

Finite-temperature full random-phase approximation of bandgap narrowing in quasi-neutral regions: Theory including nonparabolicity and application to silicon and InGaAs

Cite as: J. Appl. Phys. **130**, 015703 (2021); <https://doi.org/10.1063/5.0051055>

Submitted: 22 March 2021 . Accepted: 15 June 2021 . Published Online: 06 July 2021

 A. Schenk



View Online



Export Citation



CrossMark



Webinar
How to Characterize Magnetic
Materials Using Lock-in Amplifiers

Zurich Instruments MFLI

Zurich Instruments CRYOGENIC

Register now

Finite-temperature full random-phase approximation of bandgap narrowing in quasi-neutral regions: Theory including nonparabolicity and application to silicon and InGaAs

Cite as: J. Appl. Phys. **130**, 015703 (2021); doi: [10.1063/5.0051055](https://doi.org/10.1063/5.0051055)

Submitted: 22 March 2021 · Accepted: 15 June 2021 ·

Published Online: 6 July 2021



A. Schenk^{a)} 

AFFILIATIONS

Integrated Systems Laboratory, ETH Zurich, Gloriastr. 35, 8092 Zurich, Switzerland

^{a)}Author to whom correspondence should be addressed: schenk@iis.ee.ethz.ch

ABSTRACT

The bandgap narrowing (BGN) in quasi-neutral regions of semiconductors is calculated in a finite-temperature full random-phase approximation formalism based on a simple isotropic dispersion model including band nonparabolicity. The total quasi-particle shift (QPS) is determined by the exchange-correlation self-energy of the free carriers and the correlation energy of the interaction between carriers and ionized dopants. At cryogenic temperatures, the latter part results in giant shifts of the minority band edge in n-type semiconductors with a large ratio of valence to conduction band density of states, as often present in III–V materials. However, at room temperature, the BGN does not exceed common values. The reason for this behavior is explained analytically. Whereas the exchange-correlation energy of free carriers is known to be insensitive to band structure details, the nonparabolicity of the conduction band (CB) has a strong effect on the ionic QPS of the minority carriers in n-type III–V materials. It strongly reduces the BGN at cryogenic temperatures compared to the case of a parabolic CB. This is demonstrated numerically and also analytically for n-type InGaAs. The BGN in n-type silicon becomes independent of temperature at high concentrations, but in p-type silicon, a weak temperature dependence re-emerges above the Mott density, which also can be attributed to the ionic QPS of the minority electrons. The calculated BGN for quasi-neutral regions in silicon is in good agreement with earlier photoluminescence and more recent photo-conductance measurements.

© 2021 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/10.1063/5.0051055>

I. INTRODUCTION

Bandgap narrowing (BGN) is an important phenomenon in semiconductor devices. It is a consequence of either heavy doping or strong optical excitation and high electrical injection. High excitation/injection levels lead to plasma-induced BGN, as in photo-conductive switches, concentrator solar cells, and power devices operated in the on-state. The plasma can be neutral, e.g., in the case of laser excitation of intrinsic regions, or not, e.g., in the case of optical excitation of doped regions. Quasi-neutral regions of electronic devices (emitters of bipolar transistors, source/drain-regions of field-effect transistors) host a one-component plasma that

compensates the electrically active dopants. “Quasi” refers to the possible presence of a very small electric field when the device is biased or a small concentration of minority carriers. The many-body Coulomb interaction between free carriers as well as between carriers and ionized dopants results in self-energy and a corresponding shift of the band edges such that the bandgap shrinks. In quasi-neutral regions, one has five components: the exchange energy of the majority carriers, the free-carrier correlation energies of majority and minority carriers, and the ionic correlation energies of majority and minority carriers due to their interaction with the electrically active dopants. These contributions can be treated equally in the theoretical framework of the finite-temperature full

Random-Phase Approximation (RPA).^{1–5} All measurements of BGN involve electron–hole pairs. In quasi-neutral regions, the minority carrier band shifts because the generated/annihilated minority carrier lowers its energy due to Coulomb attraction by the mobile majority carriers and Coulomb repulsion by the sub-system of immobile ionized dopants that are the source of the majority carriers.⁶ As will be shown in this paper, the ionic correlation energy of holes in n-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ is even the largest contributor to the total BGN in this material.

Most experimental data on BGN exist for silicon but with a rather large spread, which can be attributed to the involved evaluation procedure. Absorption experiments^{7–12} require the admixture of phonon-assisted transitions and the knowledge of phonon energies and Fermi level. Problems in photoluminescence (PL) experiments^{13–16} are the nature of the initial states and the weak intensity of the spectrum. The electrical method^{17–22} uses bipolar transistors where the collector current as a function of the emitter-base voltage, the sheet resistance underneath the emitter, and the minority carrier mobility in the base must be measured. The BGN is then found from an analytical model that connects these quantities and that itself was derived with a number of approximations. For materials other than silicon, BGN data are rare, and for most materials, they are not available at all. Therefore, reliable calculations and fit models derived from them play an important role to fill this gap.

Earlier theoretical methods to derive the self-energy of the carrier–dopant interaction in the $T = 0$ -limit comprise Hartree–Fock variational calculations for a donor system distributed on a regular sub-lattice²³ using the single plasmon pole approximation²⁴ and application of second-order perturbation theory for a random system of dopants²⁵ using RPA screening. The role of multi-valley scattering in both dopant arrangements had been the topic of discussions.^{26–28} RPA is the method of choice to compute BGN for the whole temperature range, and it gains accuracy with an increase in the concentration above the Mott density. The theoretical basics can be found in Ref. 1, which also presents the temperature-dependent QPS in a neutral electron–hole plasma. The case of an extrinsic semiconductor with plasma excitation was treated in Ref. 3 where numerical results based on a random distribution of dopants were given for $T = 0$ K.

In Ref. 29, an RPA-based fit model for silicon device simulation as a function of carrier densities, doping concentration, and temperature was developed. To cover all cases (neutral plasma, quasi-neutral region, depletion region), compromises had to be made in the fitting procedure. As will be shown below, this fit model underestimates the BGN in quasi-neutral regions at very high doping concentrations. Besides the reevaluation of BGN for quasi-neutral regions in Si including nonparabolicity of the conduction band (CB), BGN in quasi-neutral $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ will be analyzed in detail. The main finding is that at a temperature of 300 K and as a consequence of strong CB nonparabolicity, the BGN in the n-type material is limited to $\approx 10\%$ of the bandgap at the highest measured concentration of electrically active doping. In the p-type material, this amount is $\approx 6\%$.

The paper is organized as follows: Section II gives a brief summary of the RPA theory and shows how nonparabolicity is included. Section III provides results for silicon: numerical curves

for the different ionic QPSs, the temperature-dependent total BGN in comparison with absorption and PL data, and the small influence of CB nonparabolicity. To better understand the ionic QPS of the minority and majority bands, they are calculated analytically as a function of doping concentration in the parabolic, $T = 0$ K-limit. Section IV presents the results for $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$: numerical free-carrier and ionic correlation energies over the whole temperature range, the comparison of parabolic and nonparabolic curves, and the total BGN in n-type and p-type materials as a function of temperature and doping. To better understand the impact of CB nonparabolicity, the ionic correlation energy of minority holes is derived analytically in the $T = 0$ K-limit. Conclusions are given in Sec. V. Appendix A contains the derivation of the full-RPA dielectric function for nonparabolic bands, its static limit, as well as the high-T/low-density and low-T/high-density limits of the static form, respectively. Appendix B demonstrates the derivation of the full-RPA correlation energy of ion–carrier interaction for nonparabolic bands in the high-T/low-density and low-T/high-density limits, respectively.

II. THEORETICAL BACKGROUND

BGN in RPA quality is calculated by the quasi-particle shift (QPS)¹ $\Delta_a(k)$ ($a = e$ for electrons, $a = h$ for holes), which is the difference between free and interacting dispersion (the vector character of momenta and coordinates is only written explicitly where needed, and the normalization volume is set to unity),

$$E_a(k) = E_a^0(k) + \Delta_a(k). \quad (1)$$

It represents the real part of the self-energy $\Delta_a(k) = \text{Re} \Sigma_a[k, E_a(k) + i0^+]$ and consists of three terms

$$\Delta_a(k) = \Delta_a^x(k) + \Delta_a^c(k) + \Delta_a^i(k), \quad (2)$$

where $\Delta_a^x(k)$ is the unscreened exchange term, $\Delta_a^c(k)$ is the correlation term of free carriers, and $\Delta_a^i(k)$ is the correlation term of the interaction between carriers and ionized dopants. The latter are assumed to be randomly arranged on regular lattice sites. The dispersion of the QPS is rather flat in the relevant energy interval between band edge and Fermi energy for all densities and temperatures.^{1,29} Therefore, the dispersive QPS $\Delta_a(k)$ can be replaced by a rigid shift Δ_a . There is no need to solve the self-consistency problem then because the only occurrence of Δ_a is in the distribution functions, where it is fixed by the given density in the quasi-neutral region. An elegant way to compute the rigid shift follows from the requirement that the QPS density should not change in the first order with respect to $\Delta_a(k) - \Delta_a$, which leads to¹

$$\Delta_a = \frac{\sum_k \partial f_a(k) / \partial \zeta_a \Delta_a(k)}{\sum_k \partial f_a(k) / \partial \zeta_a}, \quad (3)$$

with rigidly shifted bands $E_a(k) = E_a^0(k) + \Delta_a$ and Fermi functions $f_a(k)$ depending on shifted chemical potentials μ_a ,

$$f_a(k) = f(E_a^0(k) - \zeta_a), \quad \zeta_a = \mu_a - \Delta_a. \quad (4)$$

TABLE I. Parameters for silicon and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ used in the calculation.^{29,32–36}

	m_e/m_0	m_h/m_0	g_e	g_h	μ^*/m_0	α_e	α_h	γ_e (/eV)	Ry_{ex} (meV)	a_{ex} (cm)	ϵ_s
Silicon	0.321	0.346	12	4	0.1665	0.5187	0.4813	0.5	16.55	37.19×10^{-8}	11.7
InGaAs	0.043	0.236	2	4	0.0364	0.846	0.154	1.2	2.56	20.25×10^{-7}	13.9

