Multisubband plasmons: Beyond the parabolicity in the semiclassical model

Wojciech Pasek, Chris Deimert, Paul Goulain, Jean-Michel Manceau, Raffaele Colombelli, Zbig Wasilewski

To cite this version:

Wojciech Pasek, Chris Deimert, Paul Goulain, Jean-Michel Manceau, Raffaele Colombelli, et al.. Multisubband plasmons: Beyond the parabolicity in the semiclassical model. Physical Review B, 2022, 106 (11), pp.115303. 10.1103/PhysRevB.106.115303 . hal-03863311

HAL Id: hal-03863311
https://hal.science/hal-03863311v1

Submitted on 29 Nov 2022

Multi-subband plasmons: beyond the parabolicity in the semiclassical model

Wojciech J. Pasek,[1, 2, *] Chris Deimert,[1, 3] Paul Goulain,[4] Jean-Michel Manceau,[4] Raffaele Colombelli,[4] and Zbig R. Wasilewski[1, 5, 6, 7, †]

[1] *Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.*
[2]*Present address: Faculdade de Ciências Aplicadas, Universidade Estadual de Campinas, Limeira, SP, 13484-350, Brazil*
[3]*Present address: National Research Council of Canada, Ottawa K1A 0R6, Canada*
[4] *Centre de Nanosciences et de Nanotechnologies (C2N), CNRS UMR 9001, Université Paris-Saclay, 91120 Palaiseau, France*
[5] *Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.*
[6] *Waterloo Institute for Nanotechnology, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.*
[7] *Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.*

(Dated: November 28, 2022)

This work studies the impact of the non-parabolicity on the formation of multi-subband plasmons. We explore three semi-classical optical response models, and compare their predictions to temperature-dependent absorption measurements from three structures: all doped GaAs/AlGaAs quantum wells with continuously varied parabolic binding potentials. We show that qualitative improvement in the prediction of the plasmon absorption peak can only be obtained by including both the energy *and* wavefunction dependence on the in-plane wavevector. Our model, developed to include both these dependencies, uses a $\vec{k}\cdot\vec{p}$-derived current density operator (instead of the usual scalar effective mass one). The model should be readily generalizable to a wide set of nanostructures, such as asymmetric half-parabolic wells or narrow band materials nanostructures beyond the quasi-Kohn regime.

I. INTRODUCTION

Recently, harmonic continuously graded alloy semiconductor quantum wells have been examined in Ref. 1. The system exhibits a single absorption peak up to room temperature and remarkably small relative linewidths (5.6%) at liquid helium temperature. The existence of a single strong peak[a] superposition of underlying subband-pair resonators is called the multi-subband plasmon (MSP)[2]. The absorption frequency is largely independent of doping and temperature, remaining very close to target frequency 3 THz. However, while the temperature redshift is small (being basically a small order correction in the case of a nearly perfect parabolic confinement), it is in stark qualitative contrast to the blueshift predictions of the MSP formation model. As pointed out in the referenced work Ref. 1, further effort is needed to get the model details right.

A simple model of QW was given in Ref. 3, in the case of a single absorption mode coming from a single inter-subband (ISB) electron transition. This model includes the depolarisation shift[4], but it does not account for the superposition effects of different inter-subband resonators. To describe the formation of MSPs, two main classes of models have been used thus far, one quantum and one semi-classical.

The quantum model[5–8] employs a light-matter interaction Hamiltonian, second-quantized in the electric dipole gauge, describing the interaction in terms of the electric displacement field and polarization density. Using a Bogoliubov-style diagonalization procedure twice, it yields first the depolarization shift (as a self-interaction transition effect) and then the multi-subband modes. Finally, the light-matter coupling part of the Hamiltonian is given, using effective plasma frequencies and overlap factors for the MSP modes. In the quantum model for the square QWs all the ISB transitions can be assumed to have the same constant effective length[6,9]. This simplification allows the model to be extended to non-parabolic subbands, and an effective dielectric tensor can be calculated[10]. This approach mirrors much earlier work by Warburton et al. who showed an absorption concentration effect in heavily doped InAs/AlSb QWs similar to the MSP mode concentration[11]. Unfortunately, this assumption holds only for the adjacent transitions in square QWs, it does not hold for parabolic quantum wells (PQWs) or other potential shapes in general. For instance, for non-adjacent transitions in square wells, one must revert to the full diagonalization procedure of the quantum model to obtain the correct results[6].

The quantum method has been used quite successfully for square wells, and almost certainly could be extended to treat non-parabolic subbands more generally in the context of $\vec{k}\cdot\vec{p}$ theory. However, at present it is not obvious how best to tackle such an extension in a computationally feasible manner. Instead, we turn to a semi-classical approach, based on the work of Alpeggiani and Andreani, Ref. 12. Here, the "workhorse" is the non-local susceptibility tensor, as obtained in Refs. 13–15 using a Green function formalism. It has been used to study the light-matter coupling for both ISB transitions[13–16] and exciton-polaritons[17]. The derivation of the model is

---

[a] In fact, what one directly gets from our model are the effective permittivity and the absorption coefficient. For the purpose of this work, as long as there is only one absorption peak present with a single maximum (which is always the case for the results presented in this work), it is interpreted as the MSB.

given in Sec. B.

Motivated by previous experimental work, we look here at an example case of PQWs, which have the interesting property that in the superposition of underlying subband-pair resonators certain effects appear to cancel out. Let us imagine an undoped system with perfectly parabolic (in-plane) single electron energy subbands as a starting point. As the doping is increased and the conduction bands start to be occupied, two things will happen: (1) the electron-electron interaction will deform (bend) the bands resulting in a redshift in inter-subband transition energies and (2) the ISB resonators will interact resulting in a general blueshift of the MSP mode with respect to individual inter-subband modes. Those two effects should shift the energy of the renormalized MSP mode exactly oppositely, effectively canceling both contributions out. That is, the MSP transition energy should be the same as the inter-subband transition energy of the single electron. This well known effect is called the Kohn theorem[18] and it also takes place in semiconductor systems[19].[b] In the time-dependent DFT community, its generalization is known as the Harmonic Potential Theorem, see Ref. 20. In the case of a system where single electron subbands are *nearly* parabolic, like our PQW, one should expect that the MSP energy will be *nearly* unaffected, with a possible energy shift being a smaller-order effect.

Importantly, though, real-world PQWs do not conform perfectly to the Kohn theorem. Among the departures from ideal parabolicity of the electron bands we have: (1) the truncation of the PQW potential due to finite thickness of the wells, (2) the limited knowledge of the material parameters resulting in only *quasi*-ideal growth conditions, and (3) the mixing between the conduction and valence bands due to the spin-orbit interaction. These effects will result in small departures from the behaviour of an ideal parabolic potential.

PQWs are thus an interesting test case for these smaller effects. The electron interactions which would typically dominate are mostly cancelled out, leaving us in a quasi-Kohn regime where these smaller effects are more visible. Furthermore, we should expect that *in-plane* non-parabolicity due to the spin-orbit interaction (not to be confused with non-parabolicity of the PQW potential in the growth direction), will play an increasing role as the temperature is pushed towards room temperature. The higher we move above cryogenic temperatures, the more occupation we will see at $\vec{k}_{\|}$ values far from the Γ point. Therefore, we should not expect that we can assume the transition energies *nor* the wavefunctions to be constant for all $\vec{k}_{\|}$.

In principle, the linear response of the non-parabolic system can be obtained within the scope of the self-consistent time-dependent local-density approximation. For example, the finite wave-vector intersubband collective excitation spectra in wide parabolic wells were obtained at low two-dimensional electron densities, in Ref. 21. It is important, however, to develop a computationally feasible model of MSP absorption which can correctly account for these effects in non-square QWs in a way that is practical for nanodevice design, something which has thus far been lacking. In particular the computational time is an important factor in device design.

For this reason, this work is focused on correctly incorporating the small *in-plane* non-parabolicity effect into the semi-classical MSP model for PQW absorption. The experimental data (Sec. III) come from a set of three PQW samples, two of which were previously discussed in Ref. 1. The MSP energy of the samples has been measured in a multipass absorption experiment over a range of different temperatures from liquid helium to room temperature. The position of the MSP peak, and more importantly its temperature shift, is extracted with the help of the transfer matrix method. On the theoretical side, we approach the system from the point of view of the Kohn theorem, treating the "energy-locked MSP" as a unperturbed system.

We then develop the theory in stages from there, producing three models of increasing sophistication. As a first step, we employ a simple model, which takes into account only the temperature variation of the material parameters, but still works with ideally parabolic bands (Sec. II C). We refer to this as the "parabolic model." The second step takes into account the calculated non-parabolicity of the energy bands but still assumes that the corresponding wavefunctions are identical in the whole momentum plane to these at the Γ point (Sec. II D). We refer to this as the "hybrid model." The third and ultimate approach uses the in-plane wavevector-dependent energies and wavefunctions of the 8-band $\vec{k}\cdot\vec{p}$ Hamiltonian (Sec. II E). We refer to this as the 8-band $\vec{k}\cdot\vec{p}$ model.

It should be noted that the parabolic model already existed in the literature (Ref. 12), although we do provide some additional generalization to non-symmetric QW systems. The other two approaches extend this model to include in-plane non-parabolicity, with significant new theoretical additions in the 8-band $\vec{k}\cdot\vec{p}$ model. Further, it is shown that this full model is the only one able to qualitatively (and quantitatively, but with uncertainty coming from the limited knowledge of the material parameters) predict the temperature shift of the PQW systems. In particular, the in-plane parabolic model, in which the temperature comes only through the variation of the material parameters, leads in some instances to the opposite prediction: *i.e.* a blueshift instead of a redshift of the MSB energy as T increases.

---

[b] Strictly speaking, the Kohn theorem is about the cancellation of electron-electron interactions in cyclotron systems. The analogous effect in quantum well systems like ours, leaving the resonance frequency almost completely independent of the charge distribution, is described in the latter work cited. However, we have decided to nevertheless use the term *Kohn theorem* due to its prevalence in the field.

We see the usefulness and the interest of our work primarily in: (I) the successful application of the models – which were in principle derived only/primarily with parabolic dispersions in mind – to the system, where the non-parabolicity is of importance, and which is itself a breakthrough in the MBE growth techniques, and (II) in the successful "gluing together" in a coherent fashion the $\vec{k} \cdot \vec{p}$ and the ISB models, which overcomes a series of difficulties: both methodological and – primarily – of the computational complexity nature, with the minimal set of the necessary approximations/simplifications.

II. METHOD

A. System

Though our theoretical model is quite general, we consider some specific examples of parabolic quantum well (PQW) systems in $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$. We grew three PQW samples with molecular beam epitaxy (MBE), using a continuous grading technique which allows the desired composition profile to be followed precisely[22,23]. The first two samples, G0489 and G0490, are PQW arrays designed to have an absorption frequency around 3 THz. G0489 has 18 PQWs doped at $3 \times 10^{11}\,\mathrm{cm}^{-2}$ per well, and G0490 has 54 PQWs doped at $1 \times 10^{11}\,\mathrm{cm}^{-2}$ per well. The two structures are identical aside from the number of wells and doping level. The PQWs in G0489/G0490 are formed from continuously graded $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ with composition ranging in $0.02 \leq x \leq 0.30$. As shown in Fig. 1 (left), the PQWs are 103.8 nm wide, separated by 4/2/4 nm barriers of $\mathrm{Al}_{0.3}\mathrm{Ga}_{0.7}\mathrm{As}$/GaAs/$\mathrm{Al}_{0.3}\mathrm{Ga}_{0.7}\mathrm{As}$. Si doping is placed into the GaAs region of the barrier to avoid the formation of DX-centres and ensure ionization of the dopants[24]. The absorption characteristics of G0489 and G0490 were studied experimentally in Ref. 1.

The third sample, G0643, targets a lower frequency of 2.1 THz. This sample was, again, grown in continuously graded $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$, but with a lower barrier composition, ranging in $0.02 \leq x \leq 0.20$. As such, no doping wells were used in this sample. G0643 includes a periodic array of 8 PQWs doped at $1 \times 10^{11}\,\mathrm{cm}^{-2}$ per well. The wells are 130.7 nm wide, separated by 20 nm $\mathrm{Al}_{0.2}\mathrm{Ga}_{0.8}\mathrm{As}$ barriers. (Again, shown in Fig. 1.)

B. The nextnano++ simulation

From the computational perspective, the process starts with obtaining the eigenenergies $E_n(z)$ and eigenfunctions $\psi_n(z)$ of the Schrödinger-Poisson equation system with the help of the nextnano++ software, see Ref. 25. Note that in the case of the one-band parabolic model (Sec. II C), only the solution at the Γ point is needed, as the wavefunctions do not depend on $\vec{k}_{\|}$ and the dependence of the energy is trivial, while in the case of $\vec{k} \cdot \vec{p}$ simulation a custom $\vec{k}_{\|}$-mesh is defined, as described in Sec. II E. Periodic boundary conditions are used in the growth z direction. The pseudomorphic strain of the system is included in the calculation.[26]

In the cases of the G0489 and G0490 samples (the 3 THz system) the width of the periodic cell is $W_S =$ 113.8 nm (QW + separation + doping well) and the grid spacing is $\Delta z = 0.25$ nm. The number of eigenstates taken into account in the simulation is 17. The nextnano++ default parameter values were used, with the exception of the linear dependence of bandgap on composition, which was adopted after Ref. 27, slightly corrected for strain $[(E_g)^{AlAs}_{T=0} = 2.9107\ \mathrm{eV}]$, with a temperature dependent bandgap and band-offset ratio of 0.60 at zero temperature.[c] The n-Si in GaAs doping is introduced through a doping well with width of 2.0 nm per period, assuming full ionization.

For the ∼2 THz system (sample G0642/G0643) the width of the period is $W_S = 150.75$ nm and the simulation includes 14 electron eigenstates. $(E_g)^{AlAs}_{T=0} =$ 2.922 eV is used, with a linear dependence of $(E_g)_{T=0}$ on composition and band-offset ratio of 0.60 at zero temperature, as previously described. In the case of one-band and hybrid models the asymmetric delta doping profile in introduced in the simulation as-is. In the case of the $\vec{k} \cdot \vec{p}$ model, the doping profile needed to be symmetrized, resulting in $W_S = 153$ nm.

In the simulation for the 8-band $\vec{k} \cdot \vec{p}$ model, $(E_g)^{AlAs}_{T=0} =$ 2.922 eV is used; 20 electron wavefunctions and 16 hole ones are included; the doping well potential is omitted[d] and the default $S = 1$ re-scaling is used in order to avoid spurious solutions.

C. ISB formation: the parabolic model

In this section, we present the semi-classical Alpeggiani-Adreani model of the ISB formation of Ref. 12 – henceforth *AA* – which is the starting point for our analysis of the ISB formation. The model is used as-is within the parabolic approximation and is a basis for further development of the hybrid (Sec. II D) and the $\vec{k} \cdot \vec{p}$ (Sec. II E) ones. Please note that a more detailed and more general derivation is provided in Appendix, Sec. B.

We start by defining the non-local susceptibility in the growth z direction $\chi_{zz}(\omega, \mathbf{q}; z, z')$ by the following rela-

---

[c] Note that the exact band-offset ratio has remained difficult to pin down exactly experimentally. See, for example, the wide range of offsets reported in Ref.[28].

[d] The wavefunctions are generally localized entirely in the parabolic QW or in the doping well, due to their separation. However, hybridization can happen in the case of a coincidental degeneracy for a given n and $\vec{k}_{\|}$. As the $\vec{k} \cdot \vec{p}$ model model requires following which orbital is which, it is easier to remove the doping well, taking into account it had a small impact on the QW solutions anyway.

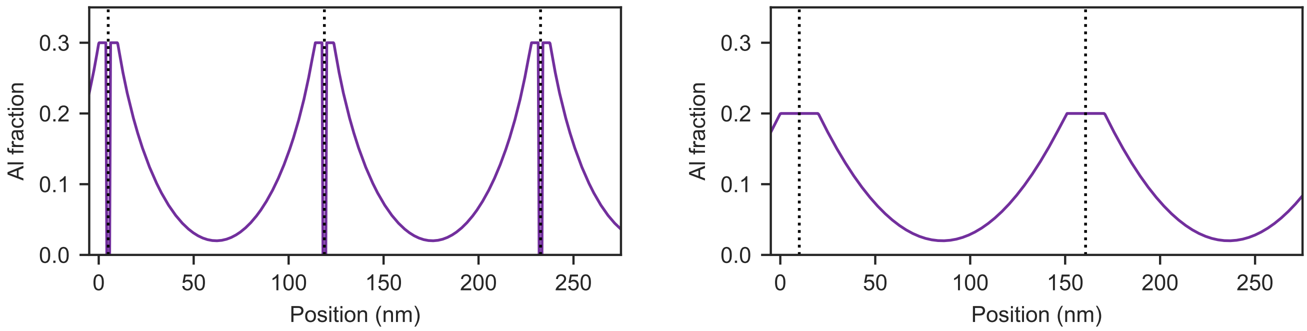

FIG. 1: Visualization of the experimental PQW systems studied here. The left panel shows the ∼ 3 THz system G0489/G0490 and the right panel shows the ∼ 2.1 THz system G0643, in the absence of doping. The solid lines show the $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ composition, x, as a function of position in the growth direction. The dotted vertical lines show the position of the Si-delta doping layer.

tion between the electric $\mathcal{E}$ and the displacement $\mathbf{D}$ fields:

$$D_z(z) = \epsilon_0\epsilon_s + \epsilon_0 \int \chi_{zz}(\omega, \mathbf{q}; z, z')\mathcal{E}_z(z')\,\mathrm{d}z' \qquad (1)$$

with ϵ_s the static permittivity of the semiconductor in the absence of doping, and

$$\chi_{zz}(\omega, \mathbf{q}; z, z') = \sum_\alpha \chi_\alpha(\omega, \mathbf{q})\xi_\alpha(z)\xi_\alpha(z') \qquad (2)$$

where $\omega, \mathbf{q}$ are the frequency and wavevector of the transition, respectively, and the index $\alpha \equiv n \to m$ stands for the transition from subband n to subband m.[e] The ξ_α's are called ISB transition current densities, and they are defined by

$$\xi_\alpha(z) = \frac{\hbar e}{2m^*(z)}\left[\psi_m(z)\partial_z\psi_n(z) - (\partial_z\psi_m(z))\psi_n(z)\right], \quad (3)$$

where $\psi_n(z)$ are the envelope eigenvectors of the one-band Schrödinger-Poisson equation system, *assumed to be real*, and m^* is the effective mass. The single-particle susceptibility $\chi_\alpha(\omega)$, in the long wavelength limit $\mathbf{q}\to 0$, is given by

$$\chi_\alpha(\omega) = -\frac{1}{\left(\omega + \frac{i\gamma^{\mathrm{IB}}}{2}\right)^2}\frac{4}{\epsilon_0\hbar\omega_\alpha s}\sum_{\vec{k}_\parallel}\Delta f_\alpha(\vec{k}_\parallel) \times \left[1 + \frac{\omega_\alpha(\vec{k}_\parallel)\omega_\alpha(0)}{\left(\omega + \frac{i\gamma^{\mathrm{IB}}}{2}\right)^2 - \omega_\alpha(\vec{k}_\parallel)^2}\right] \qquad (4)$$

where $\omega_\alpha(\vec{k}_\parallel) = E_m(\vec{k}_\parallel) - E_n(\vec{k}_\parallel)$ is the transition energy from subband $n \to m$ at in-plane wave vector $\vec{k}_\parallel$, and $\Delta f_\alpha(\vec{k}_\parallel) = f_n(\vec{k}_\parallel) - f_m(\vec{k}_\parallel)$ is the difference in Fermi occupation probability. s is the in-plane surface area.

