# Phonon-assisted tunneling in direct-bandgap semiconductors

Mazharuddin Mohammed,[1,2,a] Anne S. Verhulst,[1] Devin Verreck,[1] Maarten L. Van de Put,[3] Wim Magnus,[1,4] Bart Sorée,[1,2,4] and Guido Groeseneken[1,2]

## AFFILIATIONS

[1]Imec, Kapeldreef 75, 3001 Leuven, Belgium
[2]Department of Electrical Engineering, KU Leuven, 3001 Leuven, Belgium
[3]Department of Materials Science and Engineering, The University of Texas at Dallas, Dallas, Texas 75080, USA
[4]Department of Physics, Universiteit Antwerpen, 2020 Antwerpen, Belgium

[a]Mazharuddin.Mohammed@imec.be

## ABSTRACT

In tunnel field-effect transistors, trap-assisted tunneling (TAT) is one of the probable causes for degraded subthreshold swing. The accurate quantum-mechanical (QM) assessment of TAT currents also requires a QM treatment of phonon-assisted tunneling (PAT) currents. Therefore, we present a multi-band PAT current formalism within the framework of the quantum transmitting boundary method. An envelope function approximation is used to construct the electron-phonon coupling terms corresponding to local Fröhlich-based phonon-assisted inter-band tunneling in direct-bandgap III-V semiconductors. The PAT current density is studied in up to 100 nm long and 20 nm wide p-n diodes with the 2- and 15-band material description of our formalism. We observe an inefficient electron-phonon coupling across the tunneling junction. We further demonstrate the dependence of PAT currents on the device length, for our non-self-consistent formalism which neglects changes in the electron distribution function caused by the electron-phonon coupling. Finally, we discuss the differences in doping dependence between direct band-to-band tunneling and PAT current.

Published under license by AIP Publishing. https://doi.org/10.1063/1.5044256

## I. INTRODUCTION

The tunnel field-effect transistor (TFET) is one of the favorable candidates as an energy efficient device, promising a sub-60 mV/decade subthreshold swing (SS) for future low power technology nodes.[1–3] For Si TFETs, the large indirect-bandgap results in low drive-currents, thereby prompting the pursuit of other possible device structures and materials.[4–10] Because of the lower effective bandgap, the direct band-to-band tunneling (BTBT) in heterostructure III-V TFETs offers higher drive currents.[11] Despite promising predictions, experimental devices suffer from a degraded SS, which is presently associated with leakage currents like trap-assisted tunneling (TAT),[12–14] phonon-assisted tunneling (PAT),[15] tunneling via Auger-recombination,[16] and high doping-induced band-tail states.[17,18]

However, TAT is seen as one of the most probable causes for SS degradation in TFETs,[14,19] and therefore it requires an in-depth study in order to improve the device performance. A fundamental mechanism underlying TAT in a semiconductor device is phonon-assisted tunneling (PAT).[20,21] Moreover, recent advances have shown the need to include electric-field induced effects on the quantized trap levels into a robust TAT model.[22] Therefore, as a first step toward an accurate TAT description in a semiconductor device, the quantum-mechanical (QM) modeling of PAT is required. Such modeling of PAT has been done in small devices, but in large-size semiconductor devices, it demands enormous computational resources and so far has received little attention.[15] Hence, we present a multi-band single-phonon based PAT formalism suitable for direct-bandgap semiconductor devices and optimized for low computational efforts. The content of this paper is organized as follows. In Sec. II, we derive the PAT current density formalism and discuss the numerical implementation by applying it to an $In_{0.53}Ga_{0.47}As$ p-n diode for which the PAT current density is expected to be

observable.[23] The inefficiency of the electron-phonon coupling strength across the tunneling junction is detailed in Sec. III. Section IV discusses the impact of the device length on the PAT current density. The comparison between the BTBT and PAT current density for different doping concentrations is discussed in Sec. V. Finally, we conclude this paper in Sec. VI.

## II. FORMALISM

To allow for simulations of the PAT current in homostructure III-V devices, we follow the framework of Zener tunneling in indirect-bandgap semiconductors[24] (Sec. II A), while using a $\mathbf{k} \cdot \mathbf{p}$-based envelope function approximation to calculate the electron-phonon coupling strength (Sec. II B). We implement the formalism (Sec. II C) as an extension to an existing $\mathbf{k} \cdot \mathbf{p}$-based full-zone (30-bands) QM simulator Pharos used for calculating direct-BTBT currents (Sec. II D). Note that PAT calculations will only be shown for a 2-band and 15-band model. This implementation choice is based on Pharos' ability to simulate large device structures, while efficiently filtering the spurious solutions through spectral decomposition and demanding low computational efforts.[25]

In this article, we study the devices with an infinite size in the $y$-direction. A quantum transmitting boundary method (QTBM)[26] is employed along the transport $x$-direction, whereas confinement is present along the $z$-direction (see Fig. 1).

### A. PAT current equation

For a non-interacting boosted electron system, the many-electron Hamiltonian $\mathcal{H}_{el}$ in the framework of the second quantization is defined by

$$\widehat{\mathcal{H}}_{el} = \int d\mathbf{r} \, \widehat{\psi}^\dagger(\mathbf{r}) H_{el} \widehat{\psi}(\mathbf{r}), \tag{1}$$

where $H_{el}$ represents the one-electron Hamiltonian, which includes external electrostatic potential profiles. The electron field operators $\widehat{\psi}(\mathbf{r})$ in Eq. (1) are defined as

$$\widehat{\psi}^\dagger(\mathbf{r}) = \sum_j \hat{c}_j^\dagger \psi_j^*(\mathbf{r}); \quad \widehat{\psi}(\mathbf{r}) = \sum_j \psi_j(\mathbf{r}) \hat{c}_j, \tag{2}$$

where $\hat{c}_j^\dagger$ and $\hat{c}_j$ are the creation and annihilation operators for an electron in the $j$th state, respectively, and $\psi_j(\mathbf{r})$ is the
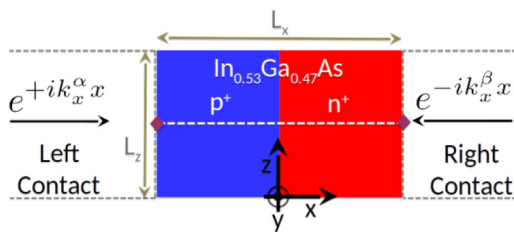


**FIG. 1.** Simulated homostructure In$_{0.53}$Ga$_{0.47}$As p-n diode configuration. Dopant profiles are abrupt and uniform. The device dimensions are $L_x$ and $L_z$ in $x$- and $z$-directions, respectively. Translational invariance is assumed in the $y$-direction. The origin of the $xyz$-coordinate system is indicated in the figure.

$j$th state one-electron wavefunction. The following anti-commutation relations apply for the fermionic electron creation and annihilation operators of Eq. (2):

$$\left\{\hat{c}_i, \hat{c}_j^\dagger\right\} = \delta_{ij}; \quad \left\{\hat{c}_i, \hat{c}_j\right\} = 0; \quad \left\{\hat{c}_i^\dagger, \hat{c}_j^\dagger\right\} = 0, \tag{3}$$

where the indices $i$ and $j$ can denote any one-electron state.

In the QTBM, an electron state consists of a mode which is injected from either of the two contacts (Fig. 1) into the active region, combined with all the corresponding reflected and transmitted modes. Here, a single-electron state $j$ is identified by the following quantum numbers: the wave vector $\mathbf{k}_j$ and spin $\sigma_j$ of the injection mode. The complete set of electron states of the non-interacting electron system is then split up into an $\alpha$-set and a $\beta$-set which correspond to the electron states identified with, respectively, the right-going and the left-going injected modes, whereby the injected modes are electron plane waves in the $x$-direction, modulated by a function in the $y$- and $z$-directions.

