Full-band Monte Carlo simulation of high-energy carrier transport in single photon avalanche diodes: Computation of breakdown probability, time to avalanche breakdown, and jitter

Denis Dolgos, Hektor Meier, Andreas Schenk, and Bernd Witzigmann

Related Articles

Ultra-low noise single-photon detector based on Si avalanche photodiode
Rev. Sci. Instrum. 82, 093110 (2011)

GaN/SiC avalanche photodiodes
Appl. Phys. Lett. 99, 131110 (2011)

Influence of the metallic contact in extreme-ultraviolet and soft x-ray diamond based Schottky photodiodes
J. Appl. Phys. 110, 054513 (2011)

Organic position sensitive photodetectors based on lateral donor-acceptor concentration gradients
APL: Org. Electron. Photonics 4, 194 (2011)

Organic position sensitive photodetectors based on lateral donor-acceptor concentration gradients
Appl. Phys. Lett. 99, 103305 (2011)

Additional information on J. Appl. Phys.

Journal Homepage: http://jap.aip.org/

Journal Information: http://jap.aip.org/about/about_the_journal

Top downloads: http://jap.aip.org/features/most_downloaded

Information for Authors: http://jap.aip.org/authors

Full-band Monte Carlo simulation of high-energy carrier transport in single photon avalanche diodes: Computation of breakdown probability, time to avalanche breakdown, and jitter

Denis Dolgos,[1,a)] Hektor Meier,[1] Andreas Schenk,[1] and Bernd Witzigmann[2]

[1]*Integrated Systems Laboratory, ETH Zurich, Gloriastrasse 35, 8092 Zurich, Switzerland*
[2]*Computational Electronics and Photonics Group, University of Kassel, Wilhelmshöher Allee 71, 34121 Kassel, Germany*

The high-energy charge transport of electrons and holes in GaAs single photon avalanche diodes with multiplication region widths of 55 nm to 500 nm is investigated by means of the full-band Monte Carlo technique incorporating computationally efficient full-band phonon scattering rates. Compared to previous works, the solution of the Boltzmann transport equation and the incorporation of the full-band structure put the evaluation of the breakdown probability, the time to avalanche breakdown, and the jitter on deeper theoretical grounds. As a main result, the breakdown probability exhibits a steeper rise versus reverse bias for smaller multiplicator sizes. The time to avalanche breakdown and jitter decrease for smaller multiplicator widths. © *2011 American Institute of Physics*. [doi:10.1063/1.3652844]

I. INTRODUCTION

Single or few photon detection has become an important feature in optoelectronic systems. The applications of single photon avalanche diodes (SPADs) comprise quantum computing,[1] quantum cryptography,[2] fundamental studies of quantum physics,[3] three-dimensional laser detection and ranging imaging,[4] free space optical communication,[5] semiconductor circuit diagnostics,[6] and fluorescence lifetime imaging in molecular biology and medicine.[7]

SPADs are reverse biased diodes operated above the breakdown voltage V_b in the Geiger-mode. The name Geiger-mode stems from the similar breakdown process in Geiger-Müller detectors for radioactive particles. For desired near infrared (NIR) photon detection, a structure with separated absorption and multiplication regions is utilized. The absorber is made of a low bandgap semiconductor whereas the multiplication layer consists of a semiconductor material with a wider bandgap to limit tunneling to an acceptable level.

A photogenerated carrier drifts into the multiplication layer and causes impact ionization. The newly impact ionized electron-hole pair can impact ionize additional pairs leading to a self-sustaining avalanche. An active or passive quenching circuit stops the avalanche feedback by regulating the applied bias below the breakdown voltage. After a hold-off time, the SPAD is biased above the breakdown voltage again. Especially devices for NIR photon counting suffer from carrier trapping and detrapping by thermal emission at defects within the multiplication layer. If the SPAD is in its armed state, a detrapped carrier can trigger a dark count named afterpulsing. Remarkable progress has been achieved for SPADs regarding their photon detection efficiency (PDE), dark count rate, jitter, and afterpulsing.[8,9]

The PDE is the product of the quantum efficiency η_q, the probability that the photoexcited carrier survives into the multiplier P_c, and the breakdown probability P_b that the carrier activates a self-sustaining avalanche:[9]

$$\text{PDE} = \eta_q P_c P_b. \tag{1}$$

The PDE depends on the electric field mainly via P_b. Concerning the PDE, a high electric field is favorable because P_b rises with the field. On the other hand, for increasing electric fields SPAD performance degrades owing to dark currents as a result of band-to-band or trap-assisted tunneling in the multiplicator. Hence, a steep rise of the breakdown probability with higher electric fields is desirable. The timing jitter originates from the different transit times of the carriers owing to the variation of the location of photon absorption, from the delayed release of trapped carriers at possible heterojunctions, and from fluctuations of the avalanche build-up time. More precisely, the jitter by virtue of the avalanche build-up originates from the randomness of the impact ionization process, the expansion of the initially point-like avalanche to the entire high field region, and from local nonuniformities of the excess bias. The avalanche build-up time is the main contribution to the timing jitter.[9]

There have been contradictory predictions concerning the dependence of the breakdown probability on the multiplication region width w. References 10 and 11 predict a rising breakdown probability for thicker multiplication regions using a history-dependent analytical impact ionization model and the recursive dead-space multiplication theory, respectively. On the other hand, Refs. 12–14 predict the opposite behavior applying the hard dead-space impact ionization model within McIntyre's extended theory, the recurrence equations by McIntyre, and the stochastic random path length model, respectively. The models of Refs. 10–14 rely on simplified impact ionization and charge transport modeling without taking scattering and the dispersion of charge

a)Electronic mail: dolgos@iis.ee.ethz.ch.

© 2011 American Institute of Physics

carriers on a microscopic level into account. Nonequilibrium effects like the dead-space, the velocity overshoot, and the nonlocal impact ionization gain importance for shrinking multiplication region sizes. The dead-space is the distance before a carrier gains enough energy for impact ionization by means of the electric field. In a velocity overshoot, the carrier velocity exceeds its stationary value substantially in a short transition region due to a rapidly changing electric field. In local impact ionization models, the rate is a function of the local electric field in contrast to nonlocal models where the ionization rate depends on the carrier energy. In simplified models, the dead-space effect is estimated analytically or taken from full-band Monte Carlo (FBMC) simulations, and the velocity overshoot and the nonlocal impact ionization are neglected. The simulation of high-energy (approximately 5 eV) carrier dynamics and kinetics renders the incorporation of the full-band structure necessary. The Boltzmann transport equation (BTE) governs the dynamics of the particles in the semiclassical transport regime. The Monte Carlo (MC) technique allows the solution of the BTE with the most comprehensive band structure description and scattering models. At present, the full-band Monte Carlo method is the most accurate device simulation method within the physics of semiclassical charge transport.[15,16] The FBMC approach for the solution of the BTE serves as benchmark for approximate methods. FBMC simulations involve a high computational burden, which forced previous research to compute SPAD breakdown characteristics to apply simplified charge transport models. However, with nowadays parallel CPU power on standard computer clusters combined with computationally efficient approaches[15] it has become feasible to gain sufficient statistics with FBMC simulations for the evaluation of breakdown probabilities and standard deviations of variables of interest. For an introduction into the MC method consult, for instance, Refs. 17 and 18.