Please note here that the broadening parameter γ^{IB} is artificially introduced to the model by evaluating the susceptibility at $\omega \to \omega + i\gamma^{\mathrm{IB}}/2$. In principle, $\gamma^{\mathrm{IB}}_{\alpha,\vec{k}_\parallel,T}$ would be the FWHM of each individual inter-subband oscillator α at $\vec{k}_\parallel$ position in the wavevector space and at temperature T. However, taking into account all this variability leads to enormous number of arbitrarily valued parameters. One could argue that, in principle, gamma could be calculated theoretically[29,30], but since certain scattering mechanisms like interface roughness will be MBE growth dependent, this is practically quite challenging. In this work, to keep things simple, a transition-independent and $\vec{k}_\parallel$-isotropic parameter is used, dropping the $\alpha, \vec{k}_\parallel$ indices. Furthermore, in the case of the parabolic model as well as in the case of the hybrid one (see Sec. II D) the γ^{IB} of individual transitions translate directly to the resulting linewidth of the ISB peak, with no coupling between this parameter and the frequency of the absorption maximum. This means that it is here an external parameter to the model and the model cannot be used to study *i.e.* the linewidth dependence on temperature, quantum well size or sample doping. Consequently, we focus on the frequency of the absorption maximum in this work. This also means that we can use any reasonable non-zero value of γ^{IB} and the T index can also be dropped. The situation is different in scope of the 8-band $\vec{k}\cdot\vec{p}$ model, see the discussion of the Eq. (26) in Sec. II E.

Assuming that the subbands are two-fold spin-

---

[e] In *AA* the corresponding notation is $\alpha \equiv n \to n'$, however there is also the α' index. While we tried to mimic the notation of that work as far as possible, for ease of the reference of the reader, we use $\alpha \equiv n \to m$ and $\alpha' \equiv n' \to m'$. This also allows us to use the $\alpha', \beta', \sigma', k'_\parallel$ set of indices in a consistent way below.

degenerate and isotropic in-plane, one obtains:

$$\chi_\alpha(\omega) = -\frac{1}{(\omega + \frac{i\gamma^{\mathrm{IB}}}{2})^2}\frac{4}{\epsilon_0 \hbar \omega_\alpha \pi}\int_0^\infty \Delta f_\alpha\left(k_\parallel\right) \times \left[1 + \frac{\omega_\alpha\left(k_\parallel\right)\omega_\alpha(0)}{\left(\omega + \frac{i\gamma^{\mathrm{IB}}}{2}\right)^2 - \omega_\alpha\left(k_\parallel\right)^2}\right] k_\parallel \, \mathrm{d}k_\parallel \,. \quad (5)$$

Solving the Maxwell's equations for *TM* waves in a layered structure, with a quantum well layer sandwiched inside a metal-insulator-metal (*MIM*) cavity[31–39] and assuming $\epsilon_{xx} = \epsilon_{yy} = \epsilon_\parallel$, one shows that the D_z field within the QW region obeys the integro-differential equation:

$$\left(\partial_z^2 + k_z^2\right) D_z = -\sum_\alpha \frac{\chi_\alpha(\omega, q_x)}{\epsilon_s}\xi_\alpha(z) \times \int \xi_\alpha(z')\left(\partial_{z'}^2 + \epsilon_{xx}k_0^2\right) D_z(z') \, \mathrm{d}z' \quad (6)$$

where k_z is defined by

$$k_z^2 = \epsilon_\parallel k_0^2 - \frac{\epsilon_\parallel}{\epsilon_s} q_\parallel^2 \quad (7)$$

with $k_0^2 = \omega^2/c^2$ and $q_\parallel$ being the in-plane wave number.

The general solution to this equation has the form

$$D_z(z) = A\cos(k_z z) + B\sin(k_z z) + q_\parallel^2 \sum_\alpha \frac{\chi_\alpha}{\epsilon_s}\left(AF_\alpha^A + BF_\alpha^B\right)\int \xi_\alpha(z') g(z, z') \, \mathrm{d}z' \quad (8)$$

where A, B are arbitrary coefficients which will be determined by the boundary conditions of the layer. $g(z, z')$ is the Green's function $\sin(k_z|z - z'|)/2k_z$ The F_α^A, F_α^B coefficients can be calculated through the following matrix relationships:

$$F_\alpha^A = C_\alpha^A - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F_{\alpha'}^A \left[I_{\alpha,\alpha'} + q_\parallel^2 D_{\alpha,\alpha'}\right] \quad (9)$$

$$F_\alpha^B = C_\alpha^B - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F_{\alpha'}^B \left[I_{\alpha,\alpha'} + q_\parallel^2 D_{\alpha,\alpha'}\right] \quad (10)$$

with

$$\begin{aligned} I_{\alpha,\alpha'} &= \int \xi_\alpha(z)\xi_{\alpha'}(z) \, \mathrm{d}z \\ C_\alpha^A &= \frac{\epsilon_\parallel}{\epsilon_s}\int \cos(k_z z)\xi_\alpha(z) \, \mathrm{d}z \\ C_\alpha^B &= \frac{\epsilon_\parallel}{\epsilon_s}\int \sin(k_z z)\xi_\alpha(z) \, \mathrm{d}z \\ D_{\alpha,\alpha'} &= -\frac{\epsilon_\parallel}{\epsilon_s}\int \xi_\alpha(z) g(z, z')\xi_{\alpha'}(z') \, \mathrm{d}z \, \mathrm{d}z' \,. \end{aligned} \quad (11)$$

From these, it is possible to calculate the transfer matrix for the QW region. Unlike in *AA*, we have not assumed symmetric quantum wells, and we have not assumed that $\epsilon_\parallel = \epsilon_s$.

Taking advantage of the long-wavelength limit ($q_x, k_0, k_z \approx 0$) allows us to define an effective local permittivity for the QW stack. Intuitively, since the electric field is approximately constant on the scale of a single QW, we are able to replace each QW period with an effective medium.

The basic approach (see Ref. 3), is to take an average of the D_z and $\mathcal{E}_z$ fields over a single QW period of length L. Then, the effective permittivity (in the z direction) is given by

$$\epsilon_{zz,\mathrm{eff}} = \frac{\langle D_z \rangle}{\epsilon_0 \langle \mathcal{E}_z \rangle} \approx \epsilon_s \left[1 - \sum_\alpha \frac{\chi_\alpha}{\epsilon_\parallel} F_\alpha^A \langle \xi_\alpha \rangle\right]^{-1} \quad (12)$$

where $\langle \cdot \rangle$ denotes an average over the entire QW region in the z direction:

$$\langle \xi_\alpha \rangle = \frac{1}{L}\int \xi_\alpha(z) \, \mathrm{d}z \quad (13)$$

where L is the thickness of the quantum well region. Furthermore, since we are using the long-wavelength limit, we can use that to simplify our equation for the F_α^A's to:

$$F_\alpha^A \approx L\langle \xi_\alpha \rangle - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F_{\alpha'}^A I_{\alpha,\alpha'} \equiv \sum_{\alpha'} C_{\alpha,\alpha'} F_{\alpha'}^A. \quad (14)$$

From the QW transition current densities, $\xi_\alpha(z)$, we solve the matrix Eq. (14) to get the F_α^A coefficients, from which we can calculate the effective permittivity $\epsilon_{zz,\mathrm{eff}}$.

D. ISB formation: the hybrid model

In the scope of the hybrid model, the dependence of the wavefunctions on $\vec{k}_\parallel$ is omitted, but the $\vec{k}_\parallel$ dispersion is partially taken into account by using a mixture of effective mass and 3-band $\vec{k} \cdot \vec{p}$ modelling. Following Warburton *et al.*, Ref. 11, the subband energy dispersion $E_n(k_\parallel)$ is given implicitly by

$$\hat{H}_h \psi_n(z) = E_n(k_\parallel)\psi_n(z), \quad (15)$$

with

$$\hat{H}_h = -\frac{\hbar^2}{2m^*(z)}k_\parallel^2 - \frac{\partial}{\partial z}\frac{\hbar^2}{2m^*(z)}\frac{\partial}{\partial z} + E_g(z) - \delta(z) \quad (16)$$

and

$$\frac{1}{m^*(z)} = \frac{E_P}{3m_e}\left(\frac{2}{E_n(k_\parallel) + \delta(z)} + \frac{1}{E_n(k_\parallel) + \Delta(z) + \delta(z)}\right) + (1 + 2F). \quad (17)$$

Here m_e is the bare electron mass. $E_g, E_P, F, \Delta, \delta$, and F are the band gap, Kane energy, F-parameter, spin-orbit splitting and valence band offset, respectively. The values of the parameters are taken from Ref. 28.

Instead of solving the Schrödinger-Poisson system self-consistently using $\hat{H}_h$ of Eq. (16), we assume that the wavefunctions calculated in the self-consistent one-band effective mass approximation are reasonably close to the actual wavefunctions. Given the effective mass wavefunctions, $\psi_n(z)$, we can then estimate the 3-band $\vec{k}\cdot\vec{p}$ dispersion by choosing $E_n(k_\|)$ so that it minimizes the overlap error

$$\int \left[\psi_n^*(z) H \psi_n(z) - E_n(k_\|) |\psi_n(z)|^2\right] \mathrm{d}z \,. \tag{18}$$

We confirmed that, while this approach does not give us the full wavefunction spinors, it at least gives a good approximation of the subband energy dispersions, as long as one is concerned with a single $\vec{k}_\|$ orientation along [100] / [010] without the spin degeneracy of orbitals being lifted.

The resulting energy dispersions are used in Eqs. (5), then (14) and finally (12).

E. ISB formation: the 8-band $\vec{k}\cdot\vec{p}$ model

The model that follows is our generalization of *AA*, which takes into account non-parabolicity by including the $\vec{k}_\|$-dependence of both the eigenenergies $E_{n,\vec{k}_\|}$ and the eigenfunctions $\psi_{n,\vec{k}_\|}$. Apart from the orders of magnitude larger computational complexity, the main challenges of the extension of the model come from these sources: (I) the simple k_z-dependent definition of the transition current densities – see Eq. (3) – does not work for the 8-component $\psi_{n,\vec{k}_\|}$, so a new definition needs to be provided, which we base on the proper definition of the probability current density, presented in Sec. II F, (II) the transition current densities are non-trivially complex[f], which renders several approximations in *AA* invalid, see Secs. D and F, (III) the simple one-band notion of an *scalar effective mass* m^* of an electron, as present *e.g.* in *AA*, cannot be used.

This section contains the presentation of the model. Where possible, the parallels and the differences between our model and *AA* are indicated and briefly discussed. For the in-depth discussion, which may require earlier reading or simultaneously following significant external material, see Secs. D, E and F in Supporting Material.

The Schrödinger-Poisson equation system is solved self-consistently with the help of nextnano++ software in the scope of the 8-band $\vec{k}\cdot\vec{p}$ model. The computational process is largely kept default, but uses the linear bandgap to composition relation plus the valence band offset re-calibration, as described previously in Sec. II B, and also the custom-defined $\vec{k}_\|$ mesh, as follows. The output: eigenenergies $E_{n,\vec{k}_\|}$ and eigenfunctions $\psi_{n,\vec{k}_\|}$ are obtained on a $\vec{k}_\|$-space mesh with 11 mesh points along each of two directions $\hat{n}$, [100] and [110], and mesh spacing $\mathrm{d}k_\| = 0.05\ \mathrm{nm}^{-1}$, starting from the Γ point.

For a nanosystem, which is itself symmetric with respect to reversing the z axis, realized in material of zinc-blende crystal structure each octant of the $k_\|$ space is equivalent. Hence, any sum over the $k_\|$ space can be reduced to sum over single octant. Due to computational complexity, we effectively represent the relevant octant by an average of its two limiting lines: $\sum_{\mathbf{k}_\|} \to \frac{1}{2}\sum_{\hat{n},k_\|}$, where $\hat{n} \in \{[100],[110]\}$.

In our model, the eigenfunctions $\psi_{n,\vec{k}_\|}$ are represented in the Bloch basis $\vec{U}$ of

$$\begin{aligned}\vec{U} = \Bigg(&\left|S,-\frac{1}{2}\right\rangle, \left|S,+\frac{1}{2}\right\rangle, \left|\frac{3}{2},+\frac{1}{2}\right\rangle, \left|\frac{3}{2},+\frac{3}{2}\right\rangle,\\ &\left|\frac{3}{2},-\frac{3}{2}\right\rangle, \left|\frac{3}{2},-\frac{1}{2}\right\rangle, \left|\frac{1}{2},-\frac{1}{2}\right\rangle, \left|\frac{1}{2},+\frac{1}{2}\right\rangle\Bigg)\,,\end{aligned} \tag{19}$$

where

$$\begin{aligned}&\left|S\frac{1}{2},-\frac{1}{2}\right\rangle = |S\downarrow\rangle\,,\quad \left|S\frac{1}{2},\frac{1}{2}\right\rangle = |S\uparrow\rangle\,,\\ &\left|\frac{3}{2},\frac{3}{2}\right\rangle = \frac{i}{\sqrt{2}}\left(|X\uparrow\rangle + i\,|Y\uparrow\rangle\right),\quad \left|\frac{3}{2},-\frac{3}{2}\right\rangle = \frac{-i}{\sqrt{2}}\left(|X\downarrow\rangle - i\,|Y\downarrow\rangle\right),\\ &\left|\frac{3}{2},\frac{1}{2}\right\rangle = \frac{-i}{\sqrt{6}}\left(|X\downarrow\rangle + i\,|Y\downarrow\rangle\right) + i\sqrt{\frac{2}{3}}\,|Z\uparrow\rangle\,,\\ &\left|\frac{3}{2},-\frac{1}{2}\right\rangle = \frac{i}{\sqrt{6}}\left(|X\uparrow\rangle - i\,|Y\uparrow\rangle\right) + i\sqrt{\frac{2}{3}}\,|Z\downarrow\rangle\,,\\ &\left|\frac{1}{2},-\frac{1}{2}\right\rangle = \frac{-i}{\sqrt{3}}\left(|X\uparrow\rangle - i\,|Y\uparrow\rangle\right) + \frac{i}{\sqrt{3}}\,|Z\downarrow\rangle\,,\\ &\left|\frac{1}{2},\frac{1}{2}\right\rangle = \frac{-i}{\sqrt{3}}\left(|X\downarrow\rangle + i\,|Y\downarrow\rangle\right) - \frac{i}{\sqrt{3}}\,|Z\uparrow\rangle\,,\end{aligned} \tag{20}$$

with the spin quantization axis oriented along the growth direction z. The $k_\|$ index refers to the corresponding mesh point of the in-plane quasi-momentum and the quantum number n refers to the spin-orbitals of the subbands as ordered by their energy around Γ point. The inter-subband transitions are described by $\alpha = (n,m)$, where $m > n$, m and n have the same spin orientation and adjacent orbitals.

It was pointed out in the last paragraph of the Sec. II C that the model is solved in the following sequence: transition current densities $\xi_\alpha(z) \to$ matrix $C_{\alpha,\alpha'} \to F_\alpha^A$ coefficients $\to$ effective permittivity $\epsilon_{zz,\mathrm{eff}}$. The same order applies in this case and the first step of the calculation is to obtain the inner functions and matrix elements of a new $\vec{k}\cdot\vec{p}$-compatible probability current operator $\hat{J}_z$ (which is itself defined in Sec. II F):

$$\begin{aligned}J^{\hat{n}}_{\alpha,1,k_\|}(z) &= \psi^{\hat{n}}_{n,k_\|}(z)\hat{J}_z\left[\psi^{\hat{n}}_{m,k_\|}(z)\right]^* - \left[\psi^{\hat{n}}_{m,k_\|}(z)\right]^*\hat{J}_z\psi^{\hat{n}}_{n,k_\|}(z),\\ J^{\hat{n}}_{\alpha,2,k_\|}(z) &= \left[J^{\hat{n}}_{\alpha,1,k_\|}(z)\right]^*\,;\ J^{\hat{n}}_{\alpha,1,k_\|} = \int J^{\hat{n}}_{\alpha,k_\|}(z)\,\mathrm{d}z\,,\end{aligned} \tag{21}$$

which are analogous to $\xi_\alpha(z)$ in *AA*. For more details on the $\xi_\alpha(z) \to J^{\hat{n}}_{\alpha,\sigma,k_\|}(z)$ substitution, see discussion in Sec. E of the Supporting Material. The σ index, which is absent in *AA*, is a result of the non-trivially complex wavefunctions – and transition current densities, as explained later in Sec. D – compare the two contributions

---

[f] By "non-trivially complex" we mean that they cannot be made real-valued by any set of phase rotations.

in Eqs. (D1) and (D2). The next step is to obtain the inter-transition current elements

$$I^{\hat{n}}_{\alpha,\sigma,k_{\|},\alpha',\sigma',k'_{\|}} = \int J^{\hat{n}}_{\alpha,\sigma,k_{\|}}(z) \left[J^{\hat{n}}_{\alpha',\sigma',k'_{\|}}(z)\right]^{*} \mathrm{d}z\,, \tag{22}$$

which are analogous to $I_{\alpha,\alpha'}$ in AA. In the two last equations $\hat{n} \in \{[100],[110]\}$ and both $k_{\|}$ and $k'_{\|}$ are oriented along $\hat{n}$.

