave vector at the intersection with the zone edge for each direction. Strong direction depen-

TABLE III. Number, C_{ii}, of possible ionization events and the average matrix element, $\langle T_{ii} \rangle$, of the ionization transitions along the three crystallographic directions. The initial electron is assumed in the lowest conduction band ($n_1=5$).

	$\mathbf{k}_1$	E_1 (eV)	C_{ii}	$\langle T_{ii} \rangle$
[100] direction	(0.0, 0.0, 0.0)	2.417	34 314	0.4559×10^{-2}
	(0.1, 0.0, 0.0)	2.269	24 934	0.4586×10^{-2}
	(0.2, 0.0, 0.0)	1.904	8462	0.3886×10^{-2}
	(0.3, 0.0, 0.0)	1.459	532	0.3252×10^{-2}
	(0.4, 0.0, 0.0)	1.021	6	0.4394×10^{-2}
	(0.5, 0.0, 0.0)	0.653	0	
[110] direction	(0.0, 0.0, 0.0)	2.417	34 314	0.4559×10^{-2}
	(0.1, 0.1, 0.0)	2.502	37 674	0.4737×10^{-2}
	(0.2, 0.2, 0.0)	2.726	44 783	0.4968×10^{-2}
	(0.3, 0.3, 0.0)	2.768	38 213	0.3095×10^{-2}
	(0.4, 0.4, 0.0)	2.517	15 252	0.4080×10^{-2}
	(0.5, 0.5, 0.0)	1.990	664	0.1892×10^{-2}
	(0.6, 0.6, 0.0)	1.395	18	0.8492×10^{-2}
	(0.7, 0.7, 0.0)	0.865	0	
[111] direction	(0.0, 0.0, 0.0)	2.417	34 314	0.4559×10^{-2}
	(0.1, 0.1, 0.1)	2.254	21 864	0.4530×10^{-2}
	(0.2, 0.2, 0.2)	1.755	2925	0.3793×10^{-2}
	(0.3, 0.3, 0.3)	1.312	0	

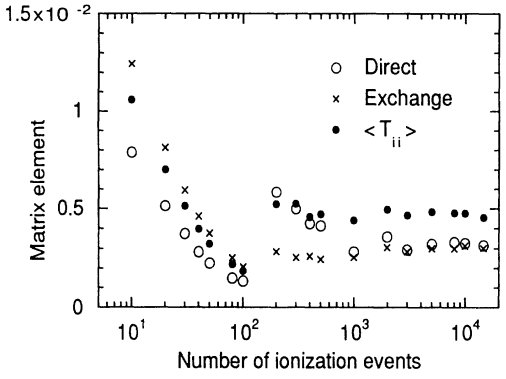

FIG. 6. Average matrix elements for the direct process, $\langle |M_D|^2 \rangle$, the exchange process, $\langle |M_E|^2 \rangle$, and $\langle T_{ii} \rangle$ as a function of the number of ionization events. The matrix elements are normalized by $[(ea)^2/(\pi V)]^2$.

dence (anisotropy) is clearly seen from the figure. It should be noted here that the anisotropy shown in Fig. 7 is associated with two factors: one due to the anisotropy implied by the band structure itself, and the other is associated with the anisotropy inherent in the ionization processes. The former factor is taken into account if one employs the isotropic ionization probability as in the Keldysh formula[7] *with* a realistic band structure along each direction. Employing the present band structure and the Keldysh formula expressed as

$$w_{\mathrm{KL}}(E)=P\frac{(E-E_{\mathrm{th}})^2}{E_{\mathrm{th}}^2}, \tag{14}$$

the ionization probability is calculated along the [100], [110], and [111] directions. Here, E is the electron energy and E_{th} the *isotropic* threshold energy given by 1.1 eV. P is a fitting parameter and is fixed by fitting $w_{\mathrm{KL}}(E)$ with the ionization probability at the Γ point of the present calculations. We have found that $P=5.79\times10^{12}$ s^{-1}. This formula has usually been used in past Monte Carlo simulations to model the ionization processes.[2-6]

Figure 8 shows the ionization probability obtained from the present calculations and from the Keldysh formula along the [100], [110], and [111] directions. The discrepancy between the two results is clearly seen in the plot, and this represents the second factor, namely that the ionization process is inherently anisotropic. In particular, the $\langle 111 \rangle$ directions show a large difference between the present *anisotropic* description and the *isotropic* Keldysh description. This results from the large difference in the ionization threshold energies used in the two cases. The Keldysh description assumes an isotropic threshold energy for all directions, while the present description imposes the direction-dependent threshold energy. Therefore, the use of *isotropic* ionization probabilities cannot be clearly justified, because the anisotropic character of the ionization processes cannot be taken into account.