Substituting Eq. (2) in Eq. (1) and including all possible state combinations while keeping in mind that all electron states identified with injected modes from the same contact are orthonormal[27] and identified with injected modes from different contacts are orthogonal, results in the following description:

$$\widehat{\mathcal{H}}_{el} = \sum_{\mathbf{k}_\alpha \sigma_\alpha} E_{\mathbf{k}_\alpha \sigma_\alpha} \hat{c}_{\mathbf{k}_\alpha \sigma_\alpha}^\dagger \hat{c}_{\mathbf{k}_\alpha \sigma_\alpha} + \sum_{\mathbf{k}_\beta \sigma_\beta} E_{\mathbf{k}_\beta \sigma_\beta} \hat{c}_{\mathbf{k}_\beta \sigma_\beta}^\dagger \hat{c}_{\mathbf{k}_\beta \sigma_\beta}$$
$$= \widehat{\mathcal{H}}_\alpha + \widehat{\mathcal{H}}_\beta, \tag{4}$$

where $E_{\mathbf{k}_\alpha \sigma_\alpha}$ and $E_{\mathbf{k}_\beta \sigma_\beta}$ are the energies for which a contact eigenvalue problem is solved in the left contact and the right contact, respectively, resulting in the determination of the injected modes. Note that a continuous range of energy values $E_{\mathbf{k}_\alpha \sigma_\alpha}$ and $E_{\mathbf{k}_\beta \sigma_\beta}$ is available. Note also that the electrons in each of the two contacts are in equilibrium and can be exchanged with an external reservoir, thereby fixing the chemical potential of these contacts.[26,28]

For the free phonon system, the second quantization Hamiltonian is described [similar to Eq. (4)] by

$$\widehat{\mathcal{H}}_{ph} = \sum_{\mathbf{q}\nu} \hbar\omega_{\mathbf{q}\nu} \hat{a}_{\mathbf{q}\nu}^\dagger \hat{a}_{\mathbf{q}\nu}, \tag{5}$$

where $\mathbf{q}$ is the phonon wave vector and the index $\nu$ specifies the phonon branch (acoustic or optical) as well as the polarization (longitudinal or transverse). The commutation rules for phonons are

$$\left[\hat{a}_{\mathbf{q}_1\nu_1}, \hat{a}_{\mathbf{q}_2\nu_2}^\dagger\right] = \delta_{\mathbf{q}_1\mathbf{q}_2}\delta_{\nu_1\nu_2}; \quad \left[\hat{a}_{\mathbf{q}_1\nu_1}, \hat{a}_{\mathbf{q}_2\nu_2}\right] = 0;$$
$$\left[\hat{a}_{\mathbf{q}_1\nu_1}^\dagger, \hat{a}_{\mathbf{q}_2\nu_2}^\dagger\right] = 0. \tag{6}$$

The second quantization Hamiltonian for the system of both non-interacting electrons and free phonons is then

$$\widehat{\mathcal{H}}_0 = \widehat{\mathcal{H}}_{el} + \widehat{\mathcal{H}}_{ph}. \tag{7}$$

Since PAT currents result from a time-dependent perturbation, as will be discussed later, we also describe the density

matrix of the electron and phonon system. The density matrix allows one to readily determine the statistical average of a time-dependent observable. In the framework of the grand canonical ensemble, the density matrix of the system of non-interacting electrons and free phonons, in which the phonon system has zero chemical potential, is given by

$$\widehat{\varrho}_0 = \frac{\exp\left[-\left(\widehat{\mathcal{H}}_0 - \mu \widehat{N}\right)/k_B T\right]}{\mathcal{Z}}, \tag{8}$$

where $k_B$ is the Boltzmann constant, T is the temperature, $\mu$ and $\widehat{N}$ correspond to the chemical potential and the electron number operator, respectively, and $\mathcal{Z}$ is the grand canonical partition function defined by $\text{Tr}\left\{\exp\left[-\left(\widehat{\mathcal{H}}_0 - \mu \widehat{N}\right)/k_B T\right]\right\}$ in which "Tr" represents the trace operation. The electron number N can be separated into the sum of the left ($N_\alpha$) and the right ($N_\beta$) contact electron numbers, which are characterized by their respective number operators and respective density matrices with chemical potentials $\mu_\alpha$ and $\mu_\beta$. Using

Eqs. (4), (5), and (7), the density matrix of Eq. (8) reduces to the direct product of the electron and phonon density matrices and is written as

$$\widehat{\varrho}_0 = \widehat{\varrho}_\alpha \otimes \widehat{\varrho}_\beta \otimes \widehat{\varrho}_{\text{ph}}. \tag{9}$$

The electron-phonon interactions are introduced, in the framework of perturbation theory, by adding an electron-phonon interaction Hamiltonian to $\widehat{\mathcal{H}}_0$ of Eq. (7), which is defined in the second quantization by

$$\widehat{H}_{\text{int}} = \sum_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}\nu} \left[g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}\nu} \hat{c}^\dagger_{\mathbf{k}_\beta \sigma_\beta} \hat{c}_{\mathbf{k}_\alpha \sigma_\alpha} \left(\hat{a}_{\mathbf{q}\nu} + \hat{a}^\dagger_{-\mathbf{q}\nu}\right)\right] + \text{h.c.}, \tag{10}$$

where $g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}\nu}$ denotes the electron-phonon coupling strength related to interband transitions mediated by electron-phonon interactions (see Sec. II B) and "h.c." stands for hermitian conjugate. The corresponding time-dependent electron-phonon interaction Hamiltonian in the interaction picture is represented by[24]

$$\widetilde{H}_{\text{int}}(t) = \sum_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}\nu} \left\{g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}\nu} \hat{c}^\dagger_{\mathbf{k}_\beta \sigma_\beta} \hat{c}_{\mathbf{k}_\alpha \sigma_\alpha} \left[\hat{a}_{\mathbf{q}\nu} \exp\left(\frac{i(E_{\mathbf{k}_\beta \sigma_\beta} - E_{\mathbf{k}_\alpha \sigma_\alpha} - \hbar\omega_{\mathbf{q}\nu})t}{\hbar}\right) + \hat{a}^\dagger_{-\mathbf{q}\nu} \exp\left(\frac{i(E_{\mathbf{k}_\beta \sigma_\beta} - E_{\mathbf{k}_\alpha \sigma_\alpha} + \hbar\omega_{\mathbf{q}\nu})t}{\hbar}\right)\right]\right\} + \text{h.c.} \tag{11}$$

The steady-state phonon-assisted current can then be calculated by taking the statistical average of the rate of change of the number of electrons in either of the contacts. The number of electrons is derived from the number operator $\widehat{N}_{\alpha(\beta)} = \sum_{\mathbf{k}_\alpha \sigma_\alpha (\mathbf{k}_\beta \sigma_\beta)} \hat{c}^\dagger_{\mathbf{k}_\alpha \sigma_\alpha (\mathbf{k}_\beta \sigma_\beta)} \hat{c}_{\mathbf{k}_\alpha \sigma_\alpha (\mathbf{k}_\beta \sigma_\beta)}$. In the framework of the density matrix approach, the current is then

$$I_{\text{ph}} = -e \lim_{t \to \infty} \left\{\frac{d}{dt}\left[\text{Tr}\left(\widehat{N}_\alpha \widetilde{\varrho}(t)\right)\right]\right\}, \tag{12}$$

with e the elementary charge and $\widetilde{\varrho}(t)$ representing the time-dependent density matrix, which is constructed by a recursive expansion of the von-Neumann equation combined with iterative time integrations.[24] The perturbative approximation consists of taking the first-order expansion of $\widetilde{\varrho}(t)$,[29] which is[24]

$$\widetilde{\varrho}^{(1)}(t) \approx \widehat{\varrho}_0 - \frac{i}{\hbar} \int_0^t dt_1 \left[\widetilde{H}_{\text{int}}(t_1), \hat{\varrho}_0\right]. \tag{13}$$

Inserting Eq. (13) in the PAT current equation (12) while removing the first term, as it corresponds to the direct-BTBT current, and applying cyclic permutations under the trace operation results in[24]

$$I_{\text{ph}} = \frac{e}{\hbar^2} \lim_{t \to \infty} \int_0^t dt_1 \text{Tr}\left(\left[\left[\widehat{N}_\alpha, \widetilde{H}_{\text{int}}(t)\right], \widetilde{H}_{\text{int}}(t_1)\right]\widehat{\varrho}_0\right). \tag{14}$$

Equation (14) is simplified in a similar way as in Vandenberghe et al.[24] For example, the double commutator in Eq. (14) results in numerous terms associated with various interband transitions, which can be brought together efficiently after identifying the non-zero terms. The trace of one such combination is

