

# Tunneling between density-of-state tails: Theory and effect on Esaki diodes

Cite as: J. Appl. Phys. **128**, 014502 (2020); <https://doi.org/10.1063/5.0008709>

Submitted: 24 March 2020 . Accepted: 16 June 2020 . Published Online: 02 July 2020

A. Schenk , and S. Sant 



View Online



Export Citation



CrossMark

Lock-in Amplifiers  
up to 600 MHz



# Tunneling between density-of-state tails: Theory and effect on Esaki diodes

Cite as: J. Appl. Phys. 128, 014502 (2020); doi: 10.1063/5.0008709

Submitted: 24 March 2020 · Accepted: 16 June 2020 ·

Published Online: 2 July 2020



A. Schenk<sup>1,a)</sup>  and S. Sant<sup>2</sup> 

## AFFILIATIONS

<sup>1</sup>Integrated Systems Laboratory, ETH Zurich, Gloriastr. 35, 8092 Zurich, Switzerland

<sup>2</sup>Infineon Technologies AG, Am Campeon 1-15, 85579 Neubiberg, Germany

<sup>a)</sup>Author to whom correspondence should be addressed: [schenk@iis.ee.ethz.ch](mailto:schenk@iis.ee.ethz.ch)

## ABSTRACT

A model for tunneling between conduction and valence band tail states in semiconductors is developed. Localized, lifetime-broadened wave functions originally proposed by Vinogradov [Fiz. Tverd. Tela 13, 3266 (1971)] facilitate the derivation of the microscopic transition rate in a homogeneous electric field of arbitrary orientation. A compact analytical form of the average macroscopic tunnel generation rate is approximately calculated assuming that the Gaussian or exponential band tail represents a ladder of closely spaced single-level densities of states. A fully analytical form yields insight into key quantities like the effective tunnel barrier, the tunneling mass, and the pre-exponential factor in comparison to band-to-band tunneling. Tail-to-tail, tail-to-band, and band-to-band tunneling rates are compared against each other over a broad range of field strengths and characteristic tail energies. The numerical implementation of the model into a commercial device simulator accounts for the inhomogeneous field in pn-junctions and excludes invalid tunnel paths. In the application to a fully characterized InGaAs *pin*-Esaki diode, all physical processes and parameters that might affect the *IV*-characteristics are carefully investigated. The value of the bandgap of In<sub>0.53</sub>Ga<sub>0.47</sub>As as a function of density, doping, and temperature is revised. It is shown that tail-induced tunneling cannot explain the strong measured valley current of the diode. Besides band-to-band tunneling, zero- and multi-phonon defect-assisted tunneling are the physical mechanisms that allow to reproduce the entire forward characteristics. Whereas tail-to-band tunneling becomes only visible for very large values of the characteristic tail energy in the heavily doped regions, tail-to-tail tunneling remains a completely negligible process.

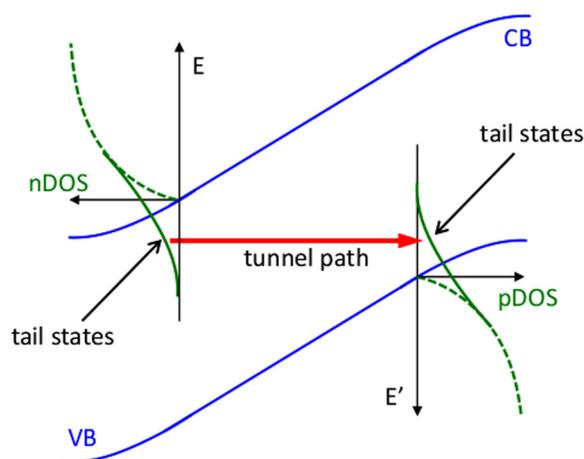
© 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0008709>

## I. INTRODUCTION

Tunnel Field Effect Transistors (TFETs) are still among the candidates for low-power switches with sub-thermionic sub-threshold swing (SS).<sup>2,3</sup> When operated in diode configuration, *IV*-characteristics with negative differential resistance (NDR) are the signature of tunneling in the forward-biased junction. In contrast to Esaki diodes,<sup>4</sup> TFETs utilize an intrinsic or lowly doped region to minimize the OFF-current of the transistor. The NDR behavior of the *pin*-diode (peak-to-valley ratio, shape, temperature dependence) provides useful information about the involved tunneling mechanisms, the density of states (DOS), and the doping profile. The latter is hard to determine precisely because of the small dimensions of nanowire (NW) TFETs and the unknown magnitude of incomplete ionization.<sup>5</sup> TFET performance is restricted by a number of non-ideality effects<sup>6</sup> including the presence of band tails.

Their detrimental impact on the TFET transfer characteristics has been the topic of recent experimental and theoretical studies.<sup>7–13</sup> The relative importance of band tails compared to other degradation mechanisms is still a matter of debate. Among the latter, Shockley–Read–Hall (SRH) generation and defect-assisted tunneling (DAT) at interfaces<sup>6</sup> and in bulk regions<sup>14</sup> are the most prominent. Another field-enhanced generation process is tunnel-assisted impact ionization,<sup>15</sup> an intrinsic limitation to the SS of TFETs.<sup>16</sup>

DOS tails extend the available states for band-to-band tunneling (BTBT) into the gap, leading to an earlier but also smoother onset of tunneling in TFETs due to the gradual increase of the joint DOS. Tails also extend the voltage range of BTBT *recombination* beyond the value where the alignment of the sharp band edges vanishes. This is illustrated in Fig. 1. The strength of the associated current depends not only on the characteristic



**FIG. 1.** Schematic band edge diagram of a semiconductor with DOS tails. Tunneling from CB tail states to VB tail states enables recombination after the alignment of the sharp band edges has disappeared under forward bias.

(band tail spreading) energy  $\eta$  and the spectral shape of the tails but also on the degree of localization of the wave functions. Origins of band tails are random doping and defects,<sup>17–20</sup> as well as thermal fluctuations of the lattice.<sup>21</sup> Pioneering work to derive the tail DOS comprises the Thomas–Fermi approximation by Kane,<sup>22</sup> the minimum counting method by Halperin and Lax,<sup>23</sup> and the optimal fluctuation method by Shkiovskii and Efros.<sup>24</sup>

In experimental studies, the tail DOS is usually approximated by an exponential [ $A\exp(-(\Delta E)/\eta)$ ] or by a Gaussian [ $A\exp(-(\Delta E)^2/\eta^2)$ ] function.<sup>25</sup> Redfield<sup>26</sup> assumed Gaussian shape and  $\eta$  in the range between 30 meV and 60 meV to interpret experimental conductivity data of compensated GaAs (doping  $\sim 3 \times 10^{18} \text{ cm}^{-3}$ ). Pankove<sup>27</sup> fitted optical absorption data of uncompensated n- and p-type GaAs (doping range  $4 \times 10^{16} \text{ cm}^{-3} - 1 \times 10^{20} \text{ cm}^{-3}$ ) to an exponential tail DOS. He obtained limits of  $\eta_c \approx 24 \text{ meV}$  at  $N_A = 10^{20} \text{ cm}^{-3}$  for the conduction band (CB) tail and of  $\eta_v \approx 20 \text{ meV}$  at  $N_D = 10^{19} \text{ cm}^{-3}$  for the valence band (VB) tail. His and the cited prior data by other authors revealed overall small values of  $\eta$  in the range between 5 meV and 10 meV up to a doping level of  $7 \times 10^{18} \text{ cm}^{-3}$ . As for bulk InAs, Dixon and Ellis<sup>28</sup> measured a shift of the absorption edge by 13 meV for  $N_A = 2.4 \times 10^{17} \text{ cm}^{-3}$  which, however, includes the rigid shift caused by doping-induced bandgap narrowing (BGN). From the slope of their  $\alpha(h\nu)$ -curve, one can infer  $\eta_c \approx 9 \text{ meV}$ . Their extracted Urbach temperature<sup>29,76</sup> of 80 K compares well with the value of 70 K found later by Malyutenko and Chernyakhovsky<sup>30</sup> from thermal emission experiments with samples of  $N_D - N_A = 1.6 \times 10^{16} \text{ cm}^{-3}$ . A value of  $\eta_c \approx 7 \text{ meV}$  can be extracted from absorption measurements at InSb samples<sup>31</sup> with  $N_A = 9.5 \times 10^{16} \text{ cm}^{-3}$ . Memisevic *et al.*<sup>11</sup> measured InAs/InGaAsSb/GaSb NW TFETs in forward-biased diode configuration and extracted  $\eta$  by fitting the negative slope of the semi-log NDR current peak in a small interval around  $V_{SD} = +0.2 \text{ V}$ . With the model assumption that the slope of the current in this range only

depends on the exponential tail DOS and that the voltage dependence of the tunnel probability can be neglected, the authors derived values of 60 meV. However, they assigned the tails to a hetero-interface  $D_{it}$ . This raises the interesting question of how a spatial localization of the tails in tunnel direction changes the picture obtained with constant  $\eta$ . The interpretation of their data is also hampered by the fact that only a few defects are present in the tiny active volume of the NW.<sup>14</sup> The continuum theory developed below is more appropriate to bulk-like homo-Esaki diodes. Such InGaAs devices were fabricated at IMEC and simulated by Bizindavyi *et al.*<sup>13</sup> In prior simulation work, except Ref. 12, the tail states were treated as extended (Bloch or plane-wave) states. In Ref. 12, we developed a theoretical model of tail-to-band tunneling (TTBT) where the tail states are localized. We argued that tunnel transitions between opposite tail states, i.e., tail-to-tail tunneling (TTTT), are negligible compared to TTBT because of the localization of the wave functions. It is the aim of the present paper to demonstrate this in great detail.

The paper is organized as follows. In Sec. II, a quantum-mechanical model for tunneling between tail states is developed. First, the microscopic transition rate between two opposing, localized tail states in a homogeneous electric field of arbitrary orientation is derived. Then, a compact analytical form of the macroscopic (average) tunnel generation rate at single-level tail states is approximately calculated. Assuming that the Gaussian and exponential tail DOS represent a ladder of closely spaced single-level DOSs, the corresponding weight functions are found to synthesize them. This leads to the TTTT generation rate in the form of a double energy integral. Further physical insight is gained by fully analytical solutions that are achievable for Gaussian tail shape in the cases of high field (HF) and very low field (VLF), respectively. At the end of Sec. II, the homogeneous field case is used to compare all tunneling rates with each other over a broad range of field strengths and characteristic tail energies  $\eta$ . Section III presents the device application of the developed model. It is shown how the TTTT rate is implemented in the commercial device simulator Sentaurus-Device (S-Device) of Synopsys<sup>44</sup> using the Dynamic Nonlocal Path (DNLP) algorithm. An in-depth simulation study of InGaAs *pin*-Esaki diodes fabricated at IMEC<sup>10,13</sup> is then performed based on the information from physical and electrical characterization. All physical processes and parameters that possibly influence the *IV*-curves are painstakingly investigated. In particular, injection-dependent bandgap narrowing (BGN) is determined with a random-phase-approximation (RPA)-based model, and the temperature dependence of the gap is revised. The electrically active doping is found by fitting the measured peak and reverse current. For this, the value of the light-hole mass is extracted from full-band calculations of the imaginary dispersion. The contributions of TTBT and TTTT are highlighted for spatially constant and doping-dependent characteristic energies of exponential DOS tails. As tail-induced tunneling cannot explain the strong measured valley current, zero- and multi-phonon DAT are included. Section IV summarizes the main findings, discusses limitations of the approach and gives an outlook on future applications. Appendix A contains the derivation of the RPA-based BGN model for  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ . Appendix B provides the ratio of the BTBT rates of Kane's two-band model and one-band effective mass approximation (EMA) model, which is needed for compatibility with the BTBT model of S-Device.

## II. QUANTUM-MECHANICAL MODEL FOR TUNNELING BETWEEN TAIL STATES

### A. Transition rate between two localized tail states in a homogeneous electric field of arbitrary orientation

We consider localized tail states with zero-field binding energies  $E_{tc,tv}$  measured from the corresponding band edges  $E_{c,v}$  situated at arbitrary positions  $\mathbf{x}_{c,v} = \{\mathbf{x}_{\perp c,\perp v}; z_{c,v}\}$ . The homogeneous electric field  $F$  is assumed to be aligned with the  $z$ -direction. To model the wave functions, the pseudo- $\delta$ -potential model<sup>1,32</sup> is used as it facilitates analytical solutions for the total potential  $-eFz + 4\pi E_{tc,tv} r_{c,v}^3 \delta(\mathbf{x})[1 + \mathbf{x} \cdot \nabla_{\mathbf{x}}]$ . When the single-band envelope<sup>33</sup> method is applied to this problem, the influence of the crystal potential can be taken into account via “effective” masses  $m_{tc,tv}$  of the localized electrons, which are related to their localization radius  $r_{c,v}$  by

$$E_{tc,tv} = \frac{\hbar^2}{2m_{tc,tv} r_{c,v}^2}. \quad (1)$$

This relation can be viewed as fitting of the parameters  $m_{tc,tv}$  to the localization radii  $r_{c,v}$ . Since tail states split from their corresponding bands, the effective masses will be comparable to the band masses, but not necessarily equal to them. Reasons for deviations are the presence of heavy and light holes, the band anisotropy, and the effect of the respective opposite band. Table I lists the localization radius  $r_{c,v}$  for various values of  $m_{tc,tv}$  at three values of  $E_{tc,tv}$ .<sup>12</sup>

The normalized ground state  $\Phi_{E00}^v$  of the envelope in the potential  $-eFz + 4\pi E_{tv} r_v^3 \delta(\mathbf{x})[1 + \mathbf{x} \cdot \nabla_{\mathbf{x}}]$  is given by<sup>1,12</sup>

$$\begin{aligned} \Phi_{E00}^v(\mathbf{x}; \mathbf{x}_v) = & \frac{\sqrt{\frac{e|F|\hbar^2}{8\pi^3(\hbar\theta_{tv})^3 m_{tv}}}}{\sqrt{\mathcal{F}(\xi_v)}\sqrt{\mathcal{F}^2(\xi_v) + \hat{\mathcal{G}}^2(\xi_v)}} \\ & \times \int_0^\infty dk_1 \int_0^\infty dk_2 \exp[i\mathbf{k} \cdot (\mathbf{x}_\perp - \mathbf{x}_{\perp v})] \\ & \times \left\{ \hat{\mathcal{G}}(\xi_v) Ai\left(\xi_v + \frac{E_{tv}^\perp}{\hbar\theta_{tv}}\right) Ai\left(\xi_v + \frac{E_{tv}^\perp + eF(z - z_v)}{\hbar\theta_{tv}}\right) \right. \\ & - \mathcal{F}(\xi_v) \left[ \Theta(z_v - z) Ai\left(\xi_v + \frac{E_{tv}^\perp}{\hbar\theta_{tv}}\right) \right. \\ & \times Bi\left(\xi_v + \frac{E_{tv}^\perp + eF(z - z_v)}{\hbar\theta_{tv}}\right) + \Theta(z - z_v) \\ & \left. \left. \times Ai\left(\xi_v + \frac{E_{tv}^\perp + eF(z - z_v)}{\hbar\theta_{tv}}\right) Bi\left(\xi_v + \frac{E_{tv}^\perp}{\hbar\theta_{tv}}\right) \right] \right\}, \quad (2) \end{aligned}$$

where  $\tilde{E} = E + E_g$  and

$$\hat{\mathcal{G}}(\xi_v) = \mathcal{G}(\xi_v) + \frac{1}{\pi} \sqrt{\frac{E_{tv}}{\hbar\theta_{tv}}}, \quad (3)$$

$$\mathcal{G}(t) = Ai'(t)Bi'(t) - tAi(t)Bi(t), \quad (4)$$

TABLE I. Localization radii  $r_c, v$  (in nm) for various values of the effective masses  $m_{tc}, tv$ .

$E_{tc}, tv = 0.01$ eV		$E_{tc}, tv = 0.025$ eV		$E_{tc}, tv = 0.05$ eV	
$m_{tc}, tv(m_0)$	$r_c, v$	$m_{tc}, tv(m_0)$	$r_c, v$	$m_{tc}, tv(m_0)$	$r_c, v$
0.001	87.34	0.001	55.24	0.001	39.06
0.01	27.62	0.01	17.47	0.01	12.35
0.025	17.47	0.025	11.05	0.025	7.81
0.05	12.35	0.05	7.81	0.05	5.52
0.1	8.73	0.1	5.52	0.1	3.91
1	2.77	1	1.75	1	1.23

$$\mathcal{F}(t) = Ai'(t)^2 - tAi(t)^2, \quad (5)$$

$$\xi_v = \frac{\tilde{E} + eFz_v}{\hbar\theta_{tv}}, \quad (6)$$

$$\hbar\theta_{tv} = \left(\frac{e^2 \hbar^2 F^2}{2m_{tv}}\right)^{1/3}.$$

Here,  $Ai$  and  $Bi$  are Airy functions of the first and second kind, respectively, and  $E_{tv}^\perp = \hbar^2 k^2 / (2m_{tv})$  denotes the transverse energy.

