Bandgap narrowing in zincblende III–V semiconductors: Finite-temperature full random-phase approximation and general analytical model ⊕

A. Schenk ✉ iD

Check for updates

CrossMark

View Online   Export Citation
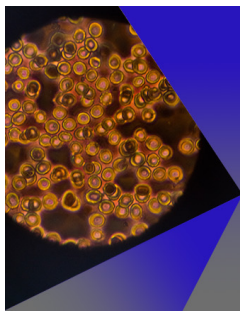
Bandgap narrowing in zincblende III–V semiconductors: Finite-temperature full random-phase approximation and general analytical model (EP)

View Online   Export Citation   CrossMark

A. Schenk[a] (iD)

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 zincblende III–V semiconductors is calculated in a finite-temperature full Random-Phase Approximation (RPA) formalism based on an isotropic dispersion model. The cases of n-type and p-type quasi-neutral regions and the case of a neutral electron–hole plasma are elaborated for the technologically important materials GaAs, AlAs, InAs, GaP, InP, GaSb, InSb, zb-GaN, zb-InN, $Al_{0.3}Ga_{0.7}As$ $GaAs_{0.5}Sb_{0.5}$, $InP_{0.69}Sb_{0.31}$, $InAs_{0.4}P_{0.6}$, $InAs_{0.4}Sb_{0.6}$, $In_{0.52}Al_{0.48}As$ $In_{0.49}Ga_{0.51}P$, $In_{0.53}Ga_{0.47}As$ $In_{0.5}Ga_{0.5}Sb$, and zb-$Ga_{0.5}In_{0.5}N$ (60 cases). In quasi-neutral regions, the correlation energy of the interaction between carriers and ionized dopants adds two terms to the total BGN. At low temperatures, inefficient screening makes the hole term dominant in n-type materials with a large ratio of the valence band to the conduction band (CB) density-of-states. The inclusion of the CB nonparabolicity is decisive here, as it prevents a diverging BGN at high concentrations. For all 60 cases, the BGN is evaluated in the temperature range from 0 to 500 K. A strong temperature dependence over the whole density range is observed in all direct n-type materials. Otherwise, the temperature dependence quickly ceases with increasing density. An analytical model of BGN without material-dependent free fit parameters is derived and compared with the full-RPA results.

I. INTRODUCTION

Bandgap narrowing (BGN) in semiconductor devices is a many-body effect that occurs in heavily doped regions and under strong optical or electrical excitation. In the first case, as in bipolar transistors (BTs) and field-effect transistors (FETs), the activated doping is compensated by a one-component plasma either in the emitter of a BT or in the source/drain of a FET. In the second case, a two-component plasma is generated, as in optically and electrically pumped laser diodes. Device characteristics can be altered by BGN, examples being the reduced current gain in a BT due to the increased emitter minority concentration or the density dependence of the laser wavelength. Modeling of BGN is facilitated by the condition of charge neutrality, which holds in quasi-neutral regions (negligible field strength) and in a neutral plasma (background doping is small compared to the plasma density). The physical origin of BGN is the many-body Coulomb interaction between free charge carriers as well as between charge carriers and ionized dopants. The resulting total self-energy shifts the band edges such that the bandgap shrinks. Four contributions have to be considered in a neutral electron–hole (eh) plasma—the exchange and correlation energies of electrons and holes, whereas five components contribute in quasi-neutral regions—the exchange energy of the majority carriers, the free-carrier correlation energies of the majority and minority carriers, and the ionic correlation energies of the majority and minority carriers. In this paper, all self-energies are calculated in the finite-temperature full Random-Phase Approximation (RPA).[1–6] RPA is increasingly accurate at higher densities where BGN becomes significant for device operation, and it is the only method to cover the whole temperature range. In n-type III–V zincblende semiconductors with small electron effective masses, the temperature dependence becomes extreme even at high densities. This is because

the ionic correlation energy of holes is the largest contributor to the total BGN here.[7] The shift of the minority carrier band is a consequence of the lowering of the energy of the generated/annihilated minority carrier due to Coulomb attraction by the mobile majority carriers and Coulomb repulsion by the sub-system of immobile ionized dopants.[8]

The basic RPA theory as well as results for the eh-plasma in various materials can be found in Ref. 1. The extrinsic semiconductor with a random distribution of dopants was treated in Ref. 4 for $T = 0$ K. Earlier calculations of the self-energy of the carrier-dopant interaction in the $T = 0$ limit were performed with the Hartree–Fock variational method for donors distributed on a regular sub-lattice[9] applying the single plasmon pole (SPP) approximation[10] and with second-order perturbation theory for randomly distributed dopants based on RPA screening.[11] The multi-valley case had been discussed for both dopant arrangements in Refs. 12–14.

Measurements of BGN are accompanied by an intricate analysis. Three types of experiments have been developed and mostly applied to silicon due to its outstanding role in technology: Absorption experiments,[15–20] photoluminescence (PL) experiments,[21–24] and electrical measurements of BTs.[25–30] In absorption experiments, phonon-assisted transitions are present besides the band-band transitions; therefore, the phonon energies must be known as well as the exact position of the Fermi energy. The evaluation of PL experiments is difficult because of the actual initial states and the weak intensity of the spectrum. The electrical method requires the measurement of collector current in the BT as a function of the emitter-base voltage, sheet resistance below the emitter, and minority carrier mobility in the base. An analytical model is then used to calculate the BGN from these quantities. BGN data for III–V materials are rare; for most of the 20 materials considered in this paper, they do not exist. Olego and Cardona[31] performed PL measurements of the BGN in p-type GaAs at 2.1 K for concentrations of $(1.6, 4, 9) \times 10^{19}$ cm^{-3}. The same material was analyzed by Titkov et al.[32] by PL spectroscopy at 4.2 K in the density range 3×10^{18}–4×10^{19} cm^{-3}. Tiwari and Wright[33] used hetero-structure BTs to measure the BGN in p-type GaAs at room temperature in the density range $(10^{18}$–$2 \times 10^{19})$ cm^{-3}. Semikolenova et al. performed absorption measurements of the BGN of n-type InAs at room temperature for tellurium concentrations between 2×10^{17} and 3×10^{18} cm^{-3}. PL spectra of p-type GaSb with densities between 3×10^{18} and 4×10^{19} cm^{-3} at 4.2 K were published by Titkov et al.[32] The BGN in n-type InP at 300 K was studied by absorption and PL experiments by Bugajski and Lewandowski,[35] covering the concentration range 10^{16}–10^{19} cm^{-3}. In their careful analysis, they considered the nonparabolicity of the conduction band (CB), band filling, and band tailing.

In view of the sparse experimental data and the complications in extracting the actual BGN, reliable calculations and analytical models derived from them are an important means to provide the electronics and opto-electronics communities with information on the density- and temperature-dependent BGN in zincblende III–V materials.

The paper is organized as follows. Sec. II gives a short outline of the RPA theory of BGN and of the band structure model that allows for both the numerical computation and the derivation of a general analytical model. All material and band-structure related parameters of the 20 zincblende III–V semiconductors are listed. The importance of nonparabolicity is demonstrated for the three materials with the smallest gap (and smallest electron effective mass). It is shown how the hole masses are calculated from Luttinger parameters and how the limitations of the band structure model lead to the allowed density maxima for each material. Available experimental upper bounds of the densities from measured limits of the activated doping are cited and found to always be smaller than the theoretical limits. Section III provides a summary of the results of the numerically calculated BGN as a function of material at various densities and temperatures. The trends in comparison to bandgap and effective masses and the differences between neutral eh-plasma, quasi-neutral n-type, and quasi-neutral p-type materials are discussed. The strong temperature dependence of BGN in n-type materials with small electron effective masses (like InSb) is contrasted with the weak one in n-type materials with large electron effective masses (like AlAs). Section IV presents the formulas of the general analytical model of BGN in quasi-neutral regions and in the neutral eh-plasma as functions of density and temperature, avoiding any material-dependent free fit parameter. The way the multiple integrals over complicated functions are solved is sketched, and the origin of the strong temperature dependence of the BGN in n-type materials is explored. It is shown that nonparabolicity increases the electron-ion correlation energy but decreases the hole-ion correlation energy. It is demonstrated how the latter effect weakens the BGN in n-type materials with small electron effective masses, preventing the bandgap from shrinking to zero in the relevant density range. Conclusions are given in Sec. V. Appendix A contains the complete presentation of the full-RPA BGN in all 20 materials as a function of density in the temperature range from 0 to 500 K. Each figure comprises quasi-neutral n-type, quasi-neutral p-type, and neutral eh-plasma and also displays the corresponding analytical curves. The density limits can be inferred from the endpoint of the density axis. Appendix B provides the derivation of the analytical expressions of the free-carrier correlation energies in quasi-neutral regions and in the neutral eh-plasma, whereas Appendix C presents the derivation of the analytical formulas of the ion-carrier correlation energies in quasi-neutral regions including the effect of nonparabolicity in n-type material, and the temperature dependence of the donor-hole correlation energy including the effect of nonparabolicity on this temperature dependence.

II. RPA THEORY AND BAND STRUCTURE MODEL

The full-RPA BGN is calculated by the quasi-particle shift (QPS)[1] $\Delta_a(k)$ (a = e for electrons, a = h for holes), which is given by the difference between interacting and free dispersion,[82]

$$\Delta_a(k) = E_a(k) - E_a^0(k), \qquad (1)$$

and equals the real part of the self-energy Σ_a,

$$\Delta_a(k) = \text{Re}\, \Sigma_a\big[k, E_a(k) + i0^+\big]. \qquad (2)$$

It consists of the three terms

$$\Delta_a(k) = \Delta_a^x(k) + \Delta_a^c(k) + \Delta_a^i(k), \qquad (3)$$

the unscreened exchange energy $\Delta_a^x(k)$, the correlation energy of free carriers $\Delta_a^c(k)$, and the correlation energy of the interaction between the carriers and ionized dopants $\Delta_a^i(k)$. A random arrangement on regular lattice sites is assumed for the dopants. Since the

dispersion of the QPS in the energy interval between band edge and Fermi energy is rather flat for the relevant densities,[1,41] the dispersive QPS $\Delta_a(k)$ can be replaced by a rigid shift Δ_a. Then a self-consistent solution of the problem is not necessary because the rigid shift drops out in the energy difference $E_a(k + q) - E_a(k)$, which becomes $E_a^0(k + q) - E_a^0(k)$, i.e., fully determined by the free dispersion. In the distribution functions, Δ_a is fixed by the given density in the quasi-neutral region or by the plasma density in the case of the neutral eh-plasma. The condition that the QPS density should not change in first order with respect to $\Delta_a(k) - \Delta_a$ results in[1],

$$\Delta_a = \frac{\sum_k \partial f_a(k)/\partial \zeta_a \, \Delta_a(k)}{\sum_k \partial f_a(k)/\partial \zeta_a}, \tag{4}$$

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

$$f_a(k) = f\left(E_a^0(k) - \zeta_a\right), \qquad \zeta_a = \mu_a - \Delta_a. \tag{5}$$

The first derivative of the Fermi–Dirac function in (4) plays the role of a weight that 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 dispersive QPS (3) is given explicitly in Refs. 1, 4, and 41. Insertion into Eq. (4) results in the following consolidated forms of the three contributors to the rigid shift:[1,4,41]

$$\Delta_a^x = -\frac{e^2}{4\pi\epsilon_0\epsilon_s\Lambda_a} F_{-1/2}(\beta\zeta_a), \tag{6}$$

$$\Delta_a^c = \frac{1}{2\beta} \int \frac{d^3q}{(2\pi)^3} \sum_\nu \left[\epsilon^{-1}(q, \Omega_\nu) - 1\right] \frac{\partial\epsilon(q, \Omega_\nu)}{\partial n_a}, \tag{7}$$

$$\Delta_a^i = -\frac{n_i}{2}\left(\frac{\partial n_a}{\partial \zeta_a}\right)^{-1} \int \frac{d^3q}{(2\pi)^3} \frac{v(q)}{\epsilon^2(q, 0)} \frac{\partial\epsilon(q, 0)}{\partial \zeta_a}. \tag{8}$$

In the correlation energies (7) and (8), $v(q) = e^2/(\epsilon_0\epsilon_s q^2)$ is the Fourier transform of the bare Coulomb potential with the static permittivity ϵ_s. $\Omega_\nu = \frac{2\pi i \nu}{\hbar\beta}$ denotes the Matsubara frequency (ν integer) and n_i the concentration of electrically active dopants. The correlation energies are governed by the RPA dielectric function,

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

where the energy difference $E_a(k + q) - E_a(k)$ in the denominator is replaced by $E_a^0(k + q) - E_a^0(k)$ (rigid shift).

In the parabolic band model, the QP density n_a in the rigid shift approximation is defined by

$$n_a = \sum_k g_a f_a(k) = g_a \Lambda_a^{-3} F_{1/2}(\beta\zeta_a), \tag{10}$$

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 $\beta = 1/k_BT$, and the Fermi energy $\zeta_a = \mu_a - \Delta_a$. The simple form of the exchange energy (6)—just a single integral (the Fermi integral)—is bound to parabolic bands $E_a^0(k) = \hbar^2k^2/2m_a$.[1,42] In contrast, the rigid-shift

expressions of the correlation energies (7) and (8) are obtained by employing (4) regardless of the specific form of the band dispersion. The exchange-correlation energy of the electron and hole plasmas is rather insensitive to band structure details as a consequence of a compensation effect.[1,43] Therefore, parabolic bands are used to evaluate Eqs. (6) and (7) in this paper. However, the nonparabolicity of the CB is crucial for the ionic correlation energy $\Delta_{h,n-type}^i$ of certain n-type III–V materials in the low-T/high-density regime[7] since, due to the small n-density-of-states (DOS), the Fermi level moves deep into the CB. The electron density n_e and the dielectric function $\epsilon(q, \Omega_\nu)$ in (8) are then calculated with the nonparabolic free dispersion,

$$E_e^{0,np}(k) = \frac{1}{2\gamma_e}\left(\sqrt{2\gamma_e\hbar^2k^2/m_e + 1} - 1\right), \tag{11}$$

where γ_e denotes the nonparabolicity parameter (the common symbol α is used for the mass ratios $\alpha_{\{e,h\}}$ in this paper). In addition, for the shift of the majority band edge in n-type material, $\Delta_{e,n-type}^i$, and for the shift of the minority band edge in p-type material, $\Delta_{e,p-type}^i$, the nonparabolic dispersion (11) is applied for electrons.

Figure 1 demonstrates the nonparabolicity effect on $\Delta_{h,n-type}^i$, which dominates the total BGN in all n-type materials with $m_e \ll m_h$ ($\alpha_h \ll 1$). It is particularly strong for the small-gap materials with strong CB-valence band (VB) coupling (Nos. 1–5 in Table I) but is also important for all other direct materials. The only exception is the indirect materials AlAs and GaP, where $m_e \approx m_h$ ($\alpha_{\{e,h\}} \approx 0.5$).

As the evaluation of (7) comes down to a sevenfold integration and that of (8) to a sixfold integration, an isotropic dispersion model is inevitable. A full numerical treatment based on a more elaborate band structure model is out of reach even in the rigid-shift approximation of BGN. Therefore, the real DOS has to be approximated by an isotropic DOS, and its range of validity must be carefully determined, which sets upper bounds on the allowed densities. The isotropic band structure model is obtained as follows: for all *direct* zincblende materials, the Γ-valley is assumed isotropic with
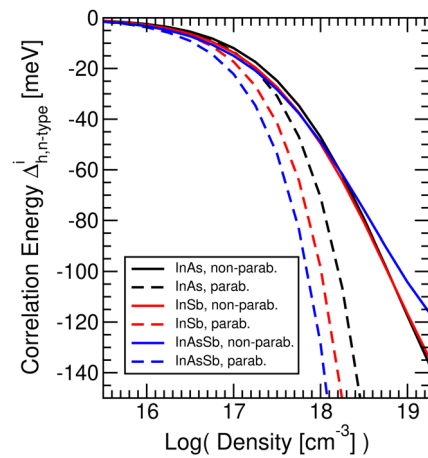

FIG. 1. Hole-ion part $\Delta_{h,n-type}^i$ of the average QPS for quasi-neutral n-type InAs, InSb, and InAs$_{0.4}$Sb$_{0.6}$ as a function of activated doping concentration at $T = 20$ K.