The weight function in (3) (the derivative of the Fermi function) filters out energies near the Fermi energy in the low-T/high-density limit and energies close to the band edge in the high-T/low-density limit. The explicit expressions for the three components of the dispersive QPS in (2) can be found in Refs. 1, 3, and 29 and are not repeated here. After insertion into Eq. (3), the rigid shifts take the compact form^{1,3}

$$\Delta_a^x = -\frac{e^2}{4\pi\epsilon_0\epsilon_s\Lambda_a} F_{-1/2}(\beta\zeta_a), \quad (5)$$

$$\Delta_a^c = \frac{1}{2\beta} \int \frac{d^3q}{(2\pi)^3} \sum_{\nu} [\epsilon^{-1}(q, \Omega_{\nu}) - 1] \frac{\partial\epsilon(q, \Omega_{\nu})}{\partial n_a}, \quad (6)$$

$$\Delta_a^i = -\frac{n_i}{2} \left(\frac{\partial n_a}{\partial \zeta_a} \right)^{-1} \int \frac{d^3q}{(2\pi)^3} \frac{v(q)}{\epsilon^2(q, 0)} \frac{\partial\epsilon(q, 0)}{\partial \zeta_a}. \quad (7)$$

In the case of parabolic bands, the QP density n_a in the rigid shift approximation is given by

$$n_a = \sum_k g_a f_a(k) = g_a \Lambda_a^{-3} F_{1/2}(\beta\zeta_a), \quad (8)$$

with the Fermi integral $F_{1/2}$, the thermal wavelength $\Lambda_a = (2\pi\hbar^2\beta/m_a)^{1/2}$, the inverse thermal energy $\beta = 1/k_B T$, and the Fermi energy $\zeta_a = \mu_a - \Delta_a$. The multi-valley conduction band, heavy- and light-hole bands, and spin summation are condensed in degeneracy factors g_a . The effective masses m_a are understood as density of states (DOS) effective masses $m_e = (m_c^2 m_l)^{1/3}$ for anisotropic CB valleys, $m_h = [(m_{lh}^{3/2} + m_{hh}^{3/2})/2]^{2/3}$ for holes. This yields a tractable isotropic dispersion model needed for analytical BGN modeling that also allows one to include nonparabolicity. More comments on the band structure model will be given in Secs. III and IV. The remaining ingredients in Eqs. (5)–(7) are the bare Coulomb potential $v(q) = e^2/(\epsilon_0\epsilon_s q^2)$ with the static permittivity ϵ_s , the Matsubara frequency $\Omega_{\nu} = \frac{2\pi i\nu}{\hbar\beta}$ (ν integer), and the concentration of electrically active dopants n_i . The correlation energies are determined by the RPA dielectric function

$$\epsilon(q, \Omega_{\nu}) = 1 - v(q) \sum_{a,k} g_a \frac{f_a(k+q) - f_a(k)}{E_a(k+q) - E_a(k) - \hbar\Omega_{\nu}}. \quad (9)$$

As the rigid QPS drops out in the energy differences, the dispersion can be replaced by the free dispersion E_a^0 . It should be noted that the single-integral form of the exchange term (5) is bound to parabolic bands $E_a^0(k) = \hbar^2 k^2/2m_a$,^{1,30} whereas (6) and (7) are obtained

by (3), regardless of the specific form of the dispersion. It is well known that band structure details have only a minor impact on the exchange-correlation energy of the electron and hole plasmas due to a compensation effect.^{1,31} Therefore, Eqs. (5) and (6) are evaluated with parabolic bands in this paper. However, as will be shown below, in the low-T/high-density limit, the term Δ_h^i can take large values in n-type III–V materials. In this case, the Fermi level is located high in the CB, where the dispersion strongly deviates from the parabolic shape. To study how much nonparabolicity changes the BGN, electron density n_e and dielectric function $\epsilon(q, \Omega_{\nu})$ in (7) are calculated with the nonparabolic free dispersion

$$E_e^{0,\text{np}}(k) = \frac{1}{2\gamma_e} \left(\sqrt{2\gamma_e \hbar^2 k^2/m_e + 1} - 1 \right), \quad (10)$$

where γ_e denotes the nonparabolicity parameter. Both dielectric function and ionic QPS can be expressed by single integrals. This is demonstrated in Appendixes A and B, respectively. The $T = 0$ K-limits are also given explicitly there as they serve to validate the tricky numerical solutions at very low but finite temperatures and to explain the high values of Δ_h^i in n-type materials with a large ratio of valence to conduction band density of states. It is convenient to use normalized quantities. All energies are normalized by the excitonic Rydberg energy $Ry_{\text{ex}} = \hbar^2/(2\mu^* a_{\text{ex}}^2)$, and all lengths by the excitonic Bohr radius $a_{\text{ex}} = \hbar^2\epsilon_s/(e^2\mu^*)$, where μ^* denotes the reduced effective mass $\mu^* = (1/m_e + 1/m_h)^{-1}$. Useful parameters are $\alpha_a = \mu^*/m_a$. All values for Si and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ are given in Table I.

III. RESULTS FOR SILICON

Application of the empirical nonparabolicity model (10) to silicon necessitates an isotropic DOS. The use of DOS effective masses $m_e = (m_c^2 m_l)^{1/3}$ to take account of the anisotropy of CB valleys is already an approximation in the correlation energies (6) and (7) with the RPA dielectric function (9). Starting with ellipsoids and transforming k and q to elliptical coordinates would not give the same expressions because of the occurrence of the Fourier transformed Coulomb potential $v(q)$. The light and heavy hole masses m_{lh} , m_{hh} are angle-averaged parameters, the light-hole band is nonparabolic, and the split-off band with its maximum at 44 meV below the VB edge is neglected. The latter simplification sets a limit to the hole density, i.e., calculated BGN values become rather uncertain above $5 \times 10^{19} \text{ cm}^{-3}$. Heavy doping also induces band tails, an effect completely ignored in this paper.

Figure 1 shows the ionic part of the average QPS Δ_e^i for neutral n- and p-type silicon, and Fig. 2 shows the ionic part of the average QPS Δ_h^i for neutral n- and p-type silicon as a function of activated doping concentration and temperature. Note that the

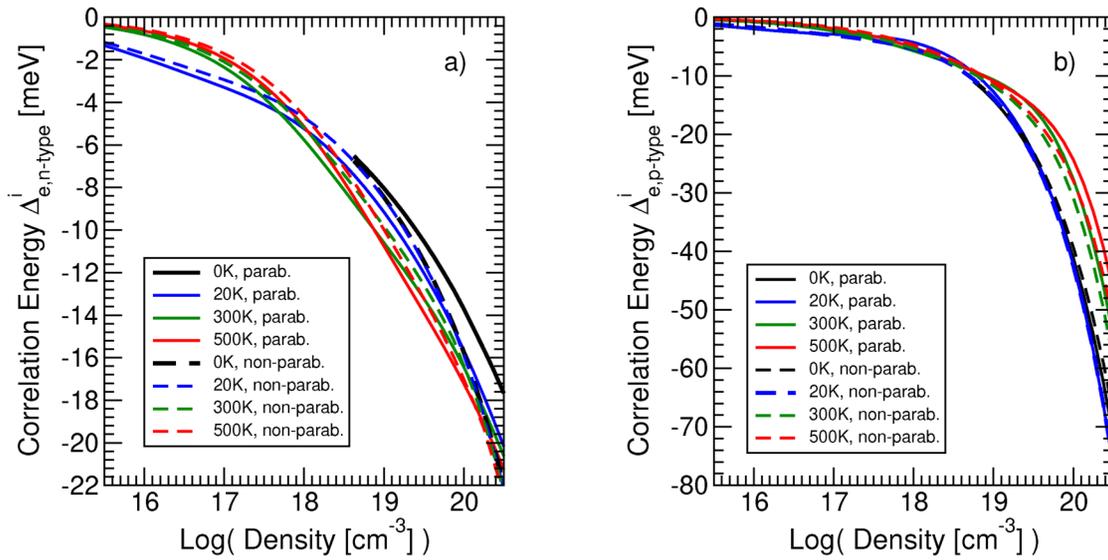


FIG. 1. (a) Electron-ion part $\Delta_{e,n-type}^i$ and (b) $\Delta_{e,p-type}^i$ of the average QPS for neutral (a) n-type and (b) neutral p-type silicon as a function of activated doping concentration and temperature.

$T = 0$ K-curves start at the Mott density³⁷ $n_{b,M} = (p_M/a_{ex}\alpha_b)^3$, using $p_M = 0.296$ for the Mott criterion.

While in a neutral electron-hole plasma the contribution of Δ_a^i to the total BGN becomes small under conditions of high excitation/injection, this is not the case in heavily doped quasi-neutral regions. The reason is that only majority carriers, and only with a fixed density n_i are available for screening there. If their DOS is

small, the screening effect is weak, and Δ_a^i of the minority carrier band can become significant. Despite the relatively large DOS of Si, this is clearly seen for p-type silicon in Fig. 1(b) and, on a smaller scale, for n-type Si in Fig. 2(a).