The normalized ground state  $\Phi_{E'00}^c$  of the envelope in the potential  $-eFz + 4\pi E_{tc} r_c^3 \delta(\mathbf{x})[1 + \mathbf{x} \cdot \nabla_{\mathbf{x}}]$  is analogous to Eq. (2),

$$\begin{aligned} \Phi_{E'00}^c(\mathbf{x}; \mathbf{x}_c) = & \frac{\sqrt{\frac{e|F|\hbar^2}{8\pi^3(\hbar\theta_{tc})^3 m_{tc}}}}{\sqrt{\mathcal{F}(\xi_c)}\sqrt{\mathcal{F}^2(\xi_c) + \hat{\mathcal{G}}^2(\xi_c)}} \\ & \times \int_0^\infty dk_1 \int_0^\infty dk_2 \exp[i\mathbf{k} \cdot (\mathbf{x}_\perp - \mathbf{x}_{\perp c})] \\ & \times \left\{ \hat{\mathcal{G}}(\xi_c) Ai\left(\xi_c + \frac{E_{tc}^\perp}{\hbar\theta_{tc}}\right) Ai\left(\xi_c + \frac{E_{tc}^\perp + eF(z - z_c)}{\hbar\theta_{tc}}\right) \right. \\ & - \mathcal{F}(\xi_c) \left[ \Theta(z_c - z) Ai\left(\xi_c + \frac{E_{tc}^\perp}{\hbar\theta_{tc}}\right) \right. \\ & \times Bi\left(\xi_c + \frac{E_{tc}^\perp + eF(z - z_c)}{\hbar\theta_{tc}}\right) + \Theta(z - z_c) \\ & \left. \left. \times Ai\left(\xi_c + \frac{E_{tc}^\perp + eF(z - z_c)}{\hbar\theta_{tc}}\right) Bi\left(\xi_c + \frac{E_{tc}^\perp}{\hbar\theta_{tc}}\right) \right] \right\}, \quad (7) \end{aligned}$$

with

$$\hat{\mathcal{G}}(\xi_c) = \mathcal{G}(\xi_c) + \frac{1}{\pi} \sqrt{\frac{E_{tc}}{\hbar\theta_{tc}}}, \quad (8)$$

$$\xi_c = -\frac{E' + eFz_c}{\hbar\theta_{tc}}, \quad (9)$$

$$\hbar\theta_{tc} = \left(\frac{e^2 \hbar^2 F^2}{2m_{tc}}\right)^{1/3}.$$

Here,  $E_{tc}^\perp = \hbar^2 k^2 / (2m_{tc})$  denotes the transverse energy. All energy variables are illustrated in Fig. 2. Note that above wave functions are proportional to the square root of the density of localized single-level states, which, in the limit of vanishing field, turns into a  $\delta$ -function of energy with a zero at the binding energy,

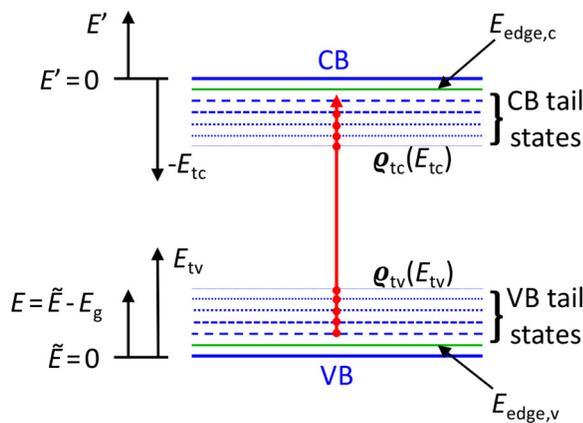
$$Q_{tc,tv}(\epsilon, E_{tc,tv}) = \frac{1}{2\pi r_{c,v}^3} \delta(\epsilon - E_{tc,tv}), \quad (10)$$

where  $\epsilon = \tilde{E}$  for states of the VB DOS and  $\epsilon = -E'$  for states of the CB DOS, respectively. As tunneling between tail states in Esaki diodes and TFETs might be an important process in the low-field range, the relevant energies  $\tilde{E}$  are close to  $E_{tv}$ , whereas the relevant energies  $E'$  are close to  $-E_{tc}$ . Due to this,  $\tilde{G} \ll \mathcal{F}$ , and the first term in the curly braces of Eqs. (2) and (7) (proportional to  $\tilde{G}$ ) can be skipped. Using

$$\begin{aligned} & \Theta(x-y)Ai(x)Bi(y) + \Theta(y-x)Bi(x)Ai(y) \\ &= -\frac{1}{\pi} \int_{-\infty}^{\infty} d\lambda \frac{\mathcal{P}}{\lambda} Ai(\lambda+x)Ai(\lambda+y), \end{aligned} \quad (11)$$

where  $\mathcal{P}$  is the Cauchy principal value, and introducing a function  $\mathcal{D}(\xi)$  by the definition

$$\mathcal{D}(\xi) = \frac{\mathcal{F}(\xi)}{\mathcal{F}^2(\xi) + \tilde{\mathcal{G}}^2(\xi)}, \quad (12)$$



**FIG. 2.** Representation of energy variables in the calculation of the macroscopic generation rate of tail-to-tail tunneling. Note that  $E_{tc}$  is positive, i.e.,  $-E_{tc}$  is a negative energy on the  $E'$ -axis.

the wave function  $\Phi_{E'00}^v$  takes the form

$$\begin{aligned} \Phi_{E'00}^v(\mathbf{x}; \mathbf{x}_v) &= \sqrt{\frac{e|F|\hbar^2 \mathcal{D}(\xi_v)}{8\pi^5 (\hbar\theta_{tv})^3 m_{tv}}} \int_0^\infty dk_1 \int_0^\infty dk_2 e^{i\mathbf{k} \cdot (\mathbf{x}_\perp - \mathbf{x}_{\perp v})} \\ &\times \int_{-\infty}^\infty d\lambda \frac{\mathcal{P}}{\lambda - \xi_v} Ai\left(\lambda + \frac{E_{tv}^\perp + eF(z - z_v)}{\hbar\theta_{tv}}\right) Ai\left(\lambda + \frac{E_{tv}^\perp}{\hbar\theta_{tv}}\right). \end{aligned} \quad (13)$$

In the same way, the wave function  $\Phi_{E'00}^c$  becomes

$$\begin{aligned} \Phi_{E'00}^c(\mathbf{x}; \mathbf{x}_c) &= \sqrt{\frac{e|F|\hbar^2 \mathcal{D}(\xi_c)}{8\pi^5 (\hbar\theta_{tc})^3 m_{tc}}} \int_0^\infty dk_1 \int_0^\infty dk_2 e^{i\mathbf{k} \cdot (\mathbf{x}_\perp - \mathbf{x}_{\perp c})} \\ &\times \int_{-\infty}^\infty d\lambda \frac{\mathcal{P}}{\lambda - \xi_c} Ai\left(\lambda + \frac{E_{tc}^\perp - eF(z - z_c)}{\hbar\theta_{tc}}\right) Ai\left(\lambda + \frac{E_{tc}^\perp}{\hbar\theta_{tc}}\right). \end{aligned} \quad (14)$$

In Eqs. (13) and (14), the functions  $\sqrt{\mathcal{D}(\xi_{v,c})}$  are proportional to the respective single-level DOSs  $\rho_{tv,tc}$  that have the zero-field limit (10) and that will be later used to define the tail DOSs in Eq. (46). In calculating the transition matrix element,

$$\begin{aligned} M_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') &= \left( \Phi_{E'00}^v(\mathbf{x}_v) | \Phi_{E'00}^c(\mathbf{x}_c) \right) \\ &= \int d^2 \mathbf{x}_\perp \int_{-\infty}^\infty dz \Phi_{E'00}^{v*}(\mathbf{x}; \mathbf{x}_v) \Phi_{E'00}^c(\mathbf{x}; \mathbf{x}_c), \end{aligned} \quad (15)$$

the space integration can be done exactly by means of

$$\int d^2 \mathbf{x}_\perp e^{i\mathbf{x}_\perp \cdot (\mathbf{k} - \boldsymbol{\kappa})} = (2\pi)^2 \delta(k_1 - \kappa_1) \delta(k_2 - \kappa_2),$$

and the auxiliary relation<sup>34</sup>

$$\int_{-\infty}^\infty d\tau Ai(x + \tau) Ai(y - \beta\tau) = \frac{1}{(1 + \beta^3)^{1/3}} Ai\left(\frac{y + \beta x}{(1 + \beta^3)^{1/3}}\right), \quad (16)$$

where  $\beta = (m_{tv}/m_{tc})^{1/3} > 0$ . The transition matrix element takes the form

$$\begin{aligned} M_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') &= \frac{\sqrt{\mathcal{D}(\xi_c)\mathcal{D}(\xi_v)}}{\beta\pi^2 \hbar\theta_{red}} \int_0^\infty dt J_0(\alpha\sqrt{t}) \\ &\times \int_{-\infty}^\infty d\epsilon \int_{-\infty}^\infty d\epsilon' \frac{\mathcal{P}}{\epsilon' - E'} \frac{\mathcal{P}}{\epsilon' + \epsilon - \tilde{E}} Ai\left(t - \frac{\epsilon' + eFz_c}{\hbar\theta_{tc}}\right) \\ &\times Ai\left(\frac{t}{\beta^2} + \frac{\epsilon + \epsilon' + eFz_v}{\hbar\theta_{tv}}\right) Ai\left(\frac{t}{\beta_c^2} + \frac{\epsilon}{\hbar\theta_{red}}\right), \end{aligned} \quad (17)$$

where  $J_0$  denotes the Bessel function of the first kind, of zero order,<sup>35</sup> and

$$\alpha = \sqrt{\frac{2m_{tc}\hbar\theta_{tc}}{\hbar^2}}r_{cv}, \quad r_{cv} = |\mathbf{x}_{\perp c} - \mathbf{x}_{\perp v}|, \quad (18)$$

$$\hbar\theta_{red} = \left(\frac{e^2\hbar^2F^2}{2m_{red}}\right)^{1/3}, \quad \beta_c = \left(\frac{m_{red}}{m_{tc}}\right)^{1/3},$$

$$\frac{1}{m_{red}} = \frac{1}{m_{tc}} + \frac{1}{m_{tv}}. \quad (19)$$

Note that the overlap in field direction ( $z$ -integral) leads to the occurrence of the reduced effective mass  $m_{red}$  in the factor that later determines the tunnel probability. The radial

distance of the “traps,”  $r_{cv}$ , enters the Bessel function, which rapidly decays with growing distance perpendicular to the field direction.

For the further calculation of the transition matrix element, it is necessary to manipulate the product of the Cauchy principal values of integrals over  $\epsilon'$  in Eq. (17). The same step must be done in the proof of the normalization of the wave functions Eqs. (2) and (7). One can show by direct computation<sup>36</sup> that

$$\frac{\mathcal{P}}{\epsilon' - E' \epsilon' + \epsilon - \tilde{E}} = \frac{\mathcal{P}}{E' - (\tilde{E} - \epsilon)} \left[ \frac{\mathcal{P}}{\epsilon' - (\tilde{E} - \epsilon)} - \frac{\mathcal{P}}{\epsilon' - E'} \right] + \pi^2 \delta(\tilde{E} - \epsilon - E') \delta(\tilde{E} - \epsilon - \epsilon'). \quad (20)$$

The transition matrix element becomes

$$M_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') = \frac{\sqrt{\mathcal{D}(\xi_c)\mathcal{D}(\xi_v)}}{\beta\pi^2\hbar\theta_{red}} \int_0^\infty dt J_0(\alpha\sqrt{t}) \left\{ \pi^2 Ai\left(t - \frac{E' + eFz_c}{\hbar\theta_{tc}}\right) Ai\left(\frac{t}{\beta^2} + \frac{\tilde{E} + eFz_v}{\hbar\theta_{tv}}\right) Ai\left(\frac{t}{\beta_c^2} + \frac{\tilde{E} - E'}{\hbar\theta_{red}}\right) + \int_{-\infty}^\infty d\epsilon \int_{-\infty}^\infty d\epsilon' \frac{\mathcal{P}}{E' - (\tilde{E} - \epsilon)} \left[ \frac{\mathcal{P}}{\epsilon' - (\tilde{E} - \epsilon)} - \frac{\mathcal{P}}{\epsilon' - E'} \right] Ai\left(t - \frac{\epsilon' + eFz_c}{\hbar\theta_{tc}}\right) Ai\left(\frac{t}{\beta^2} + \frac{\epsilon + \epsilon' + eFz_v}{\hbar\theta_{tv}}\right) Ai\left(\frac{t}{\beta_c^2} + \frac{\epsilon}{\hbar\theta_{red}}\right) \right\}. \quad (21)$$

Due to energy conservation, it is only needed for  $E = E'$  (Golden Rule). For fixed positions and given zero-field energy levels of the tail states, the possible tunnel energies  $E$  are restricted to a narrow range, because the lifetime broadening of the binding energy is rather weak in the low-field regime of interest. The functions  $\mathcal{D}(\xi_{v,c})$  [Eq. (12)] are proportional to the respective DOS  $\rho_{tv,tc}$ ,<sup>12</sup>

$$\mathcal{D}(\xi_v) = 4\pi^3 r_v^3 \sqrt{E_{tv}\hbar\theta_{tv}} \rho_{tv}(\xi_v) \rightarrow 2\pi^2 \sqrt{E_{tv}\hbar\theta_{tv}} \delta(E + E_g + eFz_v - E_{tv}) \text{ for } \hbar\theta_{tv} \ll E_g, \quad (22)$$

$$\mathcal{D}(\xi_c) = 4\pi^3 r_c^3 \sqrt{E_{tc}\hbar\theta_{tc}} \rho_{tc}(\xi_c) \rightarrow 2\pi^2 \sqrt{E_{tc}\hbar\theta_{tc}} \delta(E' + eFz_c + E_{tc}) \text{ for } \hbar\theta_{tc} \ll E_g. \quad (23)$$

They are sharp Lorentzians with peak positions close to the energy zeros of the  $\delta$ -functions in Eqs. (22) and (23) (only displaced by a small Stark effect). As the transition probability is  $\sim \mathcal{D}(\xi_v)\mathcal{D}(\xi_c)$ , the product of the two  $\delta$ -functions results in the condition

$$z_c - z_v = (E_g - E_{tv} - E_{tc})/eF. \quad (24)$$

A tunnel process between two tail states is only possible if their distance in field direction equals the tunnel length  $|z_c - z_v|$  given by Eq. (24). The corresponding tunnel barrier is the bandgap reduced by the two binding energies.

Using the low-field limit, the first term in curly braces in Eq. (21) becomes

$$\pi^2 Ai\left(t + \frac{E_{tc}}{\hbar\theta_{tc}}\right) Ai\left(\frac{t}{\beta^2} + \frac{E_{tv}}{\hbar\theta_{tv}}\right) Ai\left(\frac{t}{\beta_c^2} + \frac{E_g}{\hbar\theta_{red}}\right). \quad (25)$$

As straightened out later, this term is small compared to the second one and hence discarded in the following. It remains

$$M_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') = \frac{\sqrt{\mathcal{D}(\xi_c)\mathcal{D}(\xi_v)}}{\beta\pi^2\hbar\theta_{red}} \int_0^\infty dt J_0(\alpha\sqrt{t}) \times \int_{-\infty}^\infty d\epsilon \frac{\mathcal{P}}{\epsilon} \left[ Ai\left(\frac{t}{\beta^2} + \xi_v + \frac{\epsilon}{\hbar\theta_{tv}}\right) \int_{-\infty}^\infty d\lambda \frac{\mathcal{P}}{\lambda} \times Ai\left(\frac{t}{\beta_c^2} + \frac{E_g}{\hbar\theta_{red}} + \beta_c\lambda\right) Ai\left(t + \xi_c - \frac{\epsilon}{\hbar\theta_{tc}} + \lambda\right) + Ai\left(t + \xi_c + \frac{\epsilon}{\hbar\theta_{tc}}\right) \int_{-\infty}^\infty d\lambda \frac{\mathcal{P}}{\lambda} \times Ai\left(\frac{t}{\beta_c^2} + \frac{E_g}{\hbar\theta_{red}} + \beta_v\lambda\right) Ai\left(\frac{t}{\beta^2} + \xi_v - \frac{\epsilon}{\hbar\theta_{tv}} + \lambda\right) \right], \quad (26)$$

with

$$\beta_v = \left(\frac{m_{red}}{m_{tv}}\right)^{1/3}.$$

The principal value integrals over  $\lambda$  can only be calculated approximately<sup>12,37</sup> assuming  $\beta_{v,c}^3 \ll 1$ ,

$$\int_{-\infty}^{\infty} d\lambda \frac{\mathcal{P}}{\lambda} Ai(x + \beta_{v,c}\lambda) Ai(y + \lambda) \approx -\frac{\mathcal{P}}{y} Ai(x - \beta_{v,c}y). \quad (27)$$

With this, one obtains

$$\begin{aligned} M_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') &= \frac{\sqrt{\mathcal{D}(\xi_c)\mathcal{D}(\xi_v)}}{\beta\pi^2\hbar\theta_{red}} \int_0^\infty dt J_0(\alpha\sqrt{t}) \\ &\times \int_{-\infty}^{\infty} d\epsilon \frac{\mathcal{P}}{\epsilon} \left[ \beta \frac{\mathcal{P}}{\epsilon - \tau_c} Ai\left(\frac{t}{\beta^2} + \xi_v + \epsilon\right) \right. \\ &\times Ai\left(\frac{\beta_c t}{\beta^3} + \frac{E_g}{\hbar\theta_{red}} - \beta_c \xi_c + \beta_v \epsilon\right) + \frac{1}{\beta\epsilon - \tau_v} \\ &\left. \times Ai\left(t + \xi_c + \epsilon\right) Ai\left(\beta_c t + \frac{E_g}{\hbar\theta_{red}} - \frac{\beta_c \xi_v}{\beta} + \beta_c \epsilon\right) \right], \end{aligned} \quad (28)$$

with the abbreviations

$$\tau_c = \beta(t + \xi_c) \quad \tau_v = \frac{1}{\beta} \left( \frac{t}{\beta^2} + \xi_v \right). \quad (29)$$