It is convenient to define a new index $\beta = (\alpha,\sigma)$, where $\sigma \in \{1,2\}$ and

$$J^{\hat{n}}_{\beta,k_{\|}} \equiv J^{\hat{n}}_{\alpha,\sigma,k_{\|}}\,;\; I^{\hat{n}}_{\beta,k_{\|},\beta',k'_{\|}} \equiv I^{\hat{n}}_{\alpha,\sigma,k_{\|},\alpha',\sigma',k'_{\|}}\,, \tag{23}$$

and to introduce

$$\tilde{I}^{\hat{n}}_{\beta,k_{\|},\beta',k'_{\|}} = \frac{I^{\hat{n}}_{\beta,k_{\|},\beta',k'_{\|}}}{J^{\hat{n}}_{\beta,k_{\|}}\left(J^{\hat{n}}_{\beta',k'_{\|}}\right)^{*}}. \tag{24}$$

We want our model to be a simple extension of AA and as analogous to that model as reasonably possible. Please note that there are no $k_{\|}$-dependent terms in Eq. (14) – only the $k'_{\|}$-dependent ones are present, implicitly, through the sum in Eq. (4). In order to eliminate the $k_{\|}$ index, in the further calculation we will use the $\tilde{I}_{\alpha,\alpha',k'_{\|}}$ elements, which are averaged over the $k_{\|}$-space, with the (n,m) occupation difference serving as the weight – or alternatively the $\tilde{I}_{\beta,\beta'}$ elements, averaged symmetrically over both $k_{\|}$ and $k'_{\|}$-spaces (see Supporting Material Sec. D for the discussion of the two ways of averaging):

$$\tilde{I}^{\hat{n}}_{\beta,\beta',k'_{\|}} = \frac{\sum_{k_{\|}} \tilde{I}^{\hat{n}}_{\beta,k_{\|},\beta',k'_{\|}} \Delta f^{\hat{n}}_{\beta,k_{\|}} k_{\|}\,\mathrm{d}k_{\|}}{\sum_{k_{\|}} \Delta f^{\hat{n}}_{\beta,k_{\|}} k_{\|}\,\mathrm{d}k_{\|}}$$

$$\tilde{I}^{\hat{n}}_{\beta,\beta'} = \frac{\sum_{k'_{\|}} \Delta f^{\hat{n}}_{\beta',k'_{\|}} \tilde{I}^{\hat{n}}_{\beta,\beta',k'_{\|}} k_{\|}\,\mathrm{d}k_{\|}}{\sum_{k'_{\|}} \Delta f^{\hat{n}}_{\beta',k'_{\|}} k_{\|}\,\mathrm{d}k_{\|}}, \tag{25}$$

where $\Delta f^{\hat{n}}_{\beta,k_{\|}} = f^{\hat{n}}_{k_{\|},n} - f^{\hat{n}}_{k_{\|},m}$ and $f^{\hat{n}}_{k_{\|},n} \equiv f\left(E^{\hat{n}}_{k_{\|},n}, T\right)$ is the Fermi-Dirac distribution function. Please also note that it is the necessity of this averaging which leads us to use $\tilde{I}^{\hat{n}}_{\beta,k_{\|},\beta',k'_{\|}}$ of Eq. (24) instead of $I^{\hat{n}}_{\beta,k_{\|},\beta',k'_{\|}}$, as it turned out in practice that the dependence of the former on $k_{\|}$ is weaker than that of the latter.

The left hand side matrix, corresponding to the $C_{\alpha,\alpha'}$ in Eq. (14) of the parabolic model, is derived in Sec. D, and defined as follows:

$$\tilde{C}_{\beta,\beta'} = \delta_{\beta,\beta'} - \frac{4}{\omega^2 \epsilon_s} \sum_{\hat{n},k'_{\|}} \mathscr{X}^{\hat{n}}_{\beta',k'_{\|}} J^{\hat{n}}_{\beta',k'_{\|}} \left(J^{\hat{n}}_{\beta',k'_{\|},}\right)^{*} \times \tilde{I}^{\hat{n}}_{\beta,\beta',k'_{\|}} k'_{\|}\,\mathrm{d}k_{\|}{}'\,,$$

$$\mathscr{X}^{\hat{n}}_{\beta',k'_{\|}} = \frac{-p(\beta') f^{\hat{n}}_{k'_{\|},\nu(\beta')}}{\mathscr{E}} + \frac{\frac{1}{2}\Delta f^{\hat{n}}_{\beta',k'_{\|}}}{p(\beta')(\omega + \frac{i}{2}\gamma^{\mathrm{IB}}_{T}) - \omega^{\hat{n}}_{\beta',k'_{\|}}} \tag{26}$$

where $\omega^{\hat{n}}_{\beta',k'_{\|}} = E^{\hat{n}}_{k'_{\|},m'} - E^{\hat{n}}_{k'_{\|},n'}$ is the transition energy from subband $n' \to m'$ at in-plane wave vector $\vec{k}'_{\|}$, the $p(\beta) \equiv p(\sigma) = -1^{\sigma}$, $\nu(\beta) \equiv \nu(\sigma) = (\sigma-1)m' + (2-\sigma)n'$ and $\mathscr{E}$ can be either $E^{\hat{n}}_{k'_{\|},m'} - E^{\hat{n}}_{k'_{\|},n'}$ or $E_{\Gamma,m'} - E_{\Gamma,n'}$, depending on the version of the model (see below).

For a moment we will focus our attention on Eq. (26) and how it compares to Eq. (14). Firstly, one should take into account the straightforward renormalization of Eq. (D3), done to allow us to work with Eq. (24), which leads to Eq. (D5) and dimensionless: $\tilde{F}$ coefficients and $\tilde{C}$ matrix. Secondly, the sum over $\vec{k}'_{\|}$ space, implicitly present in Eq. (14) through the χ_{α} definition Eq. (4), in our case is taken outside the now $\vec{k}'_{\|}$-dependent J and I quantities. The $\mathscr{X}$ part corresponds to the part of χ_{α} under this summation, but again one needs to remember that the p and ν symbols, dependent on the σ index, come from non-trivially complex wavefunctions – and transition current densities, see the corresponding minus signs and occupation index changes between the two components of Eqs. (D1) and (D2) in Sec. D. Finally, the simple relation between the wavefunctions, energies and effective mass of parabolic model cannot be directly used, which prompted us to postulate the two versions of the $\mathscr{E}$ expression, see the discussion in Sec. F 4.

Let us now digress a little bit and focus on the role of the homogeneous broadening in this model. In order to do so, we need to refer to some conclusions drawn from the results we have already obtained and will be later discussed in Sec. IV. In Eq. (26) the sum over $\vec{k}_{\|}$ space is done over both $\vec{k}_{\|}$-dependent transition energies $\omega^{\hat{n}}_{\beta',k'_{\|}}$ and the $\vec{k}_{\|}$-dependent transition current densities, as a consequence of taking into account both the $\vec{k}_{\|}$-dependence of the eigenenergies and of the eigenfunctions. Thus the γ^{IB}_{T} parameter and the frequency of the absorption maximum are coupled in the 8-band $\vec{k}\cdot\vec{p}$ model, with increasing γ^{IB}_{T} leading to a small redshift in that frequency, keeping other parameters constant. This means that in principle, contrary to the parabolic and the hybrid models, the temperature dependence of γ^{IB}_{T} needs to be known. In practice however, it was confirmed that this linewidth increase → redshift dependence is small and certainly does not account for most of the redshift observed experimentally given the linewidth increase measured. In addition to that, the relation between the γ^{IB}_{T} and the resulting FWHM of the ISB peak is not longer 1:1. In this model the relation is linear with $\mathrm{FWHM}_{\mathrm{ISB}} = a\gamma^{\mathrm{IB}}_{T} + b$, where $a\neq 1$ and $b\neq 0$, which is a sign of some additional $\vec{k}_{\|}$-dependent dynamics in the combination of the individual intersubband absorption peaks while forming the ISB plasmon one. However, taking into account all of that, we decided to treat the γ^{IB}_{T} as an *approximately* free parameter.

If using the two dimensional $k_{\|}$ and $k'_{\|}$ averaging of transition current densities, one should replace $\tilde{I}^{\hat{n}}_{\beta,\beta',k'_{\|}}$ with $\tilde{I}^{\hat{n}}_{\beta,\beta'}$ in Eq. (26).

Finally, the following system of linear equations can be solved

$$\tilde{C}_{\beta,\beta'}\tilde{F}_{\beta'} = [1, 1, ..., 1] \tag{27}$$

and ε_{zz} obtained from the solution $\vec{\tilde{F}}$ is as follows:

$$\varepsilon_{zz}^{-1} = 1 + \frac{4}{L\omega^2\epsilon_s} \sum_{\beta,k_\parallel,\hat{n}} \mathscr{X}^{\hat{n}}_{\beta,k_\parallel} J^{\hat{n}}_{\beta,k_\parallel} \left(J^{\hat{n}}_{\beta,k_\parallel}\right)^* \tilde{F}_\beta k_\parallel \, \mathrm{d}k_\parallel \tag{28}$$

F. Probability current operator

In order to get the correct operator for the z-component of probability current, $\hat{J}_z$, the technique of Ref. 40 is employed on the $\vec{k}\cdot\vec{p}$ Hamiltonian defined as in Ref. 41. The derivation in Ref. 40, leads to conclusion, that the $\hat{J}_\alpha$ can be obtained the following way:

$$\hat{J}_\alpha = \frac{1}{\hbar}\frac{\partial \hat{H}}{\partial k_\alpha}. \tag{29}$$

Please note that the explicit equations for $\hat{J}_x$, and $\hat{J}_y$, Eq. (18a) and (18b) in Ref. 40, contain a few typographical errors and are inconsistent with this formula. However, its validity is confirmed e.g. in Eq. (18) of Ref. 42, which gives the very useful "sum rule":

$$\left(\hat{H}_K\right)_{jj'} = \frac{\hbar}{2}\hat{\vec{J}}_{jj'} \cdot \vec{k}, \tag{30}$$

where $\hat{H}_K$ is the kinetic part of the Hamiltonian. The simpler version of the $\vec{k}\cdot\vec{p}$ Hamiltonian:

$$\hat{H}_1 = \frac{\vec{p}^2}{2m_0} + V_0(\vec{r}) + \frac{\hbar^2\vec{k}^2}{2m_0} + \frac{\hbar}{m_0}\vec{k}\cdot\vec{p} + \frac{\hbar}{4m_0^2c^2}(\boldsymbol{\nabla}V_0)\times\vec{p}\cdot\vec{\sigma}, \tag{31}$$

is considered, omitting the $\frac{\hbar^2}{4m_0^2c^2}(\boldsymbol{\nabla}V_0) \times \vec{k}\cdot\vec{\sigma}$ term. Applying the Eq. (29) to Hamiltonian $\hat{H}_1$ spanned on the basis of Eq. 19 leads to matrix presented in Table I. The symbols S, B, P_0 stand for corresponding Kane parameters and $\gamma_1, \gamma_2, \gamma_3$ stand for *modified* Luttinger parameters. The in-plane wave components are defined: $k_\pm = k_x \pm ik_y$, so in the case of [100] direction $k_\pm = k_\parallel$ and for [110] direction: $k_\pm = \frac{1\pm i}{\sqrt{2}}k_\parallel$.

G. $\vec{k}\cdot\vec{p}$ parameters of the model

The values of the $\vec{k}\cdot\vec{p}$ parameters used in this model are presented in Table II.

The following transformation of the Luttinger γ parameters is used from the 6×6 Hamiltonian values to the 8×8 ones, according to the work of Pidgeon and Brown, Ref. 44:

$$\gamma_1^{8\times8} = \gamma_1^{6\times6} - \frac{E_P}{3E_g},\ \gamma_2^{8\times8} = \gamma_2^{6\times6} - \frac{E_P}{6E_g},$$
$$\gamma_3^{8\times8} = \gamma_3^{6\times6} - \frac{E_P}{6E_g}. \tag{32}$$

In the formula above, the energy gap E_g is dependent on temperature according to the Varshni formula, Ref. 45:

$$E_g(T) = E_g(T=0) - \frac{\alpha T^2}{\beta + T}. \tag{33}$$

Moreover, if the S parameter is set to a value different from the default S_0 in order to avoid spurious solutions, the Kane energy E_P needs to be rescaled from the default value E_P^0 as follows:

$$E_P(T) = E_P^0 - \frac{E_g(T)(E_g(T)+\Delta)}{E_g(T)+\frac{2}{3}\Delta}(S - S_0) \tag{34}$$

which makes E_P also dependent on temperature, please compare Eqs. (3.62), (3.158) and (3.159) in Ref. 26. Usually, the values of either $S = 0$ or $S = 1$ are used.

In the original AA, the probability currents are mediated by the kinetic momentum operator, which is proportional to $\hat{k}_z$ with the proportionality constant $\frac{\hbar}{m_0 m^*}$. This corresponds to the $[1,1]$ and $[2,2]$ elements of the $\hat{J}_z$ operator in Table I, *i.e.* $\frac{\hbar}{m_0}S\,\hat{k}_z$, which are the diagonal components of the $\hat{J}_z$ operator matrix for the $\left|S, \pm\frac{1}{2}\right\rangle$ Bloch states. The other non-zero elements in Table I exist due to conduction-valence band mixing. Of these, the ones that affect the conduction bands directly (first two rows / columns in Table I) are proportional to either $Bk_\pm$ or the Kane parameter $P_0 = \hbar\sqrt{\frac{E_P}{2m_0}}$. At the vicinity of the Γ point the former ones vanish, leading to the interplay between the S and P_0 as the dominating factors of the probability current dynamics.

Using the $S = 1$ renormalization in nextnano++ is necessary to avoid spurious solutions. However, it was verified that neither using $S = 1$ nor $S = S_0$ works in our ISB model, with the first one completely underestimating and the latter one completely overestimating the magnitude of the current elements. This is not unexpected, as the $S = 1$ renormalization changes the mentioned dynamic between S and P_0 in a way that may yield correct eigen-solutions for $\hat{H}$ but not for $\hat{J}_z$ as defined in Eq. (29).

Introduction of some effective parameter, which we will call A, is necessary as the correction to the $S \leftrightarrow P_0$ balance. We postulate it as a material-structural parameter, in a sense that we hope that one fitted value of A will work for a sufficiently broad class of structurally similar systems realized in the same materials. This assumption will be tested in the Sec. IV by a comparison with the experimental data and the simpler versions of the ISB formation model.

We postulate a scalar A governing the renormalization of the parameters as follows. In case of $A = 0$ the default $\vec{k}\cdot\vec{p}$ values of $S = S_0 = -2.88$ for pure GaAs and $S = S_0 = 0.04$ for pure AlAs are used (see Table II) with linear interpolation. On the other end of the spectrum, for $A = 1$, the reverse of the single-band effective mass is used: $S = \frac{1}{m^*} = \frac{1}{0.067}$ for pure GaAs and $S = \frac{1}{m^*} = \frac{1}{0.15}$ for pure AlAs, also with linear interpolation. The corresponding Kane energies, at zero temperature, are

	$\left\|S\frac{1}{2},-\frac{1}{2}\right\rangle$	$\left\|S\frac{1}{2},\frac{1}{2}\right\rangle$	$\left\|\frac{3}{2},\frac{1}{2}\right\rangle$	$\left\|\frac{3}{2},\frac{3}{2}\right\rangle$	$\left\|\frac{3}{2},-\frac{3}{2}\right\rangle$	$\left\|\frac{3}{2},-\frac{1}{2}\right\rangle$	$\left\|\frac{1}{2},-\frac{1}{2}\right\rangle$	$\left\|\frac{1}{2},\frac{1}{2}\right\rangle$
$\left\langle S\frac{1}{2},-\frac{1}{2}\right\|$	$\frac{\hbar}{m_0}S\,\hat{k}_z$	0	$\frac{1}{\sqrt{6}}\,\frac{B}{\hbar}\,k_-$	0	$-\frac{1}{\sqrt{2}}\,\frac{B}{\hbar}\,k_+$	$-\sqrt{\frac{2}{3}}\,\frac{P_0}{\hbar}$	$-\frac{1}{\sqrt{3}}\,\frac{P_0}{\hbar}$	$\frac{1}{\sqrt{3}}\,\frac{B}{\hbar}\,k_-$
$\left\langle S\frac{1}{2},\frac{1}{2}\right\|$	0	$\frac{\hbar}{m_0}S\,\hat{k}_z$	$-\sqrt{\frac{2}{3}}\,\frac{P_0}{\hbar}$	$-\frac{1}{\sqrt{2}}\,\frac{B}{\hbar}\,k_-$	0	$\frac{1}{\sqrt{6}}\,\frac{B}{\hbar}\,k_+$	$-\frac{1}{\sqrt{3}}\,\frac{B}{\hbar}\,k_+$	$\frac{1}{\sqrt{3}}\,\frac{P_0}{\hbar}$
$\left\langle\frac{3}{2},\frac{1}{2}\right\|$	$\frac{1}{\sqrt{6}}\,\frac{B}{\hbar}\,k_+$	$-\sqrt{\frac{2}{3}}\,\frac{P_0}{\hbar}$	$-\frac{\hbar\,(\gamma_1+2\gamma_2)}{m_0}\hat{k}_z$	$-\sqrt{3}\frac{\hbar\,\gamma_3}{m_0}k_+$	0	0	$\frac{3}{\sqrt{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_-$	$2\sqrt{2}\frac{\hbar\,\gamma_2}{m_0}\hat{k}_z$
$\left\langle\frac{3}{2},\frac{3}{2}\right\|$	0	$-\frac{1}{\sqrt{2}}\,\frac{B}{\hbar}\,k_+$	$-\sqrt{3}\frac{\hbar\,\gamma_3}{m_0}k_-$	$-\frac{\hbar\,(\gamma_1-2\gamma_2)}{m_0}\hat{k}_z$	0	0	0	$\sqrt{\frac{3}{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_-$
$\left\langle\frac{3}{2},-\frac{3}{2}\right\|$	$-\frac{1}{\sqrt{2}}\,\frac{B}{\hbar}\,k_-$	0	0	0	$-\frac{\hbar\,(\gamma_1-2\gamma_2)}{m_0}\hat{k}_z$	$\sqrt{3}\frac{\hbar\,\gamma_3}{m_0}k_+$	$\sqrt{\frac{3}{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_+$	0
$\left\langle\frac{3}{2},-\frac{1}{2}\right\|$	$-\sqrt{\frac{2}{3}}\,\frac{P_0}{\hbar}$	$\frac{1}{\sqrt{6}}\,\frac{B}{\hbar}\,k_-$	0	0	$\sqrt{3}\frac{\hbar\,\gamma_3}{m_0}k_-$	$-\frac{\hbar\,(\gamma_1+2\gamma_2)}{m_0}\hat{k}_z$	$-2\sqrt{2}\frac{\hbar\,\gamma_2}{m_0}\hat{k}_z$	$\frac{3}{\sqrt{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_+$
$\left\langle\frac{1}{2},-\frac{1}{2}\right\|$	$-\frac{1}{\sqrt{3}}\,\frac{P_0}{\hbar}$	$-\frac{1}{\sqrt{3}}\,\frac{B}{\hbar}\,k_-$	$\frac{3}{\sqrt{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_+$	0	$\sqrt{\frac{3}{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_-$	$-2\sqrt{2}\frac{\hbar\,\gamma_2}{m_0}\hat{k}_z$	$-\frac{\hbar\,\gamma_1}{m_0}\hat{k}_z$	0
$\left\langle\frac{1}{2},\frac{1}{2}\right\|$	$\frac{1}{\sqrt{3}}\,\frac{B}{\hbar}\,k_+$	$\frac{1}{\sqrt{3}}\,\frac{P_0}{\hbar}$	$2\sqrt{2}\frac{\hbar\,\gamma_2}{m_0}\hat{k}_z$	$\sqrt{\frac{3}{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_+$	0	$\frac{3}{\sqrt{2}}\,\frac{\hbar\,\gamma_3}{m_0}k_-$	0	$-\frac{\hbar\,\gamma_1}{m_0}\hat{k}_z$

TABLE I: The $\hat{J}_z$ z-component of probability current operator for $\hat{H}_1$ Hamiltonian.

parameter	GaAs value	AlAs value	units	source
$E_g(T=0)$	1.519	3.099	eV	Ref. 28
α	0.5405	0.885	meV/K	Ref. 28
β	204	530	K	Ref. 28
S_0	-2.88	0.04	1	Ref. 28
m^*	0.067	0.15	1	Ref. 28
E_P^0	28.8	21.1	eV	Ref. 28
Δ	0.341	0.28	eV	Ref. 28
B	3.9895	2.7955	hartree $*$ bohr^2	Ref. 43
$\gamma_1^{6\times6}$	6.98	3.76	1	Ref. 28
$\gamma_2^{6\times6}$	2.06	0.82	1	Ref. 28
$\gamma_3^{6\times6}$	2.93	1.42	1	Ref. 28

TABLE II: The values of the $\vec{k}\cdot\vec{p}$ parameters.

equal to $E_P = E_P^0$ for $A = 0$ and $E_P = 0$ for $A = 1$, see Eq. (34).