$\text{Tr}(\hat{c}^\dagger_{\mathbf{k}_\beta \sigma_\beta} \hat{c}_{\mathbf{k}_\alpha \sigma_\alpha} \hat{a}_{\mathbf{q}\nu} \hat{a}^\dagger_{\mathbf{q}'\nu'} \hat{c}^\dagger_{\mathbf{k}'_\alpha \sigma'_\alpha} \hat{c}_{\mathbf{k}'_\beta \sigma'_\beta} \widehat{\varrho}_0)$, and it is evaluated by making use of Eq. (9), while remembering that the density matrix $\widehat{\varrho}_0$ describes the system of non-interacting electrons and phonons and subsequently, applying Eqs. (3) and (6),

$$\text{Tr}(\hat{c}^\dagger_{\mathbf{k}_\beta \sigma_\beta} \hat{c}_{\mathbf{k}_\alpha \sigma_\alpha} \hat{a}_{\mathbf{q}\nu} \hat{a}^\dagger_{\mathbf{q}'\nu'} \hat{c}^\dagger_{\mathbf{k}'_\alpha \sigma'_\alpha} \hat{c}_{\mathbf{k}'_\beta \sigma'_\beta} \widehat{\varrho}_0)$$
$$= \text{Tr}(\hat{c}^\dagger_{\mathbf{k}_\beta \sigma_\beta} \hat{c}_{\mathbf{k}'_\beta \sigma'_\beta} \widehat{\varrho}_\beta)\text{Tr}(\hat{c}_{\mathbf{k}_\alpha \sigma_\alpha} \hat{c}^\dagger_{\mathbf{k}'_\alpha \sigma'_\alpha} \widehat{\varrho}_\alpha)\text{Tr}(\hat{a}_{\mathbf{q}\nu} \hat{a}^\dagger_{\mathbf{q}'\nu'} \widehat{\varrho}_{\text{ph}}) \tag{15}$$

$$= \delta_{\mathbf{k}_\beta \sigma_\beta \mathbf{k}'_\beta \sigma'_\beta} f_\beta(E_{\mathbf{k}_\beta \sigma_\beta}) \delta_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}'_\alpha \sigma'_\alpha} [1 - f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})] \delta_{\mathbf{q}\mathbf{q}'} \delta_{\nu\nu'} [\upsilon(\hbar\omega_{\mathbf{q}\nu}) + 1]. \tag{16}$$

The individual traces in Eq. (15) are the statistical averages of the single particle occupation number characterized by the Fermi-Dirac statistics $f_{(\alpha,\beta)}(E)$ for electrons and the Bose-Einstein statistics $\upsilon(E)$ for phonons, defined as

$$f_{(\alpha,\beta)}(E) = \frac{1}{1 + \exp\left[\left(E - \mu_{(\alpha,\beta)}\right)/k_B T\right]};$$
$$\upsilon(E) = \frac{1}{\exp(E/k_B T) - 1}. \tag{17}$$

Under steady state condition, the time-dependent exponential factors in the PAT current equation are approximated with an energy-conserving delta function, resulting in the following expression:[24,30]

$$I_{ph} = \frac{2\pi e}{\hbar} \sum_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta} \sum_{\mathbf{q}\nu} \left| g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}\nu} \right|^2.$$

$$\begin{bmatrix} \{f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})[1 - f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})] & \nu(\hbar\omega_{\mathbf{q}\nu}) & -f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})[1 - f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})] & [\nu(\hbar\omega_{\mathbf{q}\nu}) + 1]\} & \delta(E_{\mathbf{k}_\beta \sigma_\beta} - E_{\mathbf{k}_\alpha \sigma_\alpha} - \hbar\omega_{\mathbf{q}\nu}) \\ + & \{f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})[1 - f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})] & [\nu(\hbar\omega_{\mathbf{q}\nu}) + 1] & -f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})[1 - f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})] & \nu(\hbar\omega_{\mathbf{q}\nu})\} & \delta(E_{\mathbf{k}_\beta \sigma_\beta} - E_{\mathbf{k}_\alpha \sigma_\alpha} + \hbar\omega_{\mathbf{q}\nu}) \end{bmatrix} \quad (18)$$

The first term in Eq. (18) corresponds to the current contribution due to the excitation of an electron from an $\alpha$ state in the left contact to a $\beta$ state in the right contact, mediated by phonon absorption. The second term refers to the phonon-emission current contribution running in the opposite direction between the same $\alpha$ and $\beta$ states.

### B. Electron-phonon coupling strength

For a given electron state from the $\alpha$-set and a given electron state from the $\beta$-set, the electron-phonon coupling strength in Eq. (18) is defined by

$$g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}\nu} = M_{\mathbf{q}\nu} \int_\Omega d\mathbf{r}\, \psi^*_{\mathbf{k}_\beta \sigma_\beta}(\mathbf{r}) e^{iq_x x} e^{iq_y y} \sqrt{2}\sin(q_z z) \psi_{\mathbf{k}_\alpha \sigma_\alpha}(\mathbf{r}), \quad (19)$$

where $M_{\mathbf{q}\nu}$ refers to the bulk electron-phonon coupling strength, where each phonon wave vector $\mathbf{q}$ consists of its individual components ($\mathbf{q} \to q_x, q_y, q_z$) and $\Omega$ is the device volume. For the phonons, box normalization is assumed with periodic boundary conditions over large lengths $L_{x,ph}(\gg L_x)$ and $L_{y,ph} = L_y$ (representing a large uniform vibrational system in $x$ and $y$) and with Dirichlet boundary conditions over the confined direction $L_{z,ph} = L_z$ (representing an abrupt transition to vacuum beyond $L_{z,ph}$).[31] The phonon wavevectors $q_x$ and $q_y$ are, therefore, continuous variables running from $-\infty$ to $+\infty$, while $q_z$ is discrete with allowed values $q_z = n\pi/L_z$ with $n = \{1, 2, 3, \dots\}$. Since the materials under study are polar III-V semiconductors, the Fröhlich-interaction is the dominant electron-phonon interaction.[32] Since we solely consider the longitudinal branch of the polar optical phonons, the index $\nu$ in Eqs. (18) and (19) is disregarded for the remainder of the article. Therefore, $M_{\mathbf{q}}$ in Eq. (19) is described by the bulk polar coupling strength, in SI units given by[32]

$$M_{\mathbf{q}} = \frac{1}{|\mathbf{q}|} \sqrt{\frac{e^2 \hbar \omega_{\mathbf{q}}}{2\Omega_{\mathbf{ph}}}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon_0}\right)} = \frac{S_{\mathbf{q}}}{\sqrt{\Omega_{\mathbf{ph}}}}, \quad (20)$$

where $\hbar$ is the reduced Planck constant, $\omega_{\mathbf{q}}$ is the phonon frequency, $\Omega_{\mathbf{ph}}$ is the volume of the total vibrational system of which the device is part, and $\varepsilon_\infty$ and $\varepsilon_0$ are the high frequency and static dielectric constants, respectively.[32]

The wavefunctions corresponding to states from the $\alpha$- and $\beta$-sets are expanded using the following homostructure envelope function approximation:

$$\psi^*_{\mathbf{k}_\beta \sigma_\beta}(\mathbf{r}) = \sum_m U^*_m(\mathbf{r}) F^*_{\mathbf{k}_\beta \sigma_\beta m}(\mathbf{r});$$

$$\psi_{\mathbf{k}_\alpha \sigma_\alpha}(\mathbf{r}) = \sum_m F_{\mathbf{k}_\alpha \sigma_\alpha m}(\mathbf{r}) U_m(\mathbf{r}), \quad (21)$$

whereby $F_{\mathbf{k}_\alpha \sigma_\alpha m}(\mathbf{r})$ are slowly varying envelope functions, whose Fourier components lie within the first Brillouin zone, and $U_m(\mathbf{r})$ represents the bulk wavefunctions, thereby providing a complete set of orthonormal basis functions with lattice periodicity,[33] where $m$ represents the band index. Substituting Eq. (21) in Eq. (19) results in the following electron-phonon coupling strength expression:

$$g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}} = M_{\mathbf{q}} \sum_{mn} \int_\Omega d\mathbf{r}\, U^*_n(\mathbf{r}) F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{r}) e^{iq_x x} e^{iq_y y} \sqrt{2}\sin(q_z z)$$
$$F_{\mathbf{k}_\alpha \sigma_\alpha m}(\mathbf{r}) U_m(\mathbf{r}). \quad (22)$$