This work concentrates on the investigation of the high-energy charge transport and multiplication process in the SPAD multiplicator. The effect of tunneling is not considered. We aim to clarify the contradictory predictions of the breakdown probability versus the electric field with changing multiplication region widths by solving the semiclassical Boltzmann transport equation by means of the ensemble FBMC method. This work examines the breakdown probability, which is the main contribution to the PDE, the time to avalanche breakdown t_b, its standard deviation and main contribution to the timing jitter $\sigma = \sqrt{\langle t_b^2 \rangle - \langle t_b \rangle^2}$. The objects of investigation are GaAs PIN diodes with multiplication (intrinsic) region widths between 55 nm and 500 nm operated in the Geiger-mode at 300 K (compare with Ref. 14).

II. FULL-BAND MONTE CARLO MODEL

Surely, the ultimate goal in the simulation of carrier transport in semiconductors would be the usage of the ion potential as the only input. However, as a result of the huge computational effort, approximations have to be made while keeping the main physical features.[19] In the case of a full-band computation of breakdown characteristics of SPADs efficient modeling renders the computation feasible.

A. Full-band structure

Due to the 48-fold symmetry of the face-centered cubic lattice, it is sufficient to compute the dispersion relation and its derivatives on the 1/48th part of the first Brillouin zone, namely the irreducible wedge. We use a three-dimensional equidistant tensor grid to mesh the irreducible wedge.[20] The length of a cubic box is $l = 0.01 \cdot 2\pi/a$ leading to a total number of 87125 cubes discretizing the irreducible wedge. The one-dimensional tensor grid for the real-space discretization has a spacing of 1 nm. We compute the GaAs full-band structure on the **k**-space grid by means of the empirical pseudopotential method using parameters from Ref. 21. Figure 1 displays the dispersion relation along a special path through the Brillouin zone. The energy is expanded linearly around a grid point $\mathbf{k}_c$:

$$E(\mathbf{k}) = E(\mathbf{k}_c) + \nabla_{\mathbf{k}} E(\mathbf{k}_c)(\mathbf{k} - \mathbf{k}_c) \tag{2}$$

meaning that the gradient, and hence the velocity $\mathbf{v} = \nabla_{\mathbf{k}} E(\mathbf{k})/\hbar$, are constant within a reciprocal space discretization box. The dispersion relation has to be inverted for the selection of a final **k**-state after scattering. For a speed-up in the final box search, we use precomputed energy interval lists described in Ref. 15.

As a result of the high curvature of the Γ_6-valley, the utilized discretization is not sufficient to resolve the dispersion for low-energy transport. Therefore, we use a threshold energy of 0.4 eV for analytical nonparabolic treatment of the Γ_6-valley[22] with parameters taken from Ref. 23.

B. Particle propagation

The carriers propagate according to Newton's law[24] with a simple time-step propagation scheme.[25] A particle propagates for the minimum of the four times $\Delta t = \min(t_s, t_k, t_r, t_T)$ (Ref. 24), where t_s is the stochastically selected time for a scattering event with a random number r between 0 and 1 (Ref. 18), t_k is the flight time to reach the boundary of a cubic **k**-space box, t_r is the time till the boarder of the real space grid, and t_T is the time until the time step of the synchronous ensemble. If t_s is the minimum time, a scattering
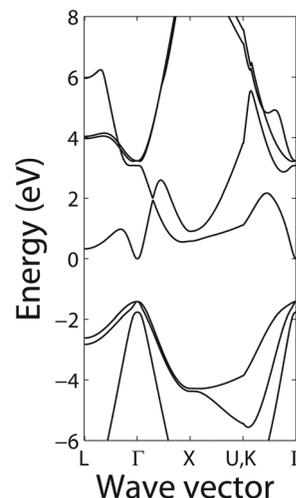

FIG. 1. Full-band structure of GaAs along special path through the Brillouin zone computed with the empirical pseudopotential method.

event and a final **k**-state are selected randomly. Otherwise, it is allowed to stop the propagation at t_k, t_r, $t_T < t_s$ and to start the algorithm again.[25] The electric field profile for the PIN structure is precomputed with a drift-diffusion model.[26] As a result of the constant electric field during the propagation, the integrated equation of motion is

$$\mathbf{k}(t) = \mathbf{k}(t_0) + \frac{q}{\hbar}\mathbf{E}\Delta t. \tag{3}$$

Owing to the small propagation time, which allows the disregard of the acceleration term,[24] and the constant velocity during the propagation time, the real space position is updated corresponding to

$$\mathbf{r}(t) = \mathbf{r}(t_0) + \mathbf{v}\Delta t. \tag{4}$$

If the particle leaves the first Brillouin zone, the carrier is transformed back by means of an Umklapp process $\mathbf{k}_{BZ} = \mathbf{k} + \mathbf{G}$ with the Umklapp vector **G**.

C. Scattering rates

All scattering rate computations are based on Fermi's golden rule.[15,17,22] One has to be aware that in the derivation of Fermi's golden rule an assumption is made that limits its usage for high electric fields. The energy conserving delta function appears for $t \gg \hbar/E$, which depicts Heisenberg's uncertainty principle. Rigorous energy conservation is not fulfilled giving rise to collision broadening.[27] The intracollisional field effect accounts for the issue of the influence of the electric field during the scattering process.[28] Works exist that include the intracollisional field effect and collision broadening into the semiclassical transport theory.[29–31] For high carrier energies, the picture of semiclassical charge transport is pushed to its limits (e.g. see Ref. 28 for assumptions made in the Boltzmann equation). On the other hand, the range of validity of the Boltzmann transport equation may have been underestimated.[22] Nevertheless, it is common practice in the community of semiclassical high-field transport to use Fermi's golden rule for scattering rate computations.[16,32,33]

The full-band Monte Carlo simulators described in literature differ mainly in the usage of the implemented scattering model. For phonon scattering the differences lie in the treatment of the deformation potential, the overlap integral, and the phonon dispersion relation. Commonly, phonons are treated to be in a thermodynamic heat bath in equilibrium. On the most basic side of the hierarchy, starting from Fermi's golden rule, the full transition rate $W_{nn'}(\mathbf{k}, \mathbf{k}')$ has to be computed from every initial $(n, \mathbf{k})$ to any final state $(n', \mathbf{k}')$ with the band index n. The deformation potentials are not fit parameters but are computed based on the full-band structure. This means that there is less freedom for calibration and the quality of the transition rates depends on the complexity of the underlying theory.[19] The full phonon dispersion is included. References 16, 34, and 35 contributed to this research-intensive and computationally heavy approach. However, the feasibility and the efficiency of device simulation ask for approximations of the full transition matrix elements. It is possible to define volumes or valleys in the reciprocal space where the deformation potentials and the matrix elements are approximately constant.[18,32–34,36] Thus, the main contribution

to the variation of the phonon scattering rates with energy is due to the density of states (DOS). This allows a computationally very efficient grouping of the phonon transition rates into packages of phonon scattering mechanisms depending on the initial and final valleys.[15] Appendix A introduces the valley allocation algorithm. The first four conduction valleys are Γ_6, L_6, X_6, for the first conduction band, and X_7 for the second conduction band. All other conduction valleys within the irreducible wedge are named c_{xy} where x is the conduction band number and where y is the valley number within x.