III. MULTIPASS ABSORPTION MEASUREMENT

As in Ref. 1, THz absorption measurements were performed in a multi-pass geometry with 45° facets [the same geometry shown in Ref. 46, Fig. 7(a)]. The samples were placed inside a continuous-flow cryostat specifically designed to fit the tight space inside the chamber of a Fourier transform infrared spectrometer. Polarized THz light from the FTIR thermal globar source (Silicon carbide) was focused at the input facet of the sample. The transmitted light was detected with a liquid-Helium cooled Silicon bolometer. Since the quantum wells only absorb TM (transverse magnetic)-polarized light, a TE (transverse electric)-polarized measurement was used as a reference. The ratio of the TM spectra to the TE spectrum provides the quantum well absorption, after correction for the source elliptical polarization. Note that, in the case of high temperatures $T > 200$ K for the G0643 sample, the signal to noise ratio was too small to unambiguously identify the frequency of the absorption maximum. (This sample had fewer quantum wells compared to the other samples, and thus yields a weaker overall absorption.)

The derivation of our mathematical model for multipass absorption through a quantum well is presented in Sec. C. The frequencies of the ISB plasmon absorption peak maxima and the corresponding FWHM obtained with the help of that model are shown in Table III.

IV. RESULTS

A. Temperature dependence of the frequency of absorption maximum for each of the samples

As explained in Sec. II G, the 8-band $\vec{k}\cdot\vec{p}$ model (see Sec. II E) uses the introduced A parameter, which needs to be fitted. Our intention was to use experimental data for one sample in order to fit this parameter and to test the performance of thus calibrated model on other two samples. We designated the G0490 sample as a fit sample as it is the system with the best performance in terms of the linewidth, while the G0489 and G0643 samples were designated as test samples (see Sec. II A for a detailed description of the samples). The parameter was fitted

temperature	G0490I – F	G0490I – L	G0490II – F	G0490II – L	G0489 – F	G0489 – L	temperature	G0643 – F	G0643 – L
10	2.939	0.110	2.935	0.164			4	2.193	0.086
78	2.891	0.117	2.888	0.182	2.845	0.355	8	2.188	0.088
100	2.876	0.141	2.873	0.214	2.833	0.379	15	2.186	0.089
150	2.850	0.239	2.842	0.287	2.812	0.435	40	2.169	0.099
200	2.832	0.371	2.822	0.405	2.790	0.642	60	2.145	0.108
250	2.811	0.535	2.803	0.555	2.776	0.732	80	2.128	0.122
293	2.791	0.715	2.777	0.711	2.779	0.813	100	2.111	0.148
							150	2.075	0.241
							200	2.050	0.273

TABLE III: The frequencies (the F columns) and the corresponding FWHM (the L columns) of the ISB plasmon absorption peak maxima in THz versus the measurement temperature in K from the multipass absorption experiment. The G0490I and G0490II columns are for the first and the second independent measurements of the G0490 sample. Please note that the data in columns G0490I and G0489 were presented first in Ref. 1

in respect to the data-set consisting of 6 frequencies of the absorption peak maxima obtained from the multipass absorption experiment (see Table III) of G0490 for $T \in (75, 300)$ K, averaged between the two independent measurements.

Because we use these data for *a posteriori* calibration, the measured FWHM values at each temperature were used as a reference for the γ_T^{IB} parameter in Eq. (26). Conversely, we do not assume to know the correct temperature dependence of γ_T^{IB} in the test cases, beyond general range of relevant values. This situation corresponds to a setup in which the model would be put to a practical use: predicting the characteristics of a proposed system, the growth of which we are only considering. To keep things simple, in the test cases we use constant γ^{IB} corresponding to experimental FWHM at $T = 78$ K, which allows us to avoid introducing additional error from a supposedly off guess, while still not putting the $\gamma_T^{\mathrm{IB}} = f(T)$ as a given.

The effective permittivity $\epsilon_{zz,\mathrm{eff}}$ was calculated for different values of A, from which the frequency of the absorption maximum was obtained. The value $A = 0.7336$ minimizes the total squared error of the predicted frequencies in respect to the measured ones in the relevant temperature range. Please note that, as has been explained in Sec. II G, $A = 0$ means that the default $\vec{k}\cdot\vec{p}$ values of S and E_P are used, while for $A = 1$, the reverse of the single-band effective mass is used: $S = \frac{1}{m^*}$ and $E_P = 0$. The fitted $A = 0.7336$ might be interpreted as a significant shift towards greater impact of the parabolic-model-like kinetic term and lower impact of the conduction-valence band mixing in the transition current densities. However, caution must be taken, as this is a correction to the Schrödinger-Poisson results that were themselves obtained with $S = 1$ renormalization.

The experimental points are marked in Fig. 2 with black circles, with (a) corresponding to the test sample. The frequencies of the absorption peak maxima for the 8-band $\vec{k}\cdot\vec{p}$ model with fitted A are presented in Fig. 2 by the blue squares. The corresponding values yielded by the parabolic (see Sec. II C) and the hybrid (see Sec. II D) models are shown by green pointing-up triangles and by red pointing-down triangles, respectively. The experimental and the 8-band $\vec{k}\cdot\vec{p}$ points are accompanied by error bars. In case of the measured values, they come from the finite frequency resolution of the experiment in Fig. 2(a,b,c). Additionally, as the sample G0490 was measured twice over the whole range of temperatures, the difference between the obtained frequencies at each temperature also contributed to the uncertainty. In case of the 8-band $\vec{k}\cdot\vec{p}$ points the bars show an estimate of the variation of the model predictions depending on the 1D vs 2D averaging of $\tilde{I}^{\hat{n}}_{\beta,k_\parallel,\beta',k'_\parallel}$ [see Eq. (25) and discussion in Sec. D] and the $\mathscr{E} = E^{\hat{n}}_{k'_\parallel,m'} - E^{\hat{n}}_{k'_\parallel,n'}$ vs $\mathscr{E} = E_{\Gamma,m'} - E_{\Gamma,n'}$ variants [see Eq. (26) and discussion in Sec. F 4].[g] Moreover, the Schrödinger-Poisson system was solved twice with the help of two versions of nextnano++ in the case of G0490 sample to control for the imperfect convergence of the self-consistent simulation. This gave a *small* additional contribution in that case.

In Fig. 2(a), it can be seen that of the three models considered, the parabolic model does the worst job in describing the experimental data. Not only are the differences between the measured and predicted frequencies of the absorption maxima the biggest at each temperature, but also their general trend vs temperature is reversed: increasing T results in a redshift as observed experimentally, but the parabolic model predicts a blueshift. With the inclusion of the non-parabolic dispersion in the scope of the hybrid model, the prediction gets better, with all the predicted peak positions moving towards lower frequencies. Moreover, the temperature trend changes to a neutral one. Finally, the data points are visibly best described by the 8-band $\vec{k}\cdot\vec{p}$ model, both on a level of

---

[g] To be specific, the predicted frequency spread at each temperature was calculated and – as they seem to variate erratically with temperature – the largest one was used for a given sample. In practice, we found that the impact of the averaging mode is negligible in comparison to the $\mathscr{E}$ one.

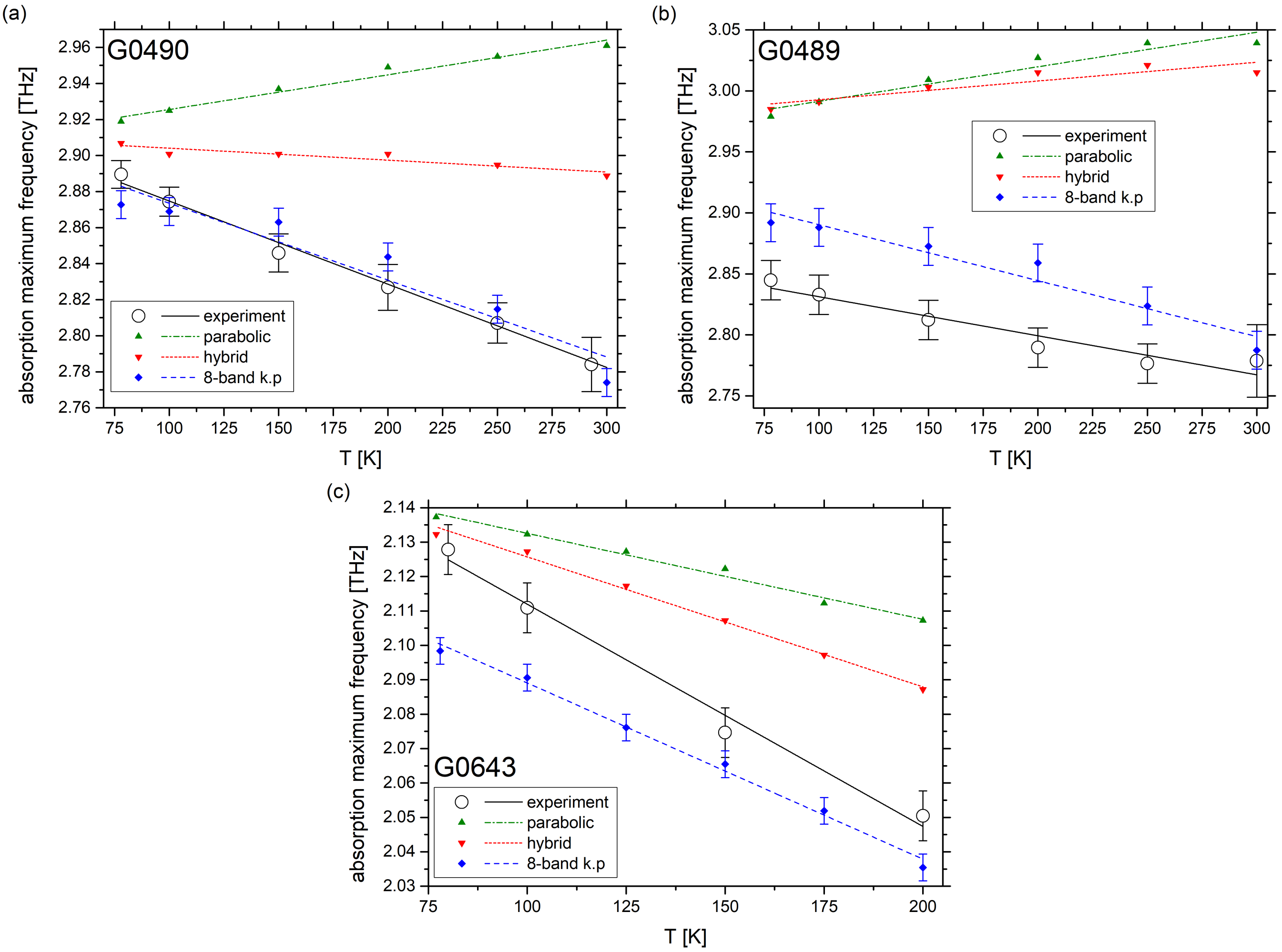

FIG. 2: The dependence of frequency of the absorption maximum on temperature for: (a) G0490, (b) G0489 and (c) G0643. The black circle points show the multipass measurement values, the green pointing-up triangles represent the parabolic model predictions, the red pointing-down triangles – the ones of the hybrid model and the blue squares – of the 8-band $\vec{k}\cdot\vec{p}$ model. The lines show the linear fits to the data points from which the average slopes were obtained (see text). The black error bars come from the experimental uncertainty estimate, while the blue ones from the differences between 1D vs 2D averaging and the $\mathscr{E}$ variants of the $\vec{k}\cdot\vec{p}$ model.

the individual frequencies, as well as in respect to the temperature redshift that this model predicts.

In order to quantify the mentioned trends, the average slope of frequency of the absorption maximum vs temperature was calculated for the experimental data as well as for the predictions of the three models of Fig. 2(a). The slopes are: $-0.46\frac{\mathrm{GHz}}{\mathrm{K}}$ for the experiment, $0.19\frac{\mathrm{GHz}}{\mathrm{K}}$ for the parabolic model, $-0.07\frac{\mathrm{GHz}}{\mathrm{K}}$ for the hybrid model and $-0.43\frac{\mathrm{GHz}}{\mathrm{K}}$ in case of the 8-band $\vec{k}\cdot\vec{p}$ model. Please note that the slope of the last model is quite close to the one of the experimental data, while the parabolic and the hybrid models yield a reversed (positive) and a neutral slope, respectively.

The validity of the $A = 0.7336$ fit itself is thus confirmed in Fig. 2(a), as there is a value of this scalar parameter which predicts reasonably well the frequencies of the absorption maxima for all temperatures in the range $T \in (75, 300)$ K – which was not a given at all in the first place. But this does not inform us about the applicability of the A as a postulated material-structural parameter, that is if it will work reasonably well for other systems of a similar structure. In order to check this, in Fig. 2(b) and (c) the frequencies of the absorption maxima are presented for samples G0489 and G0643, respectively. Please recall that the former sample is basically the same system as the fit sample but doped three times more strongly, while the latter sample has significant structural differences in terms of size, the maximal aluminium concentration used and doping – both the mode and the amount – in respect to G0490.

The corresponding results for the highly-doped G0489 are shown in Fig. 2(b). In this case the difference between the predictions of both the parabolic and the hybrid models and the experimental data is much larger, while the two models yield similar frequency values and a similar positive average slope: $0.28\frac{\mathrm{GHz}}{\mathrm{K}}$ and $0.15\frac{\mathrm{GHz}}{\mathrm{K}}$, respectively. One can see that also here the hybrid model is an improvement in respect to the parabolic one, both in the terms of individual frequencies as well as in case of the average slope, if just a slight one. The predictions of the $\vec{k}\cdot\vec{p}$ model also do not fit the experimental data as well as in the case of G0490, which is to be predicted as that was the fit system. However, at each of the temperature values considered, the error is significantly smaller that in the case of the two other models. Additionally, it is the only model yielding a negative average slope of $-0.46\frac{\mathrm{GHz}}{\mathrm{K}}$, with the experimental value of $-0.32\frac{\mathrm{GHz}}{\mathrm{K}}$ for comparison.

The hardest test for our model is the 2 THz system G0643, significantly more different from the G0490, for which the results are shown in Fig. 2(c). Here, the errors in predictions are visibly smaller in case of lower temperatures for both the parabolic and the hybrid models in comparison to the $\vec{k}\cdot\vec{p}$ model, while the situation reverses in case of higher temperatures. As no model is universally and obviously better, like it was the case in Fig. 2(a) and (b), a numerical estimate of total error needs to be used. After comparing the total absolute difference between the experimental frequencies and the ones predicted by each of the models, it turns out that the parabolic model value is 1.92 times the one of the $\vec{k}\cdot\vec{p}$ model, while the hybrid model value is 1.28 times the one of the $\vec{k}\cdot\vec{p}$ model. In case of this sample, each of the models yields a negative average slope, namely $-0.25\frac{\mathrm{GHz}}{\mathrm{K}}$, $-0.38\frac{\mathrm{GHz}}{\mathrm{K}}$ and $-0.51\frac{\mathrm{GHz}}{\mathrm{K}}$ for the parabolic, hybrid and $\vec{k}\cdot\vec{p}$ models, respectively, as compared to $-0.65\frac{\mathrm{GHz}}{\mathrm{K}}$ experimental average slope.

B. General conclusions about the models

We find that we can only achieve qualitative improvement in the prediction by using the model which includes *both* the dependence of the energy dispersion and of the wavefunction on $\vec{k}_{||}$. We see this conclusion as the main result of our work. In order to get the right slope and energies it turned out to be necessary to deviate from one or zero in the S renormalization method and to introduce a $\vec{k}\cdot\vec{p}$-derived current density operator. Unlike some simpler (e.g. few-band, parabolic or single-particle) systems, the off-Γ dynamic of the wavefunctions, and specifically their transition current elements, can in fact qualitatively change the behaviour of our multi-subband collective system. In fact, we tried to include only the off-Γ energy dispersions first and we have shown that it is not enough.

We conclude that the $\vec{k}\cdot\vec{p}$ model describes both the values and the temperature change of the absorption maxima frequency better than the other models in case of all three samples presented in Fig. 2. However, the relative performance of this model, in relation to its competitors, decreases from the fit system (G0490), to one structurally identical with a different doping (G0489), to one structurally similar but with different parameters (G0643). In the first case, Fig. 2(a), the difference between predictions of the model and the experimental data stay within or close to the experimental uncertainties. In the second case, Fig. 2(b), the error is larger, but still the predictions of the $\vec{k}\cdot\vec{p}$ model are consistently and visibly the best for any $T \in (75, 300)$ K. In the third case, Fig. 2(c), only the overall error is the smallest for the $\vec{k}\cdot\vec{p}$ model. Similarly, the average temperature slope of absorption maximum frequency for that model is always closest to the one for the experimental data, but while it is the only one in a qualitative agreement with the latter for G0490 and G0489, in the case of G0643 we can only say that its value is quantitatively the best one.