The product of the Cauchy principal values of integrals over  $\epsilon$  in Eq. (28) is again decomposed with the help of Eq. (20), and the new principal value integrals are again computed using the approximation (27). This results in

$$\begin{aligned} M_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') &= \frac{\sqrt{\mathcal{D}(\xi_c)\mathcal{D}(\xi_v)}}{\beta\pi^2\hbar\theta_{red}} \int_0^\infty dt J_0(\alpha\sqrt{t}) \\ &\times \left[ \pi^2 \delta(\tau_c) Ai\left(\xi_v - \frac{\xi_c}{\beta^2}\right) Ai\left(\frac{E_g}{\hbar\theta_{red}} - \frac{\xi_c}{\beta^2}\right) \right. \\ &+ \pi^2 \delta(\tau_v) Ai(\xi_c - \beta^2 \xi_v) Ai\left(\frac{E_g}{\hbar\theta_{red}} - \frac{\xi_v}{\beta^2}\right) \\ &+ \frac{\mathcal{P}}{\tau_c} \left( \frac{\mathcal{P}}{\tau_v + \tau_c/\beta} - \frac{\mathcal{P}}{\tau_v} \right) Ai\left(\frac{E_g}{\hbar\theta_{red}} - \beta_c \xi_c - \beta_v \xi_v\right) \\ &\left. + \frac{\mathcal{P}}{\tau_v} \left( \frac{\mathcal{P}}{\tau_c + \beta\tau_v} - \frac{\mathcal{P}}{\tau_c} \right) Ai\left(\frac{E_g}{\hbar\theta_{red}} - \beta_v \xi_v - \beta_c \xi_c\right) \right]. \end{aligned} \quad (30)$$

Due to the sharp maxima of  $\mathcal{D}(\xi_{v,c})$  at  $\xi_{v,c} \approx E_{tv,tc}/\hbar\theta_{tv,cv}$  [see Eqs. (22) and (23)],  $\tau_{v,c} > 0 \forall t$ , and the delta-functions  $\delta(\tau_{v,c})$  are always zero. Furthermore, the Cauchy principal values become obsolete. The transition matrix element takes the final form

$$\begin{aligned} M_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') &= -\frac{\sqrt{\mathcal{D}(\xi_v)\mathcal{D}(\xi_c)}}{\beta\pi^2\hbar\theta_{red}} Ai\left(\frac{E_g}{\hbar\theta_{red}} - \beta_v \xi_v - \beta_c \xi_c\right) \\ &\times \int_0^\infty dt \frac{J_0(\alpha\sqrt{t})}{\left(t + \frac{E_{tc}}{\hbar\theta_c}\right)\left(\frac{t}{\beta^2} + \xi_v\right)}. \end{aligned} \quad (31)$$

According to the envelope method,<sup>33</sup> the tunnel rate between the two tail states (dimension 1/s) is given by

$$G_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c) = \frac{(eF)^2 z_{cv}^2}{\hbar} \int_0^{E_g} d\tilde{E} \int_{-E_g}^0 dE' M_{tv,tc}^2(\mathbf{x}_v, \mathbf{x}_c; \tilde{E}, E') \delta(E - E'), \quad (32)$$

where  $z_{cv}$  is the interband transition matrix element<sup>38</sup>  $z_{cv}^2 = \hbar^2/(4m_r E_g)$  with the reduced effective mass  $m_r = m_c m_v / (m_c + m_v) = m_c m_v / m_\Sigma$ . For the completion of the band-to-band process (measurable as terminal current), it is assumed that the thermionic emission step between the tail state and its corresponding band is very fast and, therefore, not rate limiting.

Taking advantage of the low-field limits Eqs. (22) and (23), the emission rate can be easily evaluated,

$$\begin{aligned} G_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c) &= \frac{2(\hbar\theta_{red})^2 m_{red} \sqrt{\beta_v \beta_c} \sqrt{E_{tv} E_{tc}}}{\hbar m_r \beta^2 E_g} \\ &\times Ai^2\left(\frac{E_g - E_{tv} - E_{tc}}{\hbar\theta_{red}}\right) \left[ \int_0^\infty dt \frac{J_0(\alpha\sqrt{t})}{\left(t + \frac{E_{tc}}{\hbar\theta_c}\right)\left(\frac{t}{\beta^2} + \frac{E_{tv}}{\hbar\theta_v}\right)} \right]^2 \\ &\times \delta(E_g - E_{tv} - E_{tc} - eF(z_c - z_v)), \end{aligned} \quad (33)$$

where the parameter  $\alpha = \sqrt{\frac{2m_c \hbar \theta_{tc}}{\hbar^2} r_{cv}}$  contains the radial distance  $r_{cv}$ . The integral over  $t$  could now be calculated approximately; however, for the derivation of the macroscopic tunnel generation rate in Sec. II B, it is beneficial to use Eq. (33) as is.

### B. Macroscopic tunnel generation rate at single-level tail states

To obtain the macroscopic tunnel generation rate at single-level tail states, one has to multiply  $G_{tv,tc}(\mathbf{x}_v, \mathbf{x}_c)$  by the density of states at sites  $\mathbf{x}_c$  and to sum over all sites  $\mathbf{x}_v$ . Choosing  $\mathbf{x}_c$  as origin of coordinates this results in

$$G(E_{tv}, E_{tc}) = \frac{1}{2\pi r_c^3} \frac{1}{2\pi r_v^3} \int_{-\infty}^\infty dz_v 2\pi \int_0^\infty dr_{cv} r_{cv} G_{tv,tc}(\mathbf{x}_v, \mathbf{0}). \quad (34)$$

Inserting Eq. (33) in Eq. (34), the  $z_v$  integration is trivial and the  $r_{cv}$ -integral yields

$$\begin{aligned} 2\pi \int_0^\infty dr_{cv} r_{cv} J_0(\gamma\sqrt{t} r_{cv}) J_0(\gamma\sqrt{t'} r_{cv}) &= \frac{1}{\gamma^2 \sqrt{t}} \delta(\sqrt{t} - \sqrt{t'}) \\ &= \frac{2}{\gamma^2} \delta(t - t'), \end{aligned} \quad (35)$$

with  $\gamma^2 = 2m_{tc} \hbar \theta_{tc} / \hbar^2$ . The generation rate becomes

$$\begin{aligned} G(E_{tv}, E_{tc}) &= \frac{4\theta_{red} m_{red} m_{tv} m_{tc} (E_{tv} E_{tc})^2}{eF\pi^2 \hbar^4 m_r \beta E_g} Ai^2\left(\frac{E_g - E_{tv} - E_{tc}}{\hbar\theta_{red}}\right) \\ &\times \int_0^\infty dt \frac{1}{\left(t + \frac{E_{tc}}{\hbar\theta_c}\right)^2 \left(\frac{t}{\beta^2} + \frac{E_{tv}}{\hbar\theta_v}\right)^2}. \end{aligned} \quad (36)$$

The exact solution of the  $t$ -integral is

$$I(E_{\text{tv}}, E_{\text{tc}}) = \int_0^\infty dt \frac{1}{\left(t + \frac{E_{\text{tc}}}{\hbar\theta_{\text{red}}}\right)^2 \left(\frac{t}{\beta^2} + \frac{E_{\text{tv}}}{\hbar\theta_{\text{red}}}\right)^2} = \beta^4 \frac{(E_{\text{tc}}m_{\text{tc}})^2 - (E_{\text{tv}}m_{\text{tv}})^2 + 2E_{\text{tc}}E_{\text{tv}}m_{\text{tc}}m_{\text{tv}} \ln\left(\frac{E_{\text{tv}}m_{\text{tv}}}{E_{\text{tc}}m_{\text{tc}}}\right)}{E_{\text{tc}}E_{\text{tv}}m_{\text{tc}}m_{\text{tv}} \left(\frac{E_{\text{tc}}}{\hbar\theta_{\text{red}}} - \frac{m_{\text{tv}}E_{\text{tv}}}{m_{\text{tc}}\hbar\theta_{\text{red}}}\right)^3}. \quad (37)$$

As this is cumbersome and not feasible for further analytical treatment, an elegant way of simplification is to use  $E_{\text{tv}}m_{\text{tv}} = E_{\text{tc}}m_{\text{tc}}$  in  $I(E_{\text{tv}}, E_{\text{tc}})$ , which means that the localization radii are assumed to be equal:  $r_c = r_v$  [see Eq. (1)]. For this special case,

$$I(E_{\text{tv}}, E_{\text{tc}}) = \frac{\beta}{3} \left(\frac{\hbar\theta_{\text{red}}}{E_{\text{tv}}E_{\text{tc}}}\right)^{3/2}, \quad (38)$$

and the macroscopic tunnel generation rate at single-level tail states takes the final form

$$G(E_{\text{tv}}, E_{\text{tc}}) = \frac{2eF\theta_{\text{red}}m_{\text{red}}\sqrt{m_{\text{tv}}m_{\text{tc}}}\sqrt{E_{\text{tv}}E_{\text{tc}}}}{3\pi^2\hbar^2m_r} \frac{1}{E_g} Ai^2\left(\frac{E_g - E_{\text{tv}} - E_{\text{tc}}}{\hbar\theta_{\text{red}}}\right). \quad (39)$$

This can be also written as

$$G(E_{\text{tv}}, E_{\text{tc}}) = \frac{1}{3\pi^2} \frac{\theta_{\text{red}}}{r_c r_v l_t} Ai^2\left(\frac{E_g - E_{\text{tv}} - E_{\text{tc}}}{\hbar\theta_{\text{red}}}\right), \quad (40)$$

with the localization radii  $r_{c,v}$  and the interband tunnel length  $l_t = E_g/eF$ . Dimensions are determined by the electro-optical frequency  $\theta_{\text{red}}$  (1/s) and the effective volume  $r_c r_v l_t$  ( $\text{cm}^3$ ). The order of magnitude of the tunnel probability is governed by the factor  $Ai^2$  which can be replaced by its asymptotic form (semi-classical limit)

$$Ai^2(y) = \frac{1}{4\pi\sqrt{y}} e^{-4y^{3/2}}, \quad (41)$$

as long as the tunnel barrier  $E_g - E_{\text{tv}} - E_{\text{tc}}$  is much larger than  $\hbar\theta_{\text{red}}$ .

### C. DOS tail models

According to Kane's theory,<sup>22</sup> the DOS in the presence of random dopant fluctuations and crystal defects takes the form

$$\rho_{v,c}^{\text{tail}}(E) = \frac{(2m_{\text{tv,tc}})^{3/2}}{2\pi^2\hbar^3} \sqrt{\eta_{v,c}} Y_{G/\text{exp}}(E/\eta_{v,c}), \quad (42)$$

where  $\eta_{v,c}$  is the characteristic energy of the valence band (VB) and conduction band (CB) tail, respectively. In Eq. (42), the energy  $E$  counts from the respective band edge into the gap. The two standard models of DOS tails, Gaussian (G) and exponential (exp), will be used in the following. In the case of Gaussian tails the function

$Y_G(E/\eta)$  is given by

$$Y_G(x) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^x d\zeta \sqrt{x - \zeta} e^{-\zeta^2}, \quad (43)$$

which can be approximated by<sup>12</sup>

$$Y_G(x) \rightarrow \frac{e^{-x^2}}{2^{5/2}(x^{3/2} + s)}, \quad (44)$$

with<sup>12</sup>  $s = 0.566$ .

In the case of exponential tails,  $Y(E/\eta)$  becomes

$$Y_{\text{exp}}(x) = \frac{1}{2} \int_{-\infty}^x d\zeta \sqrt{x - \zeta} e^{-|\zeta|} = \frac{1}{4\sqrt{\pi}} e^{-|x|} \quad \text{for } x < 0. \quad (45)$$

### D. Tail-to-tail tunnel generation rate

The tail DOS Eq. (42) is assumed to be a ladder of closely spaced single-level DOSs defined in Eqs. (22) and (23) and will be composed with weight functions  $w(E_{\text{tv,tc}})$  as

$$\rho_{v,c}^{\text{tail}}(E) = \int_{E_{\text{edge},\{v,c\}}}^{E_g/2} dE_{\text{tv,tc}} w(E_{\text{tv,tc}}) \rho_{\text{tv,tc}}(E, E_{\text{tv,tc}}). \quad (46)$$

The energy variables used in this equation are defined in Fig. 2. The integration over tail states is restricted to  $E_{\text{edge},\{v,c\}} < E < E_g/2$ . The lower limit  $E_{\text{edge},\{v,c\}}$  separates localized states from continuum states, and it is assumed that  $0 < E_{\text{edge},\{v,c\}} < \eta_{v,c}$ . Thus,  $E_{\text{edge},\{v,c\}}$  plays the same role as the "mobility edge"<sup>39</sup> in transport.

The weight functions  $w(E_{\text{tv,tc}})$  immediately follow from equating Eqs. (42) and (46) and using the strongly localized character of the single-level DOSs,

$$w_{G,\text{exp}}(E_{\text{tv,tc}}) = \frac{\sqrt{\eta_{v,c}}}{\pi E_{\text{tv,tc}}^{3/2}} Y_{G,\text{exp}}(E_{\text{tv,tc}}/\eta_{v,c}). \quad (47)$$

The tail-to-tail tunnel generation rate (in a homogeneous field, assuming occupied initial and empty final electronic states) is given by

$$G_{G,\text{exp}} = \frac{2eF\theta_{\text{red}}m_{\text{red}}\sqrt{m_{\text{tv}}m_{\text{tc}}}}{3\pi^4\hbar^2m_rE_g} \int_{E_{\text{edge},c}}^{E_g/2} dE_{\text{tc}} \int_{E_{\text{edge},v}}^{E_g/2} dE_{\text{tv}} \frac{\sqrt{\eta_v\eta_c}}{E_{\text{tv}}E_{\text{tc}}} \times Y_{G,\text{exp}}\left(\frac{E_{\text{tv}}}{\eta_v}\right) Y_{G,\text{exp}}\left(\frac{E_{\text{tc}}}{\eta_c}\right) Ai^2\left(\frac{E_g - E_{\text{tv}} - E_{\text{tc}}}{\hbar\theta_{\text{red}}}\right). \quad (48)$$

An analytical estimate can be obtained in the case of Gaussian DOS with the method of steepest descent.<sup>12</sup> The two-dimensional integrand has a sharp global maximum due to the exponential decrease of the tail DOS and the exponential growth of the tunnel probability with increasing  $E_{\text{tv,tc}}$ . Using the WKB form of  $Ai^2$ , the two-dimensional exponent is developed up to second order around the global maximum ( $\Delta_c, \Delta_v$ ), whereas the pre-exponential factors are taken at  $E_{\text{tc}} = \Delta_c$  and  $E_{\text{tv}} = \Delta_v$ . The resulting expression

is integrated over the upper-right quadrant of the  $E_{tc}$ - $E_{tv}$ -plane without any further approximation. With the general and reasonable assumption  $\eta_{v,c} \ll E_g$ , simple analytical formulas can then be found in the cases of high field (HF) and very low field (VLF), respectively. In the first case, one obtains

$$G_G^{HF} = \frac{(eF)^2 \sqrt{m_{tv} m_{tc} m_{red}} (\eta_v \eta_c)^{3/2}}{768 \sqrt{2} \pi^5 \hbar^2 E_g m_r \Delta_v \Delta_c \sqrt{E_g - \Delta_c - \Delta_v}} \times \frac{\exp \left[ -\frac{\Delta_v^2}{\eta_c^2} - \frac{\Delta_c^2}{\eta_v^2} - \frac{4}{3} \left( \frac{E_g - \Delta_v - \Delta_c}{\hbar \theta_{red}} \right)^{3/2} \right]}{\left[ (\Delta_v / \eta_v)^{3/2} + s \right] \left[ (\Delta_c / \eta_c)^{3/2} + s \right]}, \quad (49)$$

with

$$\Delta_{v,c} \approx \frac{\sqrt{E_g \eta_{v,c}^2}}{(\hbar \theta_{red})^{3/2}} \quad (50)$$

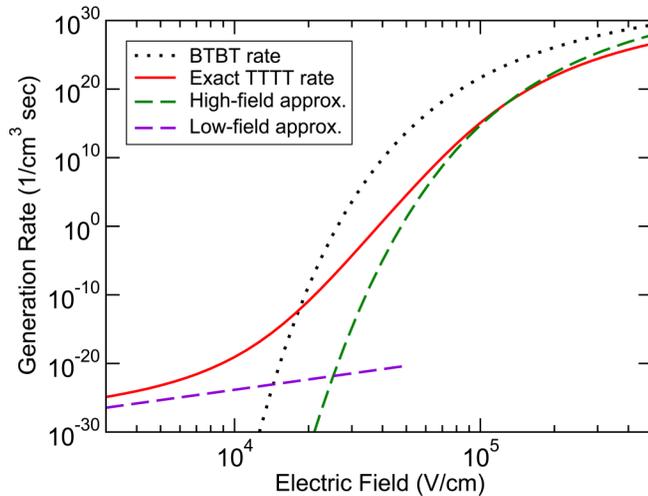
and  $s = 0.566$ . These expressions can only be used if  $F > 5 \times 10^4$  V/cm (see Fig. 3).

In the limit  $F \rightarrow 0$ , the generation rate takes the analytical form

$$G_G^{VLF} = \frac{(eF)^5 \hbar \sqrt{m_{tv} m_{tc}} (\eta_c^2 + \eta_v^2)^4}{3 \pi^5 2^9 E_g^4 m_{red} m_r (\eta_c \eta_v)^{9/2}} \exp \left( -\frac{E_g^2}{\eta_c^2 + \eta_v^2} \right), \quad (51)$$

which is only applicable if  $F < 5 \times 10^3$  V/cm (see Fig. 3).