14 July 2023 14:39:17

TABLE I. Parameters of zincblende III–V semiconductors.[44,45,69–72,74] The symbols have the following meaning: E_g(0 K) is the fundamental bandgap at zero temperature; m_e/m_0 is the density-of-states electron mass in units of the free electron mass; m_h/m_0 is the density-of-states hole mass in units of the free electron mass; $g_{e,h}$ are the band multiplicities including the spin degree of freedom; μ^*/m_0 is the reduced effective mass in units of the free electron mass; $\alpha_{e,h}$ are the mass ratios $\mu^*/m_{e,h}$; γ_e is the nonparabolicity parameter for the conduction band; Ry_{ex} is the excitonic Rydberg energy, i.e., the binding energy of the exciton (used for energy scaling); a_{ex} is the excitonic Bohr radius (used for length scaling); and ϵ_s is the static permittivity of the semiconductor in units of the vacuum dielectric constant. The excitonic quantities are very suitable reference units, even at densities where excitons do not exist.[1] Detailed explanations and corresponding formulas are given in Sec. II of the main text.

No.	Material	E_g (0 K) (eV)	m_e/m_0	m_h/m_0	g_e	g_h	μ^*/m_0	α_e	α_h	γ_e (1/eV)	Ry_{ex} (meV)	a_{ex} (nm)	ϵ_s
1	InAs$_{0.4}$Sb$_{0.6}$	0.146	0.010	0.268	2	4	0.010	0.964	0.036	6.85	0.50	88.55	16.1
2	InSb	0.235	0.0135	0.272	2	4	0.013	0.953	0.047	4.1	0.62	69.12	16.8
3	InAs	0.415	0.024	0.261	2	4	0.022	0.917	0.083	2.5	1.28	37.04	15.15
4	In$_{0.5}$Ga$_{0.5}$Sb	0.420	0.025	0.265	2	4	0.023	0.914	0.086	2.38	1.18	37.64	16.25
5	zb-InN	0.530	0.054	0.779	2	4	0.0505	0.935	0.065	4.45	2.95	15.95	15.3
6	InP$_{0.69}$Sb$_{0.31}$	0.647	0.059	0.356	2	4	0.051	0.858	0.142	1.55	3.60	14.46	13.8
7	GaAs$_{0.5}$Sb$_{0.5}$	0.809	0.054	0.316	2	4	0.046	0.854	0.146	0.86	3.07	16.41	14.3
8	GaSb	0.813	0.041	0.259	2	4	0.035	0.863	0.137	1.2	1.95	23.47	15.7
9	In$_{0.53}$Ga$_{0.47}$As	0.816	0.043	0.335	2	4	0.038	0.886	0.114	1.2	2.68	19.30	13.9
10	InAs$_{0.4}$P$_{0.6}$	0.995	0.057	0.340	2	4	0.049	0.855	0.145	1.0	3.63	14.61	13.6
11	zb-GaIn$_{0.5}$N$_{0.5}$	1.155	0.123	0.806	2	4	0.107	0.868	0.132	0.87	9.29	6.20	12.5
12	InP	1.421	0.080	0.392	2	4	0.066	0.831	0.169	0.7	5.79	9.95	12.5
13	In$_{0.52}$Al$_{0.48}$As	1.486	0.072	0.377	2	4	0.060	0.840	0.160	0.67	5.31	10.67	12.7
14	GaAs	1.519	0.067	0.374	2	4	0.057	0.848	0.152	0.66	4.86	11.47	12.9
15	Al$_{0.3}$Ga$_{0.7}$As	1.910	0.092	0.413	2	4	0.075	0.818	0.182	0.52	7.05	8.47	12.05
16	In$_{0.49}$Ga$_{0.51}$P	2.006	0.093	0.458	2	4	0.077	0.831	0.169	0.5	7.55	8.08	11.8
17	AlAs (X)	2.240	0.361	0.505	12	4	0.211	0.583	0.417	0.45	28.3	2.53	10.1
18	GaP (X)	2.350	0.378	0.522	6	4	0.219	0.580	0.420	0.43	24.2	2.68	11.1
19	zb-GaN	3.280	0.193	0.838	2	4	0.157	0.813	0.187	0.3	22.7	3.27	9.7
20	zb-AlN	4.900	0.250	1.369	2	4	0.211	0.846	0.154	0.2	35.5	2.25	9.0

an effective mass m_e taken from Refs. 45 and 44. For the two indirect materials, AlAs and GaP, the anisotropic valleys are mapped to isotropic ones with an effective DOS mass $m_e = (m_t^2 m_l)^{1/3}$. The valley degeneracy is included in the total degeneracy factor g_e. Heavy and light holes are treated as independent isotropic VB valleys, and the effective DOS mass is obtained as $m_h = [(m_{lh}^{3/2} + m_{hh}^{3/2})/2]^{2/3}$, which results in valley degeneracy factors of 2 for all materials included in the total degeneracy factor g_h. The split-off band is neglected. In some cases, the values of m_{lh} and m_{hh} were taken from Ref. 44 if based on experimental data. For most materials, however, the Luttinger parameters $\gamma_{\{1,2,3\}}$ published in the review of Vurgaftman *et al.*[45] were used to compute

$$m_{lh}/m_0 = \frac{1}{\gamma_1 + 2\bar{\gamma}}, \qquad m_{hh}/m_0 = \frac{1}{\gamma_1 - 2\bar{\gamma}}, \qquad (12)$$

with[46]

$$\bar{\gamma} = \sqrt{\frac{2\gamma_2^2 + 3\gamma_3^2}{5}}. \qquad (13)$$

Thanks to the isotropic model, the angular integrations in the correlation energies (7) and (8) and in the dielectric function (9) can be performed exactly. The numerical expense then reduces to a double integral plus the Matsubara sum for the free-carrier correlation energy and to a twofold integration for the ionic correlation energy. The Matsubara sum is transformed into an integral in the case of $T = 0$ K. For zero temperature, the multiple integrals are solved

approximately in Sec. IV to obtain a general analytical BGN model in terms of the material parameters listed in Table I.

An important aspect in the context of BGN is the possible maximum density. Applying the theory to unreasonably high densities can readily shrink the gap of certain n-type materials to zero. In quasi-neutral regions, the carrier density is equal to the electrically active doping. The intrinsic limitations of the latter have been subject to intensive research.[47,48] It has been recognized that this limitation is a property of the material and not a chemical or electronic feature of the dopant[47] (e.g., formation of inactive complexes or charge compensation due to the amphoteric nature of shallow dopants). Process-related localized defects stabilize the Fermi energy at the Fermi-level stabilization energy ~4.9 eV below the vacuum level in many III–V materials.[47,49] The amphoteric nature of these defects leads to a Fermi level pinning in the gap or in the CB, independent of the type or doping level, which explains the saturation of the electrically active doping. Experimental data for the limits of activated doping are only available for some of the 20 materials treated in this paper: InSb,[50,51] InAs,[52,53] GaSb,[54–56] InP,[57,58] GaAs,[59] GaP,[60–64] and zb-GaN.[48,65–68] These limits were adopted for the density limits since they are, in any case, smaller than the limits imposed by the band structure model. In all other cases and, in general, for the neutral eh-plasma, the maximum densities were calculated from the condition that the Fermi level must not move into the next higher CB valley (electrons) or into the split-off band (holes). The valley separation energies $E_L - E_\Gamma$ or $E_X - E_\Gamma$, and the spin–orbit energies E_{so} were adopted from Refs. 44, 45, and 69. The nonparabolic free dis-
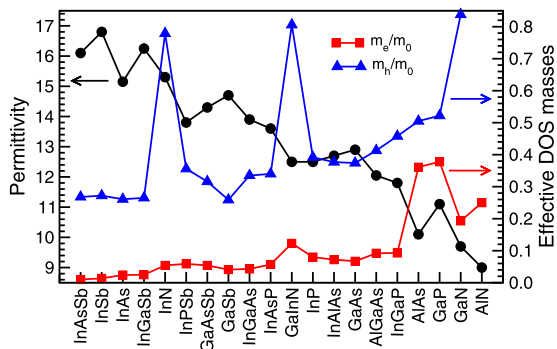
FIG. 2. Static permittivity ϵ_s and effective DOS masses m_e and m_h of zincblende III–V materials ordered by the size of their energy gap at $T = 0$ K (compare Table I). The point $m_h = 1.369\, m_0$ for zb-AlN was omitted for better scalability.
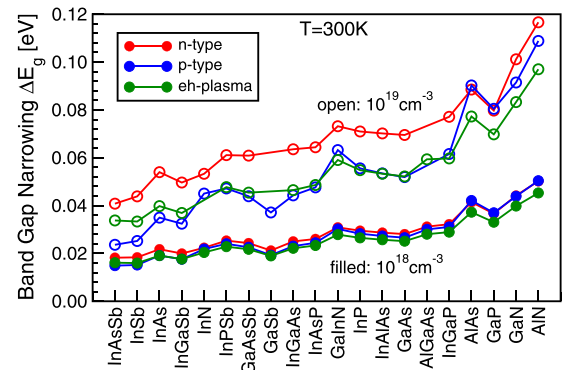

FIG. 3. Total full-RPA BGN of zincblende III–V materials at $T = 300$ K (ordered as in Fig. 2) for two concentrations (10^{18} cm^{-3}, filled circles and 10^{19} cm^{-3}, open circles) and the cases n-type (red), p-type (blue), and eh-plasma (green). Lacking data points at 10^{19} cm^{-3} are due to the density limitations described in the text.

persion (11) was used for the relation between electron density and Fermi level in n-type materials. The upper bounds of the density can be inferred from the end points of the density axis in Figs. 10–29.

If available, the nonparabolicity parameters γ_e were either directly taken from full-band calculations[72] or from the data collection of Ref. 44. In most cases, they had to be estimated by[73] $\gamma_e = 1/E_g$ (0 K), where the size of the bandgap at zero temperature was adopted from Refs. 44 and 45, except for InN, where a more recent value was used.[69,74,75] The static permittivities were taken over from Ref. 44. If they were not available for certain ternary compounds, they were calculated as the composition average of the permittivities of the binary constituents.

All material-related parameters needed for the BGN calculation are listed in Table I, where the size of the energy gap at $T = 0$ K was chosen for the ordering of the materials. The degeneracy factors g_a account for the multi-valley conduction band (AlAs, GaP), heavy and light hole bands, and spin summation. It is convenient to use normalized quantities. All energies are normalized by the excitonic Rydberg energy $Ry_{ex} = \hbar^2/(2\,\mu^* a_{ex}^2)$ and all lengths by the excitonic Bohr radius $a_{ex} = \hbar^2 \epsilon_s/(e^2 \mu^*)$, where μ^* denotes the reduced effective mass $\mu^* = (1/m_e + 1/m_h)^{-1}$. The parameters $\alpha_a = \mu^*/m_a$ serve to transform excitonic quantities back to electron/hole quantities. Note that $\alpha_e + \alpha_h = 1$.

Figure 2 shows permittivity (left ordinate) and effective DOS masses (right ordinate) as functions of material, where their ordering is the same as in Table I, i.e., according to the size of the energy gap at $T = 0$ K. The general trends are that the permittivity decreases from ~17 to ~9, and both DOS masses increase. In the case of the CB DOS mass, the indirect materials AlAs and GaP are outliers, which is related to their multi-valley nature. The nitrides are strong outliers with respect to the VB DOS mass (see the distinct peaks of the blue curve). The general trends of permittivity and DOS masses roughly explain the general trend of BGN (see Sec. III).

III. SUMMARY OF FULL-RPA RESULTS

The total full-RPA BGN as a function of material, i.e., in dependence on the increasing bandgap, is shown in Fig. 3 for 300 K and in Fig. 4 for 20 K. Two densities have been chosen for all 60 cases: 10^{18} and 10^{19} cm^{-3}. In some cases, the higher density already exceeds the

allowed limit, as discussed earlier; hence, the corresponding values were omitted.

Looking at the room-temperature results first, n-type, p-type, and eh-plasma differ only a little at the lower density (filled circles). The average trend is an increase in BGN from ~15 to ~50 meV. At the higher density, p-type and eh-plasma still behave similarly with an average increase from ~30 to ~100 meV, but the BGN in direct n-type material is significantly larger due to the small electron mass leading to reduced screening.[7] One observes an average increase from ~40 to ~120 meV. The indirect materials AlAs and GaP make an exception since their CB-DOS and VB-DOS are comparable in strength. Therefore, the BGN in n-type material is almost identical to the one in p-type material here. The trends closely follow the trends of the DOS effective masses (see Fig. 2). The maxima produced by the nitrides are reflected in the BGN curves, albeit scaled-down a lot.

Turning to the low-temperature case of 20 K in Fig. 4, the behavior of p-type material and eh-plasma is not only almost identical at the lower density of 10^{18} cm^{-3} but also very similar to the one at room temperature. To a lesser extent, this also holds at a
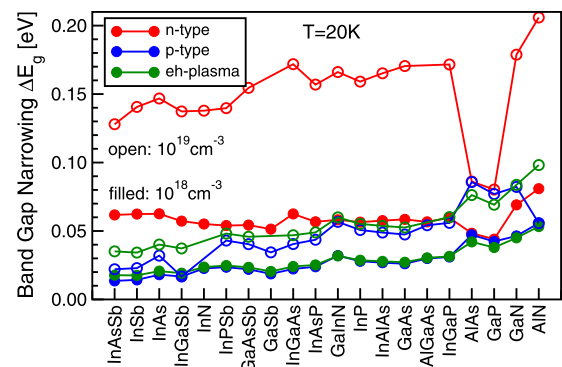

FIG. 4. Total full-RPA BGN of zincblende III–V materials at $T = 20$ K (ordered as in Fig. 2) for two concentrations (10^{18} cm^{-3}, filled circles and 10^{19} cm^{-3}, open circles) and the cases n-type (red), p-type (blue), and eh-plasma (green). Lacking data points at 10^{19} cm^{-3} are due to the density limitations described in the text.
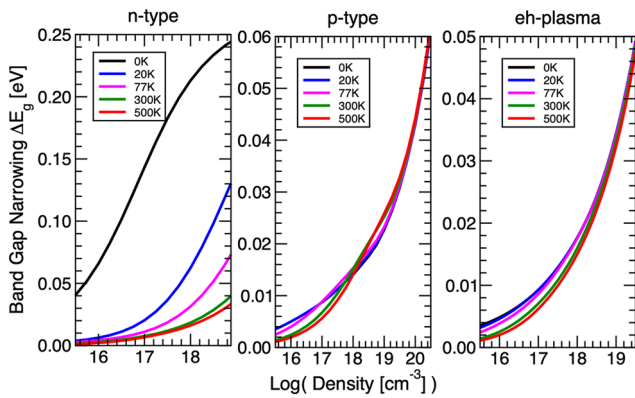
FIG. 5. Total full-RPA BGN in InSb as function of concentration for different temperatures.
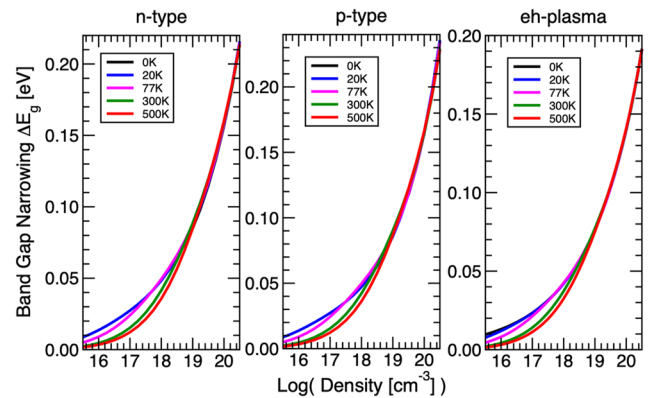

FIG. 6. Total full-RPA BGN in AlAs as function of concentration for different temperatures.

higher density of 10^{19} cm^{-3}. The behavior of n-type material, however, differs drastically from that at 300 K. At lower densities, the BGN exhibits a slightly decreasing tendency up to the indirect materials, followed by an increase for the wide-gap nitrides. Nevertheless, the BGN is relatively independent of material, taking values between ~50 and ~80 meV. In direct n-type materials, the BGN becomes dramatically increased as a consequence of the weak screening of the hole-ion interaction by majority carriers (electrons), an effect that gains importance with decreasing temperature (compare Fig. 5 and the detailed BGN representations in Appendix A). The analytical explanation for this temperature dependence will be given in Sec. IV. One can see an increasing tendency of the BGN from ~130 to ~200 meV, only interrupted by the indirect materials AlAs and GaP, where again the values are the same as in p-type materials due to the similar DOS size.