The partially qualitative character of the results is not unexpected, as many parameters of the system are not precisely known. For example, the proportion in which the difference in the energy gap between gallium and aluminum divides between the conduction and the valence bands was taken as $60:40$ at zero temperature, resulting in a valence band offset parameter which is assumed not to depend on temperature or any other variable. This may not necessarily be the case. For another instance, the Schrödinger-Poisson computation is done with periodic boundary conditions in the growth direction which would strictly speaking correspond to an infinite superlattice of quantum wells. In reality, we have a finite number of wells, which can also differ between the samples. Different well positions within the superlattice are not equivalent, due to band bending at the edges of the sample. Furthermore, while the total amount of the donor centers can be estimated from the MBE growth parameters up to about 10% uncertainty, it is not certain if all of the centers are ionized or if all of the electrons are available for the plasmon formation. Additionally, the homogeneous broadening parameter γ_{IB} was simplistically assumed to be an uniform scalar. All the uncertainties above likely weaken our ability to accurately estimate the A parameter, which is only supposed to quantify the $S \leftrightarrow P_0$ balance. Having pointed out all of that, we still have found that an overall improvement can be achieved by including the 8-band $\vec{k}$-dependent wavefunctions and dispersion relations as well as replacing the current density operator in the transition current density Eq. (3) with the one corresponding to $\vec{k}\cdot\vec{p}$ Hamiltonian. We interpret that as a success in capturing at least some of the targeted $\vec{k}_{||}$-dynamics.

Apart from that, we also conclude that the hybrid model provides a consistent improvement on the parabolic model, with only a small additional computational cost. Furthermore, we observe that the linear fit agrees with the experimental data within the uncertainty estimates and always gives a negative slope of a fraction of GHz/K order of magnitude. Consequently, we can say that our parabolic quantum well systems experience a linear redshift in the energy with rising temperature.

V. FUTURE OUTLOOK

As for the further developments of the model, the low-temperature limitation of the model applicability (see the discussion in Sec. G) could likely be improved with a more sophisticated implementation. Reducing $k_\parallel^{\max}$ for lower T might be a way to decrease $\Delta_{k_\parallel}$ without increasing the size of the mesh. One could, for example, run the nextnano++ computation, then based on the dispersion obtained, automatically or manually optimize the mesh size in each case, re-run the nextnano++ simulation with the mesh thus obtained and finally use the latter results in the ISB formation model.

More interesting, of course, would be the application of our model to other structures. Extension to a wide set of zincblende-based nanostructures should be readily achievable. For example, while the $\vec{k}\cdot\vec{p}$ model model as it is now has been simplified to wells with symmetric binding potential, it could be readily extended to asymmetric structures too. Furthermore, in GaAs/AlGaAs the primary use of this theory is for quasi-Kohn regime systems, where the not so strong $\vec{k}_\parallel$-non-parabolicity can be visible. However, this model can be extended to narrow band materials like InSb, where the inclusion of non-parabolicity due to conduction-valence mixing would be important even for non parabolic-shaped quantum wells.

While the effect of the non-parabolicity on the position of the absorption maximum is interesting in itself, this result could be also useful in further theoretical work e.g. on well-in-cavity light source based on optical parametric oscillation of ISB polaritons. Among further research goals that have yet to be realized in this area are: (I) the theory of nonlinear processes involving polaritons and (II) the microscopic model of polariton-polariton interaction. Subject (I) has been studied in the context of ISB transitions (as opposed to MSB plasmons) in Ref. 47, where a generalized Gross-Pitaevskii model was used to study intersubband polariton lasing. The work, while preserving time locality, studied the importance of non-Markovian character of the decay of cavity polariton modes to the external radiation occurring due to Bragg scattering. The nonlinear optical amplification of the pump-signal-idler polaritonic scattering type was demonstrated. Subject (II) has been addressed in Ref. 48, where the time-dependent Schrödinger-Poisson equation was solved for a system with a single-band scalar effective mass and a square well potential. The transition frequency and the electric dipole were inferred as two simple functions of the excitation level from the response of the system to a given electric field pulse drive $\epsilon(t)$.

Such works have thus far relied on the approximation that the quantum oscillator can be interpreted as an individual ISB transition, strictly connected to a single energy difference between two given QW subbands. Some collective effects are captured by assigning both a depolarization blueshift and a non-linear optical redshift. However, for cases where the more complex behaviour of MSP modes plays an important role, it could be interesting to treat the formation of MSP modes *first*, using a similar approach to the one described in the present work. Afterwards, the coupling of the quantum oscillator to the cavity modes could be applied. This could be of particular interest in the the quasi-Kohn regime, where the blueshift and the redshift strongly suppress each other. Further, using an approach like that of the present work to incorporate effects like nonparabolicity could help hone modelling for the future development of non-linear optical devices.

Appendix A: Single particle levels and occupation

The first step in the calculation process is to obtain the dispersion of subband energy levels for a single electron in the conduction band with a Schrodinger-Poisson solver. We employed the nextnano++ software. While the dispersions (and the wavefunctions) of a significant part of the Brillouin zone are taken into account in the $\vec{k}\cdot\vec{p}$ model, only the energies at the Γ point are presented in the Fig. 3(a) as a function of temperature to illustrate the general character of the system. The dots show the values from nextnano++ calculation, while the lines show the polynomial fit of the lowest possible degree as the interpolation.

In the case of the G0490 (fit sample) – the red lines – the Fermi level (green horizontal line, zero on the energy scale) lies between the energies of the first and the second subbands at low T. The former crosses the E_F at ~ 49 K, and the levels lie increasingly further above the E_F as temperature rises. In the case of G0489 (3 THz test sample) – the black lines, offset by 125 meV for clarity – the Fermi level lies between the second and the third subband energy level for $T\to 0$. The second level crosses the Fermi level at ~ 46 K, then the first one for ~ 90 K, and then the levels continue to increase in energy as T grows. For the 2 THz system (the G0643 sample) – the blue lines, offset by 250 meV – in the temperature range considered, all the levels are already significantly above the E_F.

Fig. 3(b-d) show how these dispersions translate into normalized occupation differences of the individual α resonators at the Γ point: $\frac{\Delta f_\alpha(\Gamma)}{\sum_\alpha \Delta f_\alpha(\Gamma)}$, for the three samples. In the parabolic model, the shown occupation differences are weights governing the impact of the resonator corresponding to an individual transition between the neighboring subbands, see Eq. (4). In the hybrid and $\vec{k}\cdot\vec{p}$ models, the situation is more complex, as the differences are non-trivially $\vec{k}_\parallel$-dependent, but a detailed multi-parameter analysis of this is clearly beyond the scope of the simple explanation given in this appendix.

In (b) and (c), the occupation differences for low T are dominated by one transition: the $\alpha = 1$ in the case of G0490 and $\alpha = 2$ in the case of G0489. For $T \sim 300$ K

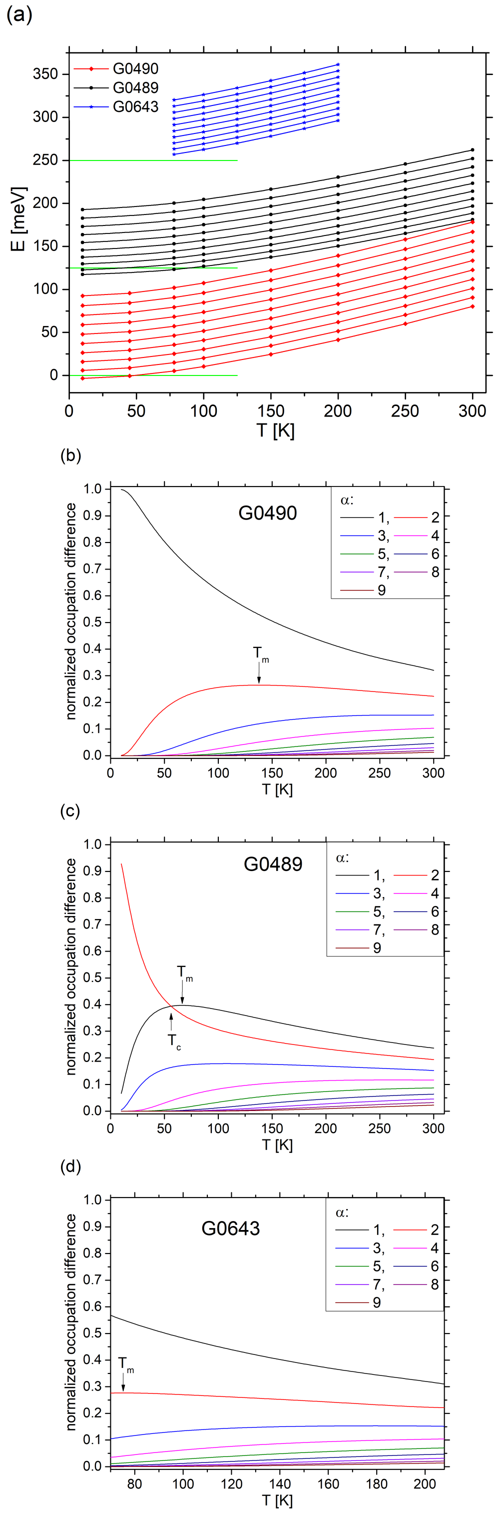

FIG. 3: (a) The energy dispersion of the subband levels vs T. Red color is for G0490, black for G0489 and blue for G0643. Symbols are the nextnano++ results and the lines are a polynomial fit. The green horizontal line is the E_F for each of the samples, which are offset for clarity. (b-d) The normalized occupation differences for individual α resonators at the Γ point vs T. The (b) is for G0490, (c) for G0489 and (d) for G0643.

however, the values become comparable, which enables multiple α to participate in the MSP formation. In the case of the 2 THz system shown in Fig. 3(d), the values for different α are relatively even more comparable in the whole temperature range considered, with the relative differences diminishing as the temperature rises. The results of Fig. 3(b,c) clearly indicate a qualitative transition between an ISB plasmon and a MSP as the temperature rises. The same effect could be suspected in the (d) case, if the temperature range considered was broader.

There is some additional dynamics in the intermediate range of temperature. For example, the crossing of the $\alpha = 1$ and $\alpha = 2$ values for G0489 in (c) at $T_c \sim 56$ K or the fact that the value that starts second largest for low T [i.e. $\alpha = 2$ in (b,d) and $\alpha = 1$ in(c)] has a maximal value at some point [$T_m \sim 138$ K in (b), $T_m \sim 67$ K in (c), $T_m \sim 75$ K in (d)] and then they start to diminish – which would probably also happen to next largest ones at sufficiently large temperature. However, in general the impact of the normalized occupation differences should not be overestimated. While it is significant, in the end it turned out that in order to explain the behavior of the system, not only including the $\vec{k}_{\|}$ non-parabolicity of the energy dispersion is needed, but also taking into account the change of the wavefunctions beyond the Γ point.

Appendix B: Electromagnetic derivations for the semiclassical model

In this section the solutions to the electromagnetic field equations for the semiclassical plasmon model will be derived, filling in some gaps of the derivation in *AA*, and also generalizing their results to *QW* that might not have inversion symmetry. Finally, an expression for a *local* effective permittivity tensor in the long-wavelength approximation is obtained.

1. Electromagnetic fields in the quantum well region

The Maxwell's equations, in the case of no sources and no magnetic materials, have the form of:

$$\nabla \times \mathcal{E} = -\mu_0 \partial_t \mathcal{H}, \tag{B1}$$

$$\nabla \times \mathcal{H} = \partial_t \mathbf{D}, \tag{B2}$$

$$\nabla \cdot \mathbf{D} = 0, \tag{B3}$$

$$\nabla \cdot \mathbf{B} = 0. \tag{B4}$$

$\mathcal{E}, \mathbf{D}, \mathcal{H}, \mathbf{B}$ are the electric field, electric displacement field, magnetic field, and magnetic flux density, respectively. In general, we would like to solve full structures such as the *MIM* cavities[31–39], however we will focus here on the *QW* region, which includes a non-local susceptibility tensor in the growth direction (z). Once we have an effective permittivity for this region, the other regions can be solved using a standard electromagnetic solver.

We assume that the QW region is time-invariant and spatially invariant in the $x-y$ plane. This will allow us to replace $\partial_t \to i\omega, \partial_x \to iq_x, \partial_y \to iq_y$ in the following, where q_x and q_y are the spatial wave numbers. (This is equivalent to taking Fourier transforms of Maxwell's equations.) For brevity of notation, we will continue to use the same field symbols, but we will assume from this point on that all fields have been Fourier transformed along these coordinates and are functions of ω, q_x, q_y. For example, $D_z(z) \equiv D_z(\omega, q_x, q_y, z) \equiv \mathcal{F}_{x,y,t}\{D_z(x,y,z,t)\}$.

Further, we can always split the field into two types of modes:[49]

- TM modes where $\mathcal{H}_z = 0$
- TE modes where $\mathcal{E}_z = 0$

The TE modes are not of interest here, since they will not couple to the quantum well and can be solved using standard techniques. We will restrict ourselves to TM modes.

For a TM mode, since $\mathcal{H}_z = 0$ by definition, we can write the magnetic field as

$$\mathcal{H} = \mathcal{H}_y(z)\hat{\mathbf{y}}. \tag{B5}$$

In principle, the TM mode could have both x and y components, but, since we have assumed invariance in the $x-y$ plane, we can always rotate our coordinate system so that $\mathcal{H}_x = 0$.[h]

For a material with a constant magnetic permeability, Eqs. (B4) and (B5) give

$$q_y \mathcal{H}_y(z) = 0 \Longrightarrow q_y = 0, \tag{B6}$$

showing that there is no spatial variation in the y direction. In other words, rotating our coordinate system so that $\mathcal{H}_x = 0$ is equivalent to rotating our coordinate system so that the wave is travelling in the $x-z$ plane.

Using Eq. (B5) we can write Eq. (B1) as

$$iq_x\mathcal{E}_y = 0, \tag{B7}$$
$$-\partial_z\mathcal{E}_y = 0, \tag{B8}$$
$$\partial_z\mathcal{E}_x - iq_x\mathcal{E}_z = -i\omega\mu_0\mathcal{H}_y \tag{B9}$$

and Eq. (B2) as

$$iq_x\mathcal{H}_y = i\omega D_z, \tag{B10}$$
$$-\partial_z\mathcal{H}_y = i\omega D_x. \tag{B11}$$

We are particularly interested in solving for the z component of the electric field, which is what couples to the QW. We can rearrange Eqs. (B9), (B10), and (B11) to obtain

$$q_x\partial_z\mathcal{E}_x - iq_x^2\mathcal{E}_z = -i\omega^2\mu_0 D_z, \tag{B12}$$
$$q_x D_x = i\partial_z D_z. \tag{B13}$$

To solve these, we need a relationship between $\mathbf{D}$ and $\mathcal{E}$. We use a non-local, anisotropic susceptibility so that

$$D_x = \epsilon_0\epsilon_{xx}\mathcal{E}_x, \tag{B14}$$
$$D_z = \epsilon_0\epsilon_s\mathcal{E}_z + \epsilon_0\int \chi_{zz}(\omega,\mathbf{q};z,z')\mathcal{E}_z(z')\,\mathrm{d}z'\,, \tag{B15}$$

where ϵ_{xx} is the in-plane permittivity and ϵ_s is the background permittivity of the undoped semiconductor. In the simplest case, we would have $\epsilon_{xx} = \epsilon_s =$ constant. However, in general, they could be different from each other and dependent on ω to account for in-plane plasma oscillations[50] and/or the frequency dependent permittivity of the background semiconductor (e.g., from optical phonons).

In the z direction, let us simplify the notation for the time being by writing

$$D_z = \epsilon_0(\epsilon_s + \chi_{zz})\mathcal{E}_z, \tag{B16}$$

with the understanding that χ_{zz} is acting as a linear operator. Plugging these in to Eq. (B12) and simplifying, (assuming $\partial_z\epsilon_{xx} = 0$) we get

$$q_x(\epsilon_s + \chi_{zz})\partial_z D_x - iq_x^2\epsilon_{xx}D_z = -i\omega^2\mu_0\epsilon_0\epsilon_{xx}(\epsilon_s+\chi_{zz})D_z. \tag{B17}$$

Then, using Eq. (B13), we get

$$(\epsilon_s + \chi_{zz})\partial_z^2 D_z - q_x^2\epsilon_{xx}D_z = -k_0^2\epsilon_{xx}(\epsilon_s+\chi_{zz})D_z, \tag{B18}$$

where $k_0^2 = \omega^2\mu_0\epsilon_0 = \frac{\omega^2}{c^2}$. Rearranging, we get

$$\left(\epsilon_s\partial_z^2 + \epsilon_{xx}\epsilon_s k_0^2 - \epsilon_{xx}q_x^2\right)D_z = -\chi_{zz}\left(\partial_z^2 + \epsilon_{xx}k_0^2\right)D_z. \tag{B19}$$

Defining

$$k_z^2 = \epsilon_{xx}k_0^2 - (\epsilon_{xx}/\epsilon_s)q_x^2 \tag{B20}$$

and writing out the χ_{zz} operator explicitly, we get the final result[i]

$$\left(\partial_z^2 + k_z^2\right)D_z = -\int \frac{\chi_{zz}(\omega,\mathbf{q};z,z')}{\epsilon_s}\left(\partial_{z'}^2 + \epsilon_{xx}k_0^2\right)D_z(z')\,\mathrm{d}z'\,. \tag{B21}$$

---

[h] Because of this, in the current section, q_x plays the role of $q_{\|}$ of Sec. II C – see below.

[i] Note: our definition of k_z reduces to that of AA when $\epsilon_{xx} = \epsilon_s$. Care must be taken when applying boundary conditions in the case $\epsilon_{xx} \neq \epsilon_s$, because in that case this k_z in the QW region will not be equal to k_z in the regions outside the well.

Finally, using Eq. (2), we obtain the result of Eq. (6) of AA:

$$\big(\partial_z^2 + k_z^2\big) D_z = -\sum_\alpha \frac{\chi_\alpha(\omega, q_x)}{\epsilon_s} \xi_\alpha(z) \times \int \xi_\alpha(z') \big(\partial_{z'}^2 + \epsilon_{xx} k_0^2\big) D_z(z') \, \mathrm{d}z' \,. \tag{B22}$$

2. A general solution

We can "guess" a solution to Eq. (B22) of the following form:

$$D_z(z) = A\cos(k_z z) + B\sin(k_z z) + q_x^2 \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} \Big(AF_\alpha^A + BF_\alpha^B\Big) \int \xi_\alpha(z') g(z, z') \, \mathrm{d}z' \tag{B23}$$

with the Green's function

$$g(z, z') = -\sin(k_z|z - z'|)/2k_z. \tag{B24}$$

This is quite similar to Eq. (8) in AA, however, their solution only works for adjacent transitions in symmetric quantum wells, where it can be assumed that $\xi_\alpha(z)$ is even with respect to z. What follows, allows to solve the equation without making such an assumption.