When the one-band effective mass approximation and the WKB limit are applied to compute the generation rate of direct



**FIG. 3.** Generation rates due to tunneling between tail states with Gaussian DOS in a homogeneous electric field. The analytical approximations for high field equation (49) and very low field equation (51) are compared with the double-integral form Eq. (48). The BTBT rate with band masses of InAs,  $m_c = 0.023 m_0$ , and  $m_v = 0.026 m_0$  is shown for comparison. Parameters:  $E_g = 0.36$  eV,  $m_{tc} = m_{tv} = 0.1 m_0$ , and  $\eta_c = \eta_v = 0.025$  eV.

BTBT (with ideal DOSs), one obtains<sup>40–42</sup>

$$G_{BTB} = \frac{(eF)^3}{64 \pi \hbar E_g^2} \exp \left[ -\frac{4}{3} \left( \frac{E_g}{\hbar \theta_r} \right)^{3/2} \right]. \quad (52)$$

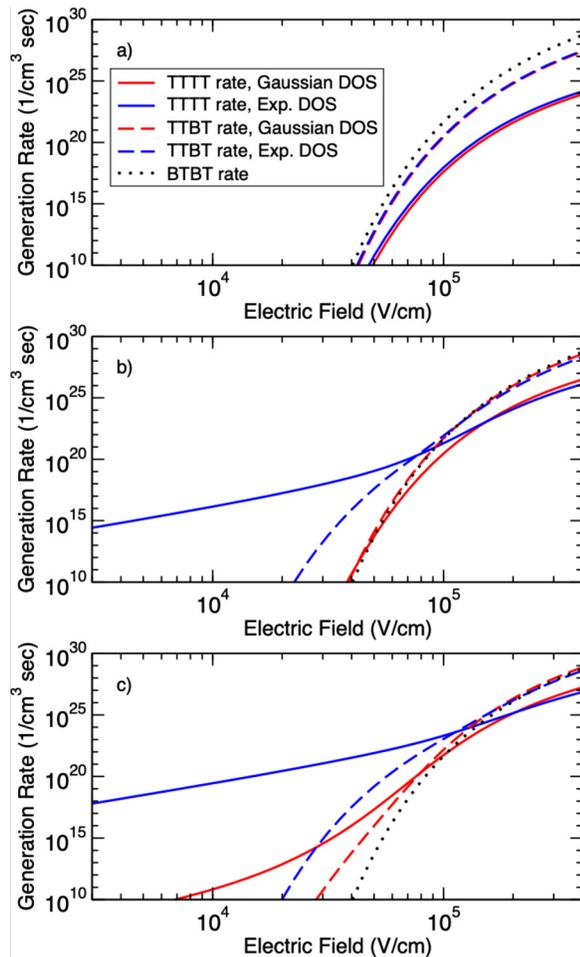
Figure 3 compares the analytical solutions Eqs. (49) and (51) with the double-integral form Eq. (48) for a Gaussian tail DOS and moderate localization of the tail states. The agreement is reasonable in the two relevant field ranges. For the chosen parameters, the TTTT rate is lower than the corresponding BTBT rate equation (52) for  $F > 2 \times 10^4$  V/cm. Generation rates below  $10^{10}$  cm<sup>-3</sup> s<sup>-1</sup> are irrelevant for device application. Thus, the HF case, Eq. (49) should be compared with the BTBT rate Eq. (52). Three differences become obvious: (i) the tunnel barrier ( $E_g$ ) is effectively reduced by  $\Delta_v + \Delta_c$  due to the energetic separation of the tail states from the ideal bands, (ii) the imaginary dispersion is determined by the reduced effective mass  $m_{red}$  [Eq. (19)] instead of the reduced band mass  $m_r$ , and (iii) the pre-exponential factor in (52) is approximately scaled by  $\frac{\sqrt{m_v m_{tc}} (\hbar \theta_{red})^6}{24 \pi^4 m_r (E_g \eta_v \eta_c)^2} \exp \left( -\frac{\Delta_v^2}{\eta_v^2} - \frac{\Delta_c^2}{\eta_c^2} \right)$ .

In Ref. 12, we argued that TTTT is always negligible with respect to TTBT (and hence to BTBT) due to the localized nature of the tail states. In order to confirm this statement, it is necessary to contrast the TTTT rate (48) with the TTBT rate<sup>12</sup> for the same parameter set. The TTBT counterpart to Eq. (48) reads in the case of tunneling from the VB tail to the CB edge [compare Ref. 12, Eq. (29)]

$$G_{G,exp}^{tv-c} = \frac{(eF)^2 \sqrt{\hbar \theta_\mu} m_c \sqrt{\mu}}{\sqrt{2} \pi^2 \hbar^2 m_r E_g} \int_{E_{edge,v}}^{E_g/2} dE_{tv} \frac{\sqrt{\eta_v}}{E_{tv}^{3/2}} Y_{G,exp} \left( \frac{E_{tv}}{\eta_v} \right) \mathcal{F} \left( \frac{E_g - E_{tv}}{\hbar \theta_\mu} \right), \quad (53)$$

with  $\mathcal{F}$  defined in Eq. (5) and  $1/\mu = 1/m_c + 1/m_{tv}$ . The comparison is shown in Fig. 4 for Gaussian (red curves) and exponential (blue curves) tail DOS, assuming symmetrical tails with three different values of the characteristic energy  $\eta$  ( $= \eta_c = \eta_v$ ). In order to facilitate the highest probability for tail tunneling, the effective masses  $m_{tc,tv}$  were set to the band masses  $m_{c,v}$ . In the case of small tails [ $\eta = 5$  meV, Fig. 4(a)], TTTT is much weaker than TTBT in the whole field strength range relevant in device application. The situation becomes different for moderate tails [ $\eta = 25$  meV, Fig. 4(b)]. Here, TTTT between Gaussian tails dominates over TTBT up to  $F \approx 4 \times 10^4$  V/cm, whereas TTTT between exponential tails dominates over TTBT up to  $F \approx 7 \times 10^4$  V/cm. In the case of strong tails [ $\eta = 50$  meV, Fig. 4(c)], TTTT between Gaussian tails dominates over TTBT up to  $F \approx 7 \times 10^4$  V/cm, whereas TTTT between exponential tails dominates over TTBT up to  $F \approx 1.5 \times 10^5$  V/cm.

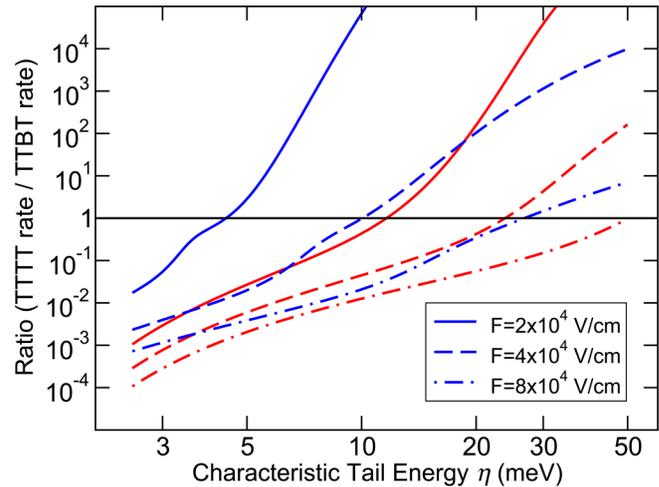
Figure 5 gives further information about the importance of TTTT. The ratio of TTTT and TTBT rate is plotted as a function of the characteristic tail energy  $\eta$  for three different values of the electric field. In general, an exponential tail DOS has a much larger effect than a Gaussian tail DOS as the former penetrates deeper into the gap. At  $F = 4 \times 10^4$  V/cm, the ratio exceeds 1 for  $\eta > 10$  meV, whereas at  $F = 8 \times 10^4$  V/cm, an  $\eta > 27$  meV is



**FIG. 4.** Comparison of the different tunneling rates as a function of the electric field for three values of the characteristic tail energy  $\eta = \eta_{c,v}$ : (a)  $\eta = 5$  meV, (b)  $\eta = 25$  meV, and (c)  $\eta = 50$  meV. InAs band masses are used in all cases, i.e.,  $m_c = m_v = 0.023 m_0$  and  $m_{tv} = m_v = 0.026 m_0$ . The TTBT rate is that for tunneling from the VB tail to the CB edge. The bandgap is  $E_g = 0.36$  eV. Red curves belong to Gaussian tail DOS and blue curves to exponential tail DOS. Equations (48), (53), and (52) were evaluated for TTTT, TTBT, and BTBT rate, respectively.

necessary. In the case of Gaussian tail DOS, the corresponding  $\eta$ -values increase to 24 meV and 50 meV, respectively.

For Esaki diodes and TFETs, it is interesting to see at which characteristic tail energy  $\eta$  TTTT starts to outnumber BTBT. Figure 6 shows the ratio of TTTT and BTBT rate as a function of  $\eta$  for three different values of the electric field in the case of Gaussian (red curves) and exponential (blue curves) tail DOS, respectively. At  $F = 8 \times 10^4$  V/cm, the ratio exceeds 1 for  $\eta > 20$  meV ( $\eta > 58$  meV), whereas at  $F = 1.6 \times 10^5$  V/cm, an  $\eta > 39$  meV ( $\eta > 70$  meV) is required for exponential (Gaussian) DOS tails. These results for homogeneous field already demonstrate that a degradation of the device performance due to TTTT can only be

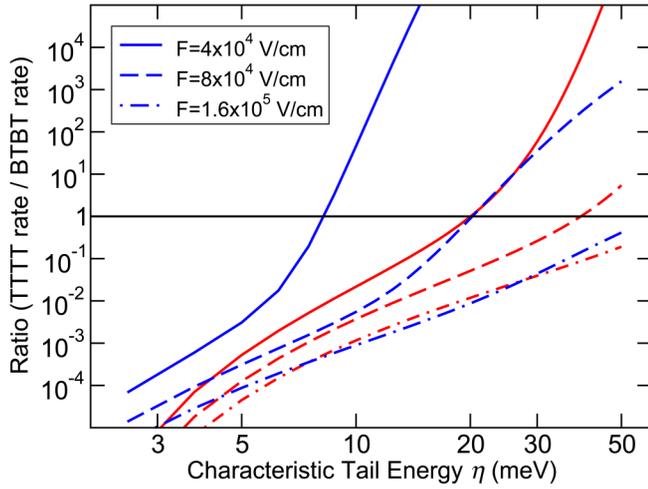


**FIG. 5.** Ratio between TTTT and TTBT rate as a function of the characteristic tail energy  $\eta = \eta_{c,v}$  for different values of the electric field as indicated in the legend. InAs band masses are used in all cases, i.e.,  $m_c = m_v = 0.023 m_0$  and  $m_{tv} = m_v = 0.026 m_0$ . The TTBT rate is that for tunneling from the VB tail to the CB edge. The bandgap is  $E_g = 0.36$  eV. Red curves belong to Gaussian tail DOS and blue curves to exponential tail DOS. Equations (48) and (53) were evaluated for TTTT and TTBT rate, respectively.

expected in the presence of very strong DOS tails. Note that the smallest possible effective masses (i.e., the values of the band masses) were used for  $m_{c,tv}$ . Stronger localization will further increase the  $\eta$ -values needed to make TTTT the dominant process.

Figures 3–6 refer to InAs, the TFET material in Ref. 12, where the impact of TTBT on the SS was studied. It is tempting to apply the TTTT model to other materials of interest and to figure out the characteristic tail energies, which make TTTT as strong as TTBT or BTBT. Figure 7(a) shows the value  $\eta^*$ , where the TTTT rate becomes equal to the TTBT rate as a function of the electric field for six materials: InAs, InGaAs, GaSb, Ge(111), GaAs, and Si(111). The solid curves are the implicit solutions of  $G_{G,exp}(\eta^*) = G_{G,exp}^{iv-c}(\eta^*)$  based on Eqs. (48) and (53) for exponential DOS tails. Figure 7(b) provides the value  $\eta^*$ , where the TTTT rate becomes equal to the BTBT rate as a function of the electric field for the same materials. Here, the solid curves are the implicit solutions of  $G_{G,exp}(\eta^*) = G_{BTB}$  based on Eq. (48) for exponential DOS tails and Eq. (52) for BTBT. For a given field strength,  $\eta^*$  increases from Si toward InAs. This increase is approximately proportional to  $1/\sqrt{E_g m_{red}}$ . Therefore, in a material with larger gap and reduced effective mass, smaller DOS tails are able to make tail-induced tunneling as strong as BTBT for the same field. However, in the device application, the electric field is related to the gap. As can be seen from Fig. 7, for a given material, the characteristic tail energy  $\eta^*$  increases nonlinearly with rising field. This means that in steeper junctions, the DOS tails are less detrimental to the TFET operation.

The described behavior can be understood analytically by the following approach. The dominating WKB exponent of the TTTT rate is developed up to first order in  $\Delta$  (Gaussian tails) or  $\eta$  (exponential tails). Then,  $\eta^*$  is approximately defined by setting the ratio



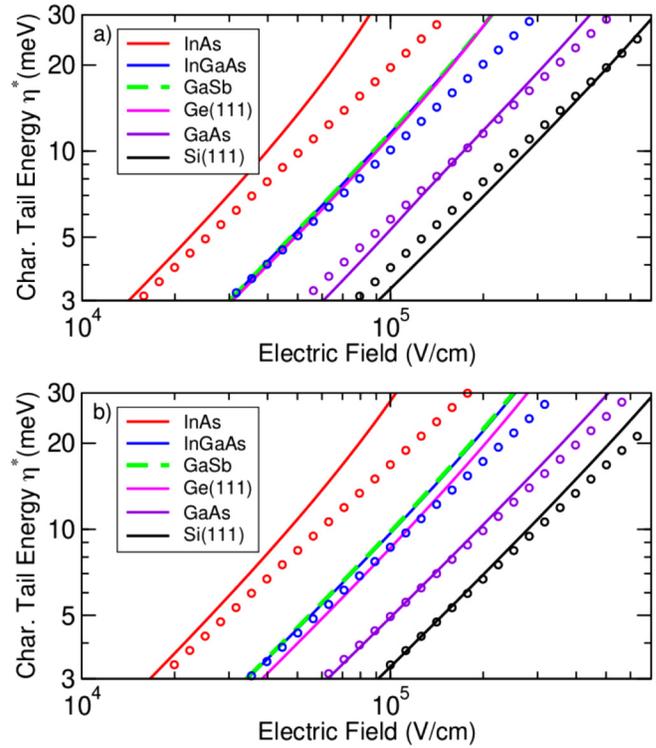
**FIG. 6.** Ratio between TTTT and BTBT rate as a function of the characteristic tail energy  $\eta = \eta_{c,v}$  for different values of the electric field as indicated in the legend. InAs band masses are used in all cases, i.e.,  $m_c = m_c = 0.023 m_0$  and  $m_v = m_v = 0.026 m_0$ . The bandgap is  $E_g = 0.36$  eV. Red curves belong to Gaussian tail DOS and blue curves to exponential tail DOS. Equations (48) and (52) were evaluated for TTTT and BTBT rate, respectively.

of the rates to unity, but treating the pre-exponential terms as constants, i.e., neglecting any field- and  $\eta$ -dependence in them. This immediately leads to

$$\eta^*(F) = \xi \frac{(\hbar\theta_{\text{red}})^{3/2}}{\sqrt{E_g}} = \text{const} \frac{F}{\sqrt{E_g m_{\text{red}}}}. \quad (54)$$

The outcome of this equation is shown by the curves with circle symbols in Fig. 7. The pre-factor  $\xi$  was fitted to 2/3 for TTTT/TTBT and 4/7 for TTTT/BTBT, respectively. The agreement with the solid curves is reasonable—even the linear field dependence fits for smaller  $\eta^*$ . The observed scaling behavior with  $1/\sqrt{E_g m_{\text{red}}}$  now becomes obvious from the last denominator in (54). It also explains the material dependence. The curves for InGaAs and GaSb are indistinguishable because both, gap and reduced effective mass, have almost the same value in these materials. In Ge(111), the value of  $\sqrt{E_g m_{\text{red}}}$  is very close to the one in InGaAs and GaSb. Si(111) has the largest and InAs the smallest  $E_g m_{\text{red}}$ -product. Hence, Eq. (54) allows a quick estimate of  $\eta^*$  just from the knowledge of gap, reduced effective mass, and internal field. It has a universal character as it holds for any semiconductor, for any of the ratios TTTT/TTBT, TTTT/BTBT, and TTBT/BTBT (with different  $\xi$ ), as well as for exponential and Gaussian DOS tails.

The models for TTTT and TTBT were developed for direct semiconductors. BTBT in Si and Ge is, however, phonon-assisted. Tunnel recombination between strongly localized CB tail states and the VB could be direct, if the spread in k-space is sufficient to enable momentum conservation without participation of a phonon.



**FIG. 7.** Characteristic tail energy  $\eta^*(F)$  that fulfills the condition (a)  $G_{G,\text{exp}}(\eta^*) = G_{G,\text{exp}}^{v-c}(\eta^*)$  and (b)  $G_{G,\text{exp}}(\eta^*) = G_{\text{BTBT}}$  in different materials as indicated in the legend. For the solid lines, Eqs. (48), (53), and (52) were evaluated for TTTT, TTBT, and BTBT rate, respectively. Symbols represent the outcome of the simple expression (54) with (a)  $\xi = 2/3$  and (b)  $\xi = 4/7$ . Band masses are used in all cases, i.e.,  $m_c = m_c$  and  $m_v = m_v$ .

This also holds for tunnel recombination between the CB edge and strongly localized VB tail states. In these cases, the inclusion of Si and Ge in Fig. 7(a) is justified. For  $\eta^*(F)$  in Fig. 7(b), which involves the BTBT rate, the inclusion of Si and Ge can also be justified, because the WKB form of the BTBT rate for indirect transitions is similar to Eq. (52) with two modifications: (i) the gap is changed to  $E_g \pm \hbar\omega_{\text{ph}}$ , where  $\hbar\omega_{\text{ph}}$  is the energy of the participating phonon and (ii) the pre-exponential factor is different, in particular, the field dependence.<sup>43</sup> These differences do not prevent the application of the above-sketched analytical treatment.