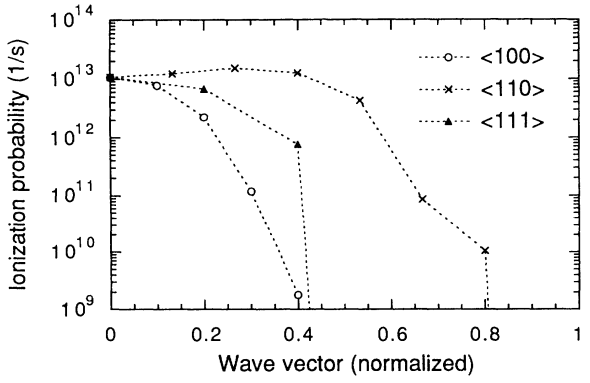

FIG. 7. Ionization probability $w_{ii}(n_1,\mathbf{k}_1)$ as a function of the wave vector $\mathbf{k}_1$ of the initial electron in state 1 along the three crystallographic directions [100], [110], and [111]. The initial electron is assumed to be in the lowest conduction band ($n_1=5$).

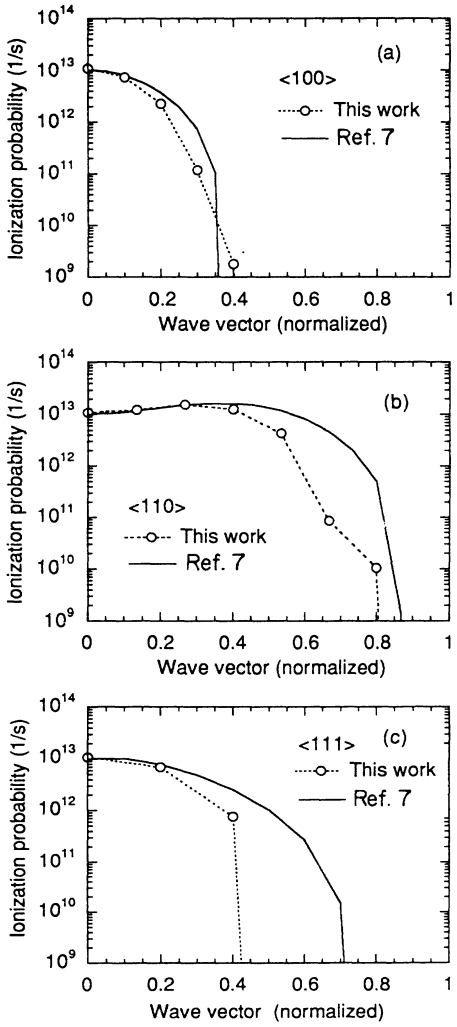

FIG. 8. Ionization probabilities obtained from the present calculations (circles with dotted curves) and from the *isotropic* Keldysh formula (solid curves) as a function of the electron wave vector along the (a) [100], (b) [110], and (c) [111] directions.

Though the present ionization probability is defined at every **k** point in the BZ, it can be expressed as a function of electron energy by averaging the ionization probability over the Si density of states, which allows us to compare our results directly with those found by others. The ionization probability is then given by

$$w_{ii}(n_1,E)=\frac{\int d^3k\, w_{ii}(n_1,\mathbf{k})\delta(E-E(n_1,\mathbf{k}))}{\int d^3k\, \delta(E-E(n_1,\mathbf{k}))} . \quad (15)$$

Figure 9 shows the ionization probability $w_{ii}(n_1,E)$ for the lowest conduction band ($n_1=5$) as a function of the electron energy, along with the ionization probabilities obtained from recent investigations.[6,17,20] The ionization probability shows the well-known "soft" behavior; that is, the probability slowly increases as the energy increases. This is again explained by the fact that Si is an indirect-band-gap material. A similar calculation for a direct- and wide-band-gap material (GaAs) is now being performed, and our preliminary result shows a rather steep increase of the ionization probability (hard threshold). Note that there are some structures in the ionization probability in Fig. 9 associated with a realistic band structure. Therefore, the ionization probability cannot be fitted by the simple Keldysh formula with a single set of parameters. This is consistent with the finding by Thoma *et al.*,[17] in which a realistic Si density of states was employed.

A large discrepancy between the present results and those of Bude *et al.*[20] is seen in the low-energy regimes. This is probably due to the different numerical methods employed in the two approaches when the ionization probability is integrated over the multidimensional wave-vector space. In our case, we explicitly carried out the integrations over the six-dimensional cubic grid. On the other hand, the Monte Carlo method was used for the integrations of Bude *et al.*[20] However, the integrands are not smooth functions of wave vector near the thresholds, but instead are strongly varying functions because of anisotropic thresholds in wave-vector space. For such functions, the Monte Carlo (random-sampling) evaluations of the integrals do not work well,[38] and this might explain the difference between the two results.