In Figs. 5 and 6, the BGN of the small-gap direct material InSb is compared with that of the wide-gap indirect material AlAs. The density limit for n-type InSb in the left part of Fig. 5 stems from the measured maximum of activated doping (7.5×10^{18} cm^{-3})[51] and results in a BGN that already slightly exceeds the value of the zero-temperature gap. Interestingly, for a limit of 4.5×10^{18} cm^{-3}, as found in a standard low-pressure growth process,[50] the gap would exactly shrink to zero at $T = 0$ K. In all other cases in Figs. 5 and 6, the upper bounds of the density are imposed by the band structure model. The striking difference between InSb and AlAs is the behavior of n-type materials. Whereas in AlAs, the BGN is similar to that in p-type and eh-plasma (note the similar BGN scale and the almost identical temperature dependence, which ceases above 10^{19} cm^{-3}), the BGN in n-type InSb exhibits a strong temperature dependence that becomes even more pronounced at higher densities. At $T = 0$ K and 10^{18} cm^{-3}, the BGN is 15 times larger than in p-type material. As outlined in Ref. 7, this is due to the small electron effective mass and the resulting weak screening of the hole-ion correlation energy. It will be shown in Sec. IV that the strong temperature dependence arises from the occupation probability of the minority (hole) band.

Appendix A comprises the presentation of BGN for all 20 materials (n-type and p-type) and for the neutral eh-plasma (in the intrinsic form of these materials) as a continuous function of density from $10^{15.5}$ cm^{-3} to the maximum at the temperatures $T = 0, 20, 77, 300,$ and 500 K. The full-RPA curves for $T = 0$ K

(black solid) start at the Mott density[76] $n_{b,M} = (p_M / a_{ex}\alpha_b)^3$, if $n_{b,M} > 10^{15.5}$ cm^{-3}, using $p_M = 0.296$ for the Mott criterion. In all p-type nitrides, $n_{h,M} > n_{i,max}$; hence, the $T = 0$ curves are completely missing. Reasons could be (i) the Mott parameter, (ii) the size of the Luttinger parameters from which the VB DOS masses are derived, (iii) the isotropization Eq. (13), or (iv) the size of the spin–orbit energy E_{so} that determines $n_{i,max}$. The latter is rather small [(6–19) meV] in the ordering of the nitrides given in Table I],[45] which leaves only a narrow energy interval compliant with the 2-valley VB model. The ratio $n_{h,M}/n_{i,max}$ increases from 5 to 12. Actually, (i) and (ii) are not independent of each other since $n_{h,M} \sim (p_M m_h)^3$. As the Mott parameter is related to the lattice constant, it decreases from zb-InN toward zb-AlN, which could partly explain the discrepancy. The analytical $T = 0$ curves are displayed over the whole density range, neglecting carrier freeze-out.

Figure 7 compares the available experimental data with the full-RPA results. The left panel shows PL[31,32] and hetero-structure BT[33] data for p-type GaAs. They fairly match the theoretical curves. In the case of p-type GaSb (right panel), the measured BGN by PL[32] is systematically larger than the RPA one. The opposite is found for n-type InP (absorption data[35]) in the high-density range. Absorption
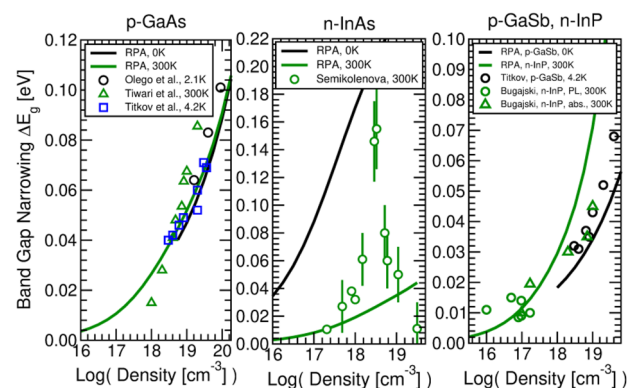

FIG. 7. Comparison of full-RPA BGN with experimental data for p-type GaAs[31–33] (left), n-type InAs[34] (middle), and p-type GaSb[32] and n-type InP (right).[35]

14 July 2023 14:39:17

measurements of the BGN in n-type InAs at room temperature[34] resulted in a sharp peak at intermediate densities. The authors interpreted this peculiar behavior by a large screening length due to the strong correlation of the tellurium dopants ("complete ordering") when their concentration falls into the interval $(2\text{–}4) \times 10^{18}$ cm^{-3}. Interestingly, full-RPA results (based on the assumption of a random impurity distribution) are in reasonable agreement both at low and very high densities.

IV. ANALYTICAL MODEL WITHOUT MATERIAL-DEPENDENT FREE FIT PARAMETERS

In this section, a fully analytical model of BGN as a function of density and temperature in terms of the material and band structure parameters of Table I is presented. The goal is not the highest accuracy but to avoid free fit parameters for each case [as performed in Refs. 77 and 78 for $T = 0$ K and local-density approximation (LDA)]. This facilitates physical insight and the possibility to quickly calculate BGN for other materials or ternary compositions, as long as they obey (or can be approximated to) the isotropic band structure model consisting of one CB valley and two VB valleys. Wurtzite materials are excluded because of their complicated VB structure caused by the crystal-field splitting.

The general approach is to combine the $T = 0$-limit $\Delta E_{g,T=0}$ with the high-temperature (Debye) limit[79] $\Delta E_{g,T\to\infty}$ by a Padé approximation[1,41] using a universal switching function,[5,41]

$$U(n, T) = (4\pi\beta)^3 n^2. \tag{14}$$

For p-type material and the neutral eh-plasma, this automatically interpolates between the high- and low-density regimes. In the case of direct n-type materials, the hole-ion correlation energy causes a strong temperature dependence at any density of interest. This is because the main origin is the occupation probability of the minority holes, which are non-degenerate no matter how strong the n-doping is. Therefore, $\Delta E_{g,T=0}$ is replaced by $\Delta E_{g,0}(T)$ with the analytical model for $\Delta_{h,n-type}^i(T)$ (see below), but the $T = 0$-limits in the case of all other energy contributions. The total BGN is, therefore, calculated by

$$\Delta E_g(n, T) = \frac{1 + U(n, T)}{1/\Delta E_{g,T\to\infty}(n, T) + U(n, T)/\Delta E_{g,0}(n, T)}. \tag{15}$$

Here, n equals the concentration n_i of activated dopants in quasi-neutral regions or the plasma density $n_e = n_h$ of the neutral eh-plasma, respectively. Note that BGN is defined as a positive quantity, i.e., the negative sum of all QPS contributions (6)–(8). The Debye limits read[1,5]

$$\Delta E_{g,T\to\infty}(n, T)$$
$$= \begin{cases} 8\pi n\beta\alpha_e/g_e + 6\sqrt{2\pi n\beta}, & n \to n_i, \quad \text{n – type,} \\ 8\pi n\beta\alpha_h/g_h + 6\sqrt{2\pi n\beta}, & n \to n_i, \quad \text{p – type,} \\ 8\pi n\beta(\alpha_e/g_e + \alpha_h/g_h) + 8\sqrt{\pi n\beta}, & n \to n_e, \quad \text{eh – pl,} \end{cases} \tag{16}$$

where the first term is the exchange energy and the second term is the sum of all correlation energies. Nonparabolicity is negligible in the low-density range.

The analytical calculation of $\Delta E_{g,0}(n, T)$ involves the following terms:

$$\Delta E_{g,0}(n, T)$$
$$= \begin{cases} -\Delta_e^x - \Delta_e^c - \Delta_h^c - \Delta_e^i - \Delta_h^i(T) & \text{for} \quad \text{n – type,} \\ -\Delta_h^x - \Delta_h^c - \Delta_e^c - \Delta_h^i - \Delta_e^i & \text{for} \quad \text{p – type,} \quad (17) \\ -\Delta_e^x - \Delta_h^x - \Delta_e^c - \Delta_h^c & \text{for} \quad \text{eh – pl.} \end{cases}$$

The expression of the exchange energy $\Delta_{\{e,h\}}^x$ is well known,[1] and numerous attempts have been made to fit power laws in terms of density to the RPA free-carrier correlation energies of the eh-plasma $\Delta_{\{e,h\}}^c$ (reviewed in Ref. 1). In Ref. 7, the ionic correlation energies $\Delta_{\{e,h\}}^i$ at $T = 0$ K were derived for n- and p-doped quasi-neutral regions of InGaAs, with a first estimate of how strongly nonparabolicity impacts $\Delta_{h,n-type}^i$. In this paper, all terms are calculated analytically starting from Eqs. (7) and (8) using the RPA dielectric function (9). A central role in the multiple integrations of complicated functions is played by the length

$$s_b = \frac{g_b}{\pi\alpha_b q_{F,b}}, \tag{18}$$

where $q_{F,b}$ is the Fermi momentum of the degenerate majority carriers,

$$q_{F,b} = \left(\frac{6\pi^2 n}{g_b}\right)^{1/3}, \tag{19}$$

and "b" is the index of the majority carrier band. Hence, s_b is proportional to the Fermi wavelength $\lambda_{F,b}$ in units of the excitonic Bohr radius, $s_b = g_b/(2\pi^2\alpha_b)\lambda_{F,b}$. The assumption $s_b \ll 1$ is sometimes used in this paper to simplify the integrands. Since $1/q_{F,b}$ is normalized by the excitonic Bohr radius a_{ex}, this assumption works better as the Fermi level moves deeper into the majority band (increasing density, weak DOS) and the excitonic Bohr radius is large, i.e., in the case of direct n-type III–Vs with small electron effective mass m_e. On the contrary, if g_b is large and also m_e, as in the case of n-type AlAs and GaP, the assumption $s_b \ll 1$ is rather poor, as $s_b = 1$ already corresponds to a density of $\sim 10^{21}$ cm^{-3}. This finally shows up in a worse agreement between analytical and numerical $\Delta E_g(n, T)$-curves. Therefore, also $s_b \gg 1$ is tested and utilized if more appropriate. The analytical form of BGN is first given in terms of the length s_b, which is more compact, and then as a function of the normalized density for more convenient use.

A. Quasi-neutral regions

In quasi-neutral regions, $\Delta E_{g,0}(n, T)$ reads in terms of the normalized length s_b

$$\Delta E_{g,0}(s_b, T) = \frac{2g_b}{\pi^2\alpha_b s_b} + \frac{2g_b}{3\pi^2\alpha_b} \ln\left(1 + \sqrt{2/s_b}\right)$$
$$+ \frac{g_b}{\sqrt{3}\pi(\alpha_b s_b)^{3/4}} + \frac{g_b[\delta_{bh} + \varphi_{np}(s_b)\delta_{be}]}{3\pi\alpha_b\sqrt{2s_b}}$$
$$+ \frac{\sqrt{2}g_b}{3\pi\alpha_a s_b^{3/2}}\left[\delta_{bh} + \frac{\varphi_T(s_b)}{\varphi_{np}^{3/2}(s_b)}\delta_{be}\right], \tag{20}$$

with

$$\varphi_{np}(s_e) = \sqrt{1 + \frac{4\gamma_e g_e^2}{\pi^2 \alpha_e s_e^2}}, \tag{21}$$

$$\varphi_T(s_e) = 1 - \frac{2}{\pi} \arctan\left[\left(\frac{2\pi^2 \alpha_e^2 s_e}{\beta \alpha_h g_e^2 \varphi_{np}(s_e)}\right)^{\frac{3}{4}}\right]. \tag{22}$$

In Eq. (20), "b" is the index of the majority carrier band ($b = e$ for n-type, $b = h$ for p-type material), and "a" is the index of the minority carrier band ($a = h$ for n-type, $a = e$ for p-type material). The ordering of the five terms is the same as in Eq. (17). Note that the third term, the minority free-carrier correlation energy, only contains the index of the majority carrier band (b), since minority carriers do not contribute to screening. Nevertheless, the effective mass of the minority carriers finally comes into play after multiplication with the normalization energy (excitonic Rydberg).

Equation (21) is the nonparabolicity correction that only applies to n-type material. Note that it increases the electron-ion correlation energy (fourth term) but decreases the hole-ion correlation energy (fifth term) with different powers. Equation (22) describes the reduction of the hole-ion correlation energy with increasing temperature at intermediate and high densities. It also applies to n-type material only and is strongly affected by nonparabolicity (note the occurrence of φ_{np} in the argument of $\arctan[\cdots]$). The zero-temperature limit ($\beta \to \infty$) is $\varphi_T = 1$. Figure 8 shows the ability of this model to reproduce the T-dependence of the hole-ion correlation energy for all densities $>10^{15.5}$ cm^{-3} over the whole temperature range. By way of example, the cases InSb (small gap), InGaAs (intermediate gap), and zb-GaN (wide gap) have been chosen. One observes a decreasing T-dependence and a changing shape of the n-dependence with a rising gap.

From Eq. (8), one can see that the temperature dependence has two origins: (i) the distribution function of minority carriers in the factor $(\partial n_h / \partial \zeta_h)^{-1}$ and (ii) the occupation probability of the majority carriers contained in the dielectric function (9). From a second-order Taylor expansion of the latter by $1/\beta$ around the Fermi
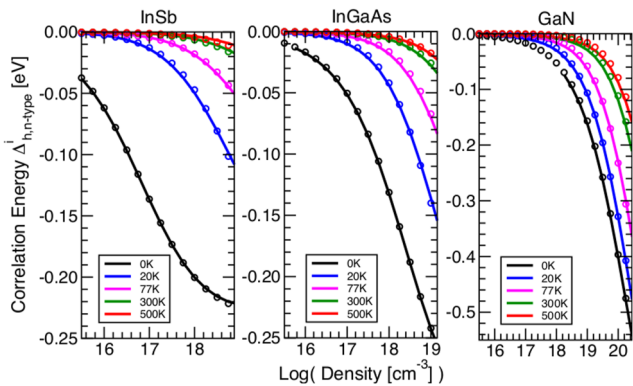

FIG. 8. Hole-ion part $\Delta_{h,n-type}^{i}$ of the average QPS for quasi-neutral n-type InSb (left), In$_{0.53}$Ga$_{0.47}$As (middle), and zb-GaN (right) as a function of density at different temperatures. Comparison of full-RPA (solid lines) with the analytical model (symbols).

energy, one obtains a quadratic T-dependence which, however, is weak and vanishes at high densities. It is negligible compared to (i), which results in the term (22). The derivation of this term is given in Appendix C.