To understand the different behavior of majority and minority carriers it is instructive to derive the ionic correlation energy in a fully analytical form. This is only possible for $T = 0$ K.

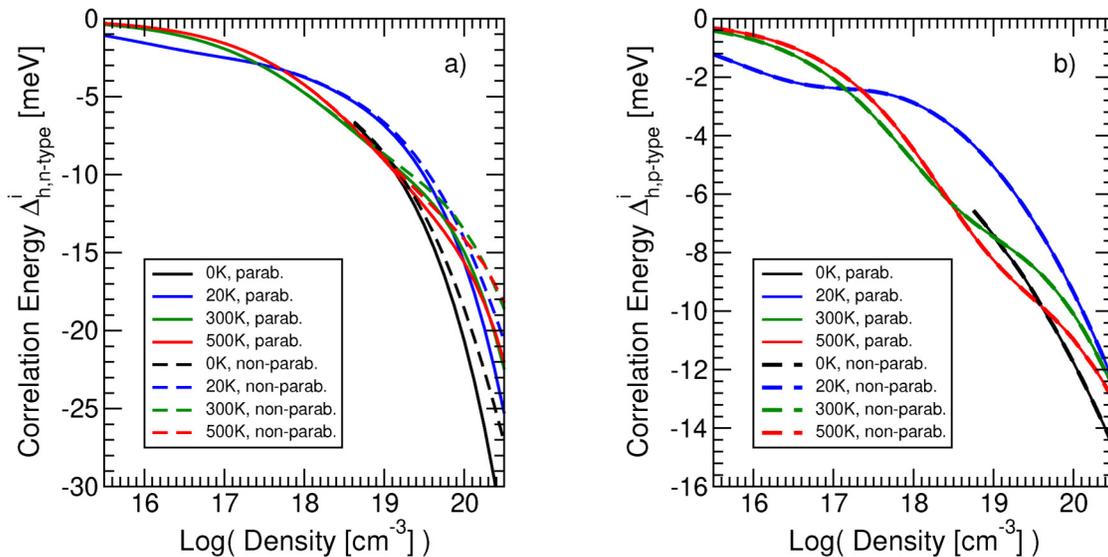


FIG. 2. (a) Hole-ion part $\Delta_{h,n-type}^i$ and (b) $\Delta_{h,p-type}^i$ of the average QPS for (a) neutral n-type and (b) neutral p-type silicon as a function of activated doping concentration and temperature.

The nonparabolicity effect is minor in Si because of the relatively small value of the nonparabolicity parameter ($\gamma_e = 0.5/\text{eV}$) and the large DOS of the CB that prevents that the Fermi level can move deep into the band with growing density. Therefore, in this section, the analytical treatment is restricted to the parabolic case. The nonparabolic case will be presented for InGaAs in Sec. IV.

The starting point is the parabolic and zero-temperature limit Eq. (B10) in Appendix B with the dielectric function Eq. (A13) in Appendix A. Inserting Eq. (A13) into Eq. (B10) yields by substituting $q = z q_{F,b}$,

$$\Delta_a^i = -\frac{4g_b}{3\pi^2} \int_0^\infty dz \frac{\frac{z}{\alpha_b} \tanh^{-1}\left(\frac{z}{1+z^2/4}\right) \delta_{ab} + \frac{4}{\alpha_a}(1-\delta_{ab})}{\left\{z^2 + s_b \left[1 + \left(\frac{1}{z} - \frac{z}{4}\right) \tanh^{-1}\left(\frac{z}{1+z^2/4}\right)\right]\right\}^2}, \quad (11)$$

with

$$s_b = \frac{g_b}{\pi \alpha_b q_{F,b}}. \quad (12)$$

Here, the Fermi momentum of the majority carriers $q_{F,b}$ is defined by

$$q_{F,b} = \left(\frac{6\pi^2 n_i}{g_b}\right)^{1/3} \quad (13)$$

and “ b ” is the index of the majority carrier band. For majority carriers [first term in (11)], an analytical approximation of the integral is difficult to find. The function $\tanh^{-1}(\dots)$ has an integrable singularity at $z = 2$ (double the amount of the Fermi momentum). The function in angular braces is bell-shaped with a maximum of 2 at $z = 0$. The shape of the total integrand strongly depends on the size of the parameter s_b . In terms of the normalized doping concentration (normalized by a_{ex}^{-3}), the parameter s_b becomes

$$s_b = \frac{g_b^{4/3}}{(6\pi^5)^{1/3} \alpha_b n_i^{1/3}}. \quad (14)$$

Therefore, one can assume $s_b \ll 1$ for very large n_i , which results in a pronounced maximum of the integrand in (11) at some $z < 1$. The function $\tanh^{-1}(\dots)$ can then be linearized and one obtains

$$\begin{aligned} \Delta_a^i &= -\frac{g_b}{3\pi} \left[\frac{1}{\sqrt{2s_b}} \frac{\delta_{ab}}{\alpha_b} + \frac{4}{(2s_b)^{3/2}} \frac{(1-\delta_{ab})}{\alpha_a} \right] \\ &= -\left[\frac{\delta_{ab}}{\sqrt{\alpha_b}} \frac{g_b^{1/3}}{3} \left(\frac{3}{4\pi}\right)^{1/6} n_i^{1/6} + \frac{(1-\delta_{ab}) \alpha_b^{3/2} 2\pi^{3/2}}{\alpha_a g_b \sqrt{3}} n_i^{1/2} \right]. \quad (15) \end{aligned}$$

The last expression is a reasonable analytical approximation of the zero-temperature limit of the ionic QPS in terms of band structure parameters (g_a, α_a), reproducing both the slope of the doping dependence at high concentrations and the different amounts of minority and majority carrier band shifts. Of particular interest is the ratio of the minority band shift in the material with the lower DOS $\Delta_{e,p\text{-type}}^i$ to the other shifts $\Delta_{h,p\text{-type}}^i, \Delta_{e,n\text{-type}}^i$, and $\Delta_{h,n\text{-type}}^i$. Since the DOS is given by $N_a = g_a \Lambda_a^{-3} = g_a / (4\pi \alpha_a \beta)^{3/2}$, it is

proportional to $g_a / \alpha_a^{3/2}$. Therefore, with the parameters of Table I, the DOS of p-type silicon is about three times smaller than that of n-type silicon. Using Eq. (15), one arrives at the following ratios:

$$\begin{aligned} \frac{\Delta_{e,p\text{-type}}^i}{\Delta_{h,p\text{-type}}^i} &= \frac{2\pi^2 \alpha_h^2}{\alpha_e g_h^{4/3}} \left(\frac{6}{\pi}\right)^{1/3} n_i^{1/6} \xrightarrow{n_i=10} 3.70 \quad (3.88), \\ \frac{\Delta_{e,p\text{-type}}^i}{\Delta_{e,n\text{-type}}^i} &= \frac{2\pi^2 \alpha_h^{3/2}}{\sqrt{\alpha_e} g_e^{1/3} g_h} \left(\frac{6}{\pi}\right)^{1/3} n_i^{1/6} \xrightarrow{n_i=10} 2.67 \quad (3.15), \quad (16) \\ \frac{\Delta_{e,p\text{-type}}^i}{\Delta_{h,n\text{-type}}^i} &= \frac{g_e}{g_h} \left(\frac{\alpha_h}{\alpha_e}\right)^{5/2} \xrightarrow{n_i=10} 2.49 \quad (2.02). \end{aligned}$$

The numbers at $n_i = 10$ correspond to a doping concentration of $1.94 \times 10^{20} \text{ cm}^{-3}$ and the numbers in braces are the numerical values extracted from the black solid curves in Figs. 1 and 2 for this density.

The total BGN in (a) n-type and (b) p-type silicon including the weak nonparabolicity effect is shown in Fig. 3. The temperature dependence starts to vanish above the Mott density, but a slight temperature effect re-emerges in p-type silicon at high densities ($n_i > 1$). It originates from the contribution of $\Delta_{e,p\text{-type}}^i$ [see Fig. 1(b)].

Figure 4 shows a comparison of the theory with optical BGN measurements in n-type silicon (a) and p-type silicon (b). Data points are from Aw *et al.*¹¹ (transmission), Balkanski *et al.*⁷ (transmission), Dumke¹³ (PL), del Alamo *et al.*¹⁰ (IR photo-response of solar cells), Lanyon *et al.*⁸ (IR photo-response of the transistor), Schmid⁹ (transmission), Yan *et al.*^{12,38} (photo-conductance), and Wagner¹⁵ (PL). Excellent agreement with the photo-conductance data by Yan *et al.* and the PL data by Wagner is found.

The black dashed curves are the result of the fit model for device simulation proposed in Ref. 29. In this model, screening had been fitted for the case of a neutral electron-hole plasma ($n_e = n_h$) to also cover the cases of high optical excitation and strong electrical injection in pn-junctions. Application to quasi-neutral regions yielded an acceptable agreement with the rather scattered experimental data, and no re-fitting was done for this case. As can be seen, the fit model underestimates the BGN in quasi-neutral regions of silicon by ~ 30 meV around a concentration of 10^{20} cm^{-3} . The reason is again the ionic QPS of minority carriers, which in a neutral plasma is significantly suppressed by the screening effect of the species that are minority carriers in quasi-neutral regions.

IV. RESULTS FOR $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$

To provide a more complete picture for InGaAs, first the free-carrier correlation energy Δ_c^i is depicted in Fig. 5. The majority carrier contribution is strongly temperature-dependent, but its magnitude remains small, in particular, in n-type InGaAs. The free-carrier QPS of minority carriers is in the order of 10 meV at $\sim 10^{19} \text{ cm}^{-3}$, the largest possible concentration of electrical active doping. This maximum concentration corresponds to $n_i = 100$ ($a_{\text{ex}}^{-3} = 1.2 \times 10^{17} \text{ cm}^{-3}$).