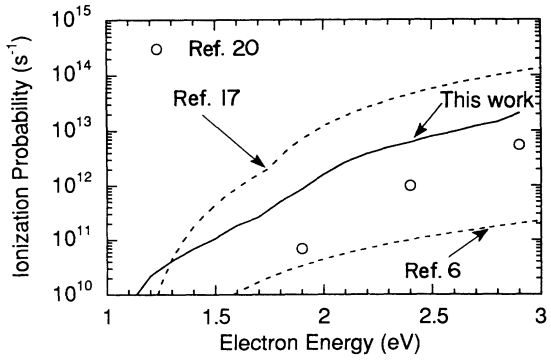

FIG. 9. Ionization probability for the initial electron in the lowest conduction band ($n_1=5$) as a function of electron energy. The wave-vector-dependent ionization probability is averaged over Si density of states. The ionization probabilities taken from Refs. 6 and 17 are also shown by dashed curves for comparison. The open circles represent the results of quantum calculations taken from Ref. 20.

We would like to stress here that the ionization probability expressed as a function of energy is somewhat misleading. When the ionization probability is expressed in terms of the electron energy, one implicitly imposes the isotropy of the ionization processes.[39] The ionization process are, however, *inherently* anisotropic, as we have shown above. Therefore, the ionization probability should be defined at every **k** point in the BZ rather than by using the electron energy. Yet, it should be noted that the anisotropic characteristics of ionization processes are not expected to be visible under *steady-state* situations because the carrier distribution is entirely spread in wave-vector space[6,16] and this would wash out any anisotropy. On the other hand, *the nearly ballistic transient transport under inhomogeneous field configurations,* in which the electron distribution in wave-vector space is expected to remain anisotropic, should be analyzed in terms of an anisotropic ionization probability like ours, because the ionization process is very sensitive to field configuration.[40]

Finally, we would like to comment on the *possible* importance of the phonon-assisted impact-ionization processes. Since the phonon-assisted processes are second-order contributions, the contributions should be small compared to the first-order contributions that we have examined here. Thus, the phonon-assisted processes have been ignored in the present work. However, the contributions from the phonon-assisted processes might become significant near threshold, where the strong restriction imposed by momentum conservation is relaxed in these processes. If this is the case, the anisotropy imposed by first-order contributions might be washed out by second-order contributions. This might explain the experimental determination that the nature of the ionization coefficients in Si is isotropic.[41] In any case, the phonon-assisted processes have not yet been clarified, and detailed investigations along the lines presented in this work are recommended.

V. CONCLUSION

We have investigated impact-ionization processes in Si by taking into account a realistic band structure obtained from an empirical pseudopotential method. The pseudo-wave-functions are used to evaluate the matrix elements of the ionization transitions consisting of the exchange process as well as the direct process, with the umklapp term associated with the periodic part of the Bloch function. The wave-vector dependence of the dielectric function is explicitly taken into account. An efficient calculational procedure has been developed and employed for the calculations of the matrix elements and the ionization probabilities. The wave-vector dependence associated with a realistic band structure of the overlap integrals,

matrix elements, ionization threshold energies, and ionization probability has been examined. It has been shown that the ionization process is inherently anisotropic and that its anisotropy (wave-vector dependence) is mainly manifested in the strong restrictions imposed by energy and momentum conservation under a realistic band structure.

ACKNOWLEDGMENTS

One of authors (N.S.) has particularly benefited from conversations with C. Jacoboni, W. L. Engl, R. Thoma, and J. Bude. The authors would like to thank K. Hirata for encouragement throughout the course of the study.

[1]See, for example, S. M. Sze, *Physics of Semiconductor Devices,* 2nd ed. (Wiley, New York, 1981).

[2]C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).

[3]C. Jacoboni and P. Lugli, *The Monte Carlo Method for Semiconductor Device Simulation* (Springer-Verlag, New York, 1989).

[4]H. Shichijo and K. Hess, Phys. Rev. B **23**, 4197 (1981).

[5]J. Y. Tang and K. Hess, J. Appl. Phys. **54**, 5139 (1983).

[6]M. V. Fischetti and S. E. Laux, Phys. Rev. B **38**, 9721 (1988).

[7]L. V. Keldysh, Zh. Eksp. Teor. Fiz. **37**, 713 (1959) [Sov. Phys.—JETP **10**, 509 (1960)].

[8]N. Sano, T. Aoki, and A. Yoshii, Appl. Phys. Lett. **55**, 1418 (1989).

[9]E. Antoncik and J. Tauc, in *Semiconductors and Semimetals,* edited by R. K. Willardson and A. C. Beer (Academic, New York, 1966), Vol. 2, p. 245.