Please note that

$$\big(\partial_z^2 + k_z^2\big) \cos(k_z z) = 0, \tag{B25}$$

$$\big(\partial_z^2 + k_z^2\big) \sin(k_z z) = 0 \tag{B26}$$

$$\text{and} \quad \big(\partial_z^2 + k_z^2\big) g(z, z') = -\delta(z - z'). \tag{B27}$$

Applying $\big(\partial_z^2 + k_z^2\big)$ to both sides of Eq. (B23) gives us

$$\big(\partial_z^2 + k_z^2\big) D_z(z) = -q_x^2 \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} \big(AF_\alpha^A + BF_\alpha^B\big) \xi_\alpha(z). \tag{B28}$$

Comparing to (B22), we see that our guess for $D_z(z)$ will be a solution if

$$AF_\alpha^A + BF_\alpha^B = \frac{1}{q_x^2} \int \xi_\alpha(z') \big(\partial_{z'}^2 + \epsilon_{xx} k_0^2\big) D_z(z') \, \mathrm{d}z' \,. \tag{B29}$$

At this point, our solution is self-referential, but it turns out that it is possible to solve for the F_α^A and F_α^B coefficients without knowing D_z first.[j]

3. Solving for the F coefficients

From Eq. (B20) we have

$$\big(\partial_z^2 + \epsilon_{xx} k_0^2\big) D_z = \frac{\epsilon_{xx}}{\epsilon_s} q_x^2 D_z + \big(\partial_z^2 + k_z^2\big) D_z, \tag{B30}$$

which, after using Eq. (B28), becomes

$$\big(\partial_z^2 + \epsilon_{xx} k_0^2\big) D_z = \frac{\epsilon_{xx}}{\epsilon_s} q_x^2 D_z - q_x^2 \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} \Big(AF_\alpha^A + BF_\alpha^B\Big) \xi_\alpha(z). \tag{B31}$$

Plugging this into Eq. (B29) gives

$$AF_\alpha^A + BF_\alpha^B = \frac{\epsilon_{xx}}{\epsilon_s} \int \xi_\alpha(z') D_z(z') \, \mathrm{d}z' - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} \big(AF_{\alpha'}^A + BF_{\alpha'}^B\big) I_{\alpha,\alpha'} \, \mathrm{d}z \,, \tag{B32}$$

where

$$I_{\alpha,\alpha'} = \int \xi_\alpha(z) \xi_{\alpha'}(z) \, \mathrm{d}z \,. \tag{B33}$$

Furthermore, by applying $(\epsilon_{xx}/\epsilon_s) \int \mathrm{d}z' \, \xi_\alpha(z')$ to Eq. (B23), we see that

$$\frac{\epsilon_{xx}}{\epsilon_s} \int \xi_\alpha(z') D_z(z') \, \mathrm{d}z' = AC_\alpha^A + BC_\alpha^B - q_x^2 \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} \big(AF_{\alpha'}^A + BF_{\alpha'}^B\big) D_{\alpha,\alpha'}, \tag{B34}$$

where we have defined for convenience

$$C_\alpha^A = \frac{\epsilon_{xx}}{\epsilon_s} \int \cos(k_z z) \xi_\alpha(z) \, \mathrm{d}z \,, \tag{B35}$$

$$C_\alpha^B = \frac{\epsilon_{xx}}{\epsilon_s} \int \sin(k_z z) \xi_\alpha(z) \, \mathrm{d}z \tag{B36}$$

$$\text{and} \quad D_{\alpha,\alpha'} = -\frac{\epsilon_{xx}}{\epsilon_s} \int \xi_\alpha(z) g(z, z') \xi_{\alpha'}(z') \, \mathrm{d}z \, \mathrm{d}z' \,. \tag{B37}$$

Plugging Eq. (B34) into Eq. (B32), we obtain

$$AF_\alpha^A + BF_\alpha^B = AC_\alpha^A + BC_\alpha^B - q_x^2 \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} \big(AF_{\alpha'}^A + BF_{\alpha'}^B\big) D_{\alpha,\alpha'} - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} \big(AF_{\alpha'}^A + BF_{\alpha'}^B\big) I_{\alpha,\alpha'}. \tag{B38}$$

Finally, we have an equation for the F_α's which does not depend self-referentially on the field $D_z(z)$. Our intention is that the A and B coefficients will depend on the boundary conditions, so we would like a solution for the F_α's that works for any and all values of A and B. As such, the A and B parts of the equation must be satisfied

---

[j] We will still be left with two unknowns A and B in the solution, but this is not surprising because we are searching for a general solution in the QW region, with unspecified boundary conditions. A and B will eventually be determined by the boundary conditions for a particular problem.

independently of each other, and we obtain the following two conditions:

$$F^A_\alpha = C^A_\alpha - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F^A_{\alpha'} \big[I_{\alpha,\alpha'} + q_x^2 D_{\alpha,\alpha'} \big], \tag{B39}$$

$$F^B_\alpha = C^B_\alpha - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F^B_{\alpha'} \big[I_{\alpha,\alpha'} + q_x^2 D_{\alpha,\alpha'} \big]. \tag{B40}$$

Note that these can be written concisely as matrix equations:

$$\mathbf{M}\mathbf{F}^A = \mathbf{C}^A, \tag{B41}$$

$$\mathbf{M}\mathbf{F}^B = \mathbf{C}^B, \tag{B42}$$

with both equations using the same $\mathbf{M}$ matrix:

$$M_{\alpha,\alpha'} = \delta_{\alpha,\alpha'} + \frac{\chi_\alpha}{\epsilon_s} \big[I_{\alpha,\alpha'} + q_x^2 D_{\alpha,\alpha'} \big]. \tag{B43}$$

In summary, we have derived matrix equations for F^A_α, F^B_α which do not depend on $D_z(z)$. We can calculate $\mathbf{M}$, $\mathbf{C}^A$, and $\mathbf{C}^B$, only requiring knowledge of χ_α and $\xi_\alpha(z)$, which can be calculated from the Schrödinger-Poisson solution for the quantum well. Then, by solving the matrix equations, we obtain the F^A_α, F^B_α coefficients, which can be used to construct the $D_z(z)$ field solution Eq. (B23). The two remaining unknowns A, B will be determined by the boundary conditions. We could, for example, construct a transfer matrix similar to AA.

Finally, note that if $\xi_\alpha(z)$ is even (as would be the case for adjacent $n \to n+1$ transitions in a symmetric QW), then we will have $\mathbf{C}^B = 0$. In that case, we will always have $\mathbf{F}^B = 0$ as a valid solution. Then, if we also assume $\epsilon_{xx} = \epsilon_s$, we see that we recover the same solution as AA in the case of inversion-symmetric quantum wells.

4. Effective permittivity

Now that we have a solution for D_z, we can try to construct an effective *local* permittivity tensor. We will do this by calculating the $\mathcal{E}_z$ field from D_z using Maxwell's equations, and then compare the two. As we will see, a local permittivity is only possible in the long-wavelength limit.

To find the $\mathcal{E}_z$ field from the D_z field, we can use Eqs. (B12), (B13), along with $D_x = \epsilon_0 \epsilon_{xx} \mathcal{E}_x$. After simplifying, we get

$$\mathcal{E}_z(z) = \frac{1}{q_x^2 \epsilon_0 \epsilon_{xx}} \big(\partial_z^2 + \epsilon_{xx} k_0^2 \big) D_z(z) \tag{B44}$$

which, using Eq. (B20), can also be written as

$$\mathcal{E}_z(z) = \frac{D_z(z)}{\epsilon_0 \epsilon_s} + \frac{1}{q_x^2 \epsilon_0 \epsilon_{xx}} \big(\partial_z^2 + k_z^2 \big) D_z(z). \tag{B45}$$

Taking into account Eq. (B28), we obtain an expression for the $\mathcal{E}_z$ field:

$$\mathcal{E}_z(z) = \frac{D_z(z)}{\epsilon_0 \epsilon_s} - \frac{1}{\epsilon_0 \epsilon_{xx}} \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} \big(A F^A_\alpha + B F^B_\alpha \big) \xi_\alpha(z). \tag{B46}$$

Comparing this to our D_z field expression Eq. (B23), it is clear (unsurprisingly) that we do not have a simple proportional relationship between the D_z and $\mathcal{E}_z$ fields with which to define a permittivity. However, we *can* define an effective permittivity in the long-wavelength limit by averaging the fields over z and taking a ratio $\langle D_z(z) \rangle / \langle \mathcal{E}_z(z) \rangle$ (a similar approach to Ref. 3). For *THz QW*, the long-wavelength limit is quite reasonable, since we have well widths $\lesssim 100\,\text{nm}$ and electromagnetic wavelengths $\gtrsim 10\,\mu\text{m}$, even after accounting for the refractive index of the semiconductor.

In the long-wavelength limit, we take $q_x, k_0 \to 0$ (which also implies $k_z \to 0$). Then our expression Eq. (B23) for the D_z field becomes simply

$$D_z(z) \approx A. \tag{B47}$$

We also have $C^B_\alpha \approx 0$ for all α, which means that $F^B_\alpha \approx 0$ as well. Our expression (B46) for the $\mathcal{E}_z$ field becomes

$$\mathcal{E}_z(z) \approx \frac{A}{\epsilon_0 \epsilon_s} - \frac{A}{\epsilon_0 \epsilon_{xx}} \sum_\alpha \frac{\chi_\alpha}{\epsilon_s} F^A_\alpha \xi_\alpha(z). \tag{B48}$$

We can define our effective permittivity (in the z direction) as

$$\epsilon_{zz,\text{eff}} = \frac{\langle D_z \rangle}{\epsilon_0 \langle \mathcal{E}_z \rangle} \approx \epsilon_s \left[1 - \sum_\alpha \frac{\chi_\alpha}{\epsilon_{xx}} F^A_\alpha \langle \xi_\alpha \rangle \right]^{-1}, \tag{B49}$$

where $\langle \cdot \rangle$ denotes an average over the entire QW region in the z direction. Furthermore, since we are using the long-wavelength limit anyway, we can also use it to simplify our equation for the F^A_α's:

$$F^A_\alpha \approx \int \xi_\alpha(z)\, \mathrm{d}z - \sum_{\alpha'} \frac{\chi_{\alpha'}}{\epsilon_s} F^A_{\alpha'} I_{\alpha,\alpha'}, \tag{B50}$$

finally reaching the Eq. (14).

It turns out that in the long wavelength limit, the problem is greatly simplified. Once we solve the matrix equation to find the F^A_α's, we can directly calculate an effective permittivity tensor which can be used in a standard electromagnetic solver.

As a final note, one must be slightly careful with the definition of L. Typically, L should be the width of the region simulated by Schrödinger-Poisson. For example, a common case might be a periodic stack of QW separated by barriers. If one performs a Schrödinger-Poisson simulation of the entire stack, then L should be the length of the entire stack. If one performs a simulation of a single well with periodic boundary conditions, then L should be the length of the period. There is some arbitrariness here, but it is important to have consistency between the Schrödinger-Poisson equations, the calculation of the effective permittivity, and in the length of material which is assigned this effective permittivity in the final electromagnetic simulation.

Appendix C: Mathematical model of multipass absorption

In this section an expression for multipass absorption through a quantum well is derived. This approach was used to fit the experimental data in Sec. IV. We consider a sample with three key regions: a substrate, an active region, and a metallic coating. For this model, we treat the metal as a perfect electric conductor, and the substrate as having a constant, real permittivity ϵ_{sub}. We model the active region with an effective complex permittivity tensor. For this work, where the active region comprises a stack of quantum wells, we assume that the growth-direction component ϵ_{zz} takes the form of

$$\frac{1}{\epsilon_{zz}(\omega)} = \frac{1}{\epsilon_s}\left(1 + \frac{\omega_P^2}{\omega^2 - \omega_0^2 + i\gamma\omega}\right) \tag{C1}$$

and that $\epsilon_{xx} = \epsilon_{yy} = \epsilon_s$. However, the electromagnetic derivation that follows is general and does not depend on the particular form of the permittivity tensor, so more sophisticated models could be used.

The main restriction here is that each individual quantum well must be small compared to the wavelength of light, so that the stack can be treated as an effective medium from the perspective of the electromagnetic field. In the case of IV, we are working with free-space wavelengths on the order of 100 µm, which will give a growth-direction wavelength of around $\lambda \approx 40$ µm in GaAs at 45° incidence. The individual quantum wells, which have a width on the order of 0.1 µm, are quite small compared to the wavelength.

It should be emphasized, though, that this does not mean the stack as a whole is negligibly thin. In the case of the 54-well stack, the total thickness of the active region is around 6 µm, which is an appreciable fraction of the wavelength. On the other extreme, the substrate thickness (∼500 µm) is quite large compared to the wavelength, and we will neglect it: effectively treating it as an incoherent reflector. As such, we assume that each reflection off the active region side of the substrate can be treated independently. Then, if we can find the power absorption spectrum for a single bounce, we can multiply to find the total absorption through the structure.

1. The single-pass reflection coefficient

Let z be the growth direction, and let the active region lie in $0 < z < L_a$ with a perfect conductor for $z < 0$. At $z > L_a$ there is substrate with permittivity ϵ_{sub}. We assume that the light arrives at the active region from far away in the substrate, so we can treat the substrate as a semi-infinite slab.

For *TM* light, let's assume the magnetic field, $\mathcal{H}$, is pointing in the y direction. Then $\mathcal{H}$ can be written in the active region $0 < z < L_a$ and substrate $z > L_a$, respectively, as

$$\mathcal{H}_{y,a} = \big(\mathcal{H}_a^- e^{-ik_{z,a}z} + \mathcal{H}_a^+ e^{+ik_{z,a}z}\big) e^{iq_{x,a}x} \tag{C2}$$

$$\text{and } \mathcal{H}_{y,s} = \Big(\mathcal{H}_s^- e^{-ik_{z,s}(z-L_a)} + \mathcal{H}_s^+ e^{+ik_{z,s}(z-L_a)}\Big) e^{iq_{x,s}x}. \tag{C3}$$

Using Maxwell's equations, we obtain

$$i\omega\epsilon_0(\epsilon \cdot \mathcal{E}) = \boldsymbol{\nabla} \times \mathcal{H} = (\partial_x \mathcal{H}_y \hat{\mathbf{z}} - \partial_z \mathcal{H}_y \hat{\mathbf{x}}), \tag{C4}$$

which leads, in the active region, to

$$\mathcal{E}_{z,a} = \frac{q_{x,a}}{\omega\epsilon_{zz}\epsilon_0}\big(\mathcal{H}_a^- e^{-ik_{z,a}z} + \mathcal{H}_a^+ e^{+ik_{z,a}z}\big) e^{iq_{x,a}x}, \tag{C5}$$

$$\mathcal{E}_{x,a} = \frac{k_{z,a}}{\omega\epsilon_{xx}\epsilon_0}\big(\mathcal{H}_a^- e^{-ik_{z,a}z} - \mathcal{H}_a^+ e^{+ik_{z,a}z}\big) e^{iq_{x,a}x} \tag{C6}$$

and in the substrate region to

$$\mathcal{E}_{z,s} = \frac{q_{x,s}}{\omega\epsilon_{\text{sub}}\epsilon_0}\Big(\mathcal{H}_s^- e^{-ik_{z,s}(z-L_a)} + \mathcal{H}_s^+ e^{+ik_{z,s}(z-L_a)}\Big) e^{iq_{x,s}x}, \tag{C7}$$

$$\mathcal{E}_{x,s} = \frac{k_{z,s}}{\omega\epsilon_{\text{sub}}\epsilon_0}\Big(\mathcal{H}_s^- e^{-ik_{z,s}(z-L_a)} - \mathcal{H}_s^+ e^{+ik_{z,s}(z-L_a)}\Big) e^{iq_{x,s}x}. \tag{C8}$$

Because of the conductor boundary condition, we must have $\mathcal{E}_{x,a} = 0$ at $z = 0$, which gives

$$\frac{k_{z,a}}{\omega\epsilon_{xx}\epsilon_0}\big(\mathcal{H}_a^- - \mathcal{H}_a^+\big) e^{iq_{x,a}x} = 0 \tag{C9}$$

or $\mathcal{H}_a^- = \mathcal{H}_a^+ := \mathcal{H}_a$. The fields in the active region are given by

$$\mathcal{H}_{y,a} = 2\mathcal{H}_a \cos(k_{z,a}z) e^{iq_{x,a}x}, \tag{C10}$$

$$\mathcal{E}_{z,a} = 2\mathcal{H}_a \frac{q_{x,a}}{\omega\epsilon_{zz}\epsilon_0} \cos(k_{z,a}z) e^{iq_{x,a}x} \tag{C11}$$

$$\text{and} \quad \mathcal{E}_{x,a} = -2i\mathcal{H}_a \frac{k_{z,a}}{\omega\epsilon_{xx}\epsilon_0} \sin(k_{z,a}z) e^{iq_{x,a}x}. \tag{C12}$$

The second boundary condition is that the tangential $\mathcal{E}$ and $\mathcal{H}$ fields must be continuous across the boundary at $z = L_a$. Firstly, this enforces that $q_{x,s} = q_{x,a}$. Secondly, it gives

$$2\mathcal{H}_a \cos(k_{z,a}L_a) = \big(\mathcal{H}_s^- + \mathcal{H}_s^+\big), \tag{C13}$$

$$-2i\mathcal{H}_a \frac{k_{z,a}}{\omega\epsilon_{xx}\epsilon_0} \sin(k_{z,a}L_a) = \frac{k_{z,s}}{\omega\epsilon_{\text{sub}}\epsilon_0}\big(\mathcal{H}_s^- - \mathcal{H}_s^+\big). \tag{C14}$$

Eliminating $\mathcal{H}_a$, we get

$$\frac{\omega\epsilon_{xx}\epsilon_0}{-ik_{z,a}} \cot(k_{z,a}L_a) = \frac{\omega\epsilon_{\text{sub}}\epsilon_0}{k_{z,s}} \frac{\mathcal{H}_s^- + \mathcal{H}_s^+}{\mathcal{H}_s^- - \mathcal{H}_s^+}, \tag{C15}$$

which becomes

$$i\frac{\epsilon_{xx}k_{z,s}}{k_{z,a}\epsilon_{\text{sub}}} \cot(k_{z,a}L_a) = \frac{1+r}{1-r}, \tag{C16}$$

where $r = \mathcal{H}_s^+/\mathcal{H}_s^-$ is the complex reflection coefficient. Solving for r in the above, we find

$$r = \frac{-k_{z,a}\epsilon_{\rm sub} + ik_{z,s}\epsilon_{xx}\cot(k_{z,a}L_a)}{k_{z,a}\epsilon_{\rm sub} + ik_{z,s}\epsilon_{xx}\cot(k_{z,a}L_a)}. \tag{C17}$$

To actually calculate r, we still need to know the wavenumbers in the substrate and active region. The wavenumbers in the substrate region must obey

$$q_{x,s}^2 + k_{z,s}^2 = \epsilon_{\rm sub}\frac{\omega^2}{c^2} = \epsilon_{\rm sub}k_0^2, \tag{C18}$$

so we can write[k]

$$q_{x,s} = \sqrt{\epsilon_{\rm sub}}k_0\sin\phi; \quad k_{z,s} = \sqrt{\epsilon_{\rm sub}}k_0\cos\phi \tag{C19}$$

where, ϕ is the incidence angle, measured relative to the growth axis.