1. Carrier-phonon scattering

Elastic acoustic phonon scattering, nonpolar optical phonon scattering, and intervalley phonon scattering may be written as[15,17,22]

$$
\begin{aligned}
W^m_{\nu\nu'}(E) &= \sum_{\mathbf{k}'} \frac{\Lambda^m_{\nu\nu'}}{V} \delta(E_{\nu'}(\mathbf{k}') - E_\nu(\mathbf{k}) \mp E_{\text{trans}}) \\
&= \Lambda^m_{\nu\nu'} \mathcal{D}_{\nu'}(E \pm E_{\text{trans}})
\end{aligned}
\tag{5}
$$

being proportional to the final density of states $\mathcal{D}_{\nu'}(E)'$ with the particle energy E before and $E' = E \pm E_{\text{trans}}$ after scattering. The transition energy is $E_{\text{trans}} = 0$ for acoustic phonons, $E_{\text{trans}} = \hbar\omega_{\text{op}}$ for optical phonons in the Einstein approximation, the optical phonon angular frequency ω_{op}, $E_{\text{trans}} = \hbar\omega_{\nu\nu'} + \Delta E_{\nu\nu'}$ for intervalley phonons with the intervalley transition phonon angular frequency $\omega_{\nu\nu'}$ and the energy separation $\Delta E_{\nu\nu'}$ between the valleys. The prefactor of scattering mechanism type m is $\Lambda^m_{\nu\nu'}$ (see Table I), where ν and ν' are the initial and final valley, respectively. We utilize the method by Gilat and Raubenheimer[37] to integrate δ-functions. The full-band DOS, including spin degeneracy, is evaluated by means of a surface integral

$$
\begin{aligned}
\mathcal{D}_\nu(E) &= \frac{1}{4\pi^3} \int_{V_\nu} \delta(E - E_{\nu'}(\mathbf{k}')) d^3k' \\
&= \frac{1}{4\pi^3} \int_{E_\nu(\mathbf{k})=E(\mathbf{k})} \frac{dA}{|\nabla_\mathbf{k} E_\nu(\mathbf{k})|} \approx \sum_i \mathcal{D}_{\nu,i}(E)
\end{aligned}
\tag{6}
$$

with the box density of states

TABLE I. Scattering rate prefactors Λ. The quantities are: Boltzmann constant k_B, lattice temperature T, longitudinal sound velocity u_l, mass density of the semiconductor material ρ, optical and intervalley phonon occupation number n_{op} and $n_{\nu\nu'}$, polar coupling constant in Fröhlich expression F, impact ionization matrix element M_{ii}, and number of unit cells N.

Scattering mechanism	Λ
Elastic acoustic phonon scattering	$\dfrac{\pi k_B T \Xi_\nu^2}{\hbar u_l^2 \rho}$
Nonpolar optical phonon scattering	$\dfrac{\pi D_\nu^2}{2\rho\omega_{\text{op}}}\left(n_{\text{op}} + \frac{1}{2} \mp \frac{1}{2}\right)$
Intervalley phonon scattering	$\dfrac{\pi D_{\nu\nu'}^2}{2\rho\omega_{\text{vv}'}}\left(n_{\nu\nu'} + \frac{1}{2} \mp \frac{1}{2}\right)$
Polar optical phonon scattering	$\dfrac{2\pi}{\hbar}e^2 F^2\left(n_{\text{op}} + \frac{1}{2} \mp \frac{1}{2}\right)$
Impact ionization	$\dfrac{2\pi}{\hbar}|M_{ii}|^2 \dfrac{V^3}{N}$

$$\mathcal{D}_{\nu,i}(E) = \frac{1}{4\pi^3} \frac{A_{\nu,i}(E)w_i}{\hbar|\mathbf{v}_{\nu,i}|} \tag{7}$$

of the ith discretization cube. Here, $A_{\nu,i}(E)$ is the area of the equi-energy surface cut with the cubic box.[37] If the energy E is not within the cube, the area is zero. Furthermore, w_i is a weighting function for the treatment of the contribution of boundary boxes in the irreducible wedge (see Appendix B). Figures 2 and 3 present the single contributions of the defined valleys to the total DOS.

The electron intervalley deformation potentials $D_{\nu\nu'}$ result from the acoustic Ξ_ν and the nonpolar optical D_ν deformation potentials[22]

$$D_{\nu\nu'} = \sqrt{|\Xi_\nu \mathbf{q}_{\nu\nu'}|^2 + (D_\nu)^2} \tag{8}$$

with the phonon wave vectors $\mathbf{q}_{\nu\nu'}$ connecting the valley minima. For the first conduction band, the intervalley phonon energies are taken from Ref. 38, and 23 meV for higher conduction bands.

Compared to group IV semiconductors, in III-V materials additionally polar optical phonon scattering appears. Approximating the overlap integral as unity, the polar optical phonon scattering rate[22] is given by

$$W_{\nu'}^{\text{pop}}(\mathbf{k}) = \sum_{\mathbf{q}} \frac{\Lambda^{\text{pop}}}{V\mathbf{q}^2} \delta(E_{n'}(\mathbf{k}') - E_n(\mathbf{k}) \mp \hbar\omega_{\text{op}}) \tag{9}$$

$$= \Lambda^{\text{pop}} \mathcal{D}_{q^{-2},\nu'}(\mathbf{k})$$

being proportional to the direction-weighted density of states

$$\mathcal{D}_{q^{-2},\nu'}(\mathbf{k}) = \frac{1}{4\pi^3} \int \frac{1}{|\mathbf{q}|^2} \delta(E(\mathbf{k}) - E_{\nu'}(\mathbf{k}') \mp \hbar\omega_{\text{op}})d^3k'$$

$$\approx \sum_i \mathcal{D}_{q^{-2},i}(\mathbf{k}) \tag{10}$$

with the direction-weighted box density of states

$$\mathcal{D}_{q^{-2},\nu,i}(\mathbf{k}) = \frac{\mathcal{D}_{\nu,i}(E(\mathbf{k}) \pm \hbar\omega_{\text{op}})}{(\mathbf{k} - \mathbf{k}_c)^2} \tag{11}$$

of the ith box. Here, $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ is the phonon wave vector. The anisotropic polar optical phonon scattering rate is averaged according to[22]

$$W(E) = \frac{1}{\mathcal{D}(E)} \sum_{\mathbf{k}} W(\mathbf{k})\delta(E(\mathbf{k}) - E) \tag{12}$$