For the crystalline semiconductors, we can replace the position vector $\mathbf{r}$ by a linear combination of the unit cell vector $\mathbf{R}_u$ and the lattice vector $\mathbf{R}$, which transforms the volume integral in the electron-phonon coupling [Eq. (22)] into

$$\mathbf{r} = \mathbf{R}_u + \mathbf{R} \Rightarrow \int_\Omega d\mathbf{r} = \sum_\mathbf{R} \int_{\Omega_c} d\mathbf{R}_u, \quad (23)$$

where $\Omega_c$ is the volume of the unit cell. By further making use of the lattice periodicity of the basis functions and assuming that the envelope functions are approximately constant over the unit cell, the basis and envelope functions can be simplified in the following way:

$$\begin{array}{lclclcl} U^*_n(\mathbf{R}_u + \mathbf{R}) & = & U^*_n(\mathbf{R}_u); & U_m(\mathbf{R}_u + \mathbf{R}) & = & U_m(\mathbf{R}_u) \\ F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{R}_u + \mathbf{R}) & \approx & F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{R}); & F_{\mathbf{k}_\alpha \sigma_\alpha m}(\mathbf{R}_u + \mathbf{R}) & \approx & F_{\mathbf{k}_\alpha \sigma_\alpha m}(\mathbf{R}). \end{array} \quad (24)$$

Substituting Eqs. (23) and (24) in Eq. (22) leads to the separation of the envelope functions from the basis functions

$$g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}} = M_{\mathbf{q}} \sum_{mn} \sum_{\mathbf{R}} F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{R})\, e^{iq_x X}\, e^{iq_y Y}\, \frac{\sqrt{2} e^{iq_z Z}}{2i} F_{\mathbf{k}_\alpha \sigma_\alpha m}(\mathbf{R}) \int_{\Omega_c} d\mathbf{R}_u U^*_n(\mathbf{R}_u)\, e^{iq_x X_u}\, e^{iq_y Y_u}\, \frac{\sqrt{2} e^{iq_z Z_u}}{2i} U_m(\mathbf{R}_u)$$

$$- M_{\mathbf{q}} \sum_{mn} \sum_{\mathbf{R}} F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{R})\, e^{iq_x X}\, e^{iq_y Y}\, \frac{\sqrt{2} e^{-iq_z Z}}{2i} F_{\mathbf{k}_\alpha \sigma_\alpha m}(\mathbf{R}) \int_{\Omega_c} d\mathbf{R}_u U^*_n(\mathbf{R}_u)\, e^{iq_x X_u}\, e^{iq_y Y_u}\, \frac{\sqrt{2} e^{-iq_z Z_u}}{2i} U_m(\mathbf{R}_u). \quad (25)$$

The phonon wave vector $\mathbf{q}$ in Eq. (22) is not necessarily restricted to the first Brillouin zone. However, a low-wave vector approximation is applied to the phonons, in line with the larger coupling strength at low wave vectors of the polar coupling. This allows to fully exploit the efficiency of the envelope function approximation. Mathematically, this implies $e^{i\mathbf{q}\cdot\mathbf{R}_u} \approx 1$ in Eq. (25). Making use of the orthonormality of the basis functions $\left(\int_{\Omega_c} d\mathbf{R}_u U_n^*(\mathbf{R}_u) U_m(\mathbf{R}_u) = \Omega_c \delta_{nm}\right)$, we find that the electron-phonon coupling strength is determined by the phonon-mediated overlap of same-band envelope functions

$$g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}} = M_{\mathbf{q}} \Omega_c \sum_n \sum_{\mathbf{R}} F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{R}) e^{iq_x X} e^{iq_y Y} \sqrt{2} \sin(q_z Z) F_{\mathbf{k}_\alpha \sigma_\alpha n}(\mathbf{R}).$$
(26)

By the inverse transformation of the lattice vector $\left(\sum_{\mathbf{R}} \to \frac{1}{\Omega_c} \int d\mathbf{r}\right)$, we regain the spatial dependence of the slowly varying envelope functions, and the electron-phonon coupling strength becomes

$$g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}} = M_{\mathbf{q}} \sum_n \int_\Omega d\mathbf{r} F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{r}) e^{iq_x x} e^{iq_y y} \sqrt{2} \sin(q_z z) F_{\mathbf{k}_\alpha \sigma_\alpha n}(\mathbf{r}),$$
(27)

where the envelope functions for a semiconductor device with dimensions $L_x$ (transport), $L_y$ (translationally invariant), and $L_z$ (confined) are of the following form:[34]

$$F^*_{\mathbf{k}_\beta \sigma_\beta n}(\mathbf{r}) = F^*_{\mathbf{k}_\beta \sigma_\beta n}(x, z) e^{-ik_y^\beta y}; \quad F_{\mathbf{k}_\alpha \sigma_\alpha n}(\mathbf{r}) = e^{ik_y^\alpha y} F_{\mathbf{k}_\alpha \sigma_\alpha n}(x, z), \quad (28)$$

since normalization of the corresponding wavefunctions to a Dirac delta function has been applied in $x$- and $y$-directions and to a Kronecker delta in the confined $z$-direction. Inserting Eq. (28) in Eq. (27), taking the absolute square and solving for the integral along the $y$-direction, results in

$$\left|g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}\right|^2 = \frac{|S_{\mathbf{q}}|^2}{\Omega_{ph}} \frac{\sin^2\left(((k_y^\alpha - k_y^\beta + q_y)/2)L_y\right)}{\left[(k_y^\alpha - k_y^\beta + q_y)/2\right]^2}$$
$$\left|\sum_n \int_{L_x} dx \int_{L_z} dz \phi^n_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}(x, z)\right|^2, \quad (29)$$

where $\phi^n_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}(x, z)$ denotes the overlap function for a given $(q_x, q_z)$ and is defined by

$$\phi^n_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}(x, z) = F^*_{\mathbf{k}_\beta \sigma_\beta n}(x, z) \, e^{iq_x x} \sqrt{2} \sin(q_z z) F_{\mathbf{k}_\alpha \sigma_\alpha n}(x, z). \quad (30)$$

The PAT current of Eq. (18) involves the summation over all possible phonon wave vectors. The phonon wave vector component along the $(x, y)$-directions can have a continuous range of values within the first Brillouin zone, therefore, the summations over $(q_x, q_y)$ are converted into integrals $\left(\sum_{q_x, q_y} \to \frac{L_{x,ph} L_{y,ph}}{4\pi^2} \int dq_x \int dq_y\right)$. The confinement along the $z$-direction splits the electron continuous bands into sub-bands and also the discrete nature of $q_z$ is retained, hence, the summation over $q_z$ is maintained. With these transformations, the summation of Eq. (29) is

$$\sum_{\mathbf{q}} \left|g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}\right|^2 = \frac{L_y}{4\pi^2} \int dq_x \int dq_y \lambda(k_y^\alpha - k_y^\beta + q_y)$$
$$\sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, q_y, q_z), \quad (31)$$

where $\lambda(k_y^\alpha - k_y^\beta + q_y)$ is the weight function defined by

$$\lambda(k_y^\alpha - k_y^\beta + q_y) = \frac{\sin\{[(k_y^\alpha - k_y^\beta + q_y)/2]L_y\}}{\left[(k_y^\alpha - k_y^\beta + q_y)/2\right]}$$
$$\text{sinc}\{[(k_y^\alpha - k_y^\beta + q_y)/2]L_y]\},$$

$$G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, q_y, q_z) = \frac{|S_{\mathbf{q}}|^2}{L_{z,ph}} \left|\sum_n \int_{L_x} dx \int_{L_z} dz \phi^n_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}(x, z)\right|^2. \quad (32)$$

In the limit $L_y \to \infty$, the first factor in the weight function $\lambda$ behaves like a delta function $2\pi\delta\left(k_y^\alpha - k_y^\beta + q_y\right)$. As a consequence, the second cardinal-sine function only matters if $k_y^\alpha - k_y^\beta + q_y = 0$, which results in a value of one

$$\sum_{\mathbf{q}} \left|g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}\right|^2 = \frac{L_y}{2\pi} \int dq_x \int dq_y \delta(k_y^\alpha - k_y^\beta + q_y) \sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, q_y, q_z). \quad (33)$$