Replacing s_b by the normalized density n using Eqs. (18) and (19), i.e., $s_b = g_b^{4/3}/[\alpha_b(6\pi^5 n)^{1/3}]$, one obtains

$$\Delta E_{g,0}(n,T) = \left(\frac{48n}{\pi g_b}\right)^{1/3} + \frac{2g_b}{3\pi^2 \alpha_b} \ln\left(1 + \frac{\sqrt{2\alpha_b}(6\pi^5 n)^{1/6}}{g_b^{2/3}}\right)$$
$$+ \left(\frac{2\pi n}{3}\right)^{1/4} + \frac{g_b^{1/3}}{3\sqrt{\alpha_b}}\left(\frac{3n}{4\pi}\right)^{1/6}[\delta_{bh} + \varphi_{np}(n)\delta_{be}]$$
$$+ \frac{2(\pi\alpha_b)^{3/2}\sqrt{n}}{\sqrt{3}\alpha_a g_b}\left[\delta_{bh} + \frac{\varphi_T(n)}{\varphi_{np}^{3/2}(n)}\delta_{be}\right], \tag{23}$$

with

$$\varphi_{np}(n) = \sqrt{1 + 4\gamma_e \alpha_e \left(\frac{6\pi^2 n}{g_e}\right)^{2/3}}, \tag{24}$$

$$\varphi_T(n) = 1 - \frac{2}{\pi} \arctan\left\{\frac{\alpha_e^{3/4}}{\left[\beta\alpha_h g_e^{2/3}\varphi_{np}(n)\right]^{\frac{3}{4}}}\left(\frac{4\pi}{3n}\right)^{\frac{1}{4}}\right\}. \tag{25}$$

In Eqs. (23)–(25), n represents the density n_i of ionized dopants (normalized by a_{ex}^{-3}). Note that BGN is normalized by the excitonic Rydberg energy Ry_{ex}, as is the thermal energy $1/\beta$ and the inverse of the nonparabolicity parameter $1/\gamma_e$.

B. Neutral electron–hole plasma

In the case of the neutral eh-plasma, both exchange energies are present, but the carrier-ion correlation energies are absent [see Eq. (17)]. The analytical model for $T = 0$ K and parabolic bands reads as a function of the normalized lengths $s_{\{e,h\}}$,

$$\Delta E_{g,0}(s_e, s_h, 0K) = \frac{2}{\pi^2}\left(\frac{g_e}{\alpha_e s_e} + \frac{g_h}{\alpha_h s_h}\right) + \frac{2g_e}{5\pi}\ln\left[1 + \left(\frac{8\alpha_e}{3}\right)^{\frac{1}{4}}\frac{1}{s_e^{3/4}}\right]$$
$$+ \frac{2g_h}{\pi}\left\{\Theta(0.85 - \alpha_e)\ln\left[1 + \left(\frac{8\alpha_h}{3}\right)^{\frac{1}{4}}\frac{1}{s_h^{3/4}}\right]\right.$$
$$\left. + \Theta(\alpha_e - 0.85)\ln\left(1 + \sqrt{\frac{2\alpha_h}{\alpha_{he}s_h}}\right)\right\}, \tag{26}$$

with

$$\alpha_{he} = \frac{\alpha_h}{1 + \left(\frac{g_e}{g_h}\right)^{\frac{2}{3}}\frac{\alpha_h}{\alpha_e}}. \tag{27}$$

The first term in (26) is the sum of the exchange energies; the second and third terms are the correlation energies of electrons and holes, respectively. These terms differ from the corresponding majority-
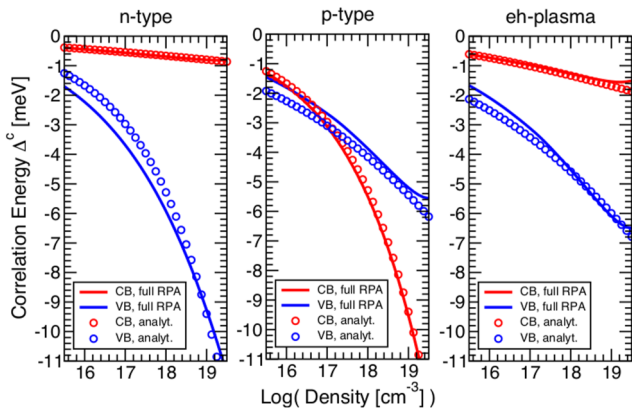
FIG. 9. Contribution of free-carrier correlation energy Δ^c to the average QPS of $In_{0.53}Ga_{0.47}As$ for quasi-neutral n-type, quasi-neutral p-type, and neutral eh-plasma as a function of density at $T = 0$ K. Comparison of full-RPA with the analytical model Eq. (28) (last term).

FIG. 11. Total BGN in InSb. Densities limited by activatable n-doping[51] (left), E_v-E_{so}-separation (middle), and E_L-E_Γ-separation (right).

carrier correlation energy of the one-component plasma [second term in Eq. (20)] because of the stronger screening by the two-component plasma. The derivation of the correlation terms with detailed explanations is presented in Appendix B.

Inserting $s_{\{e,h\}} = g_{\{e,h\}}^{4/3}/[\alpha_{\{e,h\}}(6\pi^5 n)^{1/3}]$, one obtains the model as function of plasma density n,

$$
\begin{aligned}
\Delta E_{g,0}(n,0\text{K}) = {} & \left(\frac{1}{g_e^{1/3}} + \frac{1}{g_h^{1/3}}\right)\left(\frac{48n}{\pi}\right)^{1/3} + \frac{2g_e}{5\pi}\ln\left[1 + \frac{2\alpha_e(\pi^5 n)^{\frac{1}{4}}}{g_e}\right] \\
& + \frac{2g_h}{\pi}\left\{\Theta(0.85 - \alpha_e)\ln\left[1 + \frac{2\alpha_h(\pi^5 n)^{\frac{1}{4}}}{g_h}\right] \right. \\
& \left. + \Theta(\alpha_e - 0.85)\ln\left(1 + \frac{\sqrt{2}\alpha_h(6\pi^5 n)^{1/6}}{\sqrt{\alpha_{he}}g_h^{2/3}}\right)\right\}.
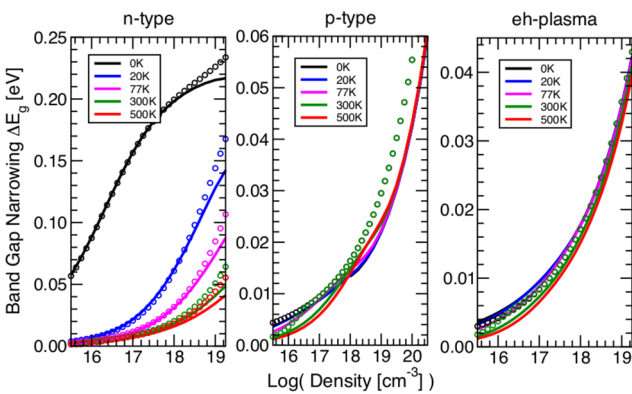\end{aligned}
\tag{28}
$$

In Fig. 9, the numerical RPA results of the band edge shifts induced by the free-carrier correlation in InGaAs at $T = 0$ K are compared with the outcomes of the analytical models. The freeze-out of carriers was disregarded to display the whole density range. The agreement is excellent for the shift of the CB edge and reasonable for the shift of the VB edge. This difference in accuracy is due to $s_e \ll s_h$ ($\alpha_e \gg \alpha_h$). Figure 9 also shows that the correlation energies in a neutral eh-plasma are very similar to the free-carrier correlation energies of the majority carriers in quasi-neutral regions. Size and slope are only slightly larger, which can be attributed to the stronger screening in a two-component plasma.

The comparison of the analytical BGN model with the numerical full-RPA is presented in Appendix A for all 60 cases (Figs. 10–29). In p-type material and in the eh-plasma, the temperature dependence is only significant at intermediate densities. For better visibility, the analytical curves were plotted just for 0 K and room temperature here. The Padé approximation leads to a relatively large deviation in the transition region,[1,41] which is unavoidable. The density and temperature dependences are correctly reproduced in all

FIG. 10. Total BGN in $InAs_{0.4}Sb_{0.6}$. Densities limited by E_L-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).
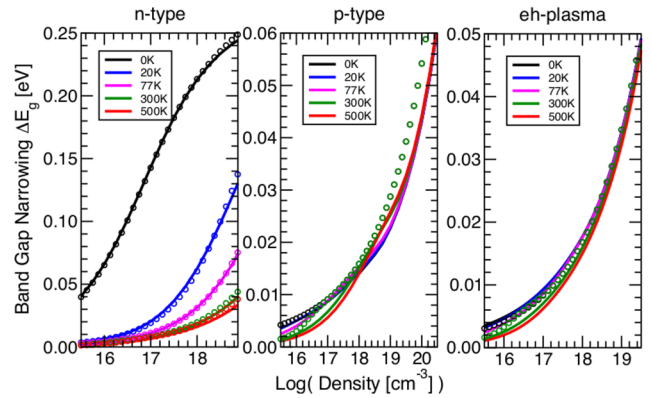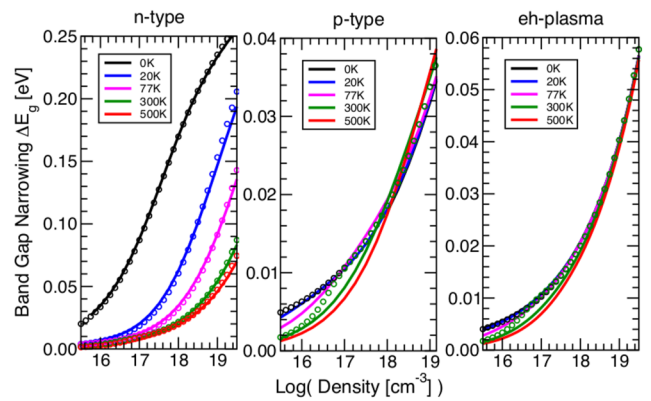

FIG. 12. Total BGN in InAs. Densities limited by activatable n-doping[52] (left), activatable p-doping[53] (middle), and E_L-E_Γ-separation (right).
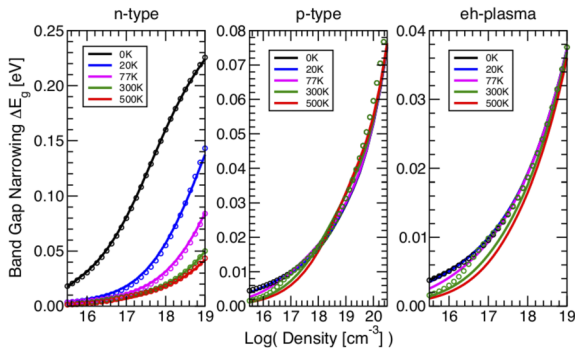
FIG. 13. Total BGN in $In_{0.5}Ga_{0.5}Sb$. Densities limited by E_L-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).

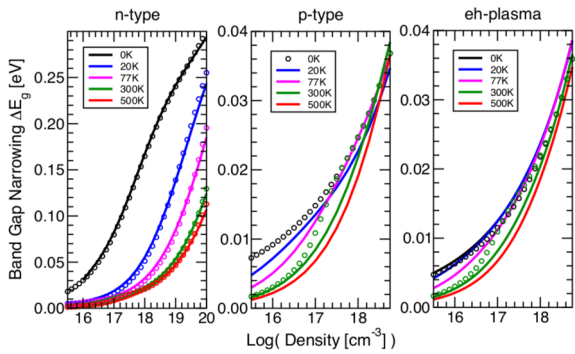**FIG. 14.** Total BGN in zb-InN. Densities limited by E_X-E_Γ-separation (left) and E_v-E_{so}-separation (middle and right).

FIG. 16. Total BGN in $GaAs_{0.5}Sb_{0.5}$. Densities limited by E_L-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).

FIG. 17. Total BGN in GaSb. Densities limited by E_L-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle). Experimental data on the limits of activatable n-doping[54–56] cover the range $(6.5 \times 10^{17}$–$3 \times 10^{18})$ cm^{-3}.

60 cases. Very good agreement is found for the direct n-type materials, except for InAsSb at high densities because of its extreme nonparabolicity. Larger deviations occur for the indirect materials AlAs and GaP, as $s_e \ll 1$ is a poor assumption here. For many p-type materials, the accuracy is satisfactory as well; exceptions are
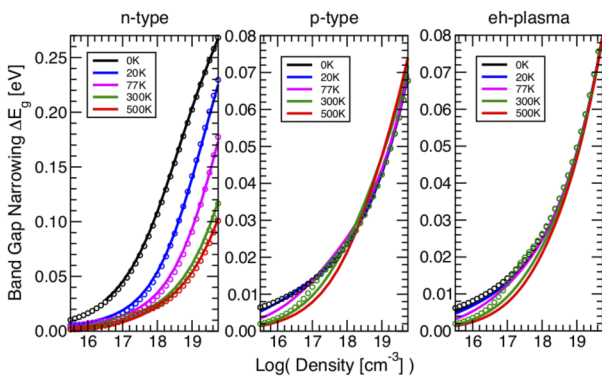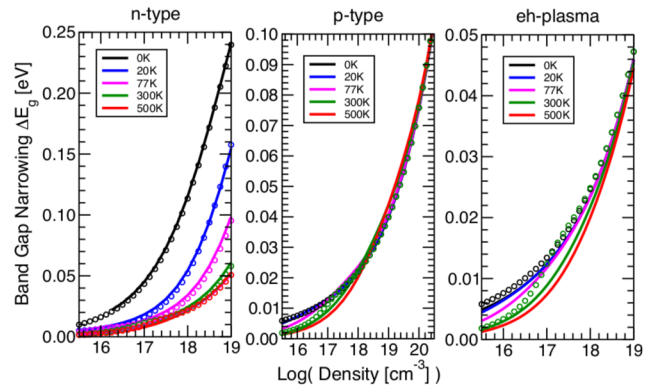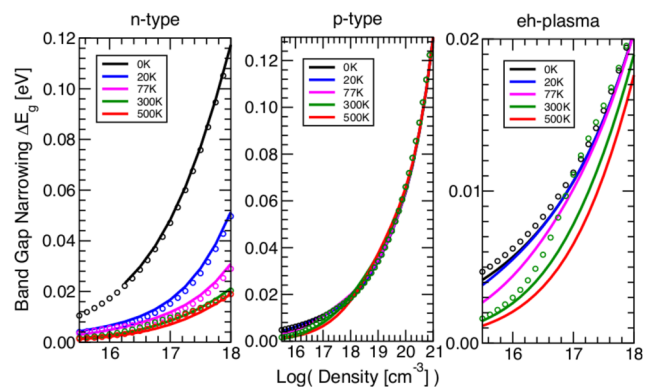
InAsSb, InSb, and the nitrides. For the neutral eh-plasma, the analytical model works best at high densities but deviates markedly at intermediate densities due to the error caused by the Padé interpolation.