### III. DEVICE APPLICATION

#### A. Implementation of tail-to-tail tunneling model using the dynamic nonlocal path (DNLP) algorithm

The analytical forms of the generation rates for TTTT have been derived for the case of a homogeneous electric field. For the real case of a finite depletion zone bounded by quasi-neutral regions, the numerical computation must exclude invalid tunnel paths and account for the continuously changing field. This is accomplished in the commercial device simulator S-Device<sup>44</sup> by the

so-called *Dynamic Nonlocal Path (DNLP) Algorithm*, where the tunnel rate is obtained by integrating the action on dynamically extracted tunnel paths. The algorithm tests if a tunnel path actually connects CB and VB states. For the adaptation of Eq. (48) to the DNLP algorithm, the energetic tunnel rates are transformed into position-dependent rates defined along a tunnel path with length  $L$  and direction  $x$  (see Fig. 8). At every pair of points at locations  $x_{t1}, x_{t2} \in \{x_1, x_1 + L\}$  satisfying  $x_{t1} < x_{t2}$ ,  $E_{tc} \rightarrow E_c(x_{t1}) - E_{tun}$ , with  $dE_{tc} \rightarrow eF(x_{t1})\Delta x_{t1}$  and  $E_{tv} \rightarrow E_{tun} - E_v(x_{t2})$ , with  $dE_{tv} \rightarrow eF(x_{t2})\Delta x_{t2}$ . In this way, Eq. (48) becomes locally defined at  $x_1$ ,

$$G_{tt}(x_1) = \sum_{x_{t1}=x_1}^{x_1+L} \sum_{x_{t2}=x_1}^{x_1+L} g_{tt}(x_{t1}, x_{t2}), \quad (55)$$

with

$$g_{tt}(x_{t1}, x_{t2}) = \frac{2eF_{av}\theta_{red}m_{red}\sqrt{m_{tv}m_{tc}}}{3\pi^4\hbar^2 m_r E_g} \times eF(x_{t1})\Delta x_{t1} eF(x_{t2})\Delta x_{t2} \frac{\sqrt{\eta_v\eta_c}}{E_{tv}E_{tc}} \times Y\left(-\frac{E_c(x_{t1}) - E_{tun}}{\eta_c}\right) Y\left(-\frac{E_{tun} - E_v(x_{t2})}{\eta_v}\right) \times A^2 \left(\frac{E_g - E_c(x_{t1}) + E_v(x_{t2})}{\hbar\theta_{red}}\right) [f_n(x_{t1}) - f_p(x_{t2})]. \quad (56)$$

Here,  $x_{t1}, x_{t2} \in \{x_1, x_1 + L\}$  are, respectively, the locations of the CB tail state and VB tail state along the tunnel path,  $\Delta x_{t1}, \Delta x_{t2}$  the discretization intervals,  $F_{av}(x_{t1}, x_{t2}) = \frac{F(x_{t1})+F(x_{t2})}{2}$  the electric field averaged over the segment of the tunnel path between  $x = x_{t1}$  and  $x = x_{t2}$  (also used in  $\hbar\theta_{red}$ ), and  $E_{tun}$  the CB energy at the beginning of the tunnel path.  $E_c(x_{tc}), E_v(x_{tv})$ , and  $F(x_{\{tc,tv\}})$  are, respectively, the CB edge, the VB edge, and the electric field at the location of the tail state.  $g_{tt}$  is the generation rate at  $\{x_{t1}, x_{t2}\}$  and  $G_{tt}$  the total generation rate along the tunnel path.

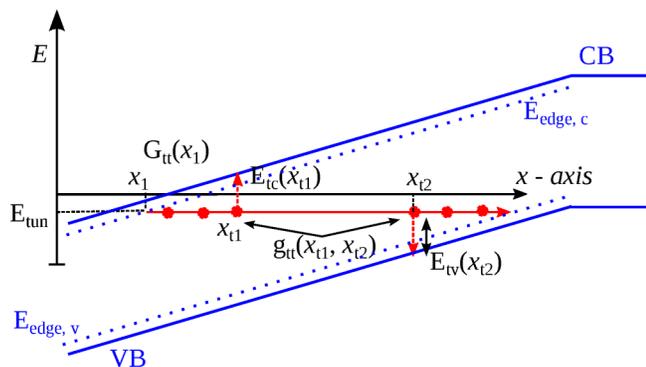


FIG. 8. Schematic representation of the tunnel path with various variables used in Eqs. (55) and (56).

The function  $f_{n/p}(x) = [\exp(E_{tun} - E_{F,n/p}(x))/k_B T + 1]^{-1}$  represents the Fermi distribution at  $x$ . The value of  $Y$  is calculated at each  $x_{\{tc,tv\}}$  using Eq. (43) or Eq. (45) for Gaussian or exponential tails, respectively. Note that the generation rate between the tail states  $g_{tt}(x_{t1}, x_{t2})$  is introduced in the generation–recombination term of the continuity equation at the vertices nearest to  $x_{t1}$  and  $x_{t2}$ . In this way, the total generation is indirectly included in the simulations, and no explicit computation of  $G_{tt}$  is required, i.e., the double integral in Eq. (48) becomes obsolete.

The analytical forms of the functions  $Y$  are used. Equation (56) has been implemented in S-Device using the Physical Model Interface (PMI) *Nonlocal Generation–Recombination*. The original DNLP BTBT model requires the effective tunnel barrier and the electron/hole effective masses as input parameters. In addition to these parameters, for the new DNLP TTTT model, one has to provide the effective masses  $m_{tc}$  and  $m_{tv}$  as well as the characteristic CB and VB tail energies  $\eta_c$  and  $\eta_v$ . For the sake of better convergence, the implementation employs an effective average electric field for the computation of the rate, which substitutes for the numerical integration of the action integral over the imaginary dispersion in the original DNLP BTBT model of S-Device. The loss in accuracy may be compensated through calibration of the parameters  $\eta_{c,v}$  and  $m_{\{tc,tv\}}$ .

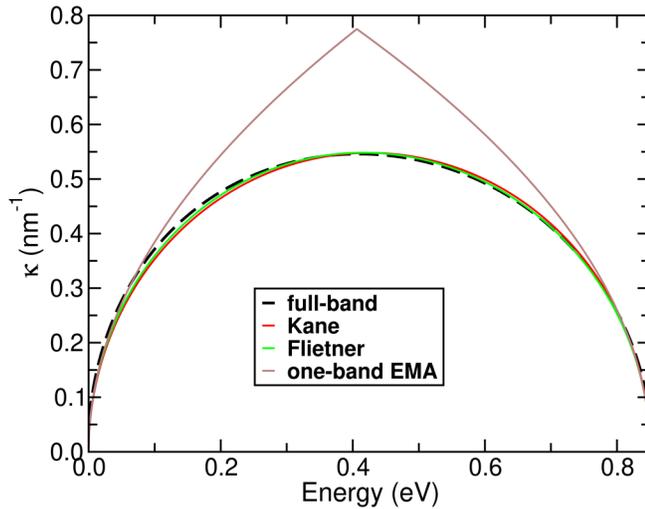
As in the original DNLP model for BTBT, the TTTT model involves the search for active tunnel paths. Once all of them are found at a given bias voltage, the TTTT rates are calculated at each discretization point using Eq. (56). Thermionic transitions between tail states and the associated band continuum are assumed to be very rapid and thus not rate-limiting. The densities of generated holes and electrons enter the Poisson equation and self-consistently impact the solution of the drift-diffusion equation system.

## B. Simulation results for InGaAs Esaki diode

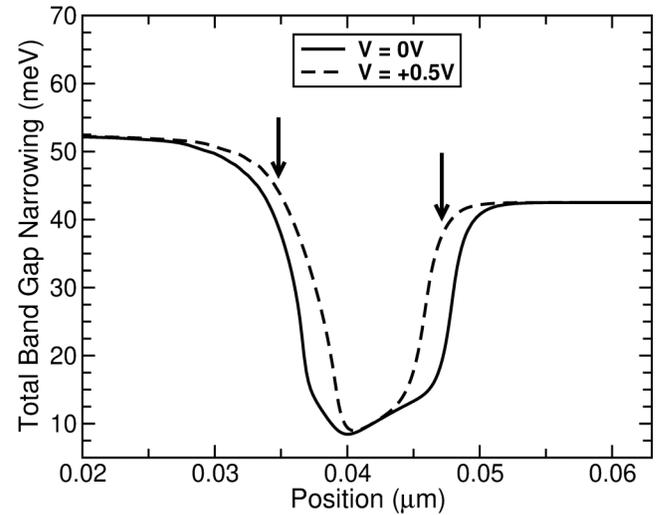
### 1. Basic physical models and parameters

In order to assess the importance of TTTT in realistic devices, InGaAs Esaki diodes fabricated and characterized at IMEC<sup>10,13</sup> were simulated with S-Device of Synopsys<sup>44</sup> using the above-described model implementation of Eq. (48). The diodes labeled “A” and “B” have cross sections of  $19.3\mu\text{m}^2$  and  $204\mu\text{m}^2$ , respectively. The BTBT-dominated parts of the  $IV$ -characteristics at 300 K and 77 K serve to determine material-dependent parameters of  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  and the electrically active doping in the *pin*-junction. The analysis starts with the evaluation of effective tunneling masses, bandgap, and imaginary dispersion. The effective electron mass is set to  $0.043 m_0$  (Ref. 45, and references therein). Experimental data for the light-hole mass<sup>45</sup> range from  $0.048 m_0$  to  $0.054 m_0$ .<sup>46,47</sup> We determine the value of the light-hole mass based on full-band calculations of the imaginary dispersion in the gap of bulk  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  with OMEN<sup>48</sup> in comparison to the corresponding dispersion from Kane’s two-band model.<sup>38,49</sup> The latter is implemented in S-Device and used here for the simulation of BTBT. As shown in Fig. 9, the best value of  $m_{lh} = 0.047 m_0$  results in a very good fit, both for Kane’s<sup>38</sup> and Flietner’s model.<sup>50</sup> In addition, this value is equal to the one suggested in Ref. 51.

The review article by Vurgaftman *et al.*<sup>45</sup> recommends to use a value of 0.816 eV for the gap at  $T = 0$  K as composite average



**FIG. 9.** Calculated imaginary dispersion  $\kappa(E)$  in the gap of  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  at  $T = 0\text{ K}$ . The full-band curve from OMEN<sup>48</sup> (black dashed) is compared with the two-band models of Kane<sup>38</sup> and Flietner<sup>50</sup> (solid red and green, respectively) and with the one-band effective mass approximation (EMA) model<sup>40</sup> (solid brown). Parameters:  $E_g = 0.85\text{ eV}$ ,  $m_c = 0.043 m_0$ , and  $m_{lh} = 0.047 m_0$ .



**FIG. 10.** Bandgap narrowing profiles across the *pin*-junction due to exchange-correlation effects. Solid curves are for equilibrium and dashed curves for a forward bias of  $+0.5\text{ V}$ . The arrows indicate the position of the maxima of the electron (left) and hole (right) BTBT recombination rates at  $+0.3\text{ V}$ .

over the experimental range between  $0.810\text{ eV}$  and  $0.821\text{ eV}$ . The measured temperature dependence<sup>52</sup> is almost linear between  $130\text{ K}$  and  $300\text{ K}$ , but exhibits a sharp saturation behavior for  $T < 100\text{ K}$ . It is impossible to fit these data with the Varshni model.<sup>53</sup> Therefore, the linear relation  $E_g(T) = E_g(0) - \alpha T$  with  $E_g(0) = 0.841\text{ eV}$  and  $\alpha = 3.363 \times 10^{-4}$  is applied in the simulation, which reproduces the needed values  $E_g(77\text{ K}) = 0.815\text{ eV}$  and  $E_g(300\text{ K}) = 0.74\text{ eV}$ . The obvious overestimation of  $E_g(0)$  has no bearing on the present analysis.

Because of the exponential dependence of the tunnel probability on the value of the gap, bandgap narrowing (BGN) must be taken into consideration. The widely used empirical model for III-V materials<sup>54</sup> would result in a vanishing gap in the n-type region of the device. Therefore, the rigid shifts of both band edges are computed with the random-phase-approximation (RPA)-based model of Ref. 55, which is available in S-Device for silicon only (called “Schenk Bandgap Narrowing Model” there). As it does not become obvious for the user of this simulator to adapt the parameters to other materials, Appendix A provides the expression of the simplified version (valid for  $T = 0\text{ K}$ ) as well as the parameters for  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ . Because the BGN effect is rather small in the Esaki diode under study (see below), the temperature dependence is neglected for simplicity. The rigid shifts represent exchange-correlation energies and depend on the free-carrier densities and the doping concentration. The free carriers have the largest contribution and make the BGN injection-dependent. This is demonstrated in Fig. 10 where the profile of the electrically active doping from Fig. 11 is used. The total BGN in the quasi-neutral n-region ( $N_D^+ = 1.4 \times 10^{19}\text{ cm}^{-3}$ ) amounts to  $52\text{ meV}$ , the BGN in the quasi-neutral p-region ( $N_A^- = 1.3 \times 10^{19}\text{ cm}^{-3}$ ) to  $43\text{ meV}$ . This is the expected order of magnitude for any material,

since BGN is relatively independent on band-structure details.<sup>56</sup> At a forward bias of  $+0.5\text{ V}$ , where diffusion starts to dominate the current, the injected carriers increase the BGN at the boundaries of the depletion region. However, the value inside drop to  $10\text{ meV}$ .

There are three physical processes that might influence the valley current at higher forward bias: Auger recombination, radiative recombination, and electron-hole scattering. Measured Auger coefficients<sup>57–59</sup>  $C$  are scattered around  $2 \times 10^{-28}\text{ cm}^{-6}\text{ s}^{-1}$ . We use  $C_{n,p} = 2.5 \times 10^{-28}\text{ cm}^{-6}\text{ s}^{-1}$  following Ref. 59. To model radiative recombination, the coefficient  $B$  in  $\tau_{\text{rad}} = (BN)^{-1}$  is set to  $0.96 \times 10^{-10}\text{ cm}^{-3}\text{ s}^{-1}$ ,<sup>60</sup> which is close to  $1.43 \times 10^{-10}\text{ cm}^{-3}\text{ s}^{-1}$  found in Ref. 57. Electron-hole scattering is simulated with Brooks–Herring screening. The default parameters of the S-Device model immediately result in the best fit to the measured *IV*-curves as can be seen from the solid curves in Fig. 13, which bend for  $V > +0.6\text{ V}$  ( $300\text{ K}$ ) and for  $V > +0.8\text{ V}$  ( $77\text{ K}$ ), respectively. Without electron-hole scattering, they would be straight and would intersect the experimental curves. The size of the diffusion current at  $300\text{ K}$  is well reproduced with SRH minority carrier lifetimes  $\tau_{n,p} = 1\text{ ns}$  and constant mobilities  $\mu_n = 3000\text{ cm}^2/\text{Vs}$  and  $\mu_p = 500\text{ cm}^2/\text{Vs}$ .

As the diode is bulk-like, a quasi-1D simulation suffices. The SIMS profiles of the silicon and beryllium implants<sup>10</sup> are presented in Fig. 11. If they are taken as electrically active doping in the simulation, the BTBT current is overestimated by a factor of 5 both in reverse and forward direction. This is a huge effect—the same would, e.g., require to change the light-hole mass by a factor of 2. Lind *et al.*<sup>61</sup> found  $1.4 \times 10^{19}\text{ cm}^{-3}$  as the upper limit of active silicon doping in  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ . We adapt this value (red solid plateau in Fig. 11) and reproduce the junction profiles by

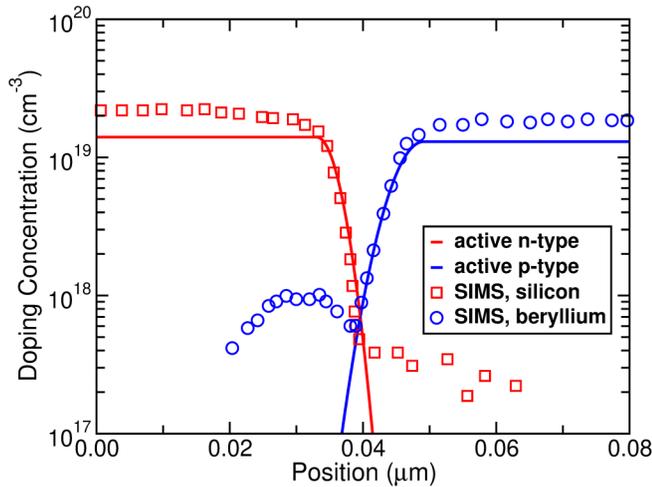


FIG. 11. Measured SIMS profiles (symbols) of the  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  Esaki diode<sup>10</sup> and electrically active doping used in the simulation (solid curves).