with the initial density of states $\mathcal{D}(E)$. Figure 4 depicts the polar optical phonon scattering rates for intraband transitions.
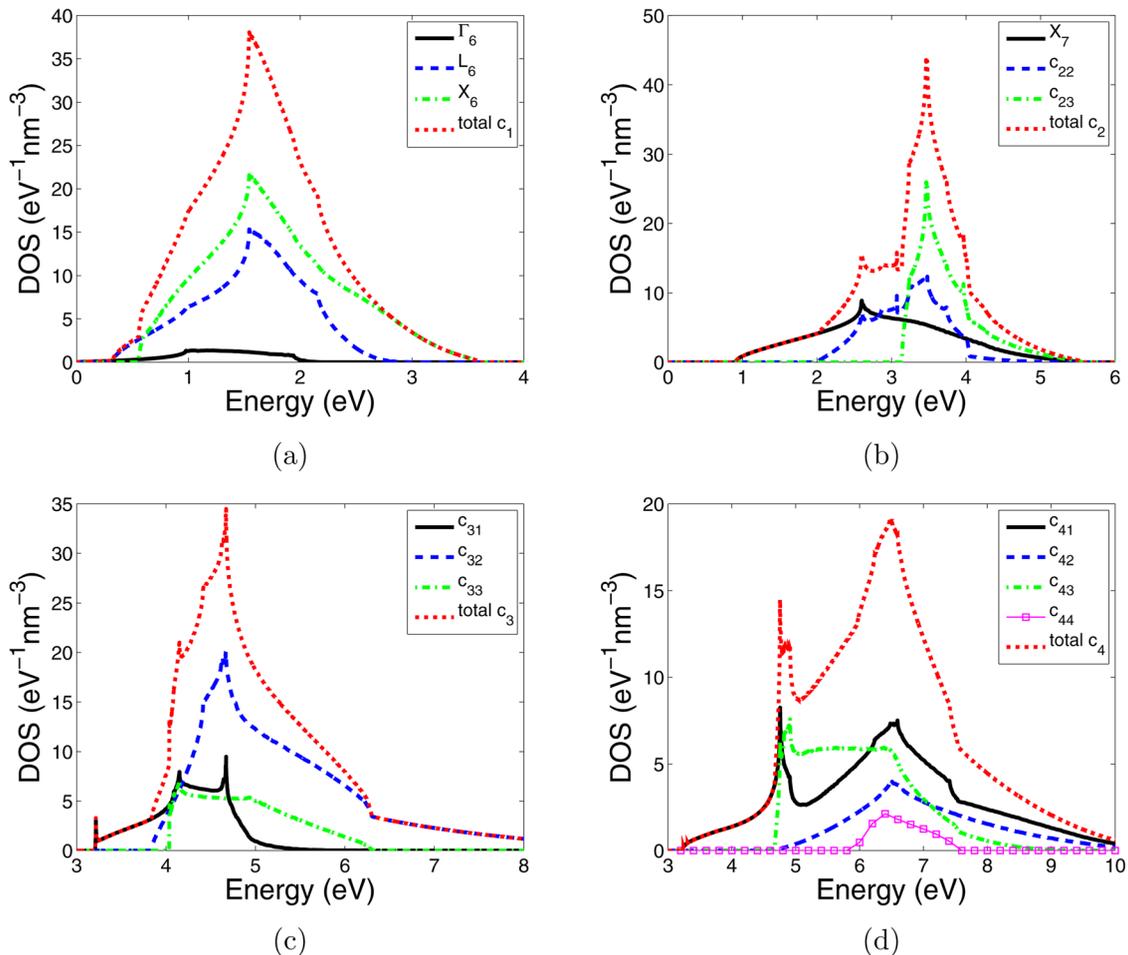

FIG. 2. (Color online) Full-band density of states of valleys of conduction bands 1 (a), 2 (b), 3 (c), and 4 (d).
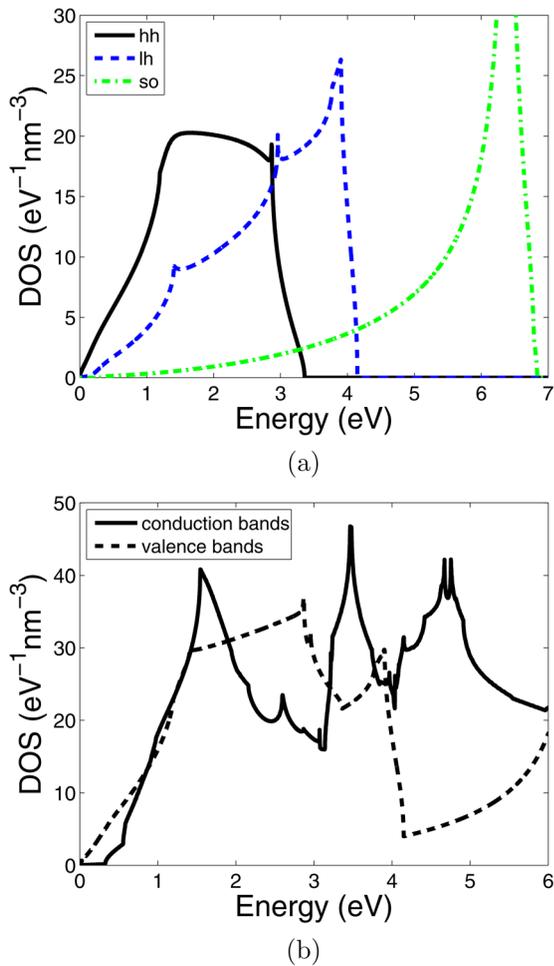
FIG. 3. (Color online) Full-band density of states of valence bands (a) and total density of states of conduction and valence bands (b).

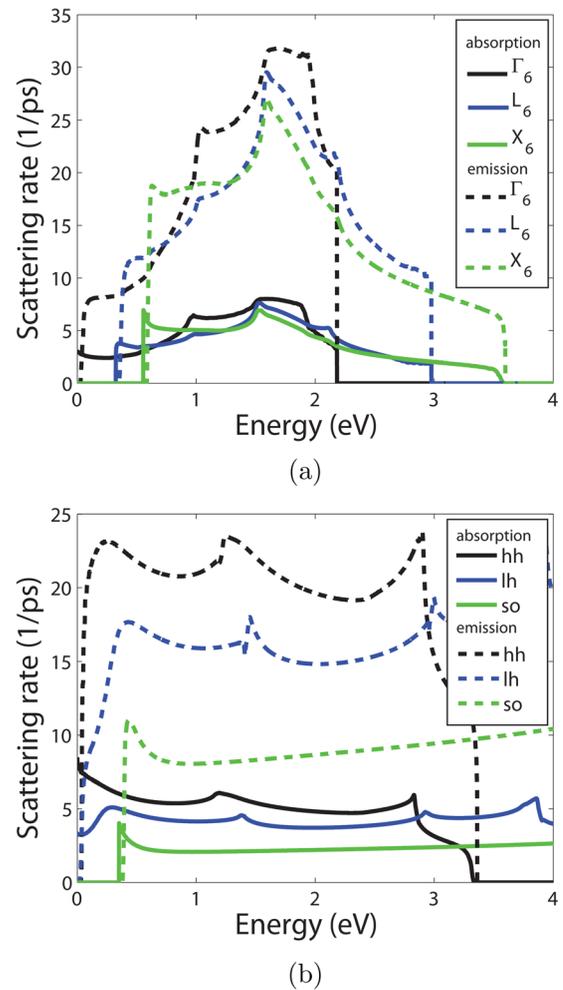FIG. 4. (Color online) Intraband polar optical phonon scattering rates for the valleys of the first conduction band (a) and valence bands (b).

2. Impact ionization

Models for the calculation of the impact ionization rates, starting from Fermi's golden rule, vary widely in literature. For example, Kane,[39] Jung *et al*.,[40] and Kuligk *et al*.[41] computed the impact ionization rate evaluating the detailed matrix elements, the dielectric function, and took momentum conservation into account. Kane introduced the random-**k** approximation (RKA),[39] which neglects momentum conservation. Sano *et al*.[42] additionally introduced an approximation that assumes mean secondary carrier energies. Some FBMC works utilized a modified Keldysh formula, which is a fit to the impact ionization rate of fundamental approaches. In the past, several groups used the Keldysh formula, which provides bad scattering rates being derived for parabolic bands.