The inner integral over $dq_y$ in Eq. (33) can be carried out trivially thanks to the delta function, leading to

$$\sum_{\mathbf{q}} \left|g_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}\right|^2 = \frac{L_y}{2\pi} \int dq_x \sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, k_y^\beta - k_y^\alpha, q_z). \quad (34)$$

### C. PAT current density for semiconductor devices

Combining the two previous sections by substituting Eq. (34) in Eq. (18), the PAT current equation is of the form

$$I_{ph} = \frac{eL_y}{\hbar} \sum_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta} \int dq_x \sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, k_y^\beta - k_y^\alpha, q_z).$$
$$\begin{bmatrix} & \{f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})[1 - f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})] & v(\hbar\omega_{\mathbf{q}}) & -f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})[1 - f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})] & [v(\hbar\omega_{\mathbf{q}}) + 1]\} & \delta(E_{\mathbf{k}_\beta \sigma_\beta} - E_{\mathbf{k}_\alpha \sigma_\alpha} - \hbar\omega_{\mathbf{q}}) \\ + & \{f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})[1 - f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})] & [v(\hbar\omega_{\mathbf{q}}) + 1] & -f_\beta(E_{\mathbf{k}_\beta \sigma_\beta})[1 - f_\alpha(E_{\mathbf{k}_\alpha \sigma_\alpha})] & v(\hbar\omega_{\mathbf{q}})\} & \delta(E_{\mathbf{k}_\beta \sigma_\beta} - E_{\mathbf{k}_\alpha \sigma_\alpha} + \hbar\omega_{\mathbf{q}}) \end{bmatrix}$$
(35)

Due to translational invariance along the $y$-direction and the planewave-like envelope functions along the transport $x$-direction, the summations of the electron wave vector along these directions are transformed into integrals. The summation of the electron quantum numbers along the $z$-direction is retained as a summation of modes $\gamma$ due to confinement along this direction

$$\sum_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta} = \frac{1}{16\pi^4} \int dk_y^\alpha \int dk_y^\beta \int dk_x^\alpha \int dk_x^\beta \sum_{\gamma_\alpha \sigma_\alpha \gamma_\beta \sigma_\beta}. \quad (36)$$

With the following $k_x^{\alpha,\beta} \to E_{k_y^\alpha, k_y^\beta}$ conversion:

$$\left. \frac{dk_x^\alpha}{dE_{k_y^\alpha}} \right|_{\gamma_\alpha \sigma_\alpha} = A_{\mathbf{k}_\alpha \sigma_\alpha}; \quad \left. \frac{dk_x^\beta}{dE_{k_y^\beta}} \right|_{\gamma_\beta \sigma_\beta} = B_{\mathbf{k}_\beta \sigma_\beta} \quad (37)$$

and applying the transformations of Eq. (36) and dividing both sides of Eq. (35) by $L_y$, the PAT current density is

$$J_{ph} = \frac{e}{16\pi^4\hbar} \int dk_y^\alpha \int dk_y^\beta \left[ \begin{array}{l} \int dE_{k_y^\alpha} \int dE_{k_y^\beta} \sum_{\gamma_\alpha \sigma_\alpha \gamma_\beta \sigma_\beta} A_{\mathbf{k}_\alpha \sigma_\alpha} B_{\mathbf{k}_\beta \sigma_\beta} \int dq_x \sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, k_y^\beta - k_y^\alpha, q_z) \\ \left\{ f_\alpha(E_{k_y^\alpha}) \left[ 1 - f_\beta(E_{k_y^\beta}) \right] \upsilon(\hbar\omega_\mathbf{q}) - f_\beta(E_{k_y^\beta}) \left[ 1 - f_\alpha(E_{k_y^\alpha}) \right] \left[ \upsilon(\hbar\omega_\mathbf{q}) + 1 \right] \right\} \delta(E_{k_y^\beta} - E_{k_y^\alpha} - \hbar\omega_\mathbf{q}) \\ + \int dE_{k_y^\alpha} \int dE_{k_y^\beta} \sum_{\gamma_\alpha \sigma_\alpha \gamma_\beta \sigma_\beta} A_{\mathbf{k}_\alpha \sigma_\alpha} B_{\mathbf{k}_\beta \sigma_\beta} \int dq_x \sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, k_y^\beta - k_y^\alpha, q_z) \\ \left\{ f_\alpha(E_{k_y^\alpha}) \left[ 1 - f_\beta(E_{k_y^\beta}) \right] \left[ \upsilon(\hbar\omega_\mathbf{q}) + 1 \right] - f_\beta(E_{k_y^\beta}) \left[ 1 - f_\alpha(E_{k_y^\alpha}) \right] \upsilon(\hbar\omega_\mathbf{q}) \right\} \delta(E_{k_y^\beta} - E_{k_y^\alpha} + \hbar\omega_\mathbf{q}) \end{array} \right]. \quad (38)$$

The inner integrals over $E_{k_y^\beta}$ in Eq. (38) are solved by using the following sifting property of the delta function

$$\int dE_{k_y^\beta} \mathcal{D}(E_{k_y^\beta}) \delta \left[ E_{k_y^\beta} - \left( E_{k_y^\alpha} \pm \hbar\omega_\mathbf{q} \right) \right] = \mathcal{D}(E_{k_y^\alpha} \pm \hbar\omega_\mathbf{q}). \quad (39)$$

This reduces the complexity of the PAT current density equation to a single energy integral over $E_{k_y^\alpha}$ and results in the final form of the PAT current density equation,

$$J_{ph} = \frac{e}{16\pi^4\hbar} \int dk_y^\alpha \int dk_y^\beta \left[ \begin{array}{l} \int dE_{k_y^\alpha} \sum_{\gamma_\alpha \sigma_\alpha \gamma_\beta \sigma_\beta} \int dq_x A_{\mathbf{k}_\alpha \sigma_\alpha} B_{\mathbf{k}_\beta \sigma_\beta}^+ \sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, k_y^\beta - k_y^\alpha, q_z) \\ \left\{ f_\alpha(E_{k_y^\alpha}) \left[ 1 - f_\beta(E_{k_y^\alpha} + \hbar\omega_\mathbf{q}) \right] \upsilon(\hbar\omega_\mathbf{q}) - f_\beta(E_{k_y^\alpha} + \hbar\omega_\mathbf{q}) \left[ 1 - f_\alpha(E_{k_y^\alpha}) \right] \left[ \upsilon(\hbar\omega_\mathbf{q}) + 1 \right] \right\} \\ + \int dE_{k_y^\alpha} \sum_{\gamma_\alpha \sigma_\alpha \gamma_\beta \sigma_\beta} \int dq_x A_{\mathbf{k}_\alpha \sigma_\alpha} B_{\mathbf{k}_\beta \sigma_\beta}^- \sum_{q_z} G_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta}(q_x, k_y^\beta - k_y^\alpha, q_z) \\ \left\{ f_\alpha(E_{k_y^\alpha}) \left[ 1 - f_\beta(E_{k_y^\alpha} - \hbar\omega_\mathbf{q}) \right] \left[ \upsilon(\hbar\omega_\mathbf{q}) + 1 \right] - f_\beta(E_{k_y^\alpha} - \hbar\omega_\mathbf{q}) \left[ 1 - f_\alpha(E_{k_y^\alpha}) \right] \upsilon(\hbar\omega_\mathbf{q}) \right\} \end{array} \right], \quad (40)$$

thereby fixing the conversion factors in the $q_x$-integral accordingly as

$$B_{\mathbf{k}_\beta \sigma_\beta}^+ = \left. \frac{dk_x^\beta}{d\left( E_{k_y^\alpha} + \hbar\omega_\mathbf{q} \right)} \right|_{\gamma_\beta \sigma_\beta}; \quad B_{\mathbf{k}_\beta \sigma_\beta}^- = \left. \frac{dk_x^\beta}{d\left( E_{k_y^\alpha} - \hbar\omega_\mathbf{q} \right)} \right|_{\gamma_\beta \sigma_\beta}. \quad (41)$$

### D. Numerical implementation

To calculate the PAT currents with Eq. (40), we first determine the envelope functions $F_{\mathbf{k}_{\alpha(\beta)} \sigma_{\alpha(\beta)} n}(\mathbf{r})$ from Pharos[25] for each energy $E_{\alpha(\beta)}$ and $k_y^{\alpha(\beta)}$. We then calculate the overlap

function densities $\phi_{\mathbf{k}_\alpha \sigma_\alpha \mathbf{k}_\beta \sigma_\beta \mathbf{q}}^n(x, z)$ and repeat the calculations for each injected mode $\gamma_\alpha$ and for all available $\gamma_\beta$ modes, $q_x$, $q_z$, $E_\alpha$, $k_y^\beta$, and $k_y^\alpha$-values, while correspondingly determining the integrations of Eq. (40). The integrations consist of 40 adaptively chosen points in energy $E_\alpha$ and 10 points in $k_y$ to cover all relevant $\alpha, \beta$-states. Additionally, the mesh points for the $q_x$-integral are adaptively discretized, from $q_x \approx 0$ till 25% of the first Brillouin zone. For the sake of numerical efficiency, the Fröhlich model for electron-phonon interaction is considered[32] with a constant energy $\hbar\omega_\mathbf{q}$ of 34 meV corresponding to the longitudinal branch of the polar optical phonon.