The wavenumbers in the active region are a bit trickier because of the anisotropy, but it can be shown (using $\boldsymbol{\nabla}\times\mathcal{E} = -i\omega\mu_0\mathcal{H}$) that

$$\frac{q_{x,a}^2}{\epsilon_{zz}} + \frac{k_{z,a}^2}{\epsilon_{xx}} = \frac{\omega^2}{c^2} = k_0^2 \tag{C20}$$

Recall from before that we have $q_{x,a} = q_{x,s}$. Solving for $k_{z,a}$ in terms of k_0, we get

$$k_{z,a}^2 = \epsilon_{xx}\left[1 - \frac{\epsilon_{\rm sub}}{\epsilon_{zz}}\sin^2\phi\right]k_0^2 \tag{C21}$$

These expressions for $k_{z,s}$ and $k_{z,a}$ can be used in Eq. (C17) to calculate the complex reflection coefficient for the active region against a gold coating.

2. Multipass absorption

In a multipass configuration (again, where we are assuming incoherent reflections off the bottom of the substrate so that each reflection is independent), the total transmission through the structure can be calculated as

$$T(\omega) = |r(\omega)|^{2N_{\rm pass}}, \tag{C22}$$

where $N_{\rm pass}$ is the number of passes through the active region (i.e., the number of reflections). This can be approximately calculated[51] using

$$N_{\rm pass} = \frac{L_{\rm sub}}{D_{\rm sub}}\cot\phi, \tag{C23}$$

where $L_{\rm sub}$ and $D_{\rm sub}$ are the length and thickness of the sample, respectively, and ϕ is the incidence angle. However, this is only an approximation. In reality, the number of passes may depend on where the light hits the facet. In general, one would need to know the distribution of light across the facet to calculate the total transmission. In practice, it is quite difficult to know this value exactly, which translates to an uncertainty in the total magnitude of the absorption.

Further, note that we have neglected the reflection off the entry and exit facets: i.e., only a certain fraction of the incoming light will actually enter the sample in the first place. In practice, however such complications are mitigated by dividing the *TM* absorption measurement by an identical *TE* measurement. This cancels out any losses which are common to both *TM* and *TE* light, isolating the effect of the active region, which is anisotropic and only absorbs *TM* light.

Appendix D: Derivation of the Eq. (27) matrix

In this section the connection between *AA* and our model using the $\vec{k}\cdot\vec{p}$-compatible definition of current operator will be discussed. The notation of the mentioned work is adopted for this derivation, with the exception of $\alpha \equiv n \to m$, and additionally defining $J_\alpha = \int \xi_\alpha(z)\,\mathrm{d}z = L\,\langle\xi_\alpha\rangle$.

Firstly, one should note that the following simplification is used in *AA*:

$$\left[\frac{1}{2}\frac{f_{n,\vec{k}} - f_{m,\vec{k}+\vec{q}}}{(\omega+i\delta) - \omega_\alpha(\vec{k},\vec{q})} - \frac{f_{m,\vec{k}+\vec{q}}}{\omega_\alpha^0}\right] + \left[\frac{1}{2}\frac{f_{n,\vec{k}} - f_{m,\vec{k}+\vec{q}}}{-(\omega+i\delta) - \omega_\alpha(\vec{k},\vec{q})} + \frac{f_{n,\vec{k}}}{\omega_\alpha^0}\right] \to \frac{f_{n,\vec{k}} - f_{m,\vec{k}+\vec{q}}}{\omega_\alpha^0}\left[1 + \frac{\omega_\alpha(\vec{k},\vec{q})\omega_\alpha^0}{\omega^2 - \omega_\alpha^2(\vec{k},\vec{q})}\right] \tag{D1}$$

However, from the linear response theory, one in fact gets the following expression (see Sec. F and also *Sec. 4.3 Current-response of a Q2DEG* of Ref. 15 plus *Sec. 3. The (001) SURFACE OF InSb* of Ref. 13 for reference)

$$\left[\frac{1}{2}\frac{f_{n,\vec{k}} - f_{m,\vec{k}+\vec{q}}}{(\omega+i\delta) - \omega_\alpha(\vec{k},\vec{q})} - \frac{f_{m,\vec{k}+\vec{q}}}{\omega_\alpha^0}\right]\xi_{\alpha,\vec{k}}(z)\xi^*_{\alpha,\vec{k}}(z') + \left[\frac{1}{2}\frac{f_{n,\vec{k}} - f_{m,\vec{k}+\vec{q}}}{-(\omega+i\delta) - \omega_\alpha(\vec{k},\vec{q})} + \frac{f_{n,\vec{k}}}{\omega_\alpha^0}\right]\xi^*_{\alpha,\vec{k}}(z)\xi_{\alpha,\vec{k}}(z'), \tag{D2}$$

which only in the case of *real* $\xi_{\alpha,\vec{k}}(z)$ can be simplified to the RHS of Eq. (D1) times $\xi_{\alpha,\vec{k}}(z)\xi_{\alpha,\vec{k}}(z')$. In our case, we have retained the expression Eq. (D2), which is the origin of the σ index in Eqs. (21), (22) and (23) – and further as as part of the β index – and also the origin of the $p(\beta) \equiv p(\sigma)$, $\nu(\beta) \equiv \nu(\sigma)$ symbols in Eqs. (26) and (28).

---

[k] Note we have implicitly assumed $q_{x,s}$ and $k_{z,s}$ to be real here, since we are assuming that the other substrate boundary is far off and thus the incoming light can be treated as a plane wave. If the substrate thickness were not very large compared to the wavelength, we could not necessarily assume that $q_{x,s}$ and $k_{z,s}$ are real, and we would need to consider waveguide modes.

Secondly, using the following set of substitutions:

$$\begin{aligned}F_\alpha &\rightarrow \tilde{F}_\alpha J_\alpha,\\ I_{\alpha,\alpha'} &\rightarrow \tilde{I}_{\alpha,\alpha'} J_\alpha J^*_{\alpha'},\\ \chi_\alpha &\rightarrow \frac{\tilde{\chi}_\alpha}{J_\alpha J^*_\alpha},\end{aligned} \tag{D3}$$

on the set of linear equations for F_α:

$$F_\alpha = J_\alpha - \sum_{\alpha'} \frac{\chi'_\alpha}{\epsilon_s} I_{\alpha,\alpha'} F_{\alpha'}, \tag{D4}$$

which is equivalent to Eq. (14), one obtains *dimensionless* equation

$$\tilde{F}_\alpha = 1 - \sum_{\alpha'} \frac{\tilde{\chi}_{\alpha'}}{\epsilon_s} \tilde{I}_{\alpha,\alpha'} \tilde{F}_{\alpha'}. \tag{D5}$$

Thirdly, one should note that as the $\tilde{I}_{\alpha,\alpha'}$ in our case depend on $\vec{k}'$, the corresponding summation in $\xi_{\alpha'}$ in Eq. (3) in Ref. 12 should be taken outside. This leads to expressions of Eqs. (22) and (26).

Finally, as the $\tilde{I}_{\alpha,\alpha'}$ will also depend on $\vec{k}$ in our model, we introduced the averaging over this index, as in Eq. (25). As the contribution in sum over $\vec{k}'$ in Eq. (26) is approximately proportional to the occupation difference between the subbands of α' transition (see *e.g.* RHS of Eq. D1), we adopted correspondingly the occupation difference between the subbands of α transition as weight for the average. The objection can be made that the averaging the transition current density over $\vec{k}$, while keeping the $\vec{k}'$ free for the sum in the Eq. (26) introduces an arbitrary unequivalence between the two dimensions. To control for this effect, we also used the two-dimensional averaging, see the second line in Eq. (25). Fortunately, it turned out that the difference between the two approaches is minimal, not visible in the scale of the figures presented in the Sec. IV.

Appendix E: Introduction of the J_z into the model

The workhorse of the ISB formation model of Ref. 12 are the transition current densities ξ_α, see Eq. (3). Please note that the $\frac{\partial}{\partial z}$ operator in the latter can be interpreted as a part of the one-band common probability current definition:

$$j_z = \frac{\hbar e i}{2m^*}\left(\psi \frac{\partial \psi^*}{\partial z} - \psi^* \frac{\partial \psi}{\partial z}\right), \tag{E1}$$

so that:

$$\xi_\alpha = i\left[j_z\right]_\alpha. \tag{E2}$$

where ξ_α in contrast to Eq. (3) does not assume that the wavefunctions are real:

$$\xi_\alpha(z) = \frac{\hbar e}{2m^*}\left[\psi^*_m(z)\frac{\partial \psi_n(z)}{\partial z} - \psi_n(z)\frac{\partial \psi^*_m(z)}{\partial z}\right], \tag{E3}$$

see Eq. (2.32) in Ref. 13 and Eq. (76) in Ref. 15. It is important to note that the latter expression is derived in the cited sources with the help of the Kubo formula and the Matsubara method (for the detailed implementation of the latter in a similar context, see *e.g.* Ref. 14). This procedure was applied to a parabolic system with a isotropic effective mass, as demonstrated *e.g.* by Eqs. (2.1)-(2.5) and (3.1)-(3.3) in Ref. 13, by Eqs. (37)-(42) in Ref. 15 and by Eqs. (1)-(8) in Ref. 14.[1] There is, in general, no easy way of including the $k_\parallel$ dependent, non-parabolic subbands spanned onto the Bloch basis of Eq. (19) into the procedure. Instead, we postulate to replace the $[j_z]_\alpha$ as per one-band definition with $J^{\hat{n}}_{\alpha,\sigma,k_\parallel}(z)$ of Eq. 21 – *i.e.* the current density using the operator $\hat{J}_z$ proper for the 8-band $\vec{k}\cdot\vec{p}$, as defined in Sec. II F, following the work of Chao and Chuang, Ref. 40.

Our approach has a couple of advantages. Firstly, the concept of a *scalar effective mass* m^* as present in Eq. (3), ill-fitting to a non-parabolic multi-band system with the $S/\gamma_1/\gamma_2/\gamma_3$ set of parameters depending on the position via composition profile (see Table II), is automatically incorporated into the definition of $\hat{J}_z$. Secondly, there is a dependence on $\vec{k}_\parallel$ – both magnitude and orientation – present in the model, both on the level of the individual wavefunctions $\psi^{\hat{n}}_{n,k_\parallel}(z)$ and the $\hat{J}_z$ operator (which explicitly depends on $\vec{k}_\parallel$).

Appendix F: Derivation of the $P_{zz}(\vec{q}, z, z', \omega)$ in Ref. 15

This section follows the derivation of the zz element of the polarization function P as presented in work of Wendler and Kändler, Ref. 15 – referenced with WK in the text and the equations of this section, using their notation. Here: $\alpha, \beta \in \{x, y, z\}$, K and K' are subband indices, $\varphi_K(z)$ is the corresponding K-subband wavefunction, n_F is the Fermi-Dirac distribution, $\Omega_{KK'}$ is the ω^0_α of AA. $P_{KK'}\left(\vec{q_\parallel}, \omega\right)$ *roughly* corresponds to χ_α while $g_{KK'}(z)$ to $\xi_\alpha(z)$. The goal of this derivation is to explain the origin of the Eq. (26) and to present the problems arising when the eigenfunctions are not real and when the simple relation between the eigenfunction and eigenenergy cannot be used.

1. Introductory steps

Let's start with Eq. (WK81):

---

[1] Strictly speaking, a kind of non-parabolicity is present in the initial considerations of Sec. 2 Ref. 13, however: (I) it is still a one-band model and (II) when the authors discuss an example of aplication of the model in Sec. 3, they *immediately* assume a parabolic system and real eigenfunctions.

$$P_{zz}^{KK'}\left(\vec{q_\parallel},\omega\right) = -\sum_{\vec{k}_\parallel}\left(\frac{\hbar^2 P_{KK'}\left(\vec{q_\parallel},\vec{k}_\parallel,\omega\right)}{4m_e^2} + \frac{\hbar e^2 n_F\left(K,\vec{k}_\parallel\right)}{A m_e^2\Omega_{KK'}}\right). \tag{F1}$$

and Eq. (77) in the same work, in the case of $\alpha = \beta = z$:

$$\begin{aligned} P_{zz}\left(\vec{q_\parallel},z,z',\omega\right) &= \sum_{KK'} P_{zz}^{KK'}\left(\vec{q_\parallel},\omega\right) g_{KK'}(z) g_{KK'}^*(z') \\ &= \sum_{KK'}\left[-\sum_{\vec{k}_\parallel}\left(\frac{\hbar^2 P_{KK'}\left(\vec{q_\parallel},\vec{k}_\parallel,\omega\right)}{4m_e^2} + \frac{\hbar e^2 n_F\left(K,\vec{k}_\parallel\right)}{A m_e^2\Omega_{KK'}}\right)\right] \\ &\quad \times g_{KK'}(z) g_{KK'}^*(z') = T_1 + T_2, \end{aligned} \tag{F2}$$

where

$$\begin{aligned} T_1 &= -\frac{\hbar^2}{4m_e^2}\sum_{KK'}\sum_{\vec{k}_\parallel} P_{KK'}\left(\vec{q_\parallel},\vec{k}_\parallel,\omega\right) g_{KK'}(z) g_{KK'}^*(z') \\ T_2 &= -\frac{\hbar e^2}{A m_e^2}\sum_{KK'}\sum_{\vec{k}_\parallel}\frac{n_F\left(K,\vec{k}_\parallel\right)}{\Omega_{KK'}} g_{KK'}(z) g_{KK'}^*(z'). \end{aligned} \tag{F3}$$

At this point, one can see that the first term is already equivalent to the first term in Eq. (WK74) if one keeps in mind that

$$F_z^{KK'}\left(\vec{q_\parallel},\vec{k}_\parallel,z\right) F_z^{KK'}\left(\vec{q_\parallel},\vec{k}_\parallel,z'\right)^* = g_{KK'}(z) g_{KK'}^*(z'), \tag{F4}$$

see Eq. (75) in the same work. Thus, the following will be an investigation on how the second term T_2 of Eq. (F2) compares to the second term of Eq. (WK74). Mirroring the derivation in *Section 3. The (001) surface of InSb* in work of Eguiluz and Maradudin Ref. 13, the derivative of the term is taken:

$$\begin{aligned} \frac{d}{dz}T_2 &= \frac{d}{dz}\left[-\frac{\hbar e^2}{A m_e^2}\sum_{KK'}\sum_{\vec{k}_\parallel}\frac{n_F\left(K,\vec{k}_\parallel\right)}{\Omega_{KK'}} g_{KK'}(z) g_{KK'}^*(z')\right] \\ &= -\frac{\hbar e^2}{A m_e^2}\sum_{KK'}\sum_{\vec{k}_\parallel}\frac{n_F\left(K,\vec{k}_\parallel\right)}{\Omega_{KK'}}\left[\frac{d}{dz}g_{KK'}(z)\right] g_{KK'}^*(z'). \end{aligned} \tag{F5}$$

Using Eq. (WK84), one gets:

$$\frac{d}{dz}T_2 = -\frac{2e^2}{A m_e}\sum_{KK'}\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\eta_{KK'}(z) g_{KK'}^*(z'). \tag{F6}$$

One can easily see, that Eqs. (WK83) and (WK85) should now be used to move towards the form that the second term takes in Eq. (74) of that work. However, the equations are investigated here in detail.

2. The case of real eigenfunctions

It is clear that, due to the completeness of φ_K basis, the Eq. (WK83) should, if fact be as follows:

$$\sum_K \varphi_K(z)\varphi_K^*(z') = \sum_K \varphi_K^*(z)\varphi_K(z') = \delta(z-z') \tag{F7}$$

and only for the real eigenfunctions, the original formulation:

$$\sum_K \varphi_K(z)\varphi_K(z') = \delta(z-z') \tag{F8}$$

can be retrieved. Moreover, the Eq. (WK85) contains the $\eta_{KK'}(z) g_{KK'}(z')$ term instead of $\eta_{KK'}(z) g_{KK'}^*(z')$ that is present in our Eq. (F6). This suggest that, contrary to what they say in their work, authors assumed φ_K's to be real when they went from Eq. (WK74) to Eqs. (WK77)-(WK81). Furthermore, as the $n_F\left(K,\vec{k}_\parallel\right)$ term is indexed by K and not K' in both Eq. (WK74) and Eq. (F6), then the sum should be over K' instead. We will start with:

$$\begin{aligned} &\sum_{K'}\eta_{KK'}(z) g_{KK'}(z') = \\ &\sum_{K'}\left[\varphi_K(z)\varphi_{K'}(z)\right]\left(\varphi_K(z')\frac{d\varphi_{K'}(z')}{dz'} - \varphi_{K'}(z')\frac{d\varphi_K(z')}{dz'}\right) \\ &\qquad = \varphi_K(z)\varphi_K(z')\sum_{K'}\varphi_{K'}(z)\frac{d\varphi_{K'}(z')}{dz'} \\ &\qquad\quad - \varphi_K(z)\frac{d\varphi_K(z')}{dz'}\sum_{K'}\varphi_{K'}(z)\varphi_{K'}(z'). \end{aligned} \tag{F9}$$

In the case of the first term, the derivative works only on the z'-dependent part of what is under the sum over K':

$$\sum_{K'}\varphi_{K'}(z)\frac{d\varphi_{K'}(z')}{dz'} = \frac{d}{dz'}\sum_{K'}\varphi_{K'}(z)\varphi_{K'}(z') = \frac{d}{dz'}\delta(z-z') \tag{F10}$$

and

$$\begin{aligned} \sum_{K'}\eta_{KK'}(z) g_{KK'}(z') &= \varphi_K(z)\varphi_K(z')\frac{d}{dz'}\delta(z-z') \\ &\quad - \varphi_K(z)\frac{d\varphi_K(z')}{dz'}\delta(z-z'). \end{aligned} \tag{F11}$$

Using the distributional derivative property of Dirac delta that $\frac{d}{dz'}\delta(z-z') = -\frac{d}{dz}\delta(z-z')$ one can further write:

$$\begin{aligned} \sum_{K'}\eta_{KK'}(z) g_{KK'}(z') &= -\varphi_K(z)\varphi_K(z')\frac{d}{dz}\delta(z-z') \\ &\quad - \varphi_K(z)\frac{d\varphi_K(z')}{dz'}\delta(z-z'). \end{aligned} \tag{F12}$$

Now, using the projective ability of the delta function $f(x)\delta(x-y) = f(y)\delta(x-y)$ twice on the second term and finally obtain:

$$\begin{aligned} &\sum_{K'}\eta_{KK'}(z) g_{KK'}(z') = \\ &\quad -\varphi_K(z)\varphi_K(z')\frac{d}{dz}\delta(z-z') - \varphi_K(z')\frac{d\varphi_K(z)}{dz}\delta(z-z') \\ &\quad = -\varphi_K(z')\left[\varphi_K(z)\frac{d}{dz}\delta(z-z') + \frac{d\varphi_K(z)}{dz}\delta(z-z')\right] \\ &\quad = -\varphi_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right]. \end{aligned} \tag{F13}$$

It can be easily shown that the minus sign difference between this result with respect to Eq. (WK85) is due to summation over K' instead of K.