FIG. 15. Total BGN in $InP_{0.69}Sb_{0.31}$. Densities limited by E_L-E_Γ-separation (left and right), and E_v-E_{so}-separation (middle).
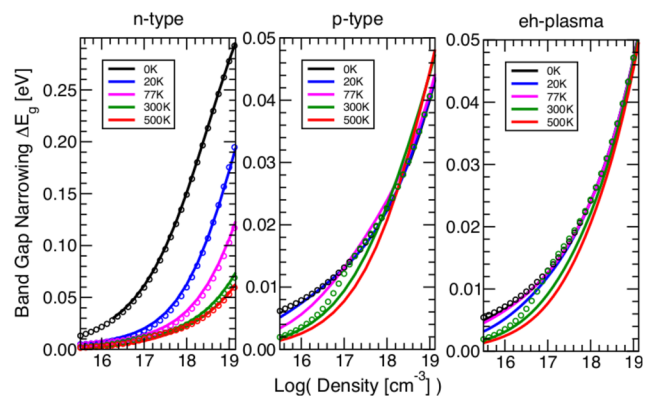

FIG. 18. Total BGN in $In_{0.53}Ga_{0.47}As$. Densities limited by activatable n-doping[80] (left), activatable p-doping[81] (middle), and E_L-E_Γ-separation (right).
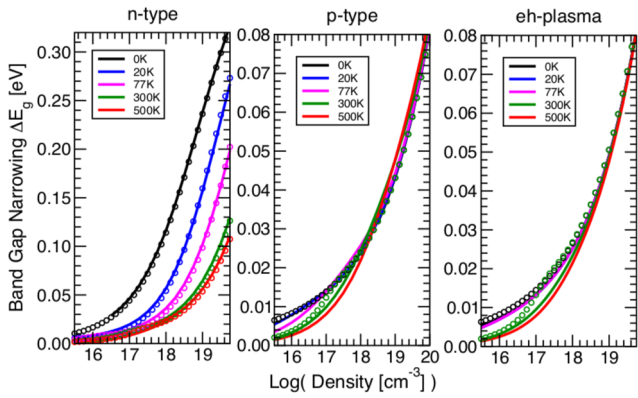
FIG. 19. Total BGN in $InAs_{0.4}P_{0.6}$. Densities limited by E_L-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).
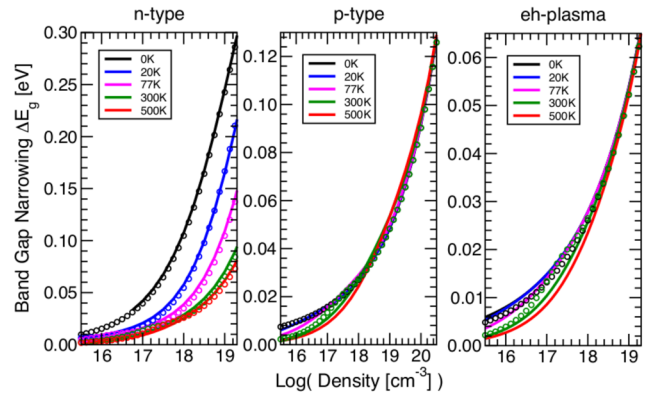

FIG. 22. Total BGN in $In_{0.52}Al_{0.48}As$. Densities limited by E_X-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).
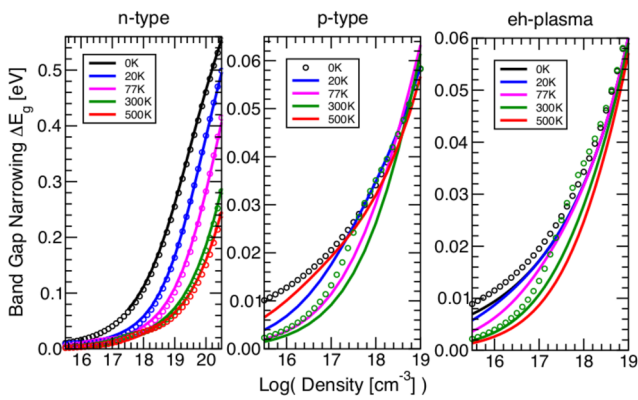
FIG. 20. Total BGN in zb-$GaIn_{0.5}N_{0.5}$. Densities limited by E_X-E_Γ-separation (left) and E_v-E_{so}-separation (middle and right).
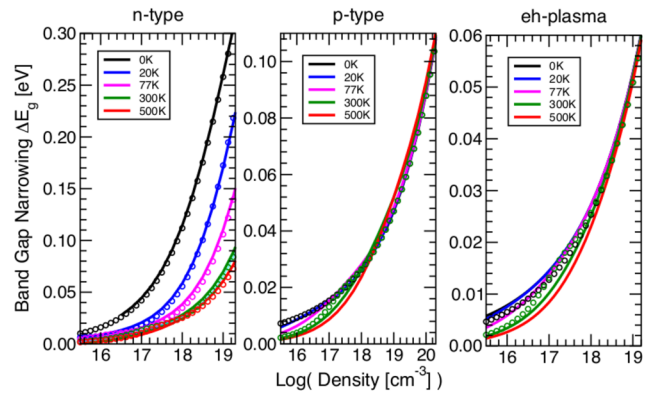
FIG. 23. Total BGN in GaAs. Densities limited by activatable n-doping[59] (left), activatable p-doping[59] (middle), and E_L-E_Γ-separation (right).
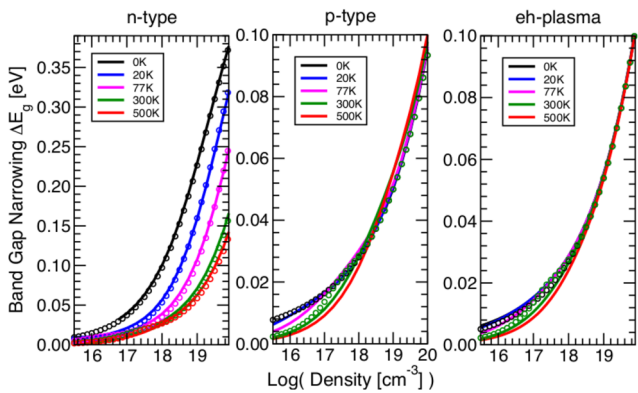
FIG. 21. Total BGN in InP. Densities limited by E_L-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle). Experimental data on the limits of activatable n-doping[57,58] cover the range $(5.5 \times 10^{19}$–$1.76 \times 10^{20})$ cm^{-3}.

FIG. 24. Total BGN in $Al_{0.3}Ga_{0.7}As$. Densities limited by E_L-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).

14 July 2023 14:39:17

FIG. 25. Total BGN in In$_{0.49}$Ga$_{0.51}$P. Densities limited by E_X-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).

FIG. 28. Total BGN in zb-GaN. Densities limited by activatable n-doping[66,68] (left), activatable p-doping[65,67] (middle), and E_v-E_{so}-separation (right).
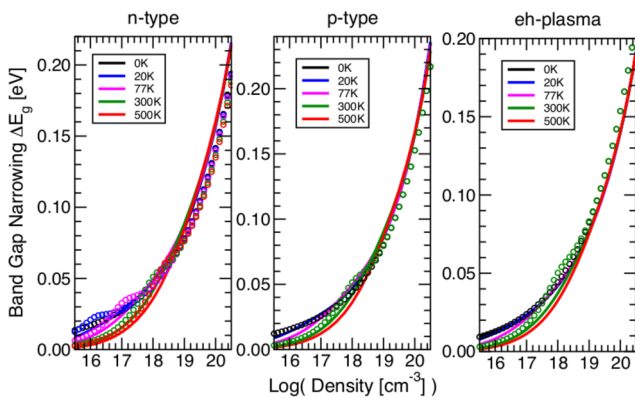
FIG. 26. Total BGN in AlAs. Densities limited by E_L-E_X-separation (left and right) and E_v-E_{so}-separation (middle).

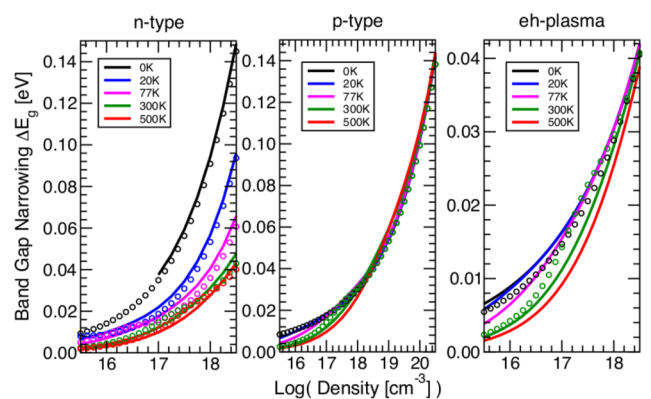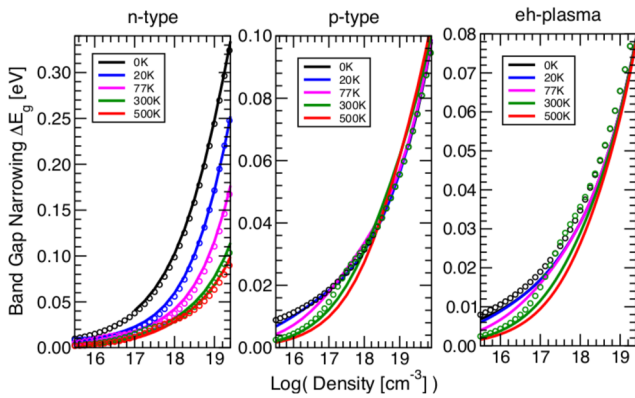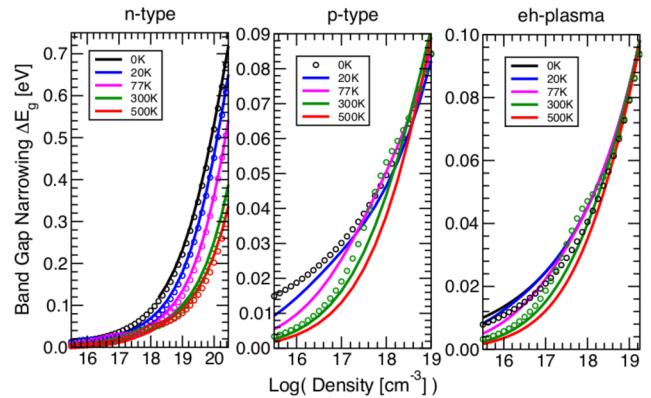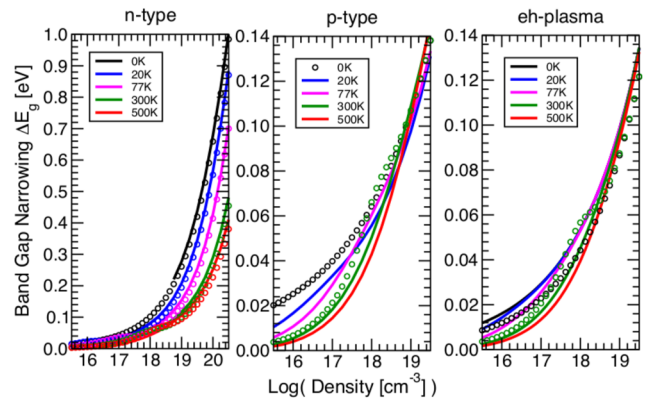**FIG. 29.** Total BGN in zb-AlN. Densities limited by E_X-E_Γ-separation (left and right) and E_v-E_{so}-separation (middle).

V. CONCLUSION

The BGN in quasi-neutral n-type and p-type regions and in a neutral electron–hole plasma of 20 technologically relevant zincblende III–V semiconductors was calculated in full-RPA as a function of density in the temperature range from 0 to 500 K. The sevenfold (plasma) and sixfold (quasi-neutral regions) integration is enabled by a careful isotropization of the band dispersion, including the important CB nonparabolicity in n-type materials. The neglect of higher CB valleys and of the split-off VB imposed by the simplified band structure model results in upper bounds for the allowed density in each case. The available experimental limits of the activated doping are always smaller than these theoretical limits. The gap can never shrink to zero, even at zero temperature. An exception is n-type InSb, where measured doping limits would lead to a vanishing gap at 0 K.

At room temperature, the trend of the BGN as a function of material (gap) follows the trend of the DOS effective masses. At a density of 10^{18} cm^{-3}, the BGN increases from ~15 meV (InAsSb) to ~50 meV (AlN) with only a little difference between n-type, p-type, and eh-plasma. At the higher density of 10^{19} cm^{-3}, the BGN

FIG. 27. Total BGN in GaP. Densities limited by activatable n-doping[61–63] (left) and E_v-E_{so}-separation (middle and right).

14 July 2023 14:39:17

of p-type and eh-plasma is still similar with an average increase from ~30 meV (InAsSb) to ~100 meV (AlN), but the BGN in direct n-type material becomes larger due to the small electron mass (reduced screening) and increases from ~40 meV (InAsSb) to ~120 meV (AlN). An exception is given by the indirect materials AlAs and GaP due to their similar CB- and VB-DOS. The nitrides are strong outliers with respect to the VB DOS mass, which becomes visible as maxima in the BGN curve.

At very low temperatures, p-type material and eh-plasma show almost identical behavior at a density of 10^{18} cm^{-3}, which is not much different from that at 300 K. At the higher density of 10^{19} cm^{-3}, p-type material and eh-plasma are still comparable, but in direct n-type materials, the BGN becomes giant as a consequence of the weak screening of the hole-ion interaction by the majority of electrons. At a density of 10^{18} cm^{-3}, the BGN is relatively independent of material, taking values between ~50 meV and ~80 meV at 20 K. At 10^{19} cm^{-3}, it increases from ~130 meV (InAsSb) to ~200 meV (AlN), again with the exception of the indirect materials AlAs and GaP.

The strong temperature dependence of the BGN in n-type materials with small CB DOS is traced back to the occupation probability of minority holes. It persists at all electron concentrations. In p-type material and eh-plasma, the temperature dependence of the BGN is much weaker; it is only significant at low and intermediate densities and always ceases at high densities.

The second major result of the paper is the derivation of an analytical BGN model without material-dependent free fit parameters, applicable to all densities and temperatures. Compared to numerical full-RPA, it reduces the central processing unit (CPU) time from days (the worst case is a very low but non-zero temperature and a high density) to less than seconds. The analytical model reveals the dependencies on band-structure and material parameters and allows for a quick calculation of the BGN in materials not explicitly treated in this paper. The only restriction here is compliance with an isotropic dispersion model (which excludes wurtzite materials). Besides other model simplifications (e.g., neglect of disorder-induced DOS tails and polaronic mass enhancement), isotropization is the key not only for the numerical full-RPA calculations but also for the analytical model. After the exact angular integrations, threefold integrals over complicated functions are approximately solved for the free-carrier correlation energies at 0 K and for the non-perturbative temperature-dependent hole-ion correlation energy, including the nonparabolicity of the CB. This is enabled by taking advantage of the ratio between Fermi wavelength and Bohr radius and by applying other approximation methods. As a result, the functional dependence of the BGN(n,p,T) on material and band structure parameters can be preserved. Comparison between analytical and numerical curves shows very good agreement in many cases, in particular for the density and temperature dependence of the BGN in n-type materials. Larger deviations occur in the case of n-type InAsSb at high densities (extreme nonparabolicity), for the indirect materials AlAs and GaP (due to their comparable CB and VB DOSs), and for some p-type materials such as InAsSb, InSb, and the nitrides.

A serious attempt has been made to express the BGN of ternary compounds by a composition average of the BGNs of their binary constituents (without or with bowing), as is common for the bandgap. However, although possible to some extent for a particular density and temperature, it fails for a wider density interval. This is probably due to the involved dependence on band filling and screening.

AUTHOR DECLARATIONS

Conflict of Interest

The author has no conflicts to disclose.

Author Contributions

A. Schenk: Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

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

APPENDIX A: BANDGAP NARROWING AS FUNCTION OF DENSITY IN THE TEMPERATURE RANGE FROM 0 TO 500 K FOR ALL 60 CASES

This appendix presents the BGN for all 20 materials (n-type and p-type) and for the neutral eh-plasma in these materials as a continuous function of density from $10^{15.5}$ cm^{-3} to the maximum at the temperatures $T = 0$, 20, 77, 300, and 500 K. The full-RPA curves for $T = 0$ K (black solid) start at the Mott density. The end point of each density axis corresponds to the allowed density maximum according to the measured limits of activated doping or to the limitations of the band-structure model. Solid lines are used for numerical full-RPA, i.e., Eqs. (6)–(8) with the RPA dielectric function Eq. (9) and densities calculated with Eq. (10) (parabolic case) or Eq. (11) (non-parabolic case), symbols are used for the analytical model outlined in Sec. IV.

APPENDIX B: ANALYTICAL APPROXIMATION OF THE FREE-CARRIER CORRELATION ENERGIES

This appendix provides the derivation of the analytical expressions of the free-carrier correlation energies in quasi-neutral regions [second and third terms in Eq. (20)] and in the neutral eh-plasma [second and third terms in Eq. (26)].