Gaussians with a decay rate of 6 nm/dec.<sup>13</sup> The possible limit for the beryllium activation remains unclear. Tell *et al.*<sup>62</sup> achieved peak values close to  $1 \times 10^{19} \text{ cm}^{-3}$  with RTA. We use the active beryllium concentration as the only free parameter to match the simulated BTBT current to the  $IV$ -data in the voltage interval  $[-0.1 \text{ V}, +0.1 \text{ V}]$ . The obtained value is  $N_{\text{A}}^- = 1.3 \times 10^{19} \text{ cm}^{-3}$  (blue solid plateau in Fig. 11). The corresponding capacitance is shown by the solid line in Fig. 12. It deviates by less than 4% from the measured capacitance.<sup>13</sup> Shifting the beryllium profile by 1 nm toward the junction yields the best fit (dashed curve). This increases the BTBT current by a factor of 1.4, which could be compensated with a larger light-hole mass. However, we refrain from such modifications and use the calibrated active doping shown in Fig. 11 as a

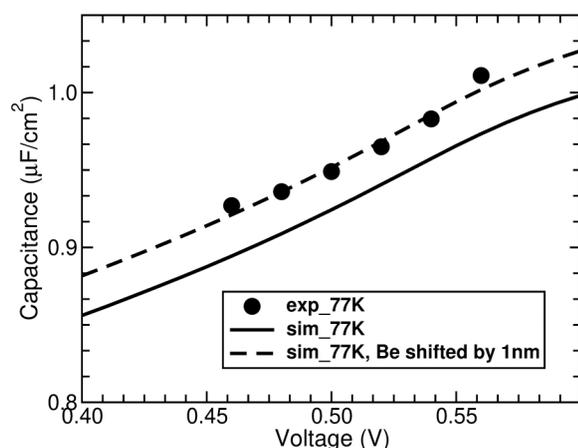


FIG. 12. Measured<sup>13</sup> (symbols) and simulated (lines) capacitance of the  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  Esaki diode at 77 K.

reliable electrostatic reference system for the further analysis. The BTBT currents (without tails) are presented in Fig. 13. For the green dashed curve, BGN was turned off. By comparing with the solid black curve (including BGN as described above), one observes that BGN is completely negligible at reverse bias. It slightly increases at low forward bias due to the injection effect. If the tunnel length of the dominant tunnel paths is defined by the distance between the sharp maxima of the electron and hole generation/recombination rates (indicated by the arrows in Fig. 10), one observes that it increases from 7 nm at  $-0.5 \text{ V}$  to 12 nm at  $+0.3 \text{ V}$ , whereas the extension of the space-charge region (SCR) (defined by the positions where the carrier concentrations equal  $5 \times 10^{18} \text{ cm}^{-3}$ ) decreases from 17 nm to 12 nm. Therefore, the effect of BGN remains negligible over the tunnel length at reverse bias. The maximum effect at forward bias is relatively small because the tunnel barrier is notably reduced only in the vicinity of the boundaries of the SCR. The difference in the diffusion branch is caused by the changed minority carrier density.

At high reverse voltage, the simulated currents become too large. The measured  $IV$ -curves (symbols) had been corrected for series resistance.<sup>13</sup> One possibility is that this correction was too conservative (however, one would roughly need twice as much series resistance for a match with the simulation). Another possible reason is the breakdown of the WKB approximation. At  $-0.5 \text{ V}$  the ratio between electro-optical energy (measure for the Franz-Keldysh tails) and the gap energy is  $\hbar\theta_{\text{red}}/E_{\text{g}} \approx 1$  and not  $\ll 1$ . Thus, the WKB approximation, linked to the Kane model, has indeed broken down. The measured reverse current at 77 K is smaller than at 300 K, but the difference is more pronounced than in the simulation, where it is solely caused by the carefully calibrated temperature dependence of the gap. The same difference naturally shows up in the simulated peak current, whereas no temperature effect can be seen in the

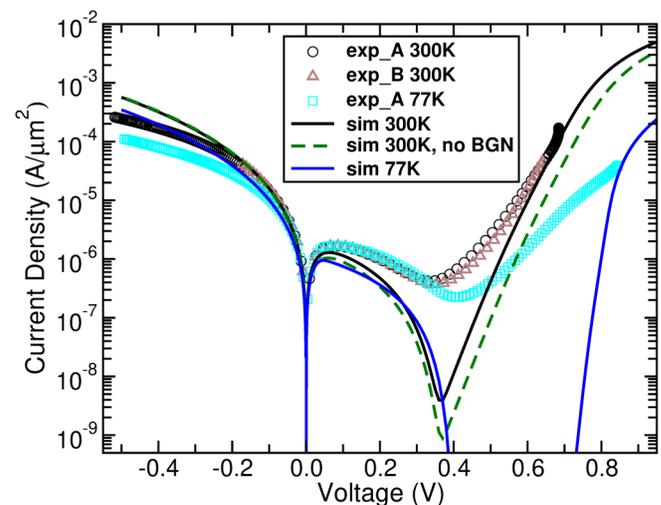


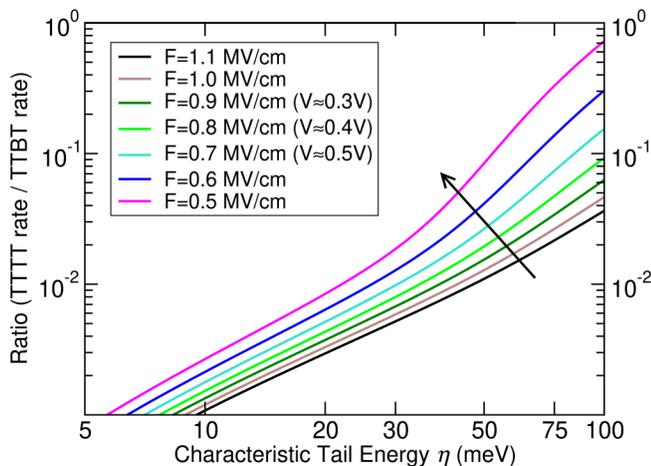
FIG. 13. Measured (symbols) and simulated (lines) currents at 300 K and 77 K. Simulations include BTBT, electron-hole scattering, Auger recombination, and radiative recombination. For the dashed curve, BGN was turned off.

measured peak currents. It should be noted that a symmetrical temperature effect at small bias was measured by Convertino *et al.*<sup>63</sup> in InAs/GaSb Esaki diodes on SOI.

In the following, it is the goal to understand the origin of the strong valley current in the voltage interval  $[+0.1 \text{ V}, +0.65 \text{ V}]$ . Before turning to the role of tails, we briefly discuss the impact of the aforementioned physical processes. It turns out that both Auger and radiative recombination are at least two orders of magnitude too weak to become visible. Also, trap-assisted Auger recombination can be safely ruled out, although it is a tempting deep-level-related mechanism that becomes stronger with increasing injection density. Electron-hole scattering only starts to “bend down” the  $IV$ -curve at about  $+0.55 \text{ V}$  at  $300 \text{ K}$ , i.e., it has no influence on its shape for lower bias. Therefore, none of these processes has an effect on the valley current. We now study the effect of DOS tails.

## 2. Contribution of tail-to-band and tail-to-tail tunneling

In the presented model, DOS tails are simplistically parameterized by characteristic energies  $\eta_{c,v}$  which have no functional dependence on position and temperature. In principal, these parameters depend on position via the doping concentration, and, if taken as representative for all effects, also on temperature because of the thermal lattice vibrations. In Ref. 12, it has been argued that TTTT is negligible compared to TTBT even if band masses are used for the effective masses  $m_{c,tv}$  that parameterize the tail states. The analytical ratio between TTTT rate and TTBT rate for exponential tail DOS as a function of the characteristic tail energy  $\eta = \eta_{c,v}$  is depicted in Fig. 14. Relevant values of the maximum electric field in the forward bias range of the  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  Esaki diode were used for the homogeneous-field approximation. The middle green curve corresponds to  $+0.4 \text{ V}$  and shows that for a

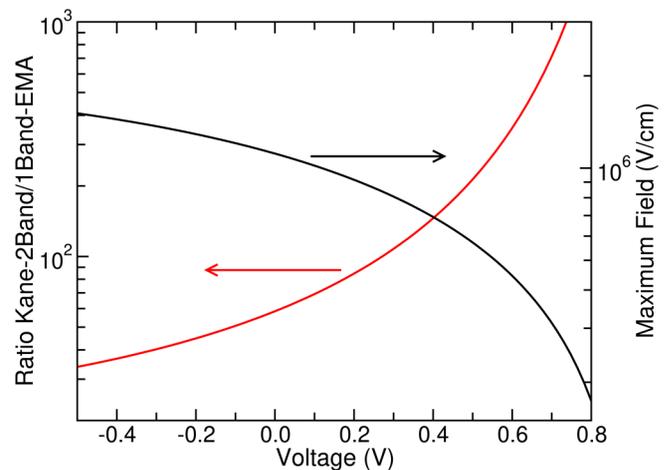


**FIG. 14.** Ratio between TTTT rate and TTBT rate for exponential tail DOS as a function of the characteristic tail energy  $\eta = \eta_{c,v}$  for relevant values of the maximum electric field in the forward bias range (indicated in the legend). Band masses are used, i.e.,  $m_c = m_v = 0.043 m_0$  and  $m_{c,v} = m_v = 0.047 m_0$ . The bandgap is  $E_g = 0.74 \text{ eV}$ . Equations (48) and (52) were evaluated for TTTT and TTBT rate, respectively.

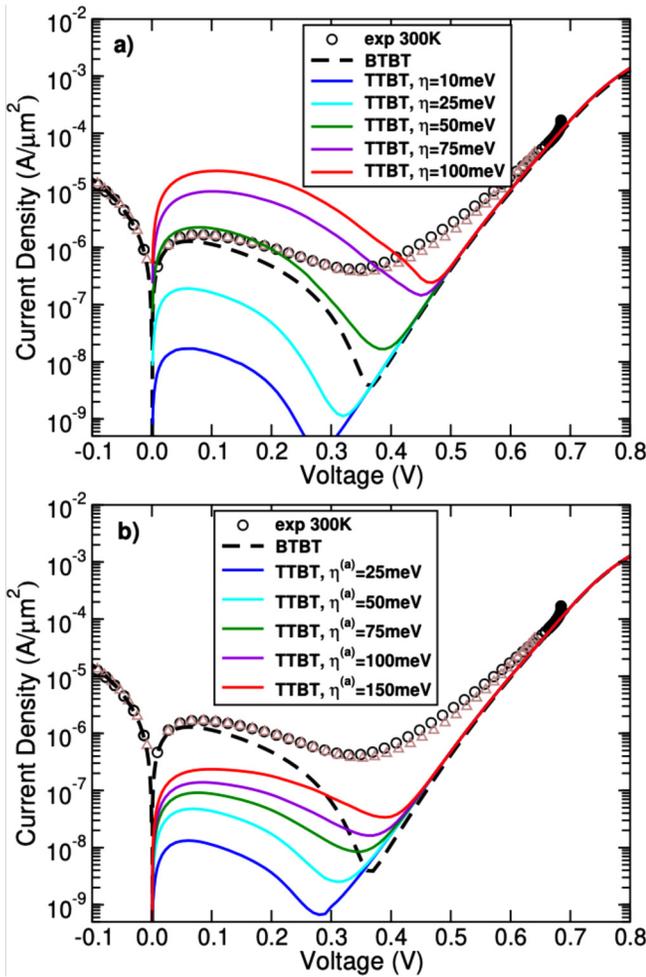
10%-effect the characteristic tail energy  $\eta$  would have to be as large as  $100 \text{ meV}$ . For  $\eta = 50 \text{ meV}$ , the TTTT fraction reduces to 2%. This will be confirmed in the more realistic device simulation below.

To be compatible with Kane’s two-band model of BTBT implemented in S-Device, the TTBT and TTTT rates are scaled up using the homogeneous-field approximation. This is demonstrated in Appendix B. The black curve in Fig. 15 is the field-voltage relation extracted from the simulation. The red curve in Fig. 15 shows the corresponding ratio of the BTBT rates of Kane’s two-band model and one-band EMA model. The strong rise of this ratio with increasing bias (decreasing field) is caused by the larger action integral in the one-band EMA model. As can be seen in Fig. 9, the integral over  $\kappa(E)$  becomes significantly larger in the one-band EMA model, and because of  $dx = dE/(eF)$ , the WKB-exponent is scaled by  $F^{-1}$ .

Figures 16(a) and 16(b) show the sum of TTBT current and diffusion current at  $300 \text{ K}$  (solid curves) in comparison to experimental data (symbols) and BTBT without tails (thick black-dashed curve). In Fig. 16(a),  $\eta$  is spatially constant, whereas in Fig. 16(b),  $\eta$  has two different values depending on the local doping concentration. In the inner of the depletion region,  $\eta^{(i)} = \eta (N_{\text{dop}} < 10^{19} \text{ cm}^{-3})$  was fixed to the small value of  $10 \text{ meV}$ , whereas in the outer of the depletion region  $\eta^{(a)} = \eta (N_{\text{dop}} > 10^{19} \text{ cm}^{-3})$  was varied up to an extreme value of  $150 \text{ meV}$ . In the first case, the maximum of the TTBT current exceeds the measured peak current for all  $\eta \geq 50 \text{ meV}$ . Hence, in this region, which contains the shortest tunnel paths and determines the peak current,  $\eta$  cannot be larger than  $50 \text{ meV}$ . The valley current is only slightly affected at this limit. All  $\eta < 30 \text{ meV}$  have practically no impact. In the second case, the increase of  $\eta^{(a)}$  does not change the peak current at all. It could only influence the slope around  $V = +0.2 \text{ V}$  (see below). Under no circumstances, TTBT can explain the strong measured valley current.



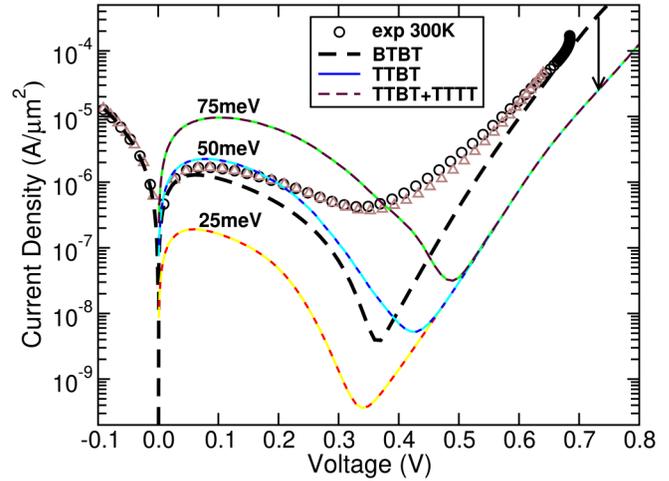
**FIG. 15.** Maximum electric field (black curve) and ratio of the BTBT rates of Kane’s two-band model and one-band EMA model (red curve).



**FIG. 16.** Measured (symbols) and simulated (lines) forward characteristics at 300 K. For comparison, the exclusive BTBT current is shown as thick black-dashed line. Parameters:  $m_c = m_v = 0.043 m_0$ ,  $m_v = m_v = 0.047 m_0$ , and  $E_g = 0.74$  eV. Solid lines represent the contribution from tail-to-band tunneling for different values of the characteristic energy of the exponential tail DOS. (a) spatially constant  $\eta = \eta_{c,v}$  and (b)  $\eta^{(a)} = \eta(N_{dop} < 10^{19} \text{ cm}^{-3}) = 10$  meV and  $\eta^{(a)} = \eta(N_{dop} > 10^{19} \text{ cm}^{-3})$  varied.

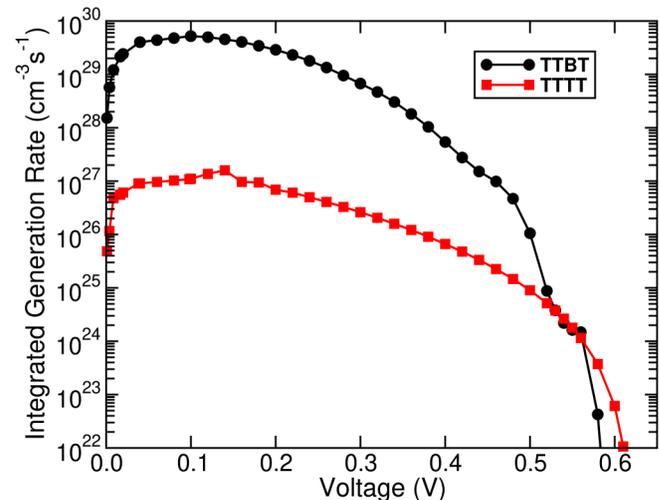
The joint contribution of TTBT + TTTT (thin dashed curves) to the total diode current is depicted in Fig. 17. For comparison, the exclusive fraction of TTBT is shown by thin solid lines. To extend the voltage range where the influence of TTTT could be possibly seen, the diffusion current was artificially suppressed by reducing the mobility until convergence ceased (indicated by the arrow). However, there is no visible difference. This confirms the conjecture in Ref. 12 that TTTT is a negligible second-order effect.