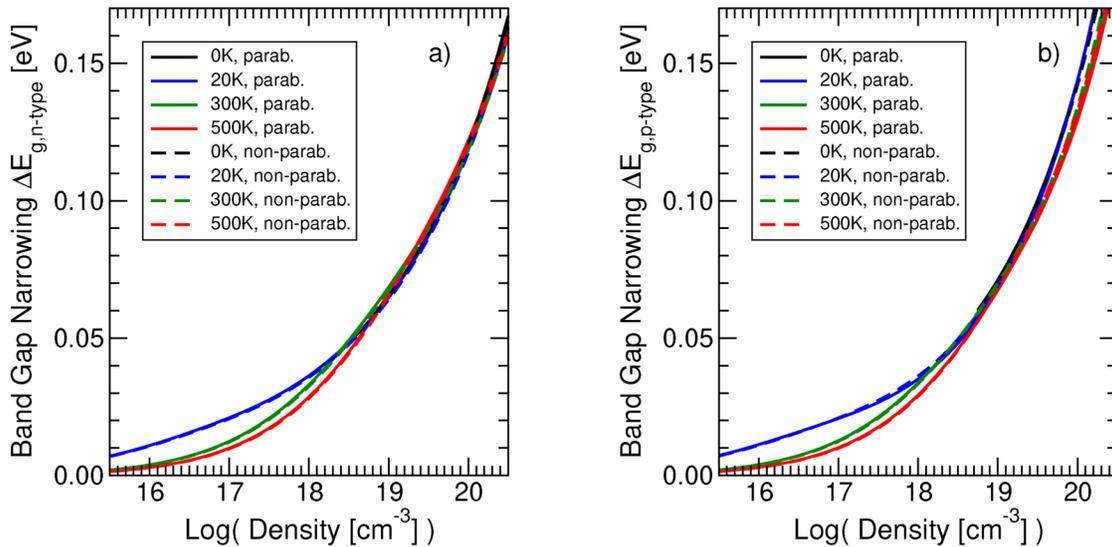


FIG. 3. Total BGN in (a) neutral n-type and (b) neutral p-type silicon as a function of activated doping concentration and temperature. Comparison of parabolic (solid lines) and nonparabolic conduction band (dashed lines).

Due to the relatively small CB DOS of InGaAs, the Fermi level can move deep into the band at high n-doping. The nonparabolic dispersion then has a strong effect on the ionic QPS. This is demonstrated in Figs. 6 and 7. For electrons in the n-type material [Fig. 6(a)], $\Delta_{e,n\text{-type}}^i$ doubles for nonparabolic CB at the highest density. As expected, in the p-type material, the nonparabolicity effect is much weaker [Fig. 6(b)]. In general, Δ_e^i is limited to 5 meV, and the temperature dependence is only relevant at intermediate concentrations.

The situation is completely different for holes. If they are minority carriers [Fig. 7(a)], $\Delta_{h,n\text{-type}}^i$ becomes giant in the low-T/high-density range, if a parabolic CB is used. It reaches 500 meV at $n_i = 100$, hundred times more than $\Delta_{h,p\text{-type}}^i$ [Fig. 7(b)]. Using Eq. (15), one finds

$$\frac{\Delta_{h,n\text{-type}}^i}{\Delta_{e,p\text{-type}}^i} = \frac{g_h}{g_e} \left(\frac{\alpha_e}{\alpha_h} \right)^{5/2} \xrightarrow{n_i=100} 141.5 \text{ (121.5)}. \quad (17)$$

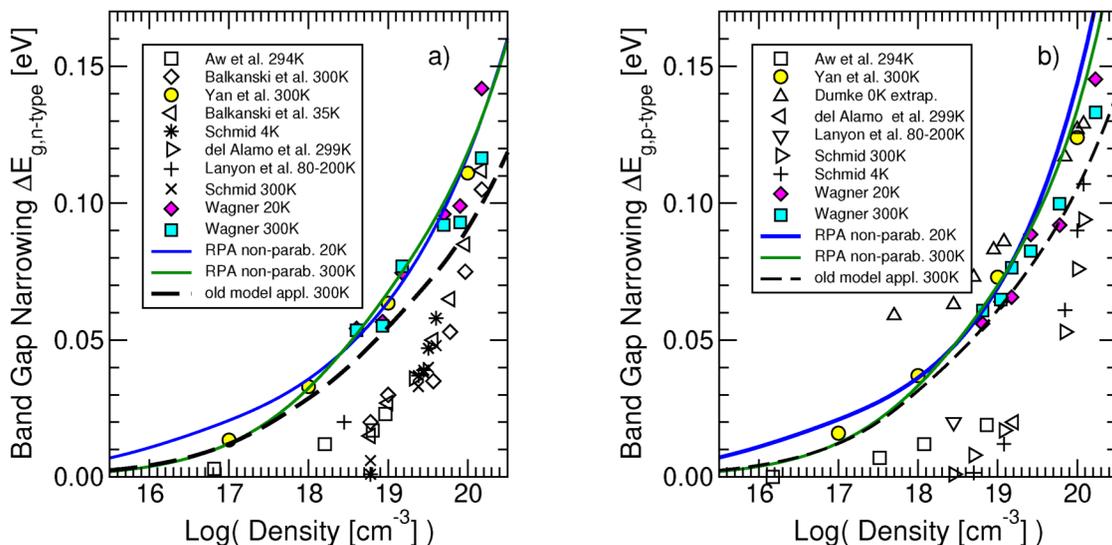


FIG. 4. Comparison with optical measurements of the BGN in (a) n-type and (b) p-type silicon.

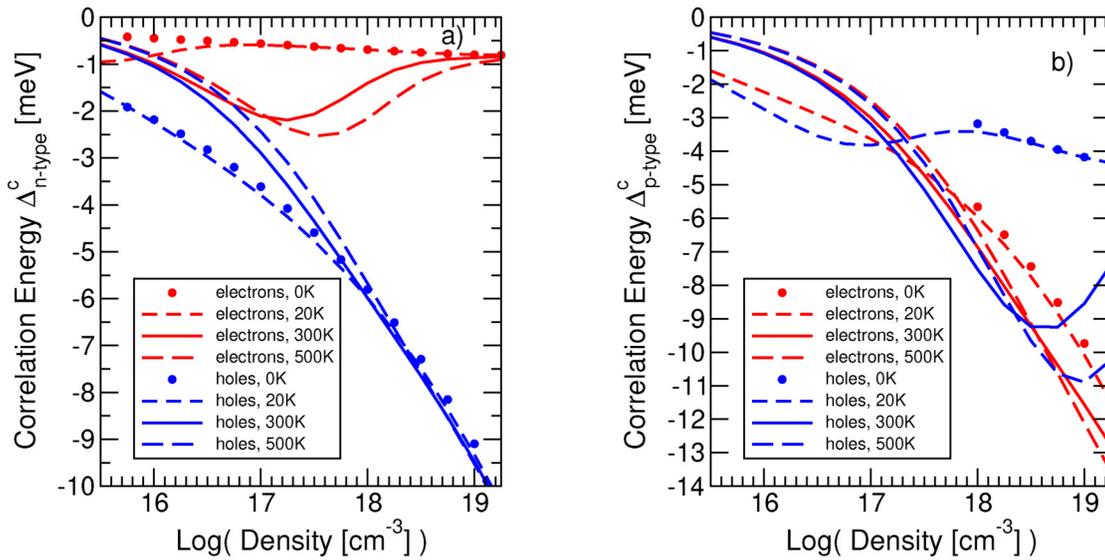


FIG. 5. Contribution of the free-carrier correlation energy to the average QPS in (a) neutral n-type and (b) neutral p-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ as a function of activated doping concentration and temperature.

The numerical result in braces was extracted from the black solid curve in Figs. 6(b) and 7(a), respectively. The ratio is equal to $N_h m_h / (N_e m_e)$, where the DOS ratio N_h / N_e is 25.7 and the mass ratio m_h / m_e 5.5. The crosses in Fig. (7)(a) are the result of the fit model proposed in Refs. 39 and 40 [also cited in the website

(Ref. 32)], which is not far from the present $T = 0$ K-curve for a parabolic CB.

With decreasing doping concentration and rising temperature the value of $\Delta_{h,n\text{-type}}^i$ quickly shrinks to “common” values. At room temperature, one finds 8 meV at $n_i = 10$ and 48 meV at $n_i = 100$.