1. Quasi-neutral regions

The zero-temperature correlation energy of free carriers is calculated by Eq. (7) in the parabolic band approximation (justification given in the main text). The integration of angles can be performed exactly. Furthermore, the limit $\beta \to \infty$ is utilized. First, in the factor $\partial \epsilon / \partial n$, the derivative of the distribution function of majority carriers can be generated by partial integration, which turns into a delta-function and removes one integral. Second, the Matsubara sum is replaced by $\sum_{\nu} \to 2\beta/\pi \int_0^\infty dt$ with $t = \pi\nu/\beta$. In the momentum integration, the variable q is normalized by the Fermi momentum of the majority carriers: $q \to zq_{F,b} = z\sqrt{\zeta_b/\alpha_b}$. This results in

$$\lim_{\beta\to\infty}\Delta_a^c = -\frac{2}{\pi^2\sqrt{\alpha_a\zeta_a}}\int_0^\infty\frac{dz}{z}\int_0^\infty dt\times\left[1-\epsilon_{T=0}^{-1}\left(z\sqrt{\zeta_b/\alpha_b},t\right)\right]$$
$$\times\left\{\tanh^{-1}\left[\frac{z}{1+z^2/4+t^2/(z^2\zeta_b^2)}\right]\delta_{ab}\right.$$
$$\left.+\left(\frac{\alpha_a\zeta_b}{\alpha_b}\right)^{3/2}\sqrt{\zeta_a}\frac{z^3}{t^2+\left(\frac{\alpha_a\zeta_b}{2\alpha_b}\right)^2z^4}\left(1-\delta_{ab}\right)\right\}. \qquad (B1)$$

The RPA dielectric function (9) becomes fully analytical at $T = 0$ K,

$$\epsilon_{T=0}\left(z\sqrt{\zeta_b/\alpha_b},t\right) = 1$$
$$+\frac{s_b}{z^2}\left\{1+\frac{1}{z}\left(1-\frac{z^2}{4}+\frac{t^2}{z^2\zeta_b^2}\right)\tanh^{-1}\left[\frac{z}{1+\frac{z^2}{4}+\frac{t^2}{z^2\zeta_b^2}}\right]\right.$$
$$\left.-\frac{t}{z\zeta_b}\left[\tan^{-1}\left(\frac{z\zeta_b}{2t}(2+z)\right)+\tan^{-1}\left(\frac{z\zeta_b}{2t}(2-z)\right)\right]\right\}\Theta(\zeta_b), \qquad (B2)$$

with the normalized length s_b defined in Eq. (18),

$$s_b = \frac{g_b}{\pi\alpha_b q_{F,b}} = \frac{g_b}{\pi\sqrt{\alpha_b\zeta_b}}. \qquad (B3)$$

Note that only majority carriers (index "b") contribute to free-carrier screening in quasi-neutral regions. Expression (B2) is obtained from Eq. (A6) in Ref. 7 in the parabolic and zero-temperature limits. It must be drastically simplified to enable the double integration of momentum (z) and frequency (t) in (B1). For that purpose, the single plasmon pole (SPP) approximation[3,10,39,40] is a good candidate. Here, it is used for a one-component plasma,

$$\epsilon_{spp}(q,\nu) = 1 + \frac{\omega_p^2}{\omega_p^2\frac{q^2}{\kappa^2}+\left(\frac{\hbar q^2}{2m_b}\right)^2-\Omega_\nu^2}, \qquad (B4)$$

where ω_p is the plasmon frequency and κ is the Thomas-Fermi momentum (the static Debye limit is $\epsilon_D = 1 + \kappa^2/q^2$). After normalization, one obtains for the case of a degenerate carrier gas at $T = 0$ K,

$$\hbar^2\omega_p^2 = 16\pi\alpha_b n_b = \frac{8g_b}{3\pi\sqrt{\alpha_b}}\zeta_b^{3/2}, \qquad (B5)$$

$$\kappa^2 = 8\pi\partial n_b/\partial\zeta_b = \frac{2g_b}{\pi\alpha_b^{3/2}}\sqrt{\zeta_b}, \qquad (B6)$$

and

$$\epsilon_{spp}\left(z\sqrt{\zeta_b/\alpha_b},t\right) = 1 + \frac{2s_b}{z^2+3\left(\frac{z^4}{4}+\frac{t^2}{\zeta_b^2}\right)}. \qquad (B7)$$

Note that ϵ_{spp} is not used in the factor $\partial\epsilon/\partial n$ of Eq. (B1) but only in the screening function $(1-1/\epsilon)$, which reads

$$1-\epsilon_{spp}^{-1}\left(z\sqrt{\zeta_b/\alpha_b},t\right) = \frac{2s_b}{2s_b+z^2+3\left(\frac{z^4}{4}+\frac{t^2}{\zeta_b^2}\right)}. \qquad (B8)$$

The correlation energy (B1) takes the form

$$\Delta_a^c = -\frac{4g_b}{\pi^3\alpha_b}\int_0^\infty dz\int_0^\infty d\tau\frac{1}{2s_b+z^2+3\left(\frac{z^4}{4}+\tau^2\right)}$$
$$\times\left\{\frac{1}{z}\tanh^{-1}\left[\frac{z^3}{z^2+\frac{z^4}{4}+\tau^2}\right]\delta_{ab}\right.$$
$$\left.+\left(\frac{\alpha_a}{\alpha_b}\right)\frac{z^2(1-\delta_{ab})}{\tau^2+\left(\frac{\alpha_a}{2\alpha_b}\right)^2z^4}\right\}, \qquad (B9)$$

where the integration variable t was changed to $\tau = t/\zeta_b$. The goal is to approximate the integrand such that the correct dependence on density (contained in s_b) is preserved as well as possible. After testing various ways, the following turned out to be the best.

a. Majority-carrier band edge

For the majority-carrier band edge (first term in curly braces), the frequency dependence in the denominator of the SPP screening function [first line in Eq. (B9)] is skipped ($\tau \to 0$). This allows us to perform the remaining τ-integration exactly, leading to

$$\Delta_{a,maj}^c \approx -\frac{2g_b\delta_{ab}}{\pi^2\alpha_b}\int_0^\infty dz\frac{z\Theta(2-z)+2\Theta(z-2)}{2s_b+z^2+3z^4/4}. \qquad (B10)$$

The numerator of the integrand is replaced by $z\Theta(2-z)+2\Theta(z-2) \to 2$, which results in

$$\Delta_{a,maj}^c \approx -\frac{4g_b\delta_{ab}}{3\pi^2\alpha_b}\int_0^\infty dz\frac{1}{\frac{2}{3}s_b+\frac{1}{3}z^2+\frac{1}{4}z^4}, \qquad (B11)$$

and after exact integration in the analytical expression

$$\Delta_{a,maj}^c \approx -\frac{2g_b\delta_{ab}}{\pi\alpha_b\sqrt{s_b}}\frac{1}{\sqrt{1+\sqrt{1-6s_b}}+\sqrt{1-\sqrt{1-6s_b}}}. \qquad (B12)$$

To obtain the density dependence in the form of a simple power law, the limits $s_b \ll 1/6$ and $s_b \gg 1/6$ are considered,

$$\Delta_{a,maj}^c \approx -\frac{g_b\delta_{ab}}{\pi\alpha_b}\times\begin{cases}\sqrt{2/s_b} & s_b\ll 1/6,\\\left(\frac{2}{3}\right)^{\frac{1}{4}}\frac{1}{s_b^{3/4}} & s_b\gg 1/6.\end{cases} \qquad (B13)$$

The first limit (high density, short Fermi wavelength) yields an $n^{1/6}$-dependence, and the second limit (low density, long Fermi wavelength) yields an $n^{1/4}$-dependence. In Fig. 30, s_e and s_h at a density of 10^{19} cm^{-3} are plotted for all materials. Neither of the two limits is actually valid, as one can see by comparison with the green dashed line (1/6). A possible simple fit to the full-RPA curves (compare Fig. 9) is the following: (i) the limit $s_b \ll 1/6$ is chosen, (ii) a logarithmic weakening[5,36,38] is performed, and (iii) the numerical pre-factor is adjusted. The result is [the second term of Eq. (20)],

$$\Delta_{a,maj}^c \approx -\frac{2g_b\delta_{ab}}{3\pi^2\alpha_b}\ln\left(1+\sqrt{2/s_b}\right), \qquad (B14)$$
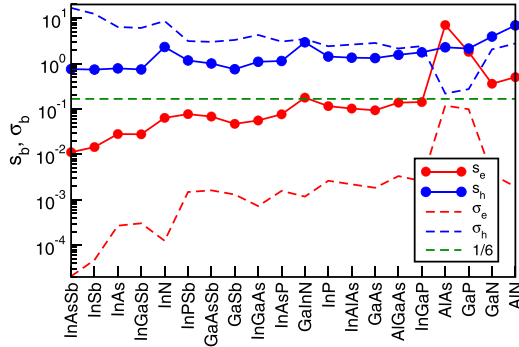
i.e., a logarithmic weakened $n^{1/6}$-dependence.

14 July 2023 14:39:17

FIG. 30. Length parameters s_b and σ_b of zincblende III–V materials (ordered as in Fig. 2) for a concentration of 10^{19} cm^{-3}.

b. Minority-carrier band edge

For the minority-carrier band edge [second term in the curly braces of Eq. (B9)], the τ-integration can be performed exactly, and one is left with the single integral,

$$
\Delta_{a,min}^{c} = -\frac{4g_b\left(1-\delta_{ab}\right)}{3\pi^2\alpha_b}
$$
$$
\times \int_0^\infty dz \frac{1}{\frac{\alpha_a z^2}{2\alpha_b}\sqrt{\frac{2}{3}s_b + \frac{z^2}{3} + \frac{z^4}{4}} + \frac{2}{3}s_b + \frac{z^2}{3} + \frac{z^4}{4}}. \quad (B15)
$$

The integrand is replaced by a simpler and integrable function that yields the correct limits $z \to 0$ and $z \to \infty$ and that closely follows the course at intermediate z,

$$
\Delta_{a,min}^{c} \approx -\frac{4g_b\left(1-\delta_{ab}\right)}{3\pi^2\alpha_b} \int_0^\infty dz \frac{1}{\frac{2}{3}s_b + \frac{z^2}{3} + \frac{z^4}{4\alpha_b}}. \quad (B16)
$$

Except for the occurrence of α_b in the last term of the denominator, this approximate integrand equals the integrand in Eq. (B11). Proceeding as for the majority-carrier band edge, one obtains

$$
\Delta_{a,min}^{c} \approx -\frac{g_b\left(1-\delta_{ab}\right)}{\pi\alpha_b} \times \begin{cases} \sqrt{2/s_b} & s_b \ll 1/6, \\ \left(\frac{2\alpha_b}{3}\right)^{\frac{1}{4}} \frac{1}{s_b^{3/4}} & s_b \gg 1/6. \end{cases} \quad (B17)
$$

A feasible power-law fit to the full-RPA curves (compare Fig. 9) is obtained as follows: (i) the limit $s_b \gg 1/6$ is chosen, and (ii) the numerical pre-factor is adjusted. The result is [the third term of Eq. (20)],

$$
\Delta_{a,min}^{c} \approx -\frac{g_b\left(1-\delta_{ab}\right)}{\sqrt{3}\pi\left(\alpha_b s_b\right)^{3/4}}, \quad (B18)
$$

i.e., an $n^{1/4}$-dependence.

It should be noted that the power-law fits for majority and minority band edges are somewhat arbitrary, as the applicability of the limits strongly depends on material and density. This might explain the many variants suggested in the past, e.g., $n^{1/6}$ in Ref. 37, $n^{1/4}$ in Ref. 38, and $n^{7/30}$ in Refs. 38 and 41.

2. Neutral electron–hole plasma

Screening by electrons *and* holes with density $n = n_e = n_h$ determines the free-carrier correlation energies in a neutral eh-plasma. Analytical modeling proceeds similarly to the case of a one-component plasma. Charge neutrality imposes a simple relation between the quasi-Fermi levels (parabolic dispersion, $T = 0$ K),

$$
\zeta_b = \left(\frac{g_a}{g_b}\right)^{2/3} \frac{\alpha_b}{\alpha_a} \zeta_a. \quad (B19)
$$

In the screening function $(1 - 1/\epsilon)$, the SPP approximation is now applied to a two-component plasma,

$$
\epsilon_{spp,eh}(q,\nu) = 1 + \frac{\omega_P^2}{\omega_P^2 \frac{q^2}{\kappa^2} + \left(\frac{\hbar q^2}{4\mu^*}\right)^2 - \Omega_\nu^2}. \quad (B20)
$$

Using relation (B19), plasmon frequency ω_P and Thomas-Fermi momentum κ at $T = 0$ K are given by

$$
\hbar^2\omega_P^2 = \frac{8}{3\pi}\left(\frac{g_e\zeta_e^{\frac{3}{2}}}{\sqrt{\alpha_e}} + \frac{g_h\zeta_h^{\frac{3}{2}}}{\sqrt{\alpha_h}}\right) = \frac{8g_e}{3\pi}\left(\frac{\zeta_e}{\alpha_e}\right)^{\frac{3}{2}} = \frac{8g_h}{3\pi}\left(\frac{\zeta_h}{\alpha_h}\right)^{\frac{3}{2}}, \quad (B21)
$$

$$
\kappa^2 = \frac{2}{\pi}\left(\frac{g_e\sqrt{\zeta_e}}{\alpha_e^{3/2}} + \frac{g_h\sqrt{\zeta_h}}{\alpha_h^{3/2}}\right)
$$
$$
= \frac{2}{\pi}\frac{g_e\sqrt{\zeta_e}}{\alpha_e^{3/2}}\left[1 + \left(\frac{g_h}{g_e}\right)^{\frac{2}{3}}\frac{\alpha_e}{\alpha_h}\right] = \frac{2}{\pi}\frac{g_h\sqrt{\zeta_h}}{\alpha_h^{3/2}}\left[1 + \left(\frac{g_e}{g_h}\right)^{\frac{2}{3}}\frac{\alpha_h}{\alpha_e}\right]. \quad (B22)
$$

The screening function reads

$$
1 - \epsilon_{spp,eh}^{-1}\left(z\sqrt{\zeta_b/\alpha_b}, t\right) = \frac{2s_b\alpha_b}{2s_b\alpha_b + z^2\alpha_{ba} + 3\left[\left(\frac{z}{2}\right)^4 + \left(\frac{t\alpha_b}{\zeta_b}\right)^2\right]}, \quad (B23)
$$

with

$$
\alpha_{ba} = \frac{\alpha_a}{1 + \left(\frac{g_a}{g_b}\right)^{\frac{2}{3}}\frac{\alpha_b}{\alpha_a}}. \quad (B24)
$$

The form of the correlation energy at zero temperature is similar to the majority-carrier term in Eq. (B1),

$$
\Delta_b^c = -\frac{2}{\pi^2\sqrt{\alpha_b\zeta_b}}\int_0^\infty \frac{dz}{z}\int_0^\infty dt\left[1 - \epsilon_{spp,eh}^{-1}\left(z\sqrt{\zeta_b/\alpha_b}, t\right)\right]
$$
$$
\times \tanh^{-1}\left[\frac{z}{1 + \left(\frac{z}{2}\right)^2 + \left(\frac{t}{z\zeta_b}\right)^2}\right]. \quad (B25)
$$

With the same approximations as for Eqs. (B10) and (B11), one obtains

$$
\Delta_b^c \approx -\frac{4g_b\alpha_b}{3\pi^2}\int_0^\infty dz \frac{1}{\frac{2}{3}s_b\alpha_b + \frac{1}{3}\alpha_{ba}z^2 + \frac{1}{16}z^4}, \quad (B26)
$$

where α_{ba} is defined in Eq. (B24). Exact integration yields the analytical expression

$$\Delta_b^c \approx -\frac{2g_b}{\pi\sqrt{s_b}}\frac{\sqrt{\alpha_b/\alpha_{ba}}}{\sqrt{1+\sqrt{1-s_b/\sigma_b}}+\sqrt{1-\sqrt{1-s_b/\sigma_b}}}, \quad (B27)$$

with

$$\sigma_b = \frac{2\alpha_{ba}^2}{3\alpha_b}. \quad (B28)$$

Again, the limits of very small and very large s_b are considered in order to simplify the model,

$$\Delta_b^c \approx -\frac{g_b}{\pi} \times \begin{cases} \sqrt{\dfrac{\alpha_b}{\alpha_{ba}}}\sqrt{2/s_b} & s_b \ll \sigma_b, \\ \left(\dfrac{8\alpha_b}{3}\right)^{\frac{1}{4}}\dfrac{1}{s_b^{3/4}} & s_b \gg \sigma_b. \end{cases} \quad (B29)$$

As can be seen from Fig. 30, the assumption $s_e \gg \sigma_e$ is fulfilled for all materials, whereas $s_h \gg \sigma_h$ is only fulfilled for the indirect materials AlAs and GaP. For the antimonides and InAs, it rather holds that $s_h \ll \sigma_h$. Inspection of the full-RPA correlation energies at a density of 10^{19} cm^{-3} shows that $\Delta_e^c \approx \Delta_h^c$ is only for the indirect materials AlAs and GaP; otherwise, the ratio Δ_h^c/Δ_e^c increases from 2 to 6 with decreasing gap. Therefore, the same simple power law for Δ_h^c cannot work in all 20 cases.