In order to prove that TTTT was turned on in the simulation, Fig. 18 presents the integrated rates of TTTT and TTBT in the *pin*-junction as a function of applied bias. The extraction was done for a spatially constant characteristic energy  $\eta = 60$  meV. TTTT



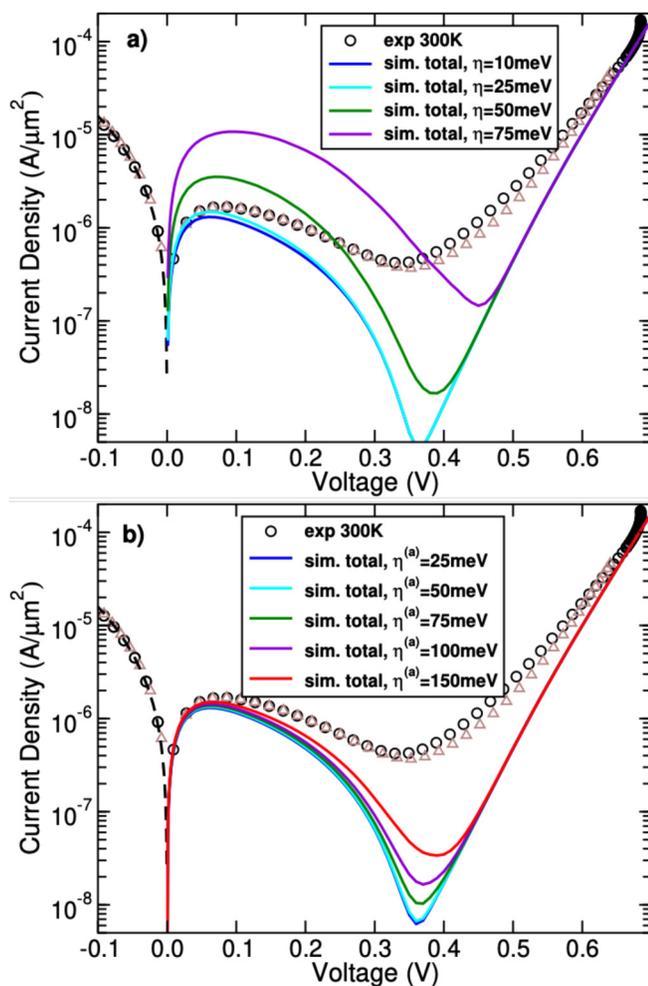
**FIG. 17.** Measured (symbols) and simulated (lines) forward characteristics at 300 K for three values of the spatially constant characteristic energies  $\eta$  of exponential DOS tails. For comparison, the exclusive BTBT current is shown as thick black-dashed line. Thin solid lines represent the contribution of tail-to-band tunneling and the thin dashed lines depict the joint contribution of tail-to-band tunneling and tail-to-tail tunneling. The diffusion current was artificially suppressed (arrow) to enlarge the visible range of tail tunneling. Parameters:  $m_c = m_v = 0.043 m_0$ ,  $m_{lv} = m_{lv} = 0.047 m_0$ , and  $E_g = 0.74$  eV.

starts to outweigh TTBT not before  $V \approx +0.55$  V. At this voltage, the tail-induced tunnel current is a few orders of magnitude smaller than the diffusion current. This happens for all reasonable values of spatially constant as well as doping-dependent  $\eta$ .



**FIG. 18.** Integrated tunnel generation rates in the  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  Esaki diode as a function of bias with spatially constant  $\eta = \eta_{c,v} = 60$  meV. TTTT exceeds TTBT beyond  $V \approx +0.55$  V. Parameters:  $m_c = m_v = 0.043 m_0$ ,  $m_{lv} = m_{lv} = 0.047 m_0$ , and  $E_g = 0.74$  eV.

Figures 19(a) and 19(b) compare the sum of all tunnel currents with the measured data for spatially constant and doping-dependent  $\eta$ , respectively. One can draw the following conclusions: (i) The effect of DOS tails on the forward characteristics cannot be ruled out, but their characteristic energy would be limited to  $\sim 30$  meV in the interior of the depletion region, since for larger values, the simulated peak current exceeds the measured peak current. (ii) Extremely strong tails at the boundaries of the *pin*-junction lead to a minor increase of the valley current but do not affect the peak current. (iii) TTTT is a negligible process for all feasible characteristic energies. (iv) Tail-induced tunneling cannot be the reason behind the strong valley current.

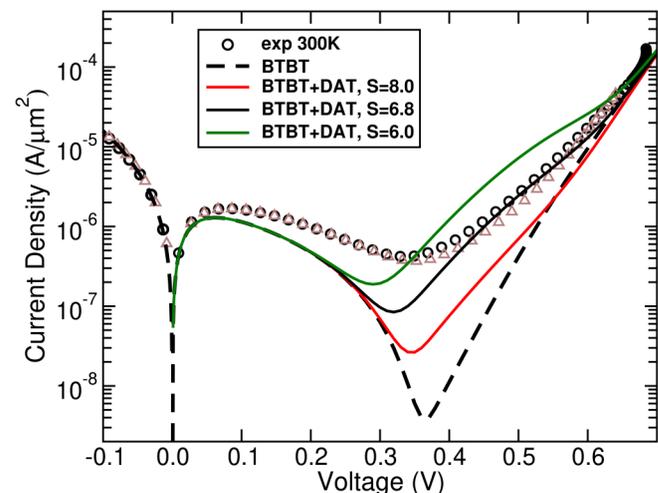


**FIG. 19.** Measured (symbols) and simulated (lines) forward characteristics at 300 K. Solid lines show the sum of all tunneling currents (BTBT, TTBT, and TTTT) for different values of the characteristic energy of the exponential DOS tails. (a) Spatially constant  $\eta = \eta_{c,v}$  and (b)  $\eta^{(i)} = \eta(N_{dop} < 10^{19} \text{ cm}^{-3}) = 10 \text{ meV}$  and  $\eta^{(a)} = \eta(N_{dop} > 10^{19} \text{ cm}^{-3})$  varied. Parameters:  $m_{lc} = m_c = 0.043 m_0$ ,  $m_{lv} = m_v = 0.047 m_0$ , and  $E_g = 0.74 \text{ eV}$ .

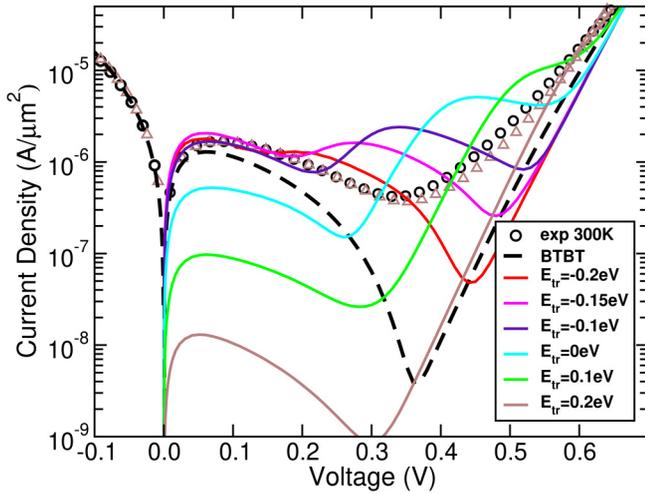
### 3. Resolving the theory-experiment discrepancy: Defect-assisted tunneling (DAT)

Deep centers are well known as the origin of leakage currents in many types of devices, e.g., tunnel FETs.<sup>64</sup> In the *pin*-junction of the  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  Esaki diode, such levels can facilitate field-enhanced multi-phonon recombination, zero-phonon defect-assisted tunneling, and resonant tunneling.<sup>65</sup> Models of the first two mechanisms are available in S-Device. They are now combined with BTBT to study the effect on the valley current. For simplicity, the spatial defect distribution is assumed to be homogeneous in the following. Figure 20 presents the results for field-enhanced multi-phonon recombination.<sup>66</sup> The model (called “Schenk Trap-assisted Tunneling (TAT)” in S-Device) assumes homogeneously distributed midgap levels with strong phonon coupling<sup>44,67</sup> and is implemented as field-enhancement of SRH lifetimes. The effective phonon energy was fixed to 30 meV, and the same tunneling masses as above were chosen. The Huang–Rhys factor<sup>68</sup>  $S$  was varied to obtain different lattice relaxation energies. The best fit turns out with  $S = 6.8$  which results in a lattice relaxation energy of  $\epsilon_r = S\hbar\omega_{\text{eff}} = 204 \text{ meV}$ . The good matching starts at  $V = +0.4 \text{ V}$ , where the DOS alignment enabling BTBT or TTBT has already ceased.

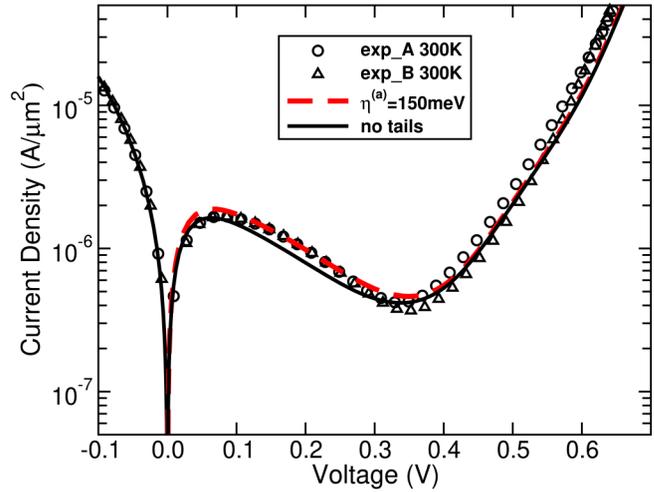
SRH lifetimes of 1 ns are an indication for the presence of defect levels at or close to midgap with a relatively high concentration. But it is also likely that other defects with levels closer to the bands exist in the measured diodes. Whereas the model of field-enhanced multi-phonon recombination is a simulation model for strong electron–phonon coupling (hence not applicable for the case of few or even zero phonons), the available zero-phonon DAT model in S-Device adapted from Ref. 69 is a simulation model for



**FIG. 20.** Measured (symbols) and simulated (lines) forward characteristics at 300 K assuming field-enhanced multi-phonon recombination as origin for the valley current. Parameters:  $\hbar\omega_{\text{eff}} = 30 \text{ meV}$ ,  $m_{lc} = m_c = 0.043 m_0$ ,  $m_{lv} = m_v = 0.047 m_0$ ,  $E_g = 0.74 \text{ eV}$ , and  $\tau_{n,p} = 1 \text{ ns}$ . Huang–Rhys factor  $S$  indicated in the legend.



**FIG. 21.** Measured (symbols) and simulated (solid lines) forward characteristics at 300 K using exclusively zero-phonon DAT via homogeneously allocated single-level defects. Parameters:  $m_c = 0.043 m_0$ ,  $m_v = 0.047 m_0$ ,  $E_g = 0.74$  eV,  $N_{tr} = 1 \times 10^{17}$  cm $^{-3}$ , and  $V_{tr} = 0.1$  nm $^3$ . Energy level  $E_{tr}$  measured from midgap indicated in the legend. The pure BTBT current is shown for comparison.

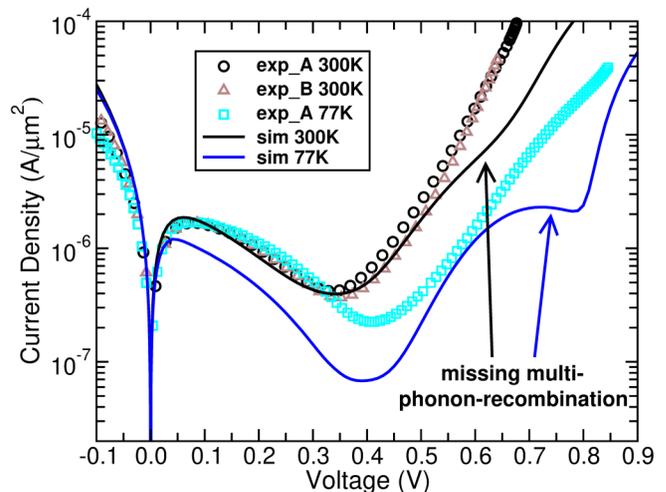


**FIG. 22.** Measured (symbols) and simulated (lines) forward characteristics at 300 K including all processes. Mutual parameters:  $m_c = 0.043 m_0$ ,  $m_v = 0.047 m_0$ , and  $E_g = 0.74$  eV. Parameters for zero-phonon DAT:  $D_{tr} = 4 \times 10^{17}$  cm $^{-3}$ eV $^{-1}$ ,  $V_{tr} = 0.3$  nm $^3$ , Gaussian trap DOS with  $E_0 = -0.2$  eV, and  $E_\sigma = 0.08$  eV. Parameters for field-enhanced multi-phonon recombination:  $S = 7$ ,  $\hbar\omega_{eff} = 30$  meV, and  $\tau_{n,p} = 1$  ns. Solid black curve: no DOS tails. Red dashed curve: including exponential DOS tails with  $m_{hc} = m_c$ ,  $m_{hv} = m_v$ , and  $\eta^{(i)} = \eta(N_{dop} < 10^{19}$  cm $^{-3}$ ) = 30 meV and  $\eta^{(a)} = \eta(N_{dop} > 10^{19}$  cm $^{-3}$ ) = 150 meV.

the opposite situation, i.e., no electron-phonon coupling. The first model is effective at larger forward bias ( $>0.4$  V), where the growing distance between the band edges favors a multi-phonon process. Close to band alignment ( $V < +0.4$  V), zero-phonon DAT becomes more probable, and defect levels inside the tunnel window located between midgap and band edge can contribute to the valley current. In Fig. 21, homogeneously distributed single-level defects were assumed and the zero-phonon DAT current was simulated for different energy levels  $E_{tr}$  measured from midgap. This current is proportional to the product of defect concentration  $N_{tr}$  and “effective trap volume”  $V_{tr}$ . The fraction of total defect volume  $N_{tr}V_{tr} = 10^{-5}$  was kept unchanged in all cases. As can be seen, energy levels in the interval between  $-0.2$  eV and  $+0.1$  eV from midgap yield current maxima covering the whole bias range from  $+0.25$  V to  $+0.5$  V.

Figure 22 presents a fit for 300 K obtained with the combination of BTBT, zero-phonon DAT, field-enhanced multi-phonon recombination, and diffusion (black solid curve). Parameters are given in the caption. In order to see an effect of DOS tails (TTBT) around  $+0.2$  V, the largest possible value of  $\eta^{(i)} = \eta(N_{dop} < 10^{19}$  cm $^{-3}$ ) = 30 meV and the huge value of  $\eta^{(a)} = \eta(N_{dop} > 10^{19}$  cm $^{-3}$ ) = 150 meV have to be used. The DOS tails make no difference at all, if  $\eta^{(a)} \leq 100$  meV.

Interestingly, zero-phonon DAT can also explain the weak temperature dependence of the valley current. Figure 23 shows a comparison for 300 K and 77 K using the same parameters. Two Gaussian defect distributions centered at  $\pm 0.16$  eV from midgap with otherwise equal parameters were assumed. Multi-phonon recombination had to be skipped because both S-Device models did not converge at 77 K. The poorer agreement for 77 K is partly due to the temperature dependence of the bandgap in the



**FIG. 23.** Measured (symbols) and simulated (lines) forward characteristics at 300 K (black) and 77 K (blue) using exclusively zero-phonon DAT via homogeneously allocated traps assuming two Gaussian distributions centered at  $E_0^{(1)} = -0.16$  eV and  $E_0^{(2)} = +0.16$  eV from midgap. Mutual parameters:  $m_c = 0.043 m_0$ ,  $m_v = 0.047 m_0$ ,  $E_g = 0.74$  eV,  $V_{tr} = 0.3$  nm $^3$ ,  $S = 0$ ,  $\hbar\omega_{eff} = 30$  meV,  $E_\sigma = 0.08$  eV, and  $D_{tr} = 4 \times 10^{17}$  cm $^{-3}$ eV $^{-1}$ .

simulation, which is not seen in the measured peak currents. It should be noted that the used defect distributions and parameters are hypothetical and they only demonstrate that zero-phonon DAT is a possible explanation for the valley current.

#### IV. CONCLUSION

A compact theoretical model of tail-to-tail tunneling in semiconductors has been developed taking the localized nature of tail states into account. The three-dimensional pseudo-delta potential<sup>1</sup> yields *s*-like states with a localization radius that is parameterized by the effective mass of the localized electron. Hence, the masses  $m_{\{c,v\}}$  are adjustable parameters that enter the tunnel probability. The potential model allows to include the field-induced lifetime broadening of the states. The microscopic transition rate between two opposing, localized tail states in a homogeneous electric field of arbitrary orientation was derived and found to be governed by the *reduced* effective mass built from  $m_{\{c,v\}}$ . These effective masses could be identified with the band masses, since tail states split from their corresponding bands, and the typical localization radii cover many unit cells. In the device application part of this paper, we chose  $m_c \rightarrow m_c$  and  $m_v \rightarrow m_{lh}$  in order to maximize the effect of tail-to-tail tunneling. As in our treatment of tail-to-band tunneling,<sup>12</sup> a ladder of closely spaced single-level DOSs of the lifetime-broadened localized states was assumed to form the tail DOS. The weight of each single-level DOS is obtained from the Gaussian or exponential shape. Therefore, two further model parameters are needed: the characteristic (band tail spreading) energies  $\eta_{\{c,v\}}$ . In the application section, they were treated as piecewise constants, but they could also be empirically modeled as a function of doping and temperature. The final expression for the macroscopic generation rate due to tail-to-tail tunneling takes the form of a double integral over the tail energies. Only for Gaussian tail shape, a fully analytical solution could be derived in the cases of high and very low electric field, respectively. This yielded insight into the effective tunnel barrier, the tunneling mass, and the reduction of the pre-exponential factor compared to band-to-band tunneling.

Both the homogeneous field case and the in-depth analysis of InGaAs *pin*-Esaki diodes revealed that tail-to-tail tunneling is a completely negligible process compared to tail-to-band tunneling due to the localized nature of the wave functions. Furthermore, tail-induced tunneling can be excluded as the reason behind the strong measured valley current. In contrast, zero-phonon defect-assisted tunneling alone can reproduce the magnitude and the temperature dependence of the latter using reasonable parameters for concentration and defect volume. In order to slightly change the negative slope in the NDR region, one has to assume that  $\eta_{\{c,v\}} > 100$  meV in the heavily doped regions of the diode. Whether such a value is feasible remains an open question, at least it contradicts all optical measurements of band tails in III-V materials. On the other hand, various approximations were necessary for the final form of the model and its numerical implementation in a commercial device simulator. Most notably are the limit of “strong” localization of the states (at length discussed in Ref. 12), the approximate treatment of the radial distance, and the numerical integration of the action. Furthermore, the developed model is a continuum model, which requires that the active volume is large

enough for a proper average over random disorder caused by doping. Nanowire TFETs and Esaki diodes might only contain a countable number of doping atoms. Then, band tails in the above-described sense lose their meaning and an atomistic simulation method like tight-binding non-equilibrium Green’s function (NEGF)<sup>8</sup> becomes indispensable.