The formalism of the PAT current density is applied to an up to 100 nm long and 20 nm wide $In_{0.53}Ga_{0.47}As$ p–n diode as shown in Fig. 1. The simulated device's mesh size is 0.1 nm along the $x$- and 0.2 nm along the $z$-direction. The device is assumed to exhibit translational symmetry along the $y$-direction.

The PAT current density based on Eq. (40) is compared with the direct-BTBT[25] for a 60 nm long and 20 nm wide $In_{0.53}Ga_{0.47}As$ p–n diode in Fig. 2. The comparison is made for both the 2-band and the 15-band model of the formalism. As expected, the PAT current density is increasing relative to the BTBT current density in forward bias. This is mainly because of the smaller available direct-BTBT tunneling window, whereas PAT offers an additional transmission path based on the gain of phonon energy. Moreover, the observed PAT currents are comparable to BTBT, which has been reported by Verhulst et al.[23] for several direct-bandgap materials. In Fig. 2, the PAT current density for the 15-band and the 2-band model is mostly identical. Note that the 15-band implementation is shown until $V_{np} = 0.4\,V$, beyond which the indirect bandgap results in complications of the numerical implementation. Owing to the faster calculation, more straightforward implementation and limited difference between the 2- and 15-band PAT currents, the 2-band model is applied for the remainder of the article.

## III. INEFFICIENT ELECTRON-PHONON COUPLING ACROSS TUNNELING JUNCTIONS

In Fig. 3, the band diagram of a 60 nm-long $In_{0.53}Ga_{0.47}As$ p–n diode is shown at the indicated cut-line of Fig. 1, where the probability density of the electron wavefunction is maximum. The envelope function densities for a 2-band model [$n = 2$ in Eq. (29)], averaged over the unit cell, are plotted along the transport $x$-direction in Fig. 4. These densities are taken at the energies $(E_\alpha, E_\beta)$ indicated in Fig. 3. The corresponding coupling density [see Eqs. (29) and (30)] is
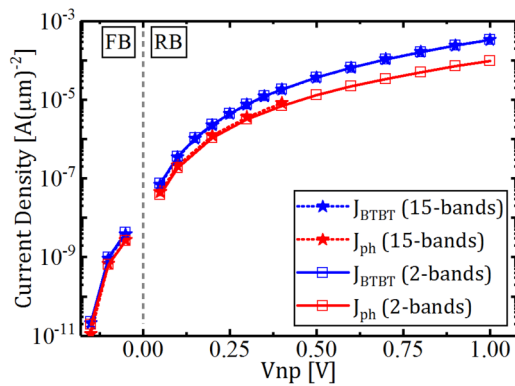


**FIG. 2.** The direct-BTBT and PAT current densities of the homostructure $In_{0.53}Ga_{0.47}As$ p–n diode of Fig. 1 as a function of applied voltage. The diode is 60 nm long and 20 nm wide and is uniformly doped with a concentration of $5 \times 10^{18}$ (at/cm$^3$). FB (RB) stands for forward (reverse) bias.
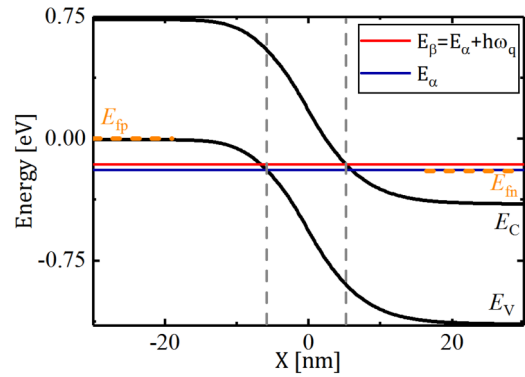


**FIG. 3.** Band diagram along the indicated cut-line of the $In_{0.53}Ga_{0.47}As$ p–n diode of Fig. 1 for a bias of $V_{np} = 0.2\,V$. The black lines correspond to valence band maximum and conduction band minimum energies. The grey-dashed vertical lines indicate the limits of the region where PAT across the junction takes place at the given energies. The orange lines represent the quasi-Fermi energy levels $E_{fp}$ and $E_{fn}$.

shown as a solid-black line in Fig. 4. A drop in coupling density, indicated by a magenta-circle in Fig. 4, implies that the electron-phonon coupling is inefficient across the tunneling junction.

This inefficiency can be explained by examining the envelope functions. The real and the imaginary components of the envelope function corresponding to each band of a 2-band model, at energies $(E_\alpha, E_\beta)$, are presented along the $x$-direction in Fig. 5. It is evident that the conduction band component (band 2) exhibits a 90°-phase shift relative to the
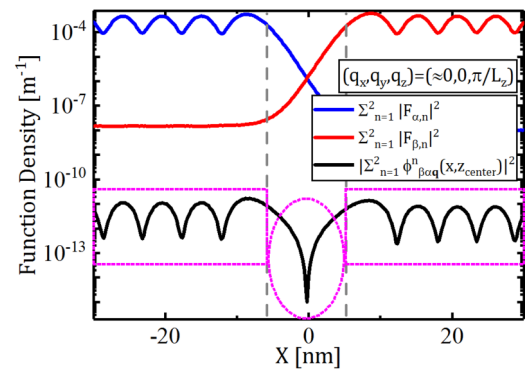


**FIG. 4.** The 2-band envelope function densities for the lowest $k_x$, $k_y$, and $k_z$ values, averaged over the unit cell at specific energies $E_\alpha$ and $E_\beta$, indicated in Fig. 3. The blue (red)-solid curve corresponds to an electron injected at $E_\alpha$ from the left contact (at $E_\beta$ from the right contact). The black-solid line represents the corresponding coupling density. The limits of the tunneling junction region, shown in Fig. 3, are repeated here and the drop of the electron-phonon coupling in this region is highlighted with a magenta-dashed ellipse. The contribution of the PAT process in the near-tunneling regions is indicated with magenta-dashed boxes.
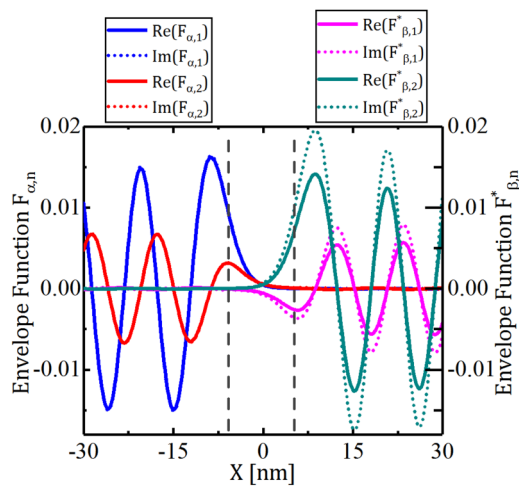
FIG. 5. The real (solid) and imaginary (dashed) components of the EFs corresponding to each band for an electron injected from each contact. The blue (red) lines correspond to the valence (conduction)-band contributions of the EF of an electron injected from the left contact. The pink (green) lines refer to the valence (conduction)-band contributions of the EF of an electron injected from the right contact. The grey-dashed vertical lines of Fig. 3 are repeated here and indicate the limits of the tunneling region.
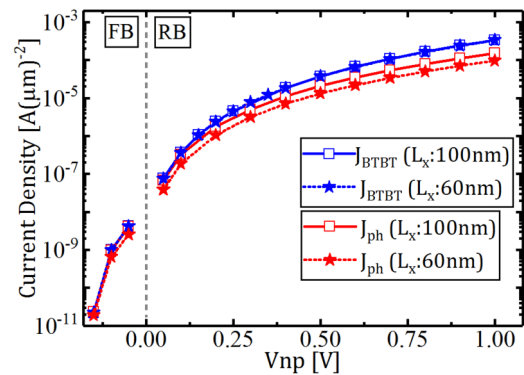