The inclusion of the momentum conservation and the computation of the Coulomb transition matrix elements render the fundamental approaches of Kane, Jung *et al*., and Kuligk *et al*. computationally expensive and their numerical implementation laborious. The transition matrix elements are approximately insensitive to changes of the primary electron energies. Momentum conservation allows a large number of possible final **k**-states due to the 48-fold symmetry of the cubic lattice. Practically, the large set of final **k**-states will

scatter randomly throughout the Brillouin zone. Additionally, the three integrals over **k**-space considerably average the details of the band structure.[39] The RKA and the constant matrix approximation (CMA) take advantage of the two properties and highly reduce the complexity of the impact ionization rate computation. The RKA and CMA reduce the nine-dimensional integration over the δ-function in reciprocal space to a two-dimensional integral over one-particle density of states in energy space. The costly computation of the Coulomb transition matrix elements is bypassed by tuning it to experimental data.[33] The application of the RKA and the CMA is an appealing possibility to compute the impact ionization rate with a manageable effort and keeping agreement with fundamental methods.[33,39,43] Compared to the approach of Sano *et al*. Kane's RKA provides information about the secondary carrier energies without taking another fit parameter into account. The additional implementation work for the RKA and CMA based rate calculation compared to the implementation of the approach of Sano *et al*. is low. The use of the modified Keldysh formula is a viable option especially for GaAs because the secondary carrier energies are provided in literature. However, we prefer Kane's RKA and CMA to compute all our scattering rates based on the same full-band structure, and therefore, keep

band structure consistency. The importance of the impact ionization rate around the threshold energy depends on the high energy tail of the carrier distribution and the ratio of the impact ionization rate to the phonon scattering rates. A carrier, being able to impact ionize, has to survive to energies above threshold emitting less phonons than the bulk of particles. The carrier distribution above the threshold energy therefore strongly depends on phonon scattering below the threshold. If the carrier has survived to impact ionization enabling energies, the occurrence of impact ionization has to be relevant compared to phonon scattering. Practically, impact ionization takes place for energies where the ratio of the impact ionization rate and the total phonon scattering rate exceeds about 10^{-2}. The energy before impact ionization is between 3 and 5 eV for electrons, i.e., high above the threshold energy. For holes the impact ionization rate seems to be important near the threshold[33] where the RKA and CMA are expected to be less applicable.[39]

The impact ionization scattering rate in the RKA and CMA is given by[39]

$$W^{\mathrm{ii}}(E) = \Lambda^{\mathrm{ii}} \mathcal{D}_{\mathrm{ii}}(E) \tag{13}$$

for an impacting electron with

$$\mathcal{D}_{\mathrm{ii}}(E_c) = \sum_{\substack{v',c',c'' \\ c'' \geq c'}} \int_0^{\hat{E}_{c'}} dE_{c'} \int_0^{\hat{E}_{v'}} dE_{v'} \mathcal{D}_{v'}(E_{v'}) \mathcal{D}_{c'}(E_{c'})$$
$$\times \mathcal{D}_{c''}(E_c - E_{c'} - E_{v'} - E_g). \tag{14}$$

Here, it is $E_{c/v} = E(\mathbf{k}_{c/v})$ and so on. The integration boundaries are $\hat{E}_{c'} = E_c - E_g$ and $\hat{E}_{v'} = E_c - E_{c'} - E_g$. Furthermore, E_g is the bandgap energy. Swapping the indices for the conduction band c and valence band v leads to the expression for a primary impacting hole. Table I summarizes the scattering mechanism prefactors Λ.[17,22] Figure 5 demonstrates the total phonon and impact ionization scattering rates for electrons and holes.

We compute the secondary carrier energies according to the secondary distribution function in the RKA. For a primary electron having the energy E_c and a secondary electron with the energy E_c' the secondary distribution function W_{ee} is given by[39]

$$W_{\mathrm{ee}}(E_c, E_c') = \frac{2}{\mathcal{D}_{\mathrm{ii}}(E_c)} \sum_{\substack{v',c',c'' \\ c'' \geq c'}} \mathcal{D}_{c'}(E_{c'})$$
$$\times \int_0^{E_c - E_{c'} - E_g} dE_{v'} \mathcal{D}_{v'}(E_{v'}) \mathcal{D}_{c''}(E_c - E_{c'} - E_{v'} - E_g). \tag{15}$$

The factor of 2 appears as a result of the indistinguishable secondary electrons. Due to computational efficiency, we use the mean value of the secondary energy distribution function:

$$\langle E_c' \rangle(E_c) = \sum_{E_c'} W_{\mathrm{ee}}(E_c, E_c') E_c'. \tag{16}$$

Figure 6 illustrates the evaluated mean secondary energies $\langle E_c' \rangle(E_c)$ for an impacting electron and the secondary elec-
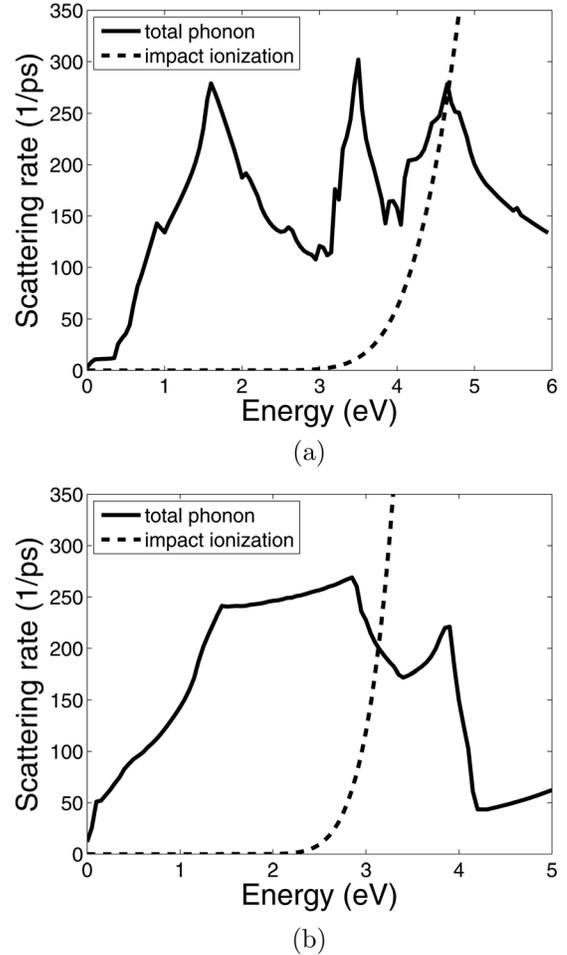

FIG. 5. Full-band scattering rates for electrons (a) and holes (b).

trons, and $\langle E_v' \rangle(E_v)$ for a primary hole and secondary holes. For a primary impacting electron the secondary carrier energies for the two electrons E_{e1}', E_{e2}' and the single hole E_h' are chosen according to

$$E_{e1}' = \langle E_c' \rangle(E_c) r, \tag{17a}$$

$$E_{e2}' = \langle E_c' \rangle(E_c) - E_{e1}', \tag{17b}$$

$$E_h' = E_c - \langle E_c' \rangle(E_c) - E_g \tag{17c}$$

with a random number r between 0 and 1. The selection of the secondary carrier energies for a primary impacting hole is treated accordingly.

The final state after scattering is chosen corresponding to Ref. 15.