To achieve a good agreement with the full-RPA curves in all 20 cases, a distinction is made based on the parameter $\alpha_e = m_h/(m_h + m_e)$: if $\alpha_e > 0.85$ (materials No. 1–11), the limit $s_h \ll \sigma_h$ is used; if $\alpha_e < 0.85$ (materials No. 12–20), the limit $s_h \gg \sigma_h$ is applied. A logarithmic weakening[5,36,38] is performed, and the numerical pre-factor is adjusted independently for electrons and holes. The result is [the second and third terms of Eq. (26)],

$$\Delta_e^c + \Delta_h^c \approx -\frac{2g_e}{5\pi}\ln\left[1+\left(\frac{8\alpha_e}{3}\right)^{\frac{1}{4}}\frac{1}{s_e^{3/4}}\right]$$
$$-\frac{2g_h}{\pi}\left\{\Theta(0.85-\alpha_e)\ln\left[1+\left(\frac{8\alpha_h}{3}\right)^{\frac{1}{4}}\frac{1}{s_h^{3/4}}\right]\right.$$
$$\left.+\Theta(\alpha_e-0.85)\ln\left(1+\sqrt{\frac{2\alpha_h}{\alpha_{he}s_h}}\right)\right\}. \quad (B30)$$

The outcome of this general analytical model is displayed for the free-carrier correlation energy in InGaAs (Fig. 9) and as part of the total BGN of the neutral eh-plasma in all materials in the right panel of Figs. 10–29.

APPENDIX C: ANALYTICAL APPROXIMATION OF THE ION-CARRIER CORRELATION ENERGIES

This appendix provides the derivation of the analytical expressions of the ion-carrier correlation energies in quasi-neutral regions including the effect of nonparabolicity in n-type material, and the temperature dependence of the donor-hole correlation energy including the effect of nonparabolicity on this temperature dependence.

The correlation energy of the ion-carrier interaction is calculated from Eq. (8). After normalization and integration over angles, it is given by

$$\Delta_a^i = -n_i\frac{2}{\pi}\left(\frac{\partial n_a}{\partial\zeta_a}\right)^{-1}\int_0^\infty dq\frac{1}{\epsilon^2(q,0)}\frac{\partial\epsilon(q,0)}{\partial\zeta_a}. \quad (C1)$$

The static RPA dielectric function reads [compare Ref. 7, Eqs. (A7)–(A9)],

$$\epsilon(q,0) = 1 + \frac{1}{\pi q^3}\sum_a\frac{g_a}{\alpha_a^2}\int_0^\infty d\kappa\kappa f_a(\kappa)\left\{2[h_a(\kappa,q)-h_a(\kappa,-q)]\right.$$
$$\left.+ h_a(\kappa,0)\ln\frac{[h_a(\kappa,0)-h_a(\kappa,q)]^2}{[h_a(\kappa,0)-h_a(\kappa,-q)]^2}\right\}, \quad (C2)$$

where

$$h_a(\kappa,q) = \sqrt{1+4\gamma_a\left(\kappa+\sqrt{\alpha_a}q\right)^2}, \quad (C3)$$

$$f_a(\kappa) = \left\{1+e^{\beta[h_a(\kappa,0)-1]/2\gamma_a-\beta\zeta_a}\right\}^{-1}, \quad (C4)$$

have to be used. For the derivative of $\epsilon(q,0)$ with respect to the Fermi energy, one needs

$$\frac{\partial f_a}{\partial\zeta_a} = -\frac{\partial f_a}{\partial\kappa}\frac{h_a(\kappa,0)}{2\kappa} = \beta f_a(\kappa)[1-f_a(\kappa)], \quad (C5)$$

under the κ-integral. To obtain the $T = 0$ K-limit, the κ-integral in Eq. (C2) is integrated by parts, where the derivative of the distribution function with respect to κ becomes a delta function,

$$\frac{\partial f_a}{\partial\kappa} = -\delta\left[\kappa-\sqrt{\zeta_a(1+\gamma_a\zeta_a)}\right]. \quad (C6)$$

Then

$$\epsilon_{T=0}(q,0) = 1 + \frac{g_b}{\pi\alpha_b^2 q^3}\int_0^{\sqrt{\tilde\zeta_b}}d\kappa\kappa\left\{2[h_b(\kappa,q)-h_b(\kappa,-q)]\right.$$
$$\left.+ h_b(\kappa,0)\ln\frac{[h_b(\kappa,0)-h_b(\kappa,q)]^2}{[h_b(\kappa,0)-h_b(\kappa,-q)]^2}\right\}\Theta(\zeta_b), \quad (C7)$$

with the abbreviation $\tilde\zeta_b = \zeta_b(1+\gamma_b\zeta_b)$ and "b" denoting the index of the majority carrier band. Only majority carriers contribute to screening in quasi-neutral regions ($\zeta_b > 0$), whereas minority carriers are frozen out. The derivative of the densities with respect to the Fermi energy becomes

$$\lim_{\beta\to\infty}\frac{\partial n_a}{\partial\zeta_a} = \frac{g_a}{4\pi^2\alpha_a^{3/2}}$$
$$\times\begin{cases}\sqrt{\tilde\zeta_a}(1+2\gamma_a\zeta_a)\Theta(\zeta_a) & \text{majority band,} \\ \dfrac{\sqrt{\pi}}{2}\dfrac{\exp(\beta\zeta_a)}{\sqrt{\beta}}\bigg|_{\beta\to\infty}\Theta(-\zeta_a) & \text{minority band.}\end{cases}$$
$$(C8)$$

For the derivative of the static dielectric function with respect to the Fermi energy $\partial\epsilon(q,0)/\partial\zeta_a$ in the limit $\beta \to \infty$, one needs

$$\lim_{\beta\to\infty}\frac{\partial f_a}{\partial\zeta_a} = -\lim_{\beta\to\infty}\frac{h(\kappa,0)}{2\kappa}\frac{\partial f_a}{\partial\kappa}$$

$$= \begin{cases} \dfrac{h(\kappa,0)}{2\kappa}\delta\!\left(\kappa - \sqrt{\tilde{\zeta}_a}\right)\Theta(\zeta_a) & \text{majority band,} \\[2ex] \beta e^{-\beta(\kappa^2-\zeta_a)}\Big|_{\beta\to\infty}\Theta(-\zeta_a) & \text{minority band,} \end{cases} \quad \text{(C9)}$$

which follows from Eq. (C5) neglecting nonparabolicity for the minority carriers. Using this in Eq. (C2) leads to

$$\lim_{\beta\to\infty}\frac{\partial\epsilon(q,0)}{\partial\zeta_a} = \frac{1}{\pi q^3}\frac{g_a}{\alpha_a^2}h_a^2\!\left(\sqrt{\tilde{\zeta}_a},0\right)$$

$$\times \left\{ \frac{h_a\!\left(\sqrt{\tilde{\zeta}_a},q\right) - h_a\!\left(\sqrt{\tilde{\zeta}_a},-q\right)}{h_a\!\left(\sqrt{\tilde{\zeta}_a},0\right)} \right.$$

$$+ \frac{1}{2}\ln\!\left(\frac{\left[h_a\!\left(\sqrt{\tilde{\zeta}_a},0\right) - h_a\!\left(\sqrt{\tilde{\zeta}_a},q\right)\right]^2}{\left[h_a\!\left(\sqrt{\tilde{\zeta}_a},0\right) - h_a\!\left(\sqrt{\tilde{\zeta}_a},-q\right)\right]^2}\right)\left.\right\}\Theta(\zeta_a)$$

$$+ \lim_{\beta\to\infty}\frac{1}{\pi q^3}\frac{g_a}{\alpha_a^2}\beta e^{\beta\zeta_a}\int_0^\infty d\kappa\,\kappa e^{-\beta\kappa^2}$$

$$\times \ln\!\left(\frac{\left[\kappa^2 - (\kappa + \sqrt{\alpha_a}q)^2\right]^2}{\left[\kappa^2 - (\kappa - \sqrt{\alpha_a}q)^2\right]^2}\right)\Theta(-\zeta_a), \quad \text{(C10)}$$

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

$$\lim_{\beta\to\infty}\beta e^{\beta\zeta_a}\int_0^\infty d\kappa\,\kappa e^{-\beta\kappa^2}\ln\!\left(\frac{\left[\kappa^2 - (\kappa + \sqrt{\alpha_a}q)^2\right]^2}{\left[\kappa^2 - (\kappa - \sqrt{\alpha_a}q)^2\right]^2}\right)$$

$$= \lim_{\beta\to\infty}2\beta e^{\beta\zeta_a}\int_0^\infty dt\,t\,e^{-t^2}\tanh^{-1}\!\left(\frac{4t}{\sqrt{\alpha_a\beta}q}\right)$$

$$= \lim_{\beta\to\infty}\frac{2\sqrt{\pi}e^{\beta\zeta_a}}{\sqrt{\alpha_a\beta}q}. \quad \text{(C11)}$$

For the second line, the relation $\ln(|x+1|/|x-1|) = 2\tanh^{-1}(1/x)$ was used. Inserting Eq. (C10) with (C11) and Eq. (C8) into Eq. (C1), the correlation energy of ion-carrier interaction in the $T = 0$ K-limit becomes

$$\lim_{\beta\to\infty}\Delta_a^i = -n_i\frac{8}{\sqrt{\alpha_a\tilde{\zeta}_a}}\int_0^\infty dq\frac{\Theta(\zeta_a)}{q^3\epsilon_{T=0}^2(q,0)}$$

$$\times \left\{ \left[h_a\!\left(\sqrt{\tilde{\zeta}_a},q\right) - h_a\!\left(\sqrt{\tilde{\zeta}_a},-q\right)\right] \right.$$

$$+ \frac{1}{2}(1 + 2\gamma_a\zeta_a)\ln\!\left(\frac{\left[1 + 2\gamma_a\zeta_a - h_a\!\left(\sqrt{\tilde{\zeta}_a},q\right)\right]^2}{\left[1 + 2\gamma_a\zeta_a - h_a\!\left(\sqrt{\tilde{\zeta}_a},-q\right)\right]^2}\right)\left.\right\}$$

$$- n_i\frac{32}{\alpha_a}\int_0^\infty dq\frac{\Theta(-\zeta_a)}{q^4\epsilon_{T=0}^2(q,0)}, \quad \text{(C12)}$$

where the static dielectric function $\epsilon_{T=0}(q,0)$ is defined in Eq. (C7) and the function h_a in Eq. (C3). Introducing normalized momentum variables of majority carriers by $z = q/q_{F,b} = z\sqrt{\alpha_b/\tilde{\zeta}_b}$ and $t = \kappa/\sqrt{\tilde{\zeta}_b}$, this can be written in the form

$$\lim_{\beta\to\infty}\Delta_a^i = -\frac{2g_b}{3\pi^2}\int_0^\infty dz\frac{\frac{z}{\alpha_b}H_b(1,z)\delta_{ab} + \frac{8}{\alpha_a}(1 - \delta_{ab})}{\left[z^2 + \frac{s_b}{z}\int_0^1 dt\,t\,H_b(t,z)\right]^2}\Theta(\zeta_b), \quad \text{(C13)}$$

with

$$s_b = \frac{g_b}{\pi\sqrt{\alpha_b\tilde{\zeta}_b}},$$

$$H_b(t,z) = 2[h_b(t,z) - h_b(t,-z)]$$

$$+ h_b(t,0)\ln\!\left(\frac{[h_b(t,0) - h_b(t,z)]^2}{[h_b(t,0) - h_b(t,-z)]^2}\right), \quad \text{(C14)}$$

$$h_b(t,z) = \sqrt{1 + 4\gamma_b\tilde{\zeta}_b(t+z)^2},$$

$$\tilde{\zeta}_b = \zeta_b(1 + \gamma_b\zeta_b),$$

after inserting (C7) into Eq. (C12).

Analytical integration necessitates the simplification of the function $H_b(t,z)$. 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 [compare Eqs. (18) and (19)],

$$s_b = \frac{g_b^{4/3}}{(6\pi^5)^{1/3}\alpha_b}\frac{1}{n_i^{1/3}}. \quad \text{(C15)}$$

Therefore, one can assume $s_b \ll 1$ for very large n_i, which results in a pronounced maximum of the integrand at some $z < 1$. The function $H_b(t,z)$ can then be linearized in z, which results in

$$H_b(t,z) \approx \frac{2z}{t}\left[h_b(t,0) + \frac{4\gamma_b\tilde{\zeta}_b t^2}{h_b(t,0)}\right]. \quad \text{(C16)}$$

With that, the t-integration in the dielectric function can be performed easily, giving

$$\frac{s_b}{z}\int_0^1 dt\,t\,H_b(t,z)$$

$$= s_b\left\{2h_b(1,0) + \left[h_b(1,0) - \frac{\sinh^{-1}\!\left(\sqrt{4\gamma_b\tilde{\zeta}_b}\right)}{\sqrt{4\gamma_b\tilde{\zeta}_b}}\right]\right\}. \quad \text{(C17)}$$

The second term is negligible compared to the first one as long as $4\gamma_b\tilde{\zeta}_b \ll 1$ can be assumed, meaning that the Fermi energy in the corresponding parabolic band must be much smaller than the bandgap. For InAsSb and InSb, this becomes critical at higher densities, as can be seen in Figs. 10 and 11. With the linearized form of $H_b(1,z)$ in the numerator of Eq. (C13) and inserting

$$\frac{s_b}{z}\int_0^1 dt\,t\,H_b(t,z) \approx 2s_b h_b(1,0), \quad \text{(C18)}$$

in the denominator, the remaining z-integration is trivial and yields

$$\lim_{\beta \to \infty} \Delta_a^i = -\frac{2g_b}{3\pi} \left\{ \frac{\frac{1}{\alpha_b}\left[h_b(1,0) + \frac{8\gamma_b\tilde{\zeta}_b}{h_b(1,0)}\right]\delta_{ab}}{\sqrt{2s_b h_b(1,0)}} \right.$$
$$\left. + \frac{\frac{4}{\alpha_a}(1 - \delta_{ab})}{[2s_b h_b(1,0)]^{3/2}} \right\} \Theta(\zeta_b). \qquad (C19)$$

A last simplification is the replacement

$$\sqrt{h_b(1,0)} + \frac{8\gamma_b\tilde{\zeta}_b}{h_b^{3/2}(1,0)} \to h_b(1,0), \qquad (C20)$$

that closely matches the numerical RPA result. Using the notation $\varphi_{np}(s_b) = h_b(1,0)$ for the nonparabolicity correction factor finally results in the ionic correlation energies in Eq. (20). Note that $\gamma_h = 0$.