The model of tail-to-tail tunneling was developed for direct materials. It might also be suitable for Ge nanowires, where elastic and inelastic tunneling are comparable.<sup>72</sup> In Si and Ge bulk-like homo-junctions, phonon-assisted band-to-band tunneling dominates. Here, tail-to-tail tunneling could be direct, if the strong localization of the tail states in real space results in a sufficient spread in *k*-space and thereby relaxes the need of a phonon for momentum conservation. In Si/InAs hetero-junctions, the imaginary dispersion in the gap was found to be continuous and smooth,<sup>73</sup> which allows to fit both effective band masses, to extract an effective “tunnel gap” for the bandgap, and to use these parameters in the model.

A renewed interest in tail-to-tail tunneling has emerged recently in the context of quantum computing and MOSFET operation at cryogenic temperatures.<sup>74</sup> The experimentally observed saturation of the sub-threshold swing toward very low temperatures could involve elastic tunneling between conduction band tail states or deeper lying trap states.<sup>75</sup> It is appealing that the saturated swing could just reflect the characteristic energy of an exponential tail DOS. Many aspects of our theory could be used to develop a mobility model that leads to the saturation of the swing at a certain temperature.

#### ACKNOWLEDGMENTS

We thank Dr. Hamilton Carrillo-Nuñez for the computation of the imaginary full-band dispersion in InGaAs with OMEN.<sup>48</sup> We are indebted to Dr. Anne Verhulst (IMEC) for providing experimental data and for many stimulating discussions.

#### APPENDIX A: BAND GAP NARROWING IN InGaAs

The random-phase-approximation (RPA)-based BGN model of Ref. 55 implemented in S-Device<sup>44</sup> for silicon as default can be applied to any material, if effective masses  $m_{\{c,lb,hb\}}$ , band multiplicities, and permittivity  $\epsilon_s$  are known. For the computation of the model parameters, lengths have to be scaled by the excitonic Bohr radius  $a_{ex}$  and energies by the excitonic Rydberg energy  $Ry_{ex}$ ,

$$a_{ex} = \frac{\hbar^2 \epsilon_s}{e^2 \mu^*}, \quad Ry_{ex} = \frac{\hbar^2}{2\mu^* a_{ex}^2}. \quad (A1)$$

The effective mass  $\mu^*$  is given by

$$\mu^* = \frac{m_c m_v}{m_c + m_v} \quad \text{with} \quad m_v^{3/2} = \frac{1}{2} \left( m_{lh}^{3/2} + m_{hh}^{3/2} \right). \quad (A2)$$

The mass  $m_v$  is the average hole DOS mass. Band and valley multiplicities, as well as spin degree of freedom, are contained in the model parameters  $g_{\{e,h\}}$ . The values of the material-dependent parameters for the case of In<sub>0.53</sub>Ga<sub>0.47</sub>As are provided in Table II.

**TABLE II.**  $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$  material-related parameters<sup>45–47,60</sup> used in the BGN model.<sup>44,55</sup>

$m_c/m_0$	$m_{lh}/m_0$	$m_{hh}/m_0$	$m_v/m_0$	$\mu^*/m_0$
0.043	0.047	0.363	0.236	0.0364

$g_c$	$g_h$	$\alpha_c$	$\alpha_h$	$Ry_{ex}$	$a_{ex}$	$\epsilon_s$
2	4	0.846	0.154	2.56 meV	$20.25 \times 10^{-7}$ cm	13.9

For simplicity, only the zero-temperature limits of the rigid band shift are considered,

$$\Delta_c = - \left[ \left( \frac{48n}{\pi g_c} \right)^{1/3} + c_e \ln(1 + d_e n_p^{p_e}) + N_{dop} \frac{0.799\alpha_c}{N_p^{3/4}} \right],$$

$$\Delta_v = - \left[ \left( \frac{48p}{\pi g_h} \right)^{1/3} + c_h \ln(1 + d_h n_p^{p_h}) + N_{dop} \frac{0.799\alpha_h}{N_p^{3/4}} \right].$$

Here,  $n_p = \alpha_c n + \alpha_h p$  and  $N_p = \alpha_c N_D^- + \alpha_h N_A^-$  with  $\alpha_c = \mu^*/m_c$  and  $\alpha_h = \mu^*/m_v$ .  $N_{dop}$  denotes the electrically active doping. The first term is the exchange energy and the second the low-temperature limit of the free-carrier correlation energy.<sup>70</sup> The parameters  $c_{\{e,h\}}$ ,  $d_{\{e,h\}}$ , and  $p_{\{e,h\}}$  had been fitted for silicon,<sup>55</sup> but can be assumed to depend only weakly on the material. They are given in Table III. The last term is the low-temperature limit of the ionic part of the correlation energy.<sup>71</sup> The BGN is given by

$$\Delta E_g = |\Delta_c| + |\Delta_v|. \quad (A3)$$

Note that in the S-Device manual<sup>44</sup> (the correctly implemented)  $N_p$  is confused with  $n_p$ .

### APPENDIX B: RELATION BETWEEN KANE'S TWO-BAND MODEL AND ONE-BAND EMA MODEL

The BTBT rate for homogeneous electric field calculated with Kane's two-band model<sup>38</sup> reads

$$G_{BTB}^{Kane} = \frac{(eF)^2}{18\pi\hbar^2} \sqrt{\frac{m_r}{E_g}} \exp\left(-\frac{\pi E_g^{\frac{3}{2}} \sqrt{m_r}}{2\hbar F}\right). \quad (B1)$$

The WKB limit of the one-band EMA model as given by Eq. (52) in Sec. II D is

$$G_{BTB}^{1B-EMA} = \frac{(eF)^3}{64\pi\hbar E_g^2} \exp\left(-\frac{4 E_g^{\frac{3}{2}} \sqrt{2m_r}}{3\hbar F}\right). \quad (B2)$$

**TABLE III.** Fit parameters used for the low-temperature limit of the free-carrier correlation energy<sup>55</sup> and<sup>44</sup>.

$c_e$	$c_h$	$d_e$	$d_h$	$p_e$	$p_h$
1.3346	1.2365	0.893	1.153	0.2333	0.2333

Thus, the ratio takes the form

$$r_{scal} = \frac{G_{BTB}^{Kane}}{G_{BTB}^{1B-EMA}} = \frac{32\sqrt{m_r} E_g^{\frac{3}{2}}}{9\hbar e F} \exp\left[\frac{4 E_g^{\frac{3}{2}} \sqrt{2m_r}}{3\hbar F} \left(1 - \frac{3\pi}{8\sqrt{2}}\right)\right], \quad (B3)$$

which is used as a scaling function in the simulation of tail-induced tunneling currents. For this, the field dependence was substituted by the  $F(V)$ -curve shown in Fig. 15. The outcome of Eq. (B3) is presented in the same figure.

### DATA AVAILABILITY

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

### REFERENCES

- W. C. Vinogradov, *Fiz. Tverd. Tela* **13**, 3266 (1971) [*Sov. Phys. Solid State* **13** (11), 2745 (1972)].
- S. Banerjee, W. Richardson, J. Coleman, and A. Chatterjee, *IEEE Electron Dev. Lett.* **8**, 347 (1987).
- A. M. Ionescu and H. Riel, *Nature* **479**, 329 (2011).
- L. Esaki, *Phys. Rev.* **109**, 603 (1958).
- M. T. Björk, H. Schmid, J. Knoch, H. Riel, and W. Riess, *Nat. Nanotechnol.* **4**, 103 (2009).
- A. Schenk, S. Sant, K. Moselund, and H. Riel, in *Proceedings of ULIS-EUROSOI* (IEEE, 2016), pp. 9–12.
- M. A. Kaye and R. Lake, *J. Appl. Phys.* **110**, 074508 (2011).
- S. Sylvia, M. Khayer, K. Alam, and R. Lake, *IEEE Trans. Electron Devices* **59**, 2996 (2012).
- S. Sylvia, K. Habib, M. Khayer, K. Alam, M. Neupane, and R. Lake, *IEEE Trans. Electron Dev.* **61**, 2208 (2014).
- Q. Smets, D. Verreck, A. S. Verhulst, R. Rooyackers, C. Merckling, M. Van De Put, E. Simoen, W. Vandervorst, N. Collaert, V. Y. Thean, B. Sorée, and G. Groeseneken, *J. Appl. Phys.* **115**, 184503 (2014).
- E. Memisevic, E. Lind, M. Hellenbrand, J. Svensson, and L.-E. Wernersson, *IEEE Electron Dev. Lett.* **38**(12), 1661 (2017).
- S. Sant and A. Schenk, *J. Appl. Phys.* **122**, 135702 (2017).
- J. Bizindavyi, A. S. Verhulst, Q. Smets, D. Verreck, B. Sorée, and G. Groeseneken, *J. Elec. Dev. Soc.* **6**, 633 (2018).
- E. Memisevic, M. Hellenbrand, E. Lind, A. R. Persson, S. Sant, A. Schenk, J. Svensson, R. Wallenberg, and L.-E. Wernersson, *Nano Lett.* **17**(7), 4373 (2017).
- M. Takeshima, *Phys. Rev. B* **12**(2), 575 (1975).
- J. Teherani, S. Agarwal, W. Chern, P. Solomon, E. Yablonovitch, and D. Antoniadis, *J. Appl. Phys.* **120**, 084507 (2016).
- A. L. Efros, *Sov. Phys. Usp.* **16**, 789 (1974).
- E. O. Kane, *Solid State Electron.* **28**, 3 (1985).
- P. A. Wolff, *Phys. Rev.* **126**, 405 (1962).
- P. Chakraborty and K. Ghatak, *Phys. Lett. A* **288**, 335 (2001).
- J. A. Guerra, A. Tejada, J. A. Töfflinger, R. Grieseler, and L. Korte, *J. Phys. D Appl. Phys.* **52**, 105303 (2019).
- E. O. Kane, *Phys. Rev.* **131**, 79 (1963).
- B. I. Halperin and M. Lax, *Phys. Rev.* **148**, 722 (1966).
- B. I. Shklovskii and A. L. Efros, *Sov. Phys. Semicond.* **4**, 249 (1970).
- J. Katahara and H. Hillhouse, *J. Appl. Phys.* **116**, 173504 (2014).
- D. Redfield, *J. Non-Crystalline Solids* **8–10**, 602 (1972).
- J. I. Pankove, *Phys. Rev.* **140**, A2059 (1965).
- R. Dixon and J. M. Ellis, *Phys. Rev.* **123**, 1560 (1961).
- F. Urbach, *Phys. Rev.* **92**, 1324 (1953).
- V. Maluytenko and V. Chernyakhovsky, *Semicond. Sci. Technol.* **9**, 1047 (1994).

- <sup>31</sup>S. W. Kurnick and J. M. Powell, *Phys. Rev.* **116**, 597 (1959).
- <sup>32</sup>G. Lucovsky, *Solid State Commun.* **3**, 299 (1965).
- <sup>33</sup>A. Schenk, R. Enderlein, and D. Suisky, *Phys. Stat. Sol. (b)* **131**, 729 (1985).
- <sup>34</sup>D. E. Aspnes, *Phys. Rev.* **147**, 554 (1966).
- <sup>35</sup>M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover Publications, New York, 1970), p. 449.
- <sup>36</sup>K. Peuker, private communication, 1981.
- <sup>37</sup>F. Bechstedt, R. Enderlein, and K. Peuker, *Phys. Stat. Sol. (b)* **68**, 43 (1975).
- <sup>38</sup>E. O. Kane, *J. Phys. Chem. Solids* **12**(2), 181 (1960).
- <sup>39</sup>N. Mott, *J. Phys. C Solid State Phys.* **20**, 3075 (1987).
- <sup>40</sup>K.-H. Hasler and H.-J. Wünsche, *Wiss. Z. Humboldt Univ. Berlin Math. Nat. R.* **35**(2), 126 (1986).
- <sup>41</sup>A. Schenk, M. Stahl, and H.-J. Wünsche, *Phys. Stat. Sol. (b)* **154**, 815 (1989).
- <sup>42</sup>H. Carrillo-Núñez, A. Ziegler, M. Luisier, and A. Schenk, *J. Appl. Phys.* **117**, 234501 (2015).
- <sup>43</sup>A. Schenk, *Solid-State Electron.* **36**(1), 19 (1993).
- <sup>44</sup>Synopsys Inc., Sentaurus Device User Guide, Version-2015.06, Mountain View, California, USA (2015).
- <sup>45</sup>I. R. Vurgaftman, R. J. Meyer, and L. R. Ram-Mohan, *J. Appl. Phys.* **89**(11), 5815 (2001).
- <sup>46</sup>C. Hermann and T. P. Pearsall, *Appl. Phys. Lett.* **38**, 450 (1981).
- <sup>47</sup>K. Alavi, R. L. Aggarwal, and S. H. Groves, *Phys. Rev. B* **21**, 1311 (1980).
- <sup>48</sup>M. Luisier, A. Schenk, and W. Fichtner, *J. Appl. Phys.* **100**, 043713 (2006).
- <sup>49</sup>E. O. Kane, *J. Appl. Phys.* **32**, 83 (1961).
- <sup>50</sup>H. Flietner, *Phys. Stat. Sol. (b)* **54**, 201 (1972).
- <sup>51</sup>A. W. Walker and M. W. Denhoff, *Appl. Phys. Lett.* **111**, 162107 (2017).
- <sup>52</sup>E. Zielinski, H. Schweizer, K. Streubel, H. Eisele, and G. Weimann, *J. Appl. Phys.* **59**, 2196 (1986).
- <sup>53</sup>Y. P. Varshni, *Physica* **34**(1), 149 (1967).
- <sup>54</sup>S. C. Jain, J. M. McGregor, and D. J. Roulston, *Appl. Phys. Lett.* **68**(7), 3747 (1990).
- <sup>55</sup>A. Schenk, *J. Appl. Phys.* **84**(7), 3684 (1998).
- <sup>56</sup>R. Zimmermann, *Many Particle Theory of Highly Excited Semiconductors* (BSB Teubner Verlagsgesellschaft, Leipzig, 1988).
- <sup>57</sup>R. K. Ahrenkiel, R. Ellingson, S. Johnston, and M. Wanlass, *Appl. Phys. Lett.* **72**(26), 3470 (1998).
- <sup>58</sup>S. Hausser, G. Fuchs, A. Hangleiter, and K. Streubel, *Appl. Phys. Lett.* **56**, 913 (1990).
- <sup>59</sup>M. E. Prise, M. R. Taghizadeh, S. D. Smith, and B. S. Wherrett, *Appl. Phys. Lett.* **45**, 652 (1984).
- <sup>60</sup>The value is provided on <http://www.ioffe.ru/SVA/NSM/Semicond/InAs>.
- <sup>61</sup>A. G. Lind, H. L. Aldridge, C. C. Bomberger, C. Hatem, J. M. O. Zide, and K. S. Jones, *ECS Trans.* **66**(7), 23 (2015).
- <sup>62</sup>B. Tell, R. F. Leheny, A. S. H. Liao, T. J. Bridges, E. G. Burkhardt, T. Y. Chang, and E. D. Beebe, *Appl. Phys. Lett.* **44**, 438 (1984).
- <sup>63</sup>C. Convertino, D. Cutaia, H. Schmid, N. Bologna, P. Paletti, A. M. Ionescu, H. Riel, and K. E. Moselund, in *2017 Joint International EUROSIOI Workshop and International Conference on Ultimate Integration on Silicon (EUROSIOI-ULIS)* (IEEE, 2017), p. 148.
- <sup>64</sup>A. Schenk, S. Sant, K. Moselund, and H. Riel, *ECS Trans.* **66**(5), 157 (2015).
- <sup>65</sup>A. Schenk, in *Advanced Physical Models for Silicon Device Simulation, Computational Microelectronics*, edited by S. Selberherr (Springer, Wien, 1998).
- <sup>66</sup>A. Schenk, *J. Appl. Phys.* **71**(7), 3339 (1992).
- <sup>67</sup>A. Schenk, *Solid-State Electron.* **35**(11), 1585 (1992).
- <sup>68</sup>K. Huang and A. Rhys, *Proc. Roy. Soc. London A* **204**, 406 (1950).
- <sup>69</sup>F. Jiménez-Molinos, F. Gámiz, A. Palma, P. Cartujo, and J. A. López-Villanueva, *J. Appl. Phys.* **91**(8), 5116 (2002).
- <sup>70</sup>M. Rösler, R. Zimmermann, and W. Richert, *Phys. Status Solidi B* **121**, 609 (1984).
- <sup>71</sup>F. Thuselt, *Phys. Lett. A* **94**, 93 (1983).
- <sup>72</sup>M. Luisier and G. Klimeck, *J. Appl. Phys.* **107**, 084507 (2010).
- <sup>73</sup>H. Carrillo-Núñez, M. Luisier, and A. Schenk, in *Proceedings of the 44th European Solid-State Device Research Conference (ESSDERC)* (IEEE, Venice, 2014), pp. 118–121.
- <sup>74</sup>A. Beckers, F. Jazaeri, and C. Enz, *IEEE Trans. Electron Devices* **65**(9), 3617 (2018).
- <sup>75</sup>A. Beckers, F. Jazaeri, and C. Enz, [arXiv:1811.09146v2](https://arxiv.org/abs/1811.09146v2) [cond-mat.mes-hall] (2019).
- <sup>76</sup>A concept that can be traced back to Urbach<sup>30</sup> who observed that the slope of the logarithm of the absorption spectrum is proportional to  $-1/k_B T$  over a wide temperature range. Due to the convolution of DOS tails and thermal occupation, the Urbach temperature is given by  $T_u \approx T \eta / (\eta + k_B T)$ .