D. Calibration

As a result of the empirical deformation potential and impact ionization ansatz, the unknowns (deformation potentials and impact ionization prefactors) have to be calibrated to experimental data.[22,33] Figures 7(a) and 7(b) show a comparison of the velocity versus electric field dependence between our FBMC simulator named *CarloS,* other FBMC programs, and experimental data. Figures 8(a) and 8(b) compare the impact ionization coefficients of electrons α and holes β simulated with *CarloS* and experiments. The bars of
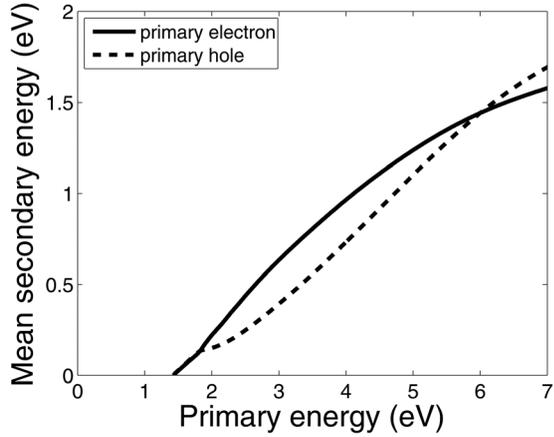
FIG. 6. Mean energy of secondary carriers of the same type as impacting ones after impact ionization.

length 2σ indicate the 68% confidence interval. The impact ionization prefactor Λ^{ii} depends exponentially on the deformation potentials. A strong carrier-phonon interaction causes a distribution function with exponentially lower high-energy tail. Therefore, to adjust the impact ionization coefficient, an exponentially higher impact ionization rate has to be chosen. The ionization coefficient depends weakly on the ionization rate but strongly on the deformation potentials.[33] Relatively small variations in the deformation potentials do not change
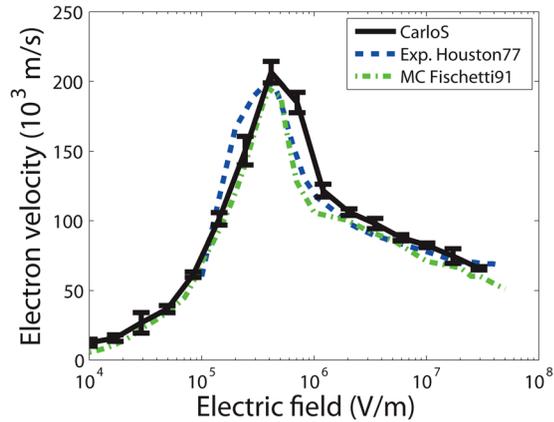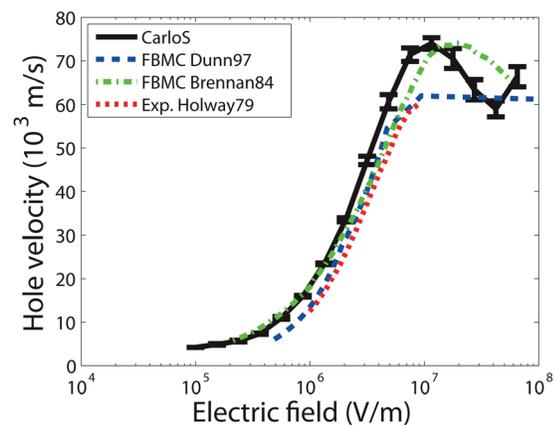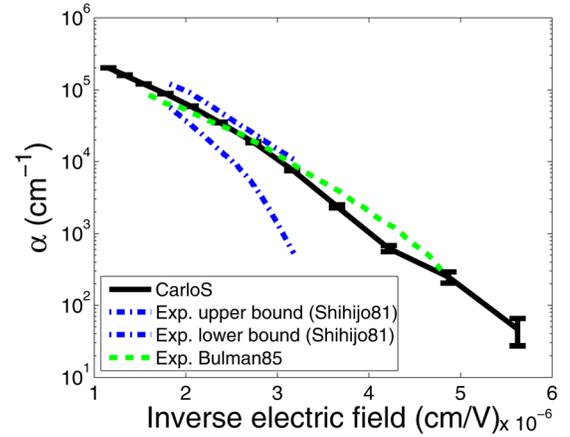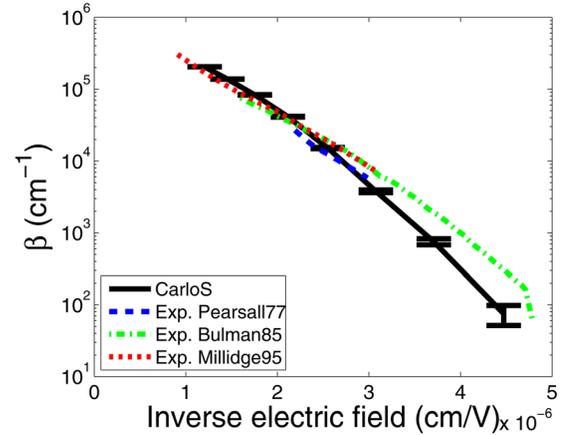

FIG. 7. (Color online) Calibration results for (a) electron and (b) hole velocities.

(a)

(b)

FIG. 8. (Color online) Calibration results for (a) electron and (b) hole impact ionization coefficients.

$v(E)$ appreciably but change the impact ionization coefficient strongly while keeping Λ^{ii} unchanged. Table II summarizes the calibrated acoustic phonon deformation potential, the nonpolar optical deformation potential, and the impact ionization prefactor for electrons and holes.

III. SIMULATION RESULTS AND DISCUSSION

We simulate a GaAs PIN diode with multiplication (intrinsic) region widths between 55 nm and 500 nm. The device operates in the Geiger-mode at 300 K. The simulation procedure consists of a single electron injection on the p-side with a kinetic energy of 10 meV in the Γ_6-valley at the time $t = 0$ ps into the GaAs PIN diode. GaAs is an electron multiplication material. Breakdown occurs when the total number of charged carriers, being generated by impact ionization, exceeds 30 within the depletion region. If a breakdown has not taken place within 500 ps, the simulation stops. The

TABLE II. Calibrated fit parameters.

Particle	Ξ (eV)	D (eV/nm)	Λ^{ii} (nm^{9}eV ps^{-1})
Electrons	5.0	19.0	0.03
Holes	5.0	50.0	1000

numerical experiments are repeated 10^4 times to gain sufficient statistical data. Using Ramo's theorem[44] for the 55 nm device and the saturation velocities of Ref. 14, our breakdown criterion corresponds to a current of approximately 5.4 μA. For this small amount of charge and high electric fields, we do not expect space-charge effects to be important.

Figure 9 presents the breakdown probability versus the reverse bias V_r and the excess bias $V_{ex} = V_r - V_b$ for different multiplication region sizes. Three regions characterize the curves. The breakdown probability slowly ascends in a small voltage interval immediately after the breakdown voltage. We define the breakdown voltage as $P_b(V_b) = 10^{-3}$ (compare with Ref. 45). With higher reverse bias, the breakdown probability increases linearly before it saturates toward unity. The smaller the multiplication region, the steeper is the rise of P_b with higher reverse bias. A steep rise is advantageous for SPADs, because it increases the photon detection efficiency for a constant excess bias. However, for the same excess bias, the electric fields are higher in thinner structures leading to an increased tunneling probability. The region of saturation is larger for increasing multiplicator widths. Compared with the simpler models of Refs. 10–14, the FBMC simulations predict a less steep rise of the breakdown probability with reverse bias and a longer bias region of saturation.