As discussed in Appendix B 1, the strong temperature dependence of the hole-ion correlation energy $\Delta_{h,n-type}^i(T)$ mainly arises from the occupation probability of holes, whereas the majority-carrier screening is only weakly T-dependent and vanishes at high concentrations. Therefore, Eq. (C1) becomes

$$\Delta_{h,n-type}^i = -n_i \frac{2}{\pi} \left(\frac{\partial n_h}{\partial \zeta_h}\right)^{-1} \int_0^\infty dq \frac{1}{\epsilon_{T=0}^2(q,0)} \frac{\partial \epsilon(q,0)}{\partial \zeta_h}. \qquad (C21)$$

Since $\gamma_h = 0$ is used throughout, nonparabolicity only needs to be taken into account in the factor $1/\epsilon_{T=0}^2(q,0)$ (screening by majority electrons). Furthermore, the T-dependent factors $(\partial n_h/\partial \zeta_h)^{-1}$ and $\partial \epsilon(q,0)/\partial \zeta_h$ can be calculated with Boltzmann statistics. From Eq. (10), one obtains after normalization

$$\left(\frac{\partial n_h}{\partial \zeta_h}\right)^{-1} = \frac{(4\pi\alpha_h)^{3/2}}{g_h}\sqrt{\beta}e^{-\beta\zeta_h}, \qquad (C22)$$

and from Eq. (C2) after partial integration

$$\frac{\partial \epsilon(q,0)}{\partial \zeta_h} = \frac{g_h}{2\pi\alpha_h^2 q^3}e^{\beta\zeta_h}\int_0^\infty d\kappa \, e^{-\beta\kappa^2}\frac{d}{d\kappa}H_h(\kappa,q), \qquad (C23)$$

with $H_h(\kappa,q)$ defined in Eq. (C14). Its parabolic version reads

$$H_h(\kappa,q) = \tanh^{-1}\left(\frac{\sqrt{\alpha_h}\kappa q}{\kappa^2 + \alpha_h q^2/4}\right). \qquad (C24)$$

Inserting

$$\frac{d}{d\kappa}H_h(\kappa,q) = -2q\sqrt{\alpha_h}\frac{\kappa^2 - \alpha_h q^2/4}{(\kappa^2 - \alpha_h q^2/4)^2}\Theta\left(\frac{q\sqrt{\alpha_h}}{2} - \kappa\right), \qquad (C25)$$

into Eq. (C23) and introducing the new integration variables $t = \kappa\sqrt{\beta}$ and $z = q\sqrt{\alpha_e}/\sqrt{\tilde{\zeta}_e}$, consequently,

$$\left(\frac{\partial n_h}{\partial \zeta_h}\right)^{-1}\frac{\partial \epsilon(q,0)}{\partial \zeta_h} = \frac{8\sqrt{\pi}\alpha_e}{z^2\tilde{\zeta}_e}\tau(z,\beta), \qquad (C26)$$

where

$$\tau(z,\beta) = \int_0^{za(\beta)} dt \frac{\beta e^{-t^2}}{a^2(\beta)z^2 - t^2}, \qquad (C27)$$

$$a^2(\beta) = \frac{\alpha_h\tilde{\zeta}_e}{4\alpha_e}\beta = \frac{\alpha_h g_e^2}{(2\pi\alpha_e s_e)^2}\beta. \qquad (C28)$$

The hole-ion correlation energy can now be written as

$$\Delta_{h,n-type}^i(T) = -16n_i\sqrt{\frac{\alpha_e}{\pi\tilde{\zeta}_e}}\int_0^\infty \frac{dz}{z^2}\frac{\tau(z,\beta)}{\epsilon_{T=0}^2(z,0)}. \qquad (C29)$$

Its zero-temperature limit[7] follows from $\lim_{\beta \to \infty}\tau(z,\beta)$ $= 2\sqrt{\pi}\alpha_e/(\alpha_h\tilde{\zeta}_e z^2)$.

In the low-T/high-density range, the denominator in (C27) can be taken at $t = 0$, and $H_e(t,z)$ can be linearized in z: $H_e(t,z) \approx 2zh_e(t,0)/t$. The dielectric function then reads $\epsilon_{T=0}(z,0) = 1 + \frac{s_e}{z^2}\int_0^1 dt \, t \, H_e(t,z) = 1 + 2s_e h_e(1,0)/z^2$, which results in

$$\Delta_{h,n-type}^i(T) = -\frac{16g_e}{3\pi^2\alpha_h}\int_0^\infty dz \frac{\text{erf}(za(\beta))}{[z^2 + 2s_e h_e(1,0)]^2}. \qquad (C30)$$

For an analytical solution of the integral in (C30), the error function is approximated as

$$\text{erf}(x) \approx \frac{2x}{\sqrt{\pi}}\Theta(1 - x) + \Theta(x - 1). \qquad (C31)$$

This results in

$$\Delta_{h,n-type}^i(T) = -\frac{\sqrt{2}g_e}{3\pi\alpha_h s_e^{3/2}\varphi_{np}^{3/2}(s_e)}$$
$$\times \left\{1 - \frac{2}{\pi}\arctan\left(\frac{1}{a(\beta)\sqrt{2s_e\varphi_{np}(s_e)}}\right)\right.$$
$$\left. + \frac{2(2 - \sqrt{\pi})}{\pi^{3/2}}\frac{a(\beta)\sqrt{2s_e\varphi_{np}(s_e)}}{1 + a^2(\beta)2s_e\varphi_{np}(s_e)}\right\}, \qquad (C32)$$

where $h_e(1,0)$ was renamed by $\varphi_{np}(s_e)$, which is the symbol for the nonparabolicity correction factor Eq. (21) in the main text. The last term in curly braces in Eq. (C32) is small compared to the second one and will be omitted. Inserting $a(\beta)$ from Eq. (C28) gives

$$\Delta_{h,n-type}^i(T) = -\frac{\sqrt{2}g_e}{3\pi\alpha_h s_e^{3/2}\varphi_{np}^{3/2}(s_e)}$$
$$\times \left\{1 - \frac{2}{\pi}\arctan\left[\left(\frac{2\pi^2\alpha_e^2 s_e}{\beta\alpha_h g_e^2\varphi_{np}(s_e)}\right)^{\frac{1}{2}}\right]\right\}. \qquad (C33)$$

In this approximation, the temperature dependence has the form $\sim\left[1 - 2\arctan\left(c\sqrt{T}/n^{1/6}\right)/\pi\right]$. Due to the assumption of very large Fermi momenta in the dielectric function, $z \ll 1$, it only matches the numerical results at low temperatures and extremely high densities, which are outside the validity range.

The opposite assumption, $z \gg 1$, would lead to $H_e(t,z) \approx 8t/z$. The dielectric function then reads $\epsilon_{T=0}(z,0) = 1 + 8s_e/(3z^4)$, resulting in a temperature and density dependence of the correlation energy given by $\sim\sqrt{n/T}$ as in the Debye limit [compare Eq. (16)].

Combining both limits with the static form of the SPP approximation (B7) leads to cumbersome expressions that are not helpful. Therefore, to cover the cases $z \ll 1$ and $z \approx 1$, a viable solution is to

replace $\arctan[(\cdots)^{1/2}]$ by $\arctan[(\cdots)^{3/4}]$ in (C33). This not only preserves the zero-temperature limit but also results in excellent agreement with the numerical curves for all n-type materials over the whole density and temperature range, as can be seen from Fig. 8 and the left panel of Figs. 10–29.

REFERENCES

[1] R. Zimmermann, *Many Particle Theory of Highly Excited Semiconductors* (BSB Teubner Verlagsgesellschaft, Leipzig, 1988).

[2] H. Stolz and R. Zimmermann, Phys. Status Solidi B **94**, 135 (1979).

[3] M. Rösler and R. Zimmermann, Phys. Status Solidi B **67**, 525 (1975).

[4] M. Rösler, F. Thuselt, and R. Zimmermann, Phys. Status Solidi B **118**, 303 (1983).

[5] M. Rösler, R. Zimmermann, and W. Richert, Phys. Status Solidi B **121**, 609 (1984).

[6] R. Zimmermann and H. Stolz, Phys. Status Solidi B **131**, 151 (1985).

[7] A. Schenk, J. Appl. Phys. **130**, 015703 (2021).

[8] B. E. Sernelius, Phys. Rev. B **33**(12), 8582–8586 (1986).

[9] G. D. Mahan, J. Appl. Phys. **51**, 2634 (1980).

[10] B. I. Lundqvist, Phys. Kondens. Mater. **6**, 193–206 (1967).

[11] K.-F. Berggren and B. E. Sernelius, Phys. Rev. B **24**, 1971 (1981).

[12] A. Selloni and S. T. Pantelides, Phys. Rev. Lett. **49**, 586 (1982).

[13] K.-F. Berggren and B. Sernelius, Phys. Rev. B **29**, 5575 (1984).

[14] S. T. Pantelides, A. Selloni, and R. Car, Solid-State Electron. **28**, 17 (1985).

[15] M. Balkanski, A. Aziza, and E. Amzallag, Phys. Status Solidi **31**, 323 (1969).

[16] H. P. D. Lanyon, A. K. McCurdy, and R. A. Tuft, in *Proceedings of the 13th IEEE Photovoltaic Specialist Conference, Washington DC, 1978* (IEEE, New York, 1978), p. 60.

[17] P. E. Schmid, Phys. Rev. B **23**, 5531 (1981).

[18] J. A. del Alamo, R. M. Swanson, and A. Lietoila, Solid-State Electron. **26**, 483 (1983).

[19] S. E. Aw, H. S. Tan, and C. K. Ong, J. Phys.: Condens. Matter **3**, 8213 (1991).

[20] D. Yan and A. Cuevas, J. Appl. Phys. **116**, 194505 (2014).

[21] W. P. Dumke, J. Appl. Phys. **54**, 3200 (1983).

[22] J. Wagner, Phys. Rev. B **32**, 1323 (1985).

[23] J. Wagner, Solid-State Electron. **30**, 1117 (1987).

[24] J. Wagner and J. A. del Alamo, J. Appl. Phys. **63**, 425 (1988).

[25] H. E. J. Wulms, IEEE J. Solid-State Circuits **12**, 143 (1977).

[26] D. D. Tang, IEEE Trans. Electron Devices **27**, 563 (1980).

[27] A. W. Wieder, IEEE Trans. Electron Devices **27**, 1402 (1980).

[28] G. E. Possin, M. S. Adler, and B. J. Baliga, IEEE Trans. Electron Devices **31**, 3 (1984).

[29] A. Neugroschel, J. S. Wang, and F. A. Lindholm, IEEE Electron Device Lett. **6**, 253 (1985).

[30] J. A. del Alamo and R. M. Swanson, IEEE Trans. Electron Devices **34**, 1580 (1987).

[31] D. Olego and M. Cardona, Phys. Rev. B **22**(2), 886 (1980).

[32] A. N. Titkov, E. I. Chaykina, E. M. Komova, and N. G. Ermakova, Sov. Phys. Semicond. **15**(2), 198–202 (1981).

[33] S. Tiwari and S. L. Wright, Appl. Phys. Lett. B **56**(6), 563–565 (1990).

[34] N. A. Semikolenova, I. M. Nesmelova, and E. N. Khabarov, Sov. Phys. Semicond. **12**(10), 1142–1193 (1978).

[35] M. Bugajski and W. Lewandowski, J. Appl. Phys. **57**(2), 521–530 (1985).

[36] M. Gell-Mann and K. A. Brueckner, Phys. Rev. **106**, 364 (1957).

[37] H.-J. Wünsche, K. Henneberger, and V. E. Khartsiev, Phys. Status Solidi B **86**, 505 (1978).

[38] W. Ebeling, W. Richert, W. D. Kraeft, and W. Stolzmann, Phys. Status Solidi B **104**, 193 (1981).

[39] T. M. Rice, Il Nuovo Cimento B **23**, 226 (1974).

[40] F. Thuselt, Phys. Lett. A **94**, 93 (1993).

[41] A. Schenk, J. Appl. Phys. **84**(7), 3684 (1998).

[42] H. E. deWitt, J. Nucl. Energy, Part C **2**, 27 (1961).

[43] P. Vashishta and R. K. Kalia, Phys. Rev. B **25**, 6492 (1982).

[44] See http://www.ioffe.ru/SVA/NSM/Semicond/InGaAs for parameters of InGaAs.

[45] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**(11), 5815 (2001).

[46] B. K. Ridley, J. Appl. Phys. **68**(9), 4667 (1990).

[47] W. Walukiewicz, Physica B **302–303**, 123 (2001).

[48] P. Pampili and P. J. Parbrook, Mater. Sci. Semicond. Process. **62**, 180 (2017).

[49] W. Walukiewicz, J. Vac. Sci. Technol. B **6**, 1257 (1988).

[50] G. I. Rekalova, U. Kebe, and L. A. Mezrina, Sov. Phys. Sem. **5**, 685 (1971).

[51] J. Jung, J. Cryst. Growth **63**, 229–231 (1983).

[52] S. R. Bank, Technical Report (University of Texas at Austin, 2015) https://apps.dtic.mil/sti/pdfs/AD1007860.pdf.

[53] K. L. Averett, S. Maimon, X. Wu, M. W. Koch, and G. W. Wicks, J. Vac. Sci. Technol. B **20**(3), 1213–1216 (2002).

[54] B. Rotelli, L. Tarricone, E. Gombia, R. Mosca, and M. Perotin, J. Appl. Phys. **81**(4), 1813–1819 (1997).

[55] A. G. Milnes and A. Y. Polyakov, Solid-State Electron. **36**(6), 803–818 (1993).

[56] A. H. Ramelan, P. Arifin, and E. Goldys, Int. J. Mater. Res. **102**(11), 1403–1407 (2011).

[57] S. Bendapudi and D. N. Bose, Bull. Mater. Sci. **13**(1–2), 75–82 (1990).

[58] F. Z. Hawrylo, Appl. Phys. Lett. **37**(11), 1038–1939 (1980).

[59] R. J. Malik, J. Nagle, M. Micovic, T. Harris, R. W. Ryan, and L. C. Hopkins, J. Vac. Sci. Technol. B **10**(2), 850–852 (1992).

[60] J. Starkiewicz and J. W. Allen, J. Phys. Chem. Solids **23**(3), 881–884 (1967).

[61] J. J. Hopfield, P. J. Dean, and D. G. Thomas, Phys. Rev. **158**(3), 748–755 (1967).

[62] T. Hara and I. Akasaki, J. Appl. Phys. **39**(1), 285–289 (1968).

[63] H. W. Seo, S. Y. Bae, J. Park, M.-i. Kang, and S. Kim, Chem. Phys. Lett. **378**, 420–424 (2003).

[64] C. Liu, J. Sun, J. Tang, and P. Yang, Nano Lett. **12**, 5407–5411 (2012).

[65] J. I. Pankove, J. T. Torvik, C.-H. Qiu, I. Grzegory, S. Porowski, P. Quigley, and B. Martin, Appl. Phys. Lett. **74**, 416 (1999).

[66] R. Kirste, M. P. Hoffmann, E. Sachet, M. Bobea, Z. Bryan, I. Bryan, C. Nenstiel, A. Hoffmann, J.-P. Maria, R. Collazo, and Z. Sitar, Appl. Phys. Lett. **103**, 242107 (2013).

[67] S. D. Burnham, G. Namkoong, D. C. Look, B. Clafin, and W. A. Doolittle, J. Appl. Phys. **104**, 024902 (2008).

[68] S. Fritze, A. Dadgar, H. Witte, M. Bügler, A. Rohrbeck, J. Bläsing, A. Hoffmann, and A. Krost, Appl. Phys. Lett. **100**, 122104 (2012).

[69] V. M. Polyakov, F. Schwierz, F. Fuchs, J. Furthmüller, and F. Bechstedt, Appl. Phys. Lett. **94**, 022102 (2009).

[70] C. Hermann and T. P. Pearsall, Appl. Phys. Lett. **38**, 450 (1981).

[71] K. Alavi, R. L. Aggarwal, and S. H. Groves, Phys. Rev. B **21**, 1311 (1980).

[72] A. Ziegler, M. Frey, L. Smith, and M. Luisier, IEEE Trans. Electron Devices **63**(5), 2050–2056 (2016).

[73] E. O. Kane, J. Phys. Chem. Solids **1**, 249 (1957).

[74] P. Rinke, M. Scheffler, A. Qteish, M. Winkelnkemper, D. Bimberg, and J. Neugebauer, Appl. Phys. Lett. **89**, 161919 (2006).

[75] J. Wu, W. Walukiewicz, K. M. Yu, J. W. Ager III, E. E. Haller, H. Lu, W. J. Schaff, Y. Saito, and Y. Nanishi, Appl. Phys. Lett. **80**(21), 3967–3969 (2002).

[76] N. F. Mott, Philos. Mag. **6**(62), 287–309 (1961).

[77] S. C. Jain, J. M. McGregor, and D. J. Roulston, Appl. Phys. Lett. **68**(7), 3747 (1990).

[78] S. C. Jain and D. J. Roulston, Solid-State Electron. **34**(5), 453–465 (1991).

[79] P. P. Debye and E. M. Conwell, Phys. Rev. **93**, 693 (1954).

[80] 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).

[81] 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).

[82] The vector character of momenta and coordinates is only written explicitly where needed, and the normalization volume is set to unity.

14 July 2023 14:39:17