FIG. 6. Dependence of 2-band direct-BTBT and PAT current densities on the device length $L_x$ along the transport direction. The simulated device is an $In_{0.53}Ga_{0.47}As$ p-n diode with varying $L_x$ and a constant 20 nm width, uniformly doped with a concentration of $5 \times 10^{18}$ (at/cm$^3$).

valence band component (band 1) of the envelope function, which can be shown to result from the different parity of the conduction and valence bands' basis function, and which is valid as long as there is limited transmission of electron wavefunction through the forbidden bandgap. This $90°$-shift in both $\alpha$ and $\beta$ states transforms the summation of the individual band's coupling $\phi_{\alpha\beta q}^n(x, z_{center})$ into their subtraction, thereby explaining the dip in Fig. 4. Note that the dip in coupling density is also observed in a 15-band model, due to the $90°$-phase shift found between the valence X and the conduction S band.[27]

## IV. PAT VERSUS DEVICE LENGTH

The finite transmission of an electron's envelope function at the given energies $(E_\alpha, E_\beta)$ into the region beyond the tunneling junction (called the near-tunneling region) is noticeable in the PAT current density as shown in Fig. 4 (indicated by magenta boxes). Consequently, this PAT process in the near-tunneling regions causes a dependence of the PAT current densities on the device length $L_x$. This is illustrated in Fig. 6, where the direct-BTBT and PAT current densities are compared for a 60 nm and 100 nm long $In_{0.53}Ga_{0.47}As$ p-n diode.

As expected, the BTBT current densities (blue) do not depend on device length. However, the PAT current densities show substantial dependence on the device length, which is due to the observable contribution of the near-tunneling regions. An accurate calculation would include the impact of the electron-phonon coupling on the electron distribution function in a self-consistent approach.[35] It is expected that

this will reduce the total PAT current increase with increasing device length presented in Fig. 6. Such a calculation, however, is beyond the scope of this article.

## V. PAT VERSUS DOPING CONCENTRATIONS

The band diagrams of a 60 nm long and 20 nm wide $In_{0.53}Ga_{0.47}As$ p-n diode, taken in the center of the z-direction, are shown for different doping concentrations in Fig. 7. The increase in tunneling length with decrease in doping concentration is apparent. In Fig. 8, the electron envelope function densities, taken at the energies $(E_\alpha, E_\beta)$ of Fig. 7, show the decrease in transmission after tunneling in the moderately doped diodes, which is due to the increase in tunneling length $L_t$. Subsequently, the coupling densities decrease with decreasing diode doping, as detailed in Fig. 9.
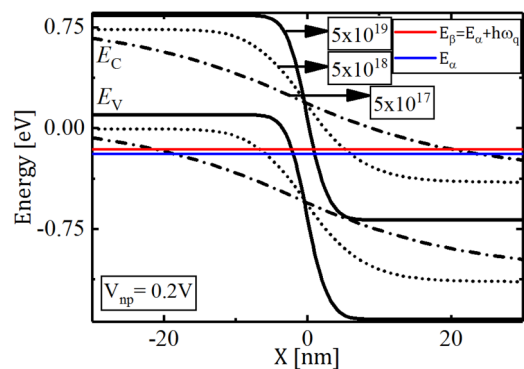


FIG. 7. Band diagrams for the different doping concentrations along the indicated cut-line of the p-n diode of Fig. 1. The solid (dot-, dashed-) lines correspond to the band-edge energies for $5 \times 10^{19}$ ($5 \times 10^{18}$, $5 \times 10^{17}$) (at/cm$^3$) doping concentrations, respectively. The blue (red)-solid line refers to the energies $E_\alpha$ ($E_\beta$).
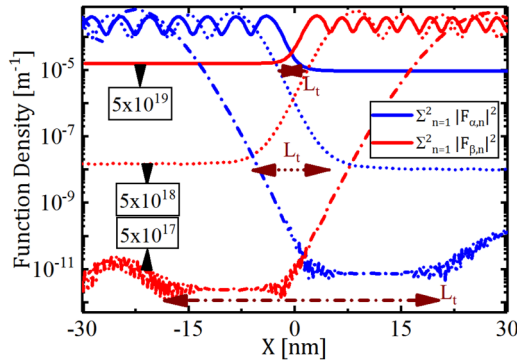
**FIG. 8.** The blue-(red-) solid (dot, dashed) curves indicate the envelope function density for the lowest $k_x$, $k_y$, and $k_z$ values at the energies $E_\alpha$ ($E_\beta$) of Fig. 7, averaged over the unit cell in a 2-band model and for doping concentrations of $5 \times 10^{19}$ ($5 \times 10^{18}$, $5 \times 10^{17}$) (at/cm$^3$), respectively. The tunneling regions are marked with horizontal arrows "$L_t$."
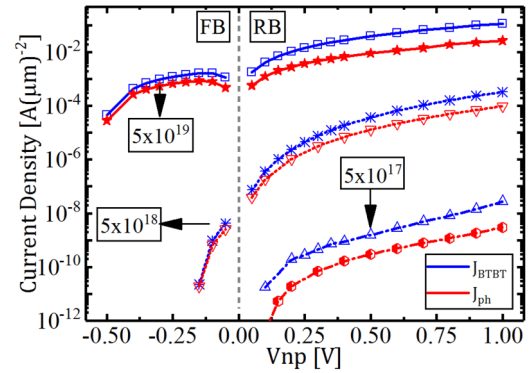


**FIG. 10.** Comparison of the 2-band direct-BTBT and PAT current densities for different doping concentrations. The blue-(red-) solid (dot, dashed) lines represent the direct-BTBT (PAT) current densities for $5 \times 10^{19}$ ($5 \times 10^{18}$, $5 \times 10^{17}$) (at/cm$^3$) doping concentrations.

The resulting PAT current density dependence on the doping concentrations is illustrated in Fig. 10.

It is observed that the PAT current density is close to the BTBT current density in forward bias for high doping concentration [$5 \times 10^{19}$ (at/cm$^3$)], and both in forward bias and at the onset of reverse bias for moderately doped p-n diodes [$5 \times 10^{18}$ (at/cm$^3$)]. For a given voltage (see Fig. 7), the effective direct-BTBT tunneling-energy window, which is determined by the doping-dependent Fermi levels, is much larger for the $5 \times 10^{19}$ diodes than for the moderately doped diodes. This implies that the PAT current in the $5 \times 10^{19}$ (at/cm$^3$) diode will only become relevant compared to BTBT at a more negative bias where this effective tunneling energy window becomes sufficiently small compared to the phonon-energy. The shift toward more positive voltages in Fig. 10 of PAT being close to BTBT for the $5 \times 10^{18}$ (at/cm$^3$) diode versus $5 \times 10^{19}$ (at/cm$^3$)
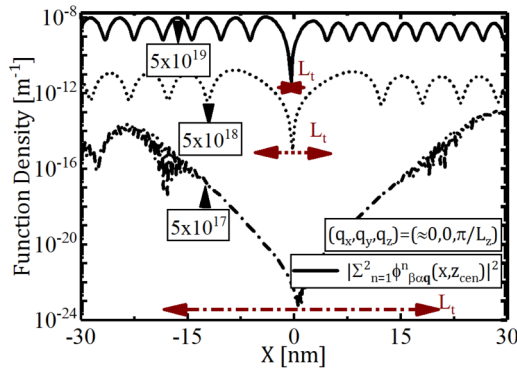


**FIG. 9.** The black- solid (dot,dashed) curves refer to the coupling density corresponding to the envelope functions of Fig. 8 of a 2-band model and for $5 \times 10^{19}$ ($5 \times 10^{18}$, $5 \times 10^{17}$) (at/cm$^3$) doping concentrations. The tunneling regions are marked with horizontal arrows "$L_t$."

diode can be ascribed to the occurrence of a smaller effective BTBT tunneling window counteracted by a lower contribution of the PAT process in the near-tunneling region (see Fig. 9). In case of the lowest doped diode [$5 \times 10^{17}$ (at/cm$^3$)], the reduction of the near-tunneling regions leads to smaller PAT relative to BTBT current densities. Hence, the combination of the effective tunneling-energy window and the length of the near-tunneling regions determine the observable doping dependence of the PAT current density.