FIG. 10. (Color online) Mean time to avalanche breakdown $\langle t_b \rangle$ and jitter σ vs. reverse bias for multiplication region widths of 55 nm, 250 nm, and 500 nm.

For decreasing multiplication width, the balance between positive feedback of the avalanche and the effective length of the gain material governs the behavior of the breakdown probability steepness.[12] For smaller *w,* and thus, for higher electric fields, the impact ionization coefficients for electrons and holes approach each other (compare Figs. 8(a) and 8(b)). This leads to a more pronounced positive feedback of impact ionization enforcing the avalanche. On the other hand, the smaller the multiplicator, the higher is the ratio between the dead-space d and w. A higher d/w corresponds to a reduction of the effective multiplication region thickness, and thus, has a negative implication on the avalanche. FBMC simulations of the high-energy charge transport reveal a dominance of the positive feedback for smaller multiplication sizes at constant excess bias. Figure 10 shows the mean time to avalanche breakdown and the jitter versus the reverse bias for different multiplication region sizes. Both, $\langle t_b \rangle$ and σ decrease for higher reverse bias. For higher electric fields, impact ionization processes take place more often, and the distribution of the time to avalanche breakdown narrows. The mean time to avalanche breakdown and jitter feature a fast decline over reverse bias in a region after the breakdown voltage before the quantities evolve into an area of small change with reverse bias. $\langle t_b \rangle$ and σ decrease with smaller multiplication region widths. Furthermore, the mean time to avalanche breakdown and jitter descend faster with a smaller multiplication region size. In Fig. 10, we attribute the peaks in the vicinity around the breakdown voltage to the small number of avalanche breakdowns leading to low statistics.

IV. CONCLUSION

In summary, we modeled the charge multiplication process in GaAs single photon avalanche diodes by means of the currently most accurate device simulation method within the physics of semiclassical charge transport and improved the computation of SPAD properties with the state-of-the-art treatment of high-energy carrier dynamics. Highly nonequilibrium effects like the dead-space, the nonlocal impact ionization, and the velocity overshoot are incorporated with the full-band Monte Carlo technique for the investigation of the
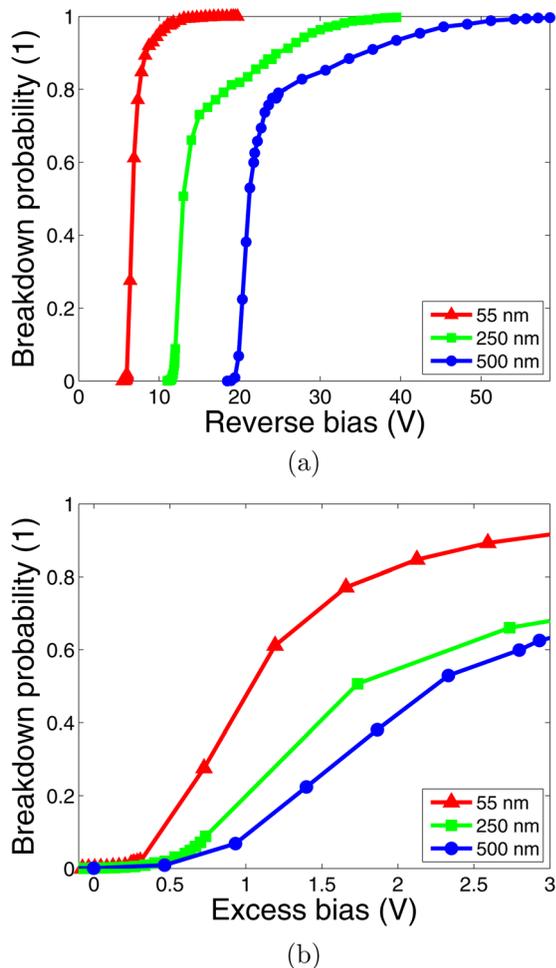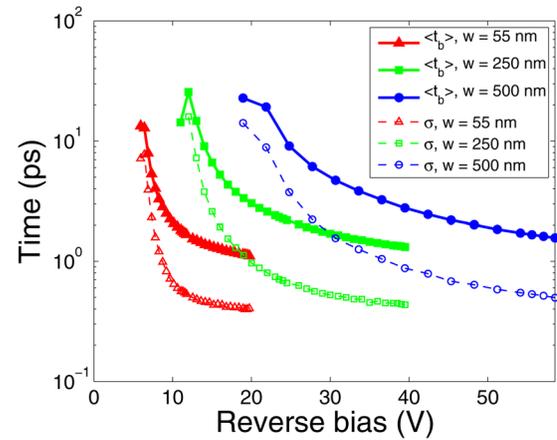

(a)

(b)

FIG. 9. (Color online) Breakdown probability vs. reverse bias (a) and breakdown probability vs. excess bias (b) for multiplication region widths of 55 nm, 250 nm, and 500 nm.

contradictory results presented in literature regarding the trend of the breakdown probability versus the reverse bias with changing multiplicator widths. We have analyzed the behavior of the breakdown probability, the time to avalanche breakdown, and the jitter for different multiplication region widths. The breakdown probability exhibits a steeper rise with reverse bias for smaller multiplicator sizes. The mean time to avalanche breakdown and jitter decrease for shorter multiplication regions. Provided that tunneling processes in smaller multiplicators are at an acceptable level, FBMC simulations suggest smaller multiplication region widths in SPADs for a given excess bias to improve the photon detection efficiency, the photon detection speed due to avalanche breakdown, and the noise owing to avalanching impact ionization processes.

APPENDIX A: VALLEY ALLOCATION

The deformation potentials are approximately constant within a valley. Therefore, the utilized scattering model asks for a partition of the **k**-space into valleys for each band. References [18] and [24] allocate the valleys according to symmetry considerations in the reciprocal space. For example in Ref. [24], the **k**-points are grouped to valleys lying within a sphere around their minima. We make use of a customized steepest descent method to allocate the **k**-space grid cubes to particular valleys. This allocation method divides the reciprocal space into groups of **k**-points according to the following algorithm. The starting point $\mathbf{k}_{start}$ is a center point of the tensor grid. The next point $\mathbf{k}_{next}$ lies in the next cubic box in the direction of the negative gradient of the current cube. If $\mathbf{k}_{next}$ leaves the irreducible wedge during the iteration loop, it is transformed back again into the irreducible wedge. After a number of iterations, the end point $\mathbf{k}_{end}$ is reached. All $\mathbf{k}_{end}$ gather around the local energy minimum of the particular valley. We allocate a cubic box to the particular valley per conduction band where the customized steepest descend method converges to its valley minimum energy. Holes possess only one local minimum per valence band rendering the valley allocation trivial. We do not expect relevant changes due to the different valley allocation approaches used in literature.

APPENDIX B: WEIGHTS OF BOUNDARY BOXES

The method of Gilat and Raubenheimer asks for the computation of the area A_0 of a cubic box cut by an equi-energy plane. Therefore, cubes lying partly outside the irreducible wedge have a different contribution compared to boxes lying completely inside the irreducible wedge. Symmetry considerations are very useful to compute the contribution of boundary boxes.[20] Consider the case where a cubic box lies completely inside the irreducible wedge. Then, the contribution to the whole Brillouin zone is $A_{BZ} = 48 A_0$. The box weight is $w = 1$. Assume that a boundary plane of the irreducible wedge cuts the cubic box such that, including the neighboring wedge, the cubic box lies again inside these two wedges. Then, the contribution to the whole Brillouin zone is $A_{BZ} = 48/2 \cdot A_0$. The box weight is $w = 1/2$. And so on.