[10]F. Capasso, in *Semiconductors and Semimetals,* edited by R. K. Willardson and A. C. Beer (Academic, New York, 1985), Vol. 22-D, p. 1.

[11]For a comprehensive review of *analytical* approaches, see D. J. Robbins, Phys. Status Solidi B **97**, 9 (1980); **97**, 387 (1980); **98**, 11 (1980).

[12]For example, E. O. Kane, in *Semiconductors and Semimetals,* edited by R. K. Willardson and A. C. Beer (Academic, New York, 1966), Vol. 1, p. 75.

[13]A. R. Beattie, R. A. Abram and P. Scharoch, Semicond. Sci. Technol. **5**, 738 (1990).

[14]S. W. Corzine, R. H. Yan, and L. A. Coldren, Appl. Phys. Lett. **57**, 2835 (1990).

[15]Y. Jiang, M. C. Teich, and W. I. Wang, Appl. Phys. Lett. **57**, 2922 (1990).

[16]N. Sano, T. Aoki, M. Tomizawa, and A. Yoshii, Phys. Rev. B **41**, 12 122 (1990).

[17]R. Thoma, H. J. Peifer, W. L. Engl, W. Quade, R. Brunetti, and C. Jacoboni, J. Appl. Phys. **69**, 2300 (1991).

[18]M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).

[19]E. O. Kane, Phys. Rev. **159**, 624 (1967).

[20]J. Bude, K. Hess, and G. J. Iafrate, in *Computational Electronics: Semiconductor Transport and Device Simulation,* edited by K. Hess, J. P. Leburton, and U. Ravaioli (Kluwer Academic, Norwell, U.K., 1991), p. 131.

[21]D. B. Laks, G. F. Neumark, A. Hangleiter, and S. T. Pantelides, Phys. Rev. Lett. **61**, 1229 (1988).

[22]The "pseudo-" wave-function is, of course, not correct near the core region, where the true wave function strongly oscillates. However, the energy range of concern here is approximately up to a few eV from the bottom of the conduction band in which the detailed form of the wave function near the core is not important.

[23]H. Nara and A. Morita, J. Phys. Soc. Jpn. **21**, 1852 (1966).

[24]It should be noted that this "umklapp" process is different from the usual umklapp processes implied by Eq. (2).

[25]S. Brand and R. A. Abram, J. Phys. C **17**, L571 (1984).

[26]See, for example, L. I. Schiff, *Quantum Mechanics* (McGraw-Hill, New York, 1968).

[27]K. Hess, *Advanced Theory of Semiconductor Devices* (Prentice-Hall, Princeton, NJ, 1987), p. 25.

[28]The basis set of more than 139 **G** vectors is not invariant under the symmetry transformations, and this fact limits the number of **G** vectors in the expansion of the periodic part of the Bloch function in Eq. (1).

[29]The value of the energy interval δE is somewhat arbitrary, and we have used the same value used by Kane in Ref. 19.

[30]S. Brand and R. A. Abram, J. Phys. C **17**, L201 (1984).

[31]M. G. Burt, S. Brand, C. Smith, and R. A. Abram, J. Phys. C **17**, 6385 (1984).

[32]M. Takeshima, Phys. Rev. B **29**, 1993 (1984).

[33]The overlap integrals I_{vc} along the [100] direction and I_{vc} between $n=4$ and 5 are very close to zero and have been omitted in the plot here.

[34]The average value of the momentum transfer is defined here as the value obtained when good convergence of the matrix elements is attained in Eq. (6). Typically, more than 20 000 terms are summed over the reciprocal-lattice vectors.

[35]R. Resta, Phys. Rev. B **16**, 2717 (1977).

[36]C. L. Anderson and C. R. Crowell, Phys. Rev. B **5**, 2267 (1972).

[37]N. Sano, M. Tomizawa, and A. Yoshii, in *Proceedings of the 1991 International Conference on Solid State Devices and Materials* (Business Center for Academic Societies–Japan, Tokyo, 1991), p. 456.

[38]The Monte Carlo method, of course, works well for such functions if one can find a *proper* weighting function.

[39]If the full band structure is incorporated for the electron's *kinematics* through, for example, Monte Carlo simulations, the resulting transport characteristics could be anisotropic. However, this simply incorporates the anisotropy associated with the band structure itself.

[40]N. Sano, M. Tomizawa, and A. Yoshii, in *Computer Aided Innovations of New Materials,* edited by M. Doyama, T. Suzuki, J. Kihara, and R. Yamamoto (Elsevier, Amsterdam, 1991), p. 745.

[41]V. M. Robbins, T. Wang, K. F. Brennan, K. Hess, and G. E. Stillman, J. Appl. Phys. **58**, 4614 (1985).