## VI. CONCLUSIONS

We present a multi-band PAT current density formalism applicable to direct-bandgap devices and implement it with high computational efficiency. We study the PAT current density in up to 100 nm long and 20 nm wide In$_{0.53}$Ga$_{0.47}$As p-n diodes with a 2- and 15-band implementation of the formalism. We observe that the PAT currents are comparable to BTBT and find that there is a limited difference in PAT current densities between the 2- and 15-band model. We discover that the electron-phonon coupling is inefficient across the tunneling junction which originates from the difference in parity between the basis functions of the coupled valence and conduction bands in the transport direction. We determine that the PAT current densities increase with device length along the transport direction owing to PAT contributions in the near-tunneling regions. This increase is an overestimate given the lack of impact on the current densities from electron redistributions based on the electron-phonon coupling. We further unravel that the combined effects of the tunneling-energy window and near-tunneling region length result in the doping dependence of the PAT current density. We expect that the PAT current density formalism can eventually be applied to study PAT in heterostructure p-n, p-i-n diodes and TFET configurations. We further expect that it can be extended to multi-phonon assisted tunneling and eventually to QM assess TAT in devices.

## REFERENCES

[1] H. Lu and A. Seabaugh, IEEE J. Electron Devices Soc. **2**, 44–49 (2014).
[2] A. M. Ionescu and H. Riel, Nature **479**, 329–337 (2011).
[3] D. Verreck, G. Groeseneken, and A. Verhulst, *Wiley Encyclopedia of Electrical and Electronics Engineering* (John Wiley & Sons, Inc., 1999).
[4] S. O. Koswatta, S. J. Koester, and W. Haensch, IEEE Trans. Electron Devices **57**, 3222–3230 (2010).
[5] C. Schulte-Braucks, R. Pandey, R. N. Sajjad, M. Barth, R. K. Ghosh, B. Grisafe, P. Sharma, N. von den Driesch, A. Vohra, G. B. Rayner, R. Loo, S. Mantl, D. Buca, C. C. Yeh, C. H. Wu, W. Tsai, D. A. Antoniadis, and S. Datta, IEEE Trans. Electron Devices **64**, 4354–4362 (2017).
[6] D. Verreck, A. S. Verhulst, M. L. Van de Put, B. Sorée, N. Collaert, A. Mocuta, A. Thean, and G. Groeseneken, IEEE Electron Device Lett. **37**, 337–340 (2016).
[7] W. Li, S. Sharmin, H. Ilatikhameneh, R. Rahman, Y. Lu, J. Wang, X. Yan, A. Seabaugh, G. Klimeck, D. Jena, and P. Fay, IEEE J. Exploratory Solid State Comput. Devices Circuits **1**, 28–34 (2015).
[8] K. Ganapathi and S. Salahuddin, IEEE Electron Device Lett. **32**, 689–691 (2011).
[9] E. Lind, E. Memišević, A. W. Dey, and L.-E. Wernersson, IEEE J. Electron Devices Soc. **3**, 96–102 (2015).
[10] U. E. Avci, D. H. Morris, and I. A. Young, IEEE J. Electron Devices Soc. **3**, 88–95 (2015).
[11] D. Mohata, S. Mookerjea, A. Agrawal, Y. Li, T. Mayer, V. Narayanan, A. Liu, D. Loubychev, J. Fastenau, and S. Datta, Appl. Phys. Exp. **4**, 024105 (2011).
[12] E. Memisevic, M. Hellenbrand, E. Lind, A. R. Persson, S. Sant, A. Schenk, J. Svensson, R. Wallenberg, and L.-E. Wernersson, Nano Lett. **17**, 4373–4380 (2017). pMID: 28613894
[13] A. S. Verhulst, D. Verreck, Q. Smets, K. H. Kao, M. L. Van de Put, R. Rooyackers, B. Sorée, A. Vandooren, K. De Meyer, G. Groeseneken, M. M. Heyns, A. Mocuta, N. Collaert, and A. V. Y. Thean, in *2014 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2014), pp. 30.2.1–30.2.4.
[14] A. Schenk, S. Sant, K. Moselund, H. Riel, E. Memisevic, and L.-E. Wernersson, in *2017 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)* (IEEE, 2017), pp. 273–276.
[15] J. Charles, P. Sarangapani, R. Golizadeh-Mojarad, R. Andrawis, D. Lemus, X. Guo, D. Mejia, J. E. Fonseca, M. Povolotskyi, T. Kubis, and G. Klimeck, J. Comput. Electron. **15**, 1123–1129 (2016).
[16] J. T. Teherani, S. Agarwal, W. Chern, P. M. Solomon, E. Yablonovitch, and D. A. Antoniadis, J. Appl. Phys. **120**, 084507 (2016).
[17] S. Sant and A. Schenk, J. Appl. Phys. **122**, 135702 (2017).
[18] J. Bizindavyi, A. S. Verhulst, Q. Smets, D. Verreck, B. Sorée, and G. Groeseneken, IEEE J. Electron Devices Soc. **6**, 633–648 (2018).
[19] Q. Smets, A. S. Verhulst, E. Simoen, D. Gundlach, C. Richter, N. Collaert, and M. M. Heyns, IEEE Trans Electron Devices **64**, 3622–3626 (2017).
[20] A. Schenk, Solid State Electron. **35**, 1585–1596 (1992).
[21] G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, IEEE Trans. Electron Devices **39**, 331–338 (1992).
[22] M. Mohammed, A. S. Verhulst, D. Verreck, M. L. Van de Put, E. Simoen, B. Sorée, B. Kaczer, R. Degraeve, A. Mocuta, N. Collaert, A. Thean, and G. Groeseneken, J. Appl. Phys. **120**, 245704 (2016).
[23] A. S. Verhulst, D. Verreck, M. A. Pourghaderi, M. Van de Put, B. Sorée, G. Groeseneken, N. Collaert, and A. V.-Y. Thean, Appl. Phys. Lett. **105**, 043103 (2014).
[24] W. Vandenberghe, B. Sorée, W. Magnus, and M. V. Fischetti, J. Appl. Phys. **109**, 124503 (2011).
[25] D. Verreck, A. S. Verhulst, M. L. Van de Put, B. Sorée, W. Magnus, A. Mocuta, N. Collaert, A. Thean, and G. Groeseneken, J. Appl. Phys. **118**, 134502 (2015).
[26] C. S. Lent and D. J. Kirkner, J. Appl. Phys. **67**, 6353–6359 (1990).
[27] D. Verreck, "Quantum mechanical transport towards the optimization of heterostructure tunnel field-effect transistors," Ph.D. thesis (Katholieke Universiteit Leuven, 2017), pp. 173–178.
[28] M. L. Van de Put, "Modeling of quantum electron transport with applications in energy filtering nanostructures," Ph.D. thesis (Universiteit Antwerpen, 2016), pp. 115–119.
[29] U. Fano, Phys. Rev. **96**, 869–873 (1954).
[30] C. Rivas, R. Lake, G. Klimeck, W. R. Frensley, M. V. Fischetti, P. E. Thompson, S. L. Rommel, and P. R. Berger, Appl. Phys. Lett. **78**, 814–816 (2001).
[31] S. Rudin and T. L. Reinecke, Phys. Rev. B **41**, 7713–7717 (1990).
[32] H. Fröhlich, Adv. Phys. **3**, 325–361 (1954).
[33] M. G. Burt, Semicond. Sci. Technol. **3**, 739 (1988).
[34] D. Verreck, A. S. Verhulst, M. L. VandePut, B. Sorée, W. Magnus, N. Collaert, A. Mocuta, and G. Groeseneken, J. Appl. Phys. **124**, 204501 (2018).
[35] H. Fröhlich and F. Mott, Proc. R. Soc. Lond. A **171**, 496–504 (1939).