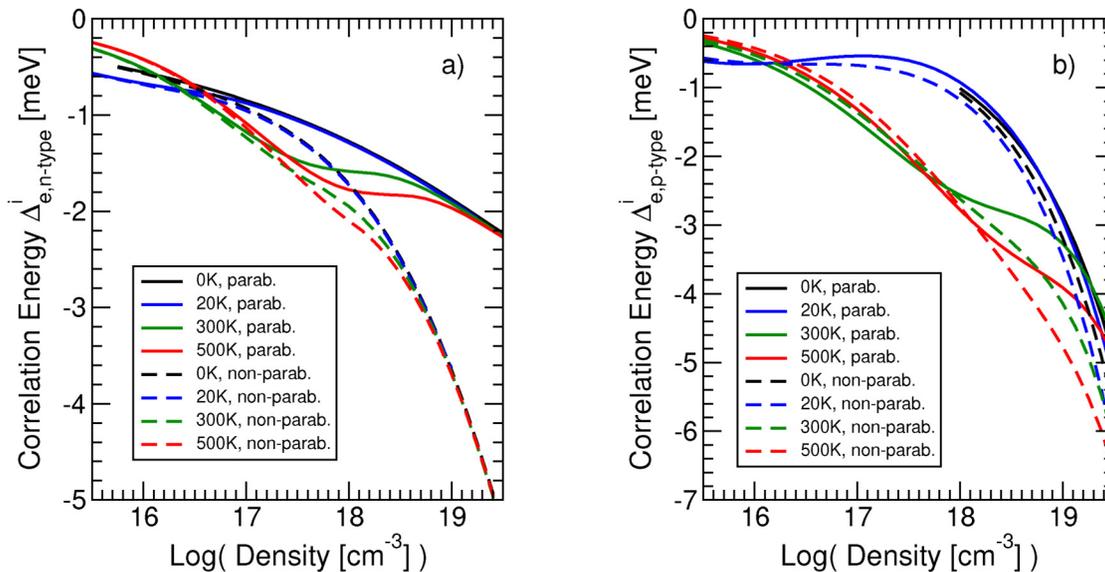


FIG. 6. (a) Electron-ion part $\Delta_{e,n\text{-type}}^i$ and (b) $\Delta_{e,p\text{-type}}^i$ of the average QPS for (a) neutral n-type and (b) neutral p-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ as a function of activated doping concentration and temperature.

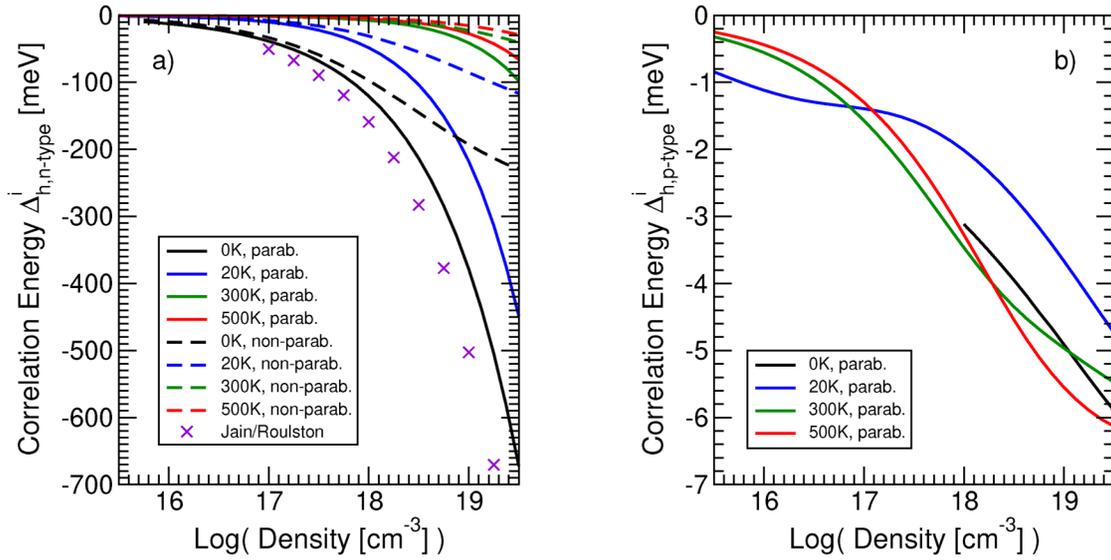


FIG. 7. (a) Hole-ion part $\Delta_{h,n-type}^i$ and (b) $\Delta_{h,p-type}^i$ of the average QPS for (a) neutral n-type and (b) neutral p-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ as a function of activated doping concentration and temperature. Crosses are the outcome from the fit model of Ref. 39.

According to Eq. (B4) $\lim_{\beta \rightarrow 0} \Delta_a^i = -\sqrt{2\pi\beta n_i}$, i.e., the value goes to zero and becomes independent of carrier type and doping species.

As can be seen in Fig. 7(a), nonparabolicity of the CB strongly decreases the value of $\Delta_{h,n-type}^i$ at low temperatures. An analytical explanation of the nonparabolicity effect is now given for $T = 0\text{K}$. In the important case of minority holes in the n-type material, one obtains instead of Eq. (11), second term,

$$\Delta_{h,n-type}^{i,np} = -\frac{16g_e}{3\pi^2\alpha_h} \int_0^\infty \frac{dz}{[z^2 + s_e \int_0^1 dt t^{\frac{1}{2}} H_e(t, z)]^2}, \quad (18)$$

with

$$s_e = \frac{g_e}{\pi\sqrt{\alpha_e\tilde{\zeta}_{e,np}}},$$

$$H_e(t, z) = 2[h_e(t, z) - h_e(t, -z)]$$

$$+ h_e(t, 0) \ln \left(\frac{[h_e(t, 0) - h_e(t, z)]^2}{[h_e(t, 0) - h_e(t, -z)]^2} \right), \quad (19)$$

$$h_e(t, z) = \sqrt{1 + 4\gamma_e\tilde{\zeta}_{e,np}(t+z)^2},$$

$$\tilde{\zeta}_{e,np} = \zeta_{e,np}(1 + \gamma_e\zeta_{e,np}).$$

which follows from Eq. (B9) inserting (A12). Note that the Fermi energy $\zeta_{e,np}$ is lower than $\zeta_{e,par}$ of the parabolic case

$$\zeta_{e,np} = \frac{1}{2\gamma_e} \left(\sqrt{1 + 4\gamma_e\zeta_{e,par}} - 1 \right) \approx \zeta_{e,par} \left(1 - \gamma_e\zeta_{e,par} \right). \quad (20)$$

The second term in the denominator of Eq. (18) is treated in the same way as before, i.e., approximating $H_e(t, z)$ for small z . This results in

$$H_e(t, z) \approx \frac{2z}{t} \sqrt{1 + 4\gamma_e\tilde{\zeta}_{e,np}t^2}, \quad (21)$$

where $\gamma_e\tilde{\zeta}_{e,np} \ll 1$ had to be assumed. The ionic correlation energy of minority holes (18) then becomes

$$\Delta_{h,n-type}^{i,np} = -\frac{\sqrt{2\pi}\alpha_e^{3/4}}{3\alpha_h\sqrt{g_e}} \left(\frac{\zeta_{e,np}}{1 + \gamma_e\zeta_{e,np}} \right)^{3/4}. \quad (22)$$

The ratio between nonparabolic and parabolic shift is computed using relation (20),

$$\frac{\Delta_{h,n-type}^{i,np}}{\Delta_{h,n-type}^{i,par}} = \left(\frac{\zeta_{e,np}}{\zeta_{e,par}} \right)^{3/4} \frac{1}{(1 + \gamma_e\zeta_{e,np})^{3/4}}$$

$$\approx \left(\frac{1 - \gamma_e\zeta_{e,par}}{1 + \gamma_e\zeta_{e,par}} \right)^{3/4}. \quad (23)$$

Expressing the Fermi energy by the doping concentration $\zeta_{e,par} = (6\pi^2\alpha_e^{3/2}n_i/g_e)^{2/3}$ gives

$$\frac{\Delta_{h,n-type}^{i,np}}{\Delta_{h,n-type}^{i,par}} = \left[\frac{1 - \gamma_e(6\pi^2\alpha_e^{3/2}n_i/g_e)^{2/3}}{1 + \gamma_e(6\pi^2\alpha_e^{3/2}n_i/g_e)^{2/3}} \right]^{3/4}$$

$$\xrightarrow{n_i=100} 0.41 \quad (0.42). \quad (24)$$

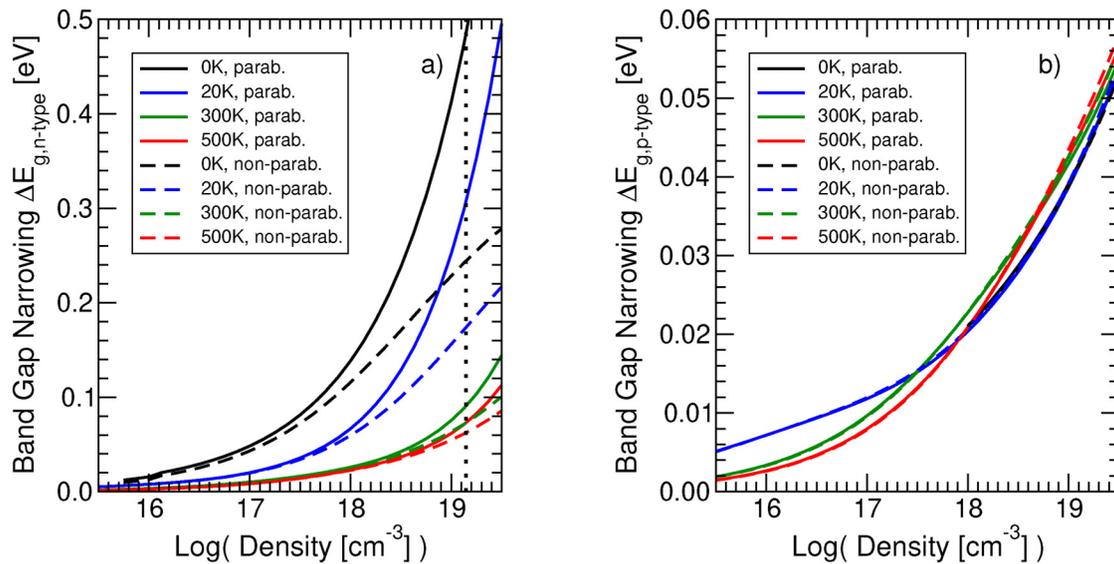


FIG. 8. Total BGN in (a) neutral n-type and (b) neutral p-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ as a function of activated doping concentration and temperature. Comparison of parabolic (solid lines) and nonparabolic conduction band (dashed lines). The dotted line indicates the published limit of activated n-doping.⁴¹

The number in braces is the numerical value extracted from the solid and dashed black curve in Fig. 7(a). Note that this 60%-reduction at $n_i = 100$ is caused by the factor multiplying γ_e , which is proportional to $(n_i/N_c)^{2/3}$, i.e., by the small DOS of the electrons in InGaAs.