[1] E. Knill, R. Laflamme, and G. J. Milburn, Nature **409**, 46 (2001).
[2] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. **74**, 145 (2002).
[3] J. Rarity and P. Tapster, Phys. Rev. Lett. **64**, 2495 (1990).
[4] B. F. Aull, Proc. SPIE **5353**, 105 (2004).
[5] V. W. S. Chan, J. Lightwave Technol. **24**, 4750 (2006).
[6] A. Lacaita, F. Zappa, S. Bigliardi, and M. Manfredi, IEEE Trans. Electron Devices **40**, 577 (1993).
[7] R. Cubeddu, D. Comelli, C. D'Andrea, P. Taroni, and G. Valentini, J. Phys. D **35**, R61 (2002).
[8] S. Cova, M. Ghioni, A. Lotito, I. Rech, and F. Zappa, J. Mod. Opt. **51**, 1267 (2004).
[9] M. Itzler, X. Jiang, M. Entwistle, K. Slomkowski, A. Tosi, F. Acerbi, F. Zappa, and S. Cova, J. Mod. Opt. **58**, 174 (2011).
[10] S. Wang, F. Ma, X. Li, G. Karve, X. Zheng, and J. C. Campbell, Appl. Phys. Lett. **82**, 1971 (2003).
[11] D. Ramirez, M. Hayat, G. Karve, J. Campbell, S. Torres, B. Saleh, and M. Teich, IEEE J. Quantum Electron. **42**, 137 (2006).
[12] J. Ng, C. Tan, J. David, and G. Rees, "Theoretical study of breakdown probabilities in single photon avalanche diodes," The 16th Annual Meeting of the IEEE, Lasers and Electro-Optics Society, 27–28, Oct. 2003 (LEOS 2003), pp. 773–774.
[13] M. Hayat, U. Sakoglu, J. Campbell, B. Saleh, and M. Teich, IEEE J. Quantum Electron. **39**, 179 (2003).
[14] S. L. Tan, D. S. Ong, and H. K. Yow, J. Appl. Phys. **102**, 044506 (2007).
[15] C. Jungemann, S. Keith, M. Bartels, and B. Meinerzhagen, IEICE Trans. Electron. E82C, 870 (1999).
[16] K. Hess, *Monte Carlo Device Simulation: Full Band and Beyond* (Kluwer Academic, Dordrecht, 1991).
[17] C. Jacoboni and L. Reggiani, Rev. Mod. Phys. **55**, 645 (1983).
[18] C. Jungemann and B. Meinerzhagen, *Hierarchical Device Simulation. The Monte Carlo Perspective* (Springer, New York, 2003).
[19] U. Ravaioli, Semicond. Sci. Technol. **13**, 1 (1998).
[20] P. M. Marcus, J. F. Janak, and A. R. Williams, *Computational Methods in Band Theory* (Plenum, New York, 1971).
[21] J. Chelikowsky and M. Cohen, Phys. Rev. B **14**, 556 (1976).
[22] M. Fischetti and S. Laux, Phys. Rev. B **38**, 9721 (1988).
[23] G. M. Dunn, G. J. Rees, J. P. R. David, S. A. Plimmer, and D. C. Herbert, Semicond. Sci. Technol. **12**, 111 (1997).
[24] F. M. Bufler, *Full-Band Monte Carlo Simulation of Nanoscale Strained-Silicon MOSFETs* (Hartung-Gorre, Konstanz, 2003).
[25] F. M. Bufler, A. Schenk, and W. Fichtner, Math. Comput. Simul. **62**, 323 (2003).
[26] Synopsys, Inc., *Sentaurus Device User Guide, version Y-2006.06*, Mountain View, California, 2006.
[27] E. Schöll, *Theory of Transport Properties of Semiconductor Nanostructures* (Chapman & Hall, London, 1998).
[28] D. K. Ferry, J. R. Barker, and C. Jacoboni, *Physics of Nonlinear Transport in Semiconductors* (Plenum, New York, 1980).
[29] J. Y. Tang and K. Hess, J. Appl. Phys. **54**, 5139 (1983).
[30] M. V. Fischetti, D. J. DiMaria, S. D. Brorson, T. N. Theis, and J. R. Kirtley, Phys. Rev. B **31**, 8124 (1985).
[31] L. Reggiani, P. Lugli, and A.-P. Jauho, J. Appl. Phys. **64**, 3072 (1988).
[32] A. Abramo, L. Baudry, R. Brunetti, R. Castagne, M. Charef, F. Dessenne, P. Dollfus, R. Dutton, W. Engl, R. Fauquembergue, C. Fiegna, M. Fischetti, S. Galdin, N. Goldsman, M. Hackel, C. Hamaguchi, K. Hess, K. Hennacy, P. Hesto, J. Higman, T. Iizuka, C. Jungemann, Y. Kamakura, H. Kosina, T. Kunikiyo, S. Laux, C. Maziar, H. Mizuno, H. Peifer, S. Ramaswamy, N. Sano, P. Scrobohaci, S. Selberherr, M. Takenaka, K. Taniguchi, J. Thobel, R. Thoma, K. Tomizawa, M. Tomizawa, T. Vogelsang, P. Yoder, and A. Yoshii, IEEE Trans. Electron Devices **41**, 1646 (1994).
[33] M. V. Fischetti and S. E. Laux, "Monte Carlo simulation of electron transport in Si: The first 20 years," Proceedings of ESSDERC (26th European Solid State Device Research Conference, 1996), pp. 813-820.
[34] T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, and C. Hamaguchi, J. Appl. Phys. **75**, 297 (1994).
[35] P. D. Yoder, V. D. Natoli, and R. M. Martin, J. Appl. Phys. **73**, 4378 (1993).
[36] C. Jungemann, S. Keith, F. M. Bufler, and B. Meinerzhagen, Electr. Eng. **79**, 99 (1996).
[37] G. Gilat and L. J. Raubenheimer, Phys. Rev. **144**, 390 (1966).
[38] M. Fischetti, IEEE Trans. Electron Devices **38**, 634 (1991).
[39] E. Kane, Phys. Rev. **159**, 624 (1967).

[40]H. K. Jung, K. Taniguchi, and C. Hamaguchi, J. Appl. Phys. **79**, 2473 (1996).

[41]A. Kuligk, N. Fitzer, and R. Redmer, Phys. Rev. B **71**, 085201 (2005).

[42]N. Sano and A. Yoshii, J. Appl. Phys. **77**, 2020 (1995).

[43]M. V. Fischetti, N. Sano, S. E. Laux, and K. Natori, "Full-band Monte Carlo simulation of high-energy transport and impact ionization of electrons and holes in Ge, Si, and GaAs," Proceedings of the International Conference on Simulation of Semiconductor Processes and Devices, 1996, pp. 43–44.

[44]S. Ramo, "Currents induced by electron motion," *Proceedings of the IRE*, Vol. 27, pp. 584–585 (1939).

[45]S. Mun, C. Tan, S. Dimler, L. Tan, J. Ng, Y. Goh, and J. David, IEEE J. Quantum Electron. **45**, 566 (2009).