The question arises how reliable this result is in view of the simplistic isotropic two-band dispersion model. Compared to silicon, the situation is more favorable here. The direct CB can be considered as isotropic, and the split-off hole band is more than 0.3 eV away from the VB edge. For $\Delta_{h,n\text{-type}}^i$, screening is solely due to electrons. As holes are minorities, the Fermi level is close to or in the CB; hence, the valence bands (VBs) are just probed over an energy interval $k_B T$ which becomes narrow at low temperatures. This makes nonparabolicity of the hole bands negligible. The remaining error is then caused by disregarding their warping. But this only affects the pre-factor as can be explicitly seen from the last line in Eq. (B9) showing that $\Delta_{h,n\text{-type}}^i(\beta \rightarrow \infty)$ is proportional to m_{hh} . Therefore, an error estimate can be obtained from the variation of m_{hh} over the warped iso-energy surface. Citing a three-parameter warping model for GaAs (Ref. 44 and references therein), one finds $(0.67 \dots 1.73)m_{hh}$, where m_{hh} is the average over azimuthal and polar angles. The error bounds for $\Delta_{h,n\text{-type}}^i$ are hence similar. However, provided one could perform the fourfold angle integration numerically (together with the two remaining integrations), the angle dependence would be efficiently averaged out. Based on this, effects of the real band structure cannot be expected to nullify the main findings of this chapter: the very large value of $\Delta_{h,n\text{-type}}^i$ at low temperatures and its significant reduction due to nonparabolicity of the CB and increasing temperature.

The total BGN in quasi-neutral $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ is shown in Fig. 8. In the n-type material, it is completely dominated by the

behavior of $\Delta_{h,n\text{-type}}^i$. However, the nonparabolicity of the CB decreases the zero-temperature value to 237 meV at $n_i = 100$ ($1.2 \times 10^{19} \text{ cm}^{-3}$). The corresponding value in the p-type material is 41 meV. At room temperature, the values are 69 and 45 meV, respectively. The dotted line indicates the published limit of electrical activation in the n-type material⁴¹ ($1.4 \times 10^{19} \text{ cm}^{-3}$). A similar value seems to hold for p-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$.⁴² Hence, the above BGN values can be roughly considered as upper limits too.

V. CONCLUSION

The analysis of BGN in quasi-neutral regions of Si and $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ based on the finite-temperature RPA formalism including band nonparabolicity led to the following conclusions.

In n-type silicon, the gap shrinks by 119 meV at an active donor concentration of $1 \times 10^{20} \text{ cm}^{-3}$. In p-type silicon, at the same density of active acceptors, the gap shrinks by 144 meV at 20 K and by 131 meV at 300 K, respectively. Both the difference in magnitude between doping types and the temperature dependence in the p-type material have the same origin—the ionic QPS of the minority carrier band (electrons) in the material with the smaller DOS (p-Si). A smaller DOS results in weaker screening.

In $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, this effect becomes extraordinary at low temperatures. The 26 times smaller DOS in the parabolic model and the five to six times smaller mass of electrons make screening in the n-type material so weak that the ionic QPS of the minority carrier band (holes) takes giant values (450 meV at $T = 0 \text{ K}$ at a donor density of $1.4 \times 10^{19} \text{ cm}^{-3}$, i.e., more than half of the bandgap). However, this result is due to the assumption of a parabolic CB. The strong degeneracy enforces to include nonparabolicity which lowers the above value to 204 meV. A nonparabolic band

effectively implies a larger DOS and a larger average mass to be used in a parabolic model, thus more screening. However, with increasing temperature, the ionic QPS of minority holes quickly shrinks to small values (29 meV at $T = 300$ K) due to the $\sim T^{-1/2}$ -dependence in the high-T limit. For the BGN, this means a limit of 72 meV at room temperature. In p-type $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, the DOS is large and screening is strong. The contribution from the ionic QPS of minority electrons is just ~ 4 meV, and the total BGN amounts to 46 meV. This is accompanied by a very weak temperature dependence (43 meV at 20 K).

The nonparabolicity of the CB must be taken into account in all cases where the DOS/mass of holes is much larger than that of electrons, as often encountered in III-V materials. The paper has demonstrated that this is possible—numerically for all temperatures, and even analytically for the ionic QPS at $T = 0$ K, if a simple isotropic two-band model is used, where the multi-valley CB and the heavy and light-hole bands are incorporated by DOS effective masses. There is hope that a fully analytical $T = 0$ K-limit can also be found for the free-carrier correlation energies. Padé interpolation^{1,29} then would enable a BGN model as a function of doping and temperature essentially free of material-specific fit parameters.

A serious problem is the impact of all the band structure effects, which are not considered in this work, like anisotropy, warping, contribution of split-off band, DOS tails, and polaronic mass enhancement. Even though a full numerical treatment based on an elaborate band structure model seems to be out of reach even in the rigid-shift approximation of BGN, some aspects could be tested on a smaller scale, e.g., in the screening function.

ACKNOWLEDGMENTS

I am obliged to Dr. Pietro Altermatt (Trina Solar) for making me aware of Refs. 12 and 38.

APPENDIX A: RPA DIELECTRIC FUNCTION FOR NONPARABOLIC BANDS

This appendix provides the derivation of the full-RPA dielectric function for nonparabolic bands, its static limit, as well as the high-T/low-density and low-T/high-density limits of the static form, respectively.

1. Frequency-dependent form at arbitrary density and temperature

With the Fourier transform of the bare Coulomb potential $v(q) = 8\pi a_{\text{ex}} R y_{\text{ex}} / q^2$ and the nonparabolic free dispersion (10), the RPA dielectric function (9) takes the form

$$\epsilon(q, \Omega_\nu) = 1 + \frac{1}{(\pi q)^2} \sum_a g_a \int d^3k f_a(k) \times \left[\frac{1}{E_a^{0,\text{np}}(k+q) - E_a^{0,\text{np}}(k) - \hbar\Omega_\nu} - \frac{1}{E_a^{0,\text{np}}(k) - E_a^{0,\text{np}}(k-q) - \hbar\Omega_\nu} \right], \quad (\text{A1})$$

with

$$E_a^{0,\text{np}}(k \pm q) = \frac{1}{2\gamma_a} \left[\sqrt{4\gamma_a \alpha_a (k^2 + q^2 \pm 2kq \cos \Theta)} - 1 \right].$$

Energies are normalized by $R y_{\text{ex}}$, wave numbers by a_{ex}^{-1} . The k_z -axis is aligned with \mathbf{q} , hence $\cos \Theta = \mathbf{k} \cdot \mathbf{q} / (|\mathbf{k}||\mathbf{q}|)$. Setting $\tau(t) = 8\gamma_a \alpha_a k q t$ with $t = \cos \Theta$, one obtains

$$\epsilon(q, \Omega_\nu) = 1 + \frac{2}{\pi q^2} \sum_a g_a \int_0^\infty dk k^2 f_a(k) \frac{2\gamma_a}{\tau(1)} \int_{-\tau(1)}^{\tau(1)} d\tau \times \left(\frac{1}{\sqrt{\frac{\tau(1)}{2} \left(\frac{k}{q} + \frac{q}{k} \right) + \tau + 1} - \sqrt{\frac{\tau(1)k}{2q} + 1} - 2\gamma_a \hbar \Omega_\nu} + \frac{1}{\sqrt{\frac{\tau(1)}{2} \left(\frac{k}{q} + \frac{q}{k} \right) - \tau + 1} - \sqrt{\frac{\tau(1)k}{2q} + 1} + 2\gamma_a \hbar \Omega_\nu} \right). \quad (\text{A2})$$

The τ -integral

$$I_\tau = \int_{-\tau(1)}^{\tau(1)} d\tau \left(\frac{1}{\sqrt{a + \tau} - b_+} + \frac{1}{\sqrt{a - \tau} - b_-} \right), \quad (\text{A3})$$

where

$$a = \frac{\tau(1)}{2} \left(\frac{k}{q} + \frac{q}{k} \right) + 1,$$

$$b_\pm = \sqrt{\frac{\tau(1)k}{2q} + 1} \pm 2\gamma_a \hbar \Omega_\nu$$

can be solved using

$$\int d\tau \frac{1}{\sqrt{a + \tau} - b} = 2\sqrt{a + \tau} + 2b \ln(b - \sqrt{a + \tau}),$$

which gives

$$I_\tau = 4 \left(\sqrt{a + \tau(1)} - \sqrt{a - \tau(1)} \right) + 2b_+ \ln \frac{b_+ - \sqrt{a + \tau(1)}}{b_+ - \sqrt{a - \tau(1)}} + 2b_- \ln \frac{b_- - \sqrt{a + \tau(1)}}{b_- - \sqrt{a - \tau(1)}}. \quad (\text{A4})$$

Inserting into Eq. (A2) yields

$$\epsilon(q, \Omega_\nu) = 1 + \frac{1}{\pi q^3} \sum_a \frac{g_a}{\alpha_a} \int_{-\infty}^\infty dk k f_a(k) \{ 2h_a(k, q) + [h_a(k, 0) + g_{v,a}] \ln [h_a(k, 0) - h_a(k, q) + g_{v,a}] + [h_a(k, 0) - g_{v,a}] \ln [h_a(k, 0) - h_a(k, q) - g_{v,a}] \}, \quad (\text{A5})$$

with the auxiliary functions

$$h_a(k, q) = \sqrt{1 + 4\gamma_a \alpha_a (k + q)^2},$$

$$g_{v,a} = 2\gamma_a \hbar \Omega_v = \frac{4\pi\gamma_a v}{\beta} i.$$

Note that $h_a(-k, q) = h_a(k, -q)$ and $h_a(k, 0) = h_a(-k, 0)$. In the case of parabolic bands, the k -integral in Eq. (A5) is integrated by parts to remove the logarithmic singularities.¹ This is impossible for nonparabolic bands at $T > 0$. Therefore, the final frequency-dependent form for numerical integration is

$$\epsilon(q, \Omega_v) = 1 + \frac{2}{\pi q^3} \sum_a \frac{g_a}{\alpha_a} \int_{-\infty}^{\infty} dk \kappa f_a(k) \left\{ h_a(k, q) + \frac{1}{2} [h_a(k, 0) + g_{v,a}] \ln [(h_a(k, 0) - h_a(k, q))^2 + |g_{v,a}|^2] - |g_{v,a}| \arctg \left[\frac{|g_{v,a}|}{h_a(k, 0) - h_a(k, q)} \right] \right\}. \quad (\text{A6})$$

2. Static form at arbitrary density and temperature

For the ionic part of the correlation energy, only the static limit $g_{v,a} \rightarrow 0$ of Eq. (A6) is needed. It can be written as

$$\epsilon(q, 0) = 1 + \frac{1}{\pi q^3} \sum_a \frac{g_a}{\alpha_a^2} \int_0^{\infty} d\kappa \kappa f_a(\kappa) \times \left\{ 2[h_a(\kappa, q) - h_a(\kappa, -q)] + h_a(\kappa, 0) \ln \frac{[h_a(\kappa, 0) - h_a(\kappa, q)]^2}{[h_a(\kappa, 0) - h_a(\kappa, -q)]^2} \right\}, \quad (\text{A7})$$

where

$$h_a(\kappa, q) = \sqrt{1 + 4\gamma_a (\kappa + \sqrt{\alpha_a q})^2}, \quad (\text{A8})$$

$$f_a(\kappa) = \left\{ 1 + e^{\beta[h_a(\kappa, 0) - 1]/2\gamma_a - \beta\zeta_a} \right\}^{-1} \quad (\text{A9})$$

have to be used.

3. Debye limit and $T = 0$ K-limit of the static form

The Debye limit⁴³ of the static RPA dielectric function follows from Eq. (A7) by $q \rightarrow 0$ and using Boltzmann statistics (high-temperature, long-wavelength limit),

$$\epsilon_D(q, 0) = 1 + \frac{2}{\pi q^2} \sum_a \frac{g_a e^{\beta\zeta_a}}{\alpha_a^{3/2}} \int_0^{\infty} d\kappa e^{-\beta[h_a(\kappa, 0) - 1]/2\gamma_a} \times \left[\frac{4\gamma_a \kappa^2}{h_a(\kappa, 0)} + h_a(\kappa, 0) \right]$$

$$= 1 + \frac{1}{\pi q^2} \sum_a \frac{g_a e^{\beta\zeta_a}}{\alpha_a^{3/2}} \int_0^{\infty} dt e^{-\beta t} \times \left[4\gamma_a \sqrt{t(1 + \gamma_a t)} + \frac{(2\gamma_a t + 1)^2}{\sqrt{t(1 + \gamma_a t)}} \right].$$

This is exactly the same as

$$\epsilon_D(q, 0) = 1 + \frac{8\pi}{q^2} \sum_a \frac{\partial n_a}{\partial \zeta_a}, \quad (\text{A10})$$

which is the expression in the case of parabolic bands. Therefore, the form of the Debye limit is independent of the band dispersion as expected.

To obtain the $T = 0$ K-limit, the κ -integral in Eq. (A7) is integrated by parts where the derivative of the distribution function with respect to κ becomes a delta function

$$\frac{\partial f_a}{\partial \kappa} = -\delta \left[\kappa - \sqrt{\zeta_a(1 + \gamma_a \zeta_a)} \right]. \quad (\text{A11})$$

Then,

$$\epsilon_{T=0}(q, 0) = 1 + \frac{1}{\pi q^3} \sum_a \frac{g_a}{\alpha_a^2} \int_0^{\sqrt{\zeta_a(1 + \gamma_a \zeta_a)}} d\kappa \kappa \times \left\{ 2[h_a(\kappa, q) - h_a(\kappa, -q)] + h_a(\kappa, 0) \ln \frac{[h_a(\kappa, 0) - h_a(\kappa, q)]^2}{[h_a(\kappa, 0) - h_a(\kappa, -q)]^2} \right\} \Theta(\zeta_a). \quad (\text{A12})$$

Only the majority carriers of the quasi-neutral regions contribute ($\zeta_a > 0$), whereas the minority carriers are frozen out.

From (A12) the parabolic limit ($\gamma_a \rightarrow 0$) is easily retrieved,

$$\lim_{\gamma_a \rightarrow 0} \epsilon_{T=0}(q, 0) = 1 + \frac{1}{\pi q^3} \sum_a \frac{g_a}{\alpha_a^2} \times \left\{ \sqrt{\alpha_a \zeta_a} q + (\zeta_a - \alpha_a q^2/4) \tanh^{-1} \left(\frac{\sqrt{\alpha_a \zeta_a} q}{\zeta_a + \alpha_a q^2/4} \right) \right\} \Theta(\zeta_a). \quad (\text{A13})$$

APPENDIX B: CORRELATION ENERGY OF ION-CARRIER INTERACTION

This appendix provides the derivation of the full-RPA correlation energy of ion-carrier interaction for nonparabolic bands in the high- T /low-density and low- T /high-density limits, respectively.

1. Arbitrary density and temperature

The correlation energy of ion-carrier interaction is calculated by Eq. (7). After normalization and integration over angles, it reads

$$\Delta_a^i = -n_i \frac{2}{\pi} \left(\frac{\partial n_a}{\partial \zeta_a} \right)^{-1} \int_0^\infty dq \frac{1}{\epsilon^2(q, 0)} \frac{\partial \epsilon(q, 0)}{\partial \zeta_a}. \quad (B1)$$

For the derivative of the dielectric function (A7) with respect to the Fermi energy, one has to apply

$$\frac{\partial f_a}{\partial \zeta_a} = -\frac{\partial f_a}{\partial \kappa} \frac{h_a(\kappa, 0)}{2\kappa} = \beta f_a(\kappa) [1 - f_a(\kappa)] \quad (B2)$$

under the κ -integral.

2. Debye limit and $T = 0$ K-limit

In the high-T/low-density regime, the effect of nonparabolicity is negligible, and one obtains from Eq. (7) with the dielectric function (A10), the well-known result^{1,29}

$$\Delta_a^i = -n_i \frac{\sqrt{2\pi\beta}}{\sqrt{n_e + n_h}}. \quad (B3)$$

In quasi-neutral regions $n_e + n_h = n_i$, which simplifies Eq. (B3) to

$$\Delta_a^i = -\sqrt{2\pi\beta} n_i. \quad (B4)$$

At $T = 0$ K, one must distinguish between majority and minority bands. The derivative of the densities with respect to the Fermi energy becomes

$$\lim_{\beta \rightarrow \infty} \frac{\partial n_a}{\partial \zeta_a} = \frac{g_a}{4\pi^2 \alpha_a^3} \times \begin{cases} \sqrt{\zeta_a} (1 + 2\gamma_a \zeta_a) \Theta(\zeta_a) & \text{majority band,} \\ \frac{\sqrt{\pi} \exp(\beta \zeta_a)}{\sqrt{\beta}} \Big|_{\beta \rightarrow \infty} \Theta(-\zeta_a) & \text{minority band,} \end{cases} \quad (B5)$$

where the abbreviation $\tilde{\zeta}_a = \zeta_a (1 + \gamma_a \zeta_a)$ was introduced. For the derivative of the static dielectric function with respect to the Fermi energy $\partial \epsilon(q, 0) / \partial \zeta_a$ in the limit $\beta \rightarrow \infty$, one needs

$$\lim_{\beta \rightarrow \infty} \frac{\partial f_a}{\partial \zeta_a} = -\lim_{\beta \rightarrow \infty} \frac{h(\kappa, 0)}{2\kappa} \frac{\partial f_a}{\partial \kappa} = \begin{cases} \frac{h(\kappa, 0)}{2\kappa} \delta\left(\kappa - \sqrt{\zeta_a}\right) \Theta(\zeta_a) & \text{majority band,} \\ \beta e^{-\beta(\kappa^2 - \zeta_a)} \Big|_{\beta \rightarrow \infty} \Theta(-\zeta_a) & \text{minority band,} \end{cases} \quad (B6)$$

which follows from Eq. (B2) and neglecting nonparabolicity for the

minority carriers. Using this in Eq. (A7) leads to

$$\begin{aligned} \lim_{\beta \rightarrow \infty} \frac{\partial \epsilon(q, 0)}{\partial \zeta_a} &= \frac{1}{\pi q^3} \frac{g_a}{\alpha_a^2} h_a^2(\sqrt{\zeta_a}, 0) \\ &\times \left\{ \frac{h_a(\sqrt{\zeta_a}, q) - h_a(\sqrt{\zeta_a}, -q)}{h_a(\sqrt{\zeta_a}, 0)} \right. \\ &+ \frac{1}{2} \ln \left(\frac{[h_a(\sqrt{\zeta_a}, 0) - h_a(\sqrt{\zeta_a}, q)]^2}{[h_a(\sqrt{\zeta_a}, 0) - h_a(\sqrt{\zeta_a}, -q)]^2} \right) \left. \right\} \Theta(\zeta_a) \\ &+ \lim_{\beta \rightarrow \infty} \frac{1}{\pi q^3} \frac{g_a}{\alpha_a^2} \beta e^{\beta \zeta_a} \int_0^\infty d\kappa \kappa e^{-\beta \kappa^2} \\ &\times \ln \left(\frac{[\kappa^2 - (\kappa + \sqrt{\alpha_a} q)^2]^2}{[\kappa^2 - (\kappa - \sqrt{\alpha_a} q)^2]^2} \right) \Theta(-\zeta_a), \end{aligned} \quad (B7)$$

where in the term $\sim \Theta(-\zeta_a)$, nonparabolicity was again neglected. The κ -integral can be calculated analytically,

$$\begin{aligned} \lim_{\beta \rightarrow \infty} \beta e^{\beta \zeta_a} \int_0^\infty d\kappa \kappa e^{-\beta \kappa^2} \ln \left(\frac{[\kappa^2 - (\kappa + \sqrt{\alpha_a} q)^2]^2}{[\kappa^2 - (\kappa - \sqrt{\alpha_a} q)^2]^2} \right) \\ = \lim_{\beta \rightarrow \infty} 2\beta e^{\beta \zeta_a} \int_0^\infty dt t e^{-t^2} \tanh^{-1} \left(\frac{4t}{\sqrt{\alpha_a} \beta q} \right) \\ = \lim_{\beta \rightarrow \infty} \frac{2\sqrt{\pi} e^{\beta \zeta_a}}{\sqrt{\alpha_a} \beta q}. \end{aligned} \quad (B8)$$

For the second line, the relation $\ln(|x+1|/|x-1|) = 2 \tanh^{-1}(1/x)$ was used. Inserting Eq. (B7) with (B8) and Eq. (B5) into Eq. (B1), the correlation energy of ion-carrier interaction in the $T = 0$ K-limit finally becomes

$$\begin{aligned} \lim_{\beta \rightarrow \infty} \Delta_a^i &= -n_i \frac{8}{\sqrt{\alpha_a} \tilde{\zeta}_a} \int_0^\infty dq \frac{\Theta(\zeta_a)}{q^3 \epsilon_{T=0}^2(q, 0)} \\ &\times \left\{ [h_a(\sqrt{\zeta_a}, q) - h_a(\sqrt{\zeta_a}, -q)] \right. \\ &+ \frac{1}{2} (1 + 2\gamma_a \zeta_a) \ln \left(\frac{[1 + 2\gamma_a \zeta_a - h_a(\sqrt{\zeta_a}, q)]^2}{[1 + 2\gamma_a \zeta_a - h_a(\sqrt{\zeta_a}, -q)]^2} \right) \left. \right\} \\ &- n_i \frac{32}{\alpha_a} \int_0^\infty dq \frac{\Theta(-\zeta_a)}{q^4 \epsilon_{T=0}^2(q, 0)}, \end{aligned} \quad (B9)$$

where the static dielectric function $\epsilon_{T=0}(q, 0)$ is defined in Eq. (A12) and the function h_a in Eq. (A8).

From (B9), the parabolic limit ($\gamma_a \rightarrow 0$) is easily retrieved

$$\lim_{\gamma_a \rightarrow 0} \Delta_a^i = -8n_i \int_0^\infty \frac{dq}{q^3} \epsilon_{T=0}^{-2}(q, 0) \times \left\{ \frac{1}{\sqrt{\alpha_a \zeta_a}} \tanh^{-1} \left(\frac{\sqrt{\alpha_a \zeta_a} q}{\zeta_a + \alpha_a q^2/4} \right) \Theta(\zeta_a) + \frac{4}{\alpha_a q} \Theta(-\zeta_a) \right\}. \quad (\text{B10})$$

DATA AVAILABILITY

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

REFERENCES

- ¹R. Zimmermann, *Many Particle Theory of Highly Excited Semiconductors* (BSB Teubner Verlagsgesellschaft, Leipzig, 1988).
- ²H. Stolz and R. Zimmermann, *Phys. Stat. Sol. B* **94**, 135 (1979).
- ³M. Rösler, F. Thuselt, and R. Zimmermann, *Phys. Stat. Sol. B* **118**, 303 (1983).
- ⁴M. Rösler, R. Zimmermann, and W. Richert, *Phys. Stat. Sol. B* **121**, 609 (1984).
- ⁵R. Zimmermann and H. Stolz, *Phys. Stat. Sol. B* **131**, 151 (1985).
- ⁶B. E. Sernelius, *Phys. Rev. B* **33**(12), 8582–8586 (1986).
- ⁷M. Balkanski, A. Aziza, and E. Amzallag, *Phys. Stat. Sol. B* **31**, 323 (1969).
- ⁸H. P. D. Lanyon, A. K. McCurdy, and R. A. Tuft, *Proceedings of the 13th IEEE Photovoltaic Specialist Conference, Washington D.C.* (IEEE New York, 1978), p. 60.
- ⁹P. E. Schmid, *Phys. Rev. B* **23**, 5531 (1981).
- ¹⁰J. A. del Alamo, R. M. Swanson, and A. Lietoila, *Solid State Electron.* **26**, 483 (1983).
- ¹¹S. E. Aw, H. S. Tan, and C. K. Ong, *J. Phys. C* **3**, 8213 (1991).
- ¹²D. Yan and A. Cuevas, *J. Appl. Phys.* **116**, 194505 (2014).
- ¹³W. P. Dumke, *J. Appl. Phys.* **54**, 3200 (1983).
- ¹⁴J. Wagner, *Phys. Rev. B* **32**, 1323 (1985).
- ¹⁵J. Wagner, *Solid State Electron.* **30**, 1117 (1987).
- ¹⁶J. Wagner and J. A. del Alamo, *J. Appl. Phys.* **63**, 425 (1988).
- ¹⁷H. E. J. Wulms, *IEEE J. Solid-State Circuits* **12**, 143 (1977).
- ¹⁸D. D. Tang, *IEEE Trans. Electron Devices* **27**, 563 (1980).
- ¹⁹A. W. Wieder, *IEEE Trans. Electron Devices* **27**, 1402 (1980).
- ²⁰G. E. Possin, M. S. Adler, and B. J. Baliga, *IEEE Trans. Electron Devices* **31**, 3 (1984).
- ²¹A. Neugroschel, J. S. Wang, and F. A. Lindholm, *IEEE Electron Device Lett.* **6**, 253 (1985).
- ²²J. A. del Alamo and R. M. Swanson, *IEEE Trans. Electron Devices* **34**, 1580 (1987).
- ²³G. D. Mahan, *J. Appl. Phys.* **51**, 2634 (1980).
- ²⁴B. I. Lundqvist, *Phys. Kond. Mat.* **6**, 193 (1967), 206.
- ²⁵K.-F. Berggren and B. E. Sernelius, *Phys. Rev. B* **24**, 1971 (1981).
- ²⁶A. Selloni and S. T. Pantelides, *Phys. Rev. Lett.* **49**, 586 (1982).
- ²⁷K.-F. Berggren and B. E. Sernelius, *Phys. Rev. B* **29**, 5575 (1984).
- ²⁸S. T. Pantelides, A. Selloni, and R. Car, *Solid State Electron.* **28**, 17 (1985).
- ²⁹A. Schenk, *J. Appl. Phys.* **84**(7), 3684 (1998).
- ³⁰H. E. deWitt, *J. Nucl. Energy, Part C Plasma Phys.* **2**, 27 (1961).
- ³¹P. Vashishta and R. K. Kalia, *Phys. Rev. B* **25**, 6492 (1982).
- ³²See <http://www.ioffe.ru/SVA/NSM/Semicond/InGaAs> for BGN of InGaAs as function of donor concentration taken from Ref. 39.
- ³³I. R. Vurgaftman, R. J. Meyer, and L. R. Ram-Mohan, *J. Appl. Phys.* **89**(11), 5815 (2001).
- ³⁴C. Hermann and T. P. Pearsall, *Appl. Phys. Lett.* **38**, 450 (1981).
- ³⁵K. Alavi, R. L. Aggarwal, and S. H. Groves, *Phys. Rev. B* **21**, 1311 (1980).
- ³⁶A. Ziegler, M. Frey, L. Smith, and M. Luisier, *IEEE Trans. Electron Devices* **63**390(5), 2050–2056 (2016).
- ³⁷N. F. Mott, *Phil. Mag.* **6**(62), 287–309 (1961).
- ³⁸D. Yan and A. Cuevas, *J. Appl. Phys.* **114**, 044508 (2013).
- ³⁹S. C. Jain, J. M. McGregor, and D. J. Roulston, *Appl. Phys. Lett.* **68**(7), 3747 (1990).
- ⁴⁰S. C. Jain and D. J. Roulston, *Solid-State Electron.* **34**(5), 453–465 (1991).
- ⁴¹A. G. Lind, H. L. Aldridge, C. C. Bomberger, C. Hatem, J. M. O. Zide, and K. S. Jones, *ECS Trans.* **66**(7), 23 (2015).
- ⁴²B. 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).
- ⁴³P. P. Debye and E. M. Conwell, *Phys. Rev.* **93**, 693 (1954).
- ⁴⁴N. Nintunze and M. A. Osman, *Semicond. Sci. Technol.* **10**, 11 (1995).