Combining Eq. (F6) with Eq. (F13) yields:

$$\begin{aligned}\frac{d}{dz}T_2 &= -\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(\sum_{K'}\eta_{KK'}(z)g_{KK'}(z')\right)\\ &= -\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(-\varphi_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right]\right)\\ &= \frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(\varphi_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right]\right)\\ &= \frac{d}{dz}\left[\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\varphi_K(z')\varphi_K(z)\delta(z-z')\right] \end{aligned}\tag{F14}$$

and, after using the projective property again, one arrives at:

$$\begin{aligned}\frac{d}{dz}T_2 &= -\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(\sum_{K'}\eta_{KK'}(z)g_{KK'}(z')\right)\\ &= -\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(-\varphi_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right]\right)\\ &= \frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(\varphi_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right]\right)\\ &= \frac{d}{dz}\left[\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)|\varphi_K(z)|^2\delta(z-z')\right]. \end{aligned}\tag{F15}$$

Having omitted the integration constant, which could be argued to be zero (see the corresponding discussion in Sec. 3 of Ref. 13), one obtains:

$$T_2 = \delta(z-z')\frac{2e^2}{Am_e}\sum_K\left(|\varphi_K(z)|^2\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\right), \tag{F16}$$

which is equal to *minus* the corresponding term in Eq. (WK74) for $\alpha = \beta = z$.

In fact, the derivation is done the other way in WK, so the correct version of Eq. (F2) should be:

$$P_{zz}\left(\vec{q_\parallel}, z, z', \omega\right) = T_1 - T_2, \tag{F17}$$

and Eq. (F1) should take the form of:

$$P_{zz}^{KK'}\left(\vec{q_\parallel},\omega\right) = -\sum_{\vec{k}_\parallel}\left(\frac{\hbar^2 P_{KK'}\left(\vec{q_\parallel},\vec{k}_\parallel,\omega\right)}{4m_e^2} - \frac{\hbar e^2 n_F\left(K,\vec{k}_\parallel\right)}{Am_e^2\Omega_{KK'}}\right). \tag{F18}$$

This is the minus sign of the "omega terms" in the parabolic model of Eqs. (3) and (15) in Ref. 12 of the $\vec{q}_\parallel \to 0$ limit of the $\chi(\vec{q}_\parallel,\omega)$.

3. The case of complex eigenfunctions

If the φ functions are not real, then the relations of Eq. (F7) can only be used, not of Eq. (F8). In addition, the $\eta_{KK'}(z)g_{KK'}(z')$ term will be replaced with the $\eta_{KK'}(z)g^*_{KK'}(z')$ one, see Eq. (F6). With this in mind, starting with:

$$\begin{aligned}&\sum_{K'}\eta_{KK'}(z)g^*_{KK'}(z') = \\ &\sum_{K'}\left[\varphi_K(z)\varphi^*_{K'}(z)\right]\left(\varphi^*_K(z')\frac{d\varphi_{K'}(z')}{dz'} - \varphi_{K'}(z')\frac{d\varphi^*_K(z')}{dz'}\right)\\ &\qquad = \varphi_K(z)\varphi^*_K(z')\sum_{K'}\varphi^*_{K'}(z)\frac{d\varphi_{K'}(z')}{dz'}\\ &\qquad\quad - \varphi_K(z)\frac{d\varphi^*_K(z')}{dz'}\sum_{K'}\varphi^*_{K'}(z)\varphi_{K'}(z') \end{aligned}\tag{F19}$$

and similarly to the case of real wavefunctions, one gets:

$$\sum_{K'}\varphi^*_{K'}(z)\frac{d\varphi_{K'}(z')}{dz'} = \frac{d}{dz'}\sum_{K'}\varphi^*_{K'}(z)\varphi_{K'}(z') = \frac{d}{dz'}\delta(z-z') \tag{F20}$$

and

$$\begin{aligned}\sum_{K'}\eta_{KK'}(z)g^*_{KK'}(z') &= \varphi_K(z)\varphi^*_K(z')\frac{d}{dz'}\delta(z-z')\\ &\quad - \varphi_K(z)\frac{d\varphi^*_K(z')}{dz'}\delta(z-z'). \end{aligned}\tag{F21}$$

With $\frac{d}{dz'}\delta(z-z') = -\frac{d}{dz}\delta(z-z')$ one obtains:

$$\begin{aligned}\sum_{K'}\eta_{KK'}(z)g^*_{KK'}(z') &= -\varphi_K(z)\varphi^*_K(z')\frac{d}{dz}\delta(z-z')\\ &\quad - \varphi_K(z)\frac{d\varphi^*_K(z')}{dz'}\delta(z-z'). \end{aligned}\tag{F22}$$

Until this moment the derivation was paralleling the previous (real) case. However, using the projective ability correspondingly, yields:

$$\begin{aligned}\sum_{K'}\eta_{KK'}(z)g^*_{KK'}(z') &= -\varphi_K(z)\varphi^*_K(z')\frac{d}{dz}\delta(z-z')\\ &\quad - \varphi_K(z')\frac{d\varphi^*_K(z)}{dz}\delta(z-z') \end{aligned}\tag{F23}$$

and the conjugations of the first and the second term on RHS do not match. It is easy to notice, that the root of the problem is that the projective ability does not swap the conjugation of the accompanying function, which renders the next would-be step "illegal". It needs to be underlined, that this means there is no possibility of writing the Eq. (WK74) in a form of Eq. (WK77), on which hinges the whole crucial idea of separation of $\chi(\omega)$ from $\xi(z/z')$ in nonlocal susceptibility of AA.

To proceed further somehow, we will for now pretend that the conjugations in the first and the second term agree. One could propose two versions, going with either conjugation ordering of the first or of the second term. They are, respectively:

$$\begin{aligned}&(I)\sum_{K'}\eta_{KK'}(z)g^*_{KK'}(z') = -\varphi^*_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right],\\ &(II)\sum_{K'}\eta_{KK'}(z)g^*_{KK'}(z') = -\varphi_K(z')\frac{d}{dz}\left[\varphi^*_K(z)\delta(z-z')\right]. \end{aligned}\tag{F24}$$

Going with (I) yields:

$$\begin{aligned}\frac{d}{dz}T_2 &= -\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(\sum_{K'}\eta_{KK'}(z)g^*_{KK'}(z')\right)\\ &= -\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(-\varphi^*_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right]\right)\\ &= \frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\left(\varphi^*_K(z')\frac{d}{dz}\left[\varphi_K(z)\delta(z-z')\right]\right)\\ &= \frac{d}{dz}\left[\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\varphi^*_K(z')\varphi_K(z)\delta(z-z')\right] \quad \text{(F25)}\end{aligned}$$

and, after using the projective property again, one arrives at:

$$\frac{d}{dz}T_2 = \frac{d}{dz}\left[\frac{2e^2}{Am_e}\sum_K\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)|\varphi_K(z)|^2\delta(z-z')\right], \quad \text{(F26)}$$

so:

$$T_2 = \delta(z-z')\frac{2e^2}{Am_e}\sum_K\left(|\varphi_K(z)|^2\sum_{\vec{k}_\parallel} n_F\left(K,\vec{k}_\parallel\right)\right), \quad \text{(F27)}$$

which is the same result, as the one for real wavefunctions. It can be easily shown that, going with the second version of the approximation also leads the same result.

4. The relation between the energy, mass and current elements

The relation Eq. (WK84)

$$\frac{d}{dz}g_{KK'}(z) = \frac{2m_e\Omega_{KK'}}{\hbar}\eta_{KK'}(z), \quad \text{(F28)}$$

can be easily shown to follow from the explicit form of the Schrödinger equation of Eq. (WK40) and the definitions of $\eta_{KK'}$ and $g_{KK'}$. The same relation was adopted in *AA*.

The origin of this $m_e \leftrightarrow \Omega_{KK'}$ and $g_{KK'}(z) \leftrightarrow \eta_{KK'}(z)$ relation is the explicit one-band Hamiltonian, from which the Schrödinger equation for 8-band $\vec{k}\cdot\vec{p}$ would be very different. Thus a lot of additional mixing terms would appear, which most probably prohibit the crucial $\chi(\omega)$ vs $\xi(z/z')$ separation anyway. There is no direct way to address this problem and even if one retains the relation of Eq. (F28) as-is, it is not obvious is it better to use the $E_{\Gamma,m'} - E_{\Gamma,n'}$ (as in *AA*) or $E^{\hat{n}}_{\vec{k}'_\parallel,m'} - E^{\hat{n}}_{\vec{k}'_\parallel,n'}$ (including the dependence of energy on $\vec{k}_\parallel$) in place of $\Omega_{KK'}$. For this reason, we decided to use two versions of the relation as approximations while checking the difference between them as a rough estimate of error, with $\Omega_{KK'} \to \mathscr{E}$ in Eq. (26).

Please also note that the relevant m_e symbols present in WK, in our work are being embedded as a part of J_z definition for $\vec{k}\cdot\vec{p}$.

Appendix G: The low-temperature limit of the 8-band $\vec{k}\cdot\vec{p}$ model usability

It is very probable that the reader would ask why only limited temperature ranges were shown in Fig. 2 and taken into account in this analysis, as compared to the measured values in Table III. In the case of $T < 78$ K temperatures for all the three samples, they are out of the scope of applicability of our 8-band $\vec{k}\cdot\vec{p}$ model. This is a consequence of the $\vec{k}_\parallel$ mesh we have used in both the nextnano++ and the ISB formation model, see Secs. II B and II E, respectively.

Please note that there are four parameters describing the mesh. First one is the number of $\vec{k}_\parallel$ directions taken into account. Thankfully, due to the symmetry of the crystal lattice and for a system with a symmetry point in the growth z direction, all the $\vec{k}_\parallel$ variability can be described by a single octant, let us say between the [100] and [110] directions. However, the issue of how many orientations $N^O_{\vec{k}_\parallel}$ of $\vec{k}_\parallel$ take into account in that octant remains. The second parameter is the maximal magnitude of the in-plane wavevector $k^{\max}_\parallel$ taken into account. Here, one wants to be sure that the part of the Brillouin zone that was cut off is sufficiently far away from the Γ point that it contains no meaningful contribution. The third parameter is the mesh spacing $\Delta_{k_\parallel}$, defining the *granularity* of the mesh. The fourth parameter is the total mesh size, which is given by the other ones as $N^O_{\vec{k}_\parallel}\frac{k^{\max}_\parallel}{\Delta_{k_\parallel}}$. Please note that the computational complexity of the model is approximately proportional to the second power of the mesh size.

With the computational complexity in mind, we already have taken the path of least resistance with $N^O_{\vec{k}_\parallel} = 2$. After verifying that the dispersions and wavefunctions of the [100] and [110] are relatively the most different in the octant, while the intermediate orientations have intermediate dispersions and wavefunctions, we only took into account these two cases. Additionally, we limited the model to 20 conduction band subbands (10 orbitals × 2 spin orientations) and took only the transitions between orbitals adjacent in energy into account.[m] With all this, however, the mesh as defined and used by us is already at the limit of practical usability.

Now, what is happening in the lower temperatures will be discussed. As $T \to 0$ the Fermi-Dirac distribution $f_{n,\vec{k}_\parallel}$ becomes more rigid. Consequently, the occupation differences of adjacent orbital subbands $\Delta f_{\alpha,\vec{k}_\parallel}$ change significantly over progressively smaller $\vec{k}_\parallel$ ranges. At some

---

[m] If the typical energy separation of adjacent orbitals is $\omega_{n,n+1}$, then the typical energy separation of next-nearest neighbors will be about $\omega_{n,n+2} \approx 2\omega_{n,n+1}$, which means that they would only make a small, if not totally negligible contribution, to the ISB plasmon formation.

point, the granularity of the mesh becomes too big to describe the dynamics of the ISB plasmon formation and the model falls apart. We made an estimate based of the calculated 8-band $\vec{k} \cdot \vec{p}$ dispersion relation for the [100] direction of the G0490 sample, that we model cannot be expected to work properly for $T \ll 78$ K.

In principle, the lower temperature limit $T_{\min}$ of the model could be *somewhat* pushed, but the difficulty with this push raises strongly non-linearly with lower $T_{\min}$, putting the liquid helium range out of the question. This non-linearity comes from two sources: (I) the way how Fermi-Dirac function reacts to reducing temperature, (II) the upward drift of the Fermi level with reducing temperature. As to the second point, in the estimate mentioned above, the Fermi level crosses the first subband at the Γ point at about $T = 53$ K. Because now some inter-subband transitions cross the Fermi level, when the derivative magnitude of the Fermi-Dirac function is the biggest (in fact approaching infinity for $T \to 0$), the model would need a minuscule $\Delta_{k_\parallel}$ to work.

Please note that, while the measurement for G0643 was performed with more or less equidistant temperatures, the ones for G0490 and G0489 have one point in liquid-helium temperature range ($T = 10$ K) and the next one is already the liquid nitrogen temperature. We consider the second type of measurement more typical, because if the measurement is done with the use of liquid helium anyway, there seems to be no reason not to do it in very low temperature. We had just accepted that the scope of applicability of our model is the liquid-nitrogen temperatures and above as opposed to the liquid-helium temperature range.

---

[*] Electronic address: `wjpasek@unicamp.br`

[†] Electronic address: `zbig.wasilewski@uwaterloo.ca`

[1] C. Deimert, P. Goulain, J. M. Manceau, W. Pasek, T. Yoon, A. Bousseksou, N. Y. Kim, R. Colombelli, and Z. R. Wasilewski, Physical Review Letters **125**, 097403 (2020), URL `https://link.aps.org/doi/10.1103/PhysRevLett.125.097403`.

[2] A. Delteil, A. Vasanelli, Y. Todorov, C. F. Palma, M. R. St-Jean, G. Beaudoin, I. Sagnes, and C. Sirtori, Physical review letters **109**, 246808 (2012).

[3] M. Załużny and C. Nalewajko, Phys. Rev. B **59**, 13043 (1999), URL `https://link.aps.org/doi/10.1103/PhysRevB.59.13043`.

[4] T. Ando, A. B. Fowler, and F. Stern, Rev. Mod. Phys. **54**, 437 (1982), URL `https://link.aps.org/doi/10.1103/RevModPhys.54.437`.

[5] Y. Todorov and C. Sirtori, Phys. Rev. B **85**, 045304 (2012), URL `https://link.aps.org/doi/10.1103/PhysRevB.85.045304`.

[6] G. Pegolotti, A. Vasanelli, Y. Todorov, and C. Sirtori, Phys. Rev. B **90**, 035305 (2014), URL `https://link.aps.org/doi/10.1103/PhysRevB.90.035305`.

[7] Y. Todorov, Phys. Rev. B **89**, 075115 (2014), URL `https://link.aps.org/doi/10.1103/PhysRevB.89.075115`.

[8] Y. Todorov, Phys. Rev. B **91**, 125409 (2015), URL `https://link.aps.org/doi/10.1103/PhysRevB.91.125409`.

[9] Y. Todorov, L. Tosetto, A. Delteil, A. Vasanelli, C. Sirtori, A. M. Andrews, and G. Strasser, Phys. Rev. B **86**, 125314 (2012), URL `https://link.aps.org/doi/10.1103/PhysRevB.86.125314`.

[10] B. Askenazi, A. Vasanelli, A. Delteil, Y. Todorov, L. Andreani, G. Beaudoin, I. Sagnes, and C. Sirtori, New Journal of Physics **16**, 043029 (2014), URL `http://stacks.iop.org/1367-2630/16/i=4/a=043029`.

[11] R. J. Warburton, C. Gauer, A. Wixforth, J. P. Kotthaus, B. Brar, and H. Kroemer, Physical Review B **53**, 7903 (1996), URL `https://link.aps.org/doi/10.1103/PhysRevB.53.7903`.

[12] F. Alpeggiani and L. C. Andreani, Physical Review B **90**, 115311 (2014), URL `https://link.aps.org/doi/10.1103/PhysRevB.90.115311`.

[13] A. Eguiluz and A. Maradudin, Annals of Physics **113**, 29 (1978), ISSN 0003-4916, URL `http://www.sciencedirect.com/science/article/pii/000349167890249X`.

[14] L. Wendler and R. Pechstedt, Physica Status Solidi (b) **138**, 197 (1986), ISSN 1521-3951, URL `https://doi.org/10.1002/pssb.2221380121`.

[15] L. Wendler and E. Kändler, Physica Status Solidi (b) **177**, 9 (1993).

[16] M. Załużny, Physica B+C **128**, 171 (1985), ISSN 0378-4363, URL `https://www.sciencedirect.com/science/article/pii/0378436385901007`.

[17] F. Tassone, F. Bassani, and L. C. Andreani, Il Nuovo Cimento D **12**, 1673 (1990), ISSN 0392-6737, URL `https://doi.org/10.1007/BF02451267`.

[18] W. Kohn, Phys. Rev. **123**, 1242 (1961), URL `https://link.aps.org/doi/10.1103/PhysRev.123.1242`.

[19] L. Brey, N. F. Johnson, and B. I. Halperin, Phys. Rev. B **40**, 10647 (1989), URL `https://link.aps.org/doi/10.1103/PhysRevB.40.10647`.

[20] J. F. Dobson, Phys. Rev. Lett. **73**, 2244 (1994), URL `https://link.aps.org/doi/10.1103/PhysRevLett.73.2244`.

[21] I. K. Marmorkos and S. Das Sarma, Phys. Rev. B **48**, 1544 (1993), URL `https://link.aps.org/doi/10.1103/PhysRevB.48.1544`.

[22] C. Deimert and Z. R. Wasilewski, Journal of Crystal Growth **514**, 103 (2019), ISSN 0022-0248, URL `http://www.sciencedirect.com/science/article/pii/S0022024819301277`.

[23] C. Deimert and Z. R. Wasilewski, Journal of Vacuum Science & Technology A **39**, 043407 (2021), URL `https://avs.scitation.org/doi/abs/10.1116/6.0001019`.

[24] M. J. Manfra, Annu. Rev. Condens. Matter Phys. **5**, 347 (2014), ISSN 1947-5454, URL `https://doi.org/10.1146/annurev-conmatphys-031113-133905`.

[25] S. Birner, T. Zibold, T. Andlauer, T. Kubis, M. Sabathil, A. Trellakis, and P. Vogl, IEEE Transactions on Electron Devices **54**, 2137 (2007).

[26] S. Birner, *Modeling of semiconductor nanostructures and semiconductor-electrolyte interfaces* (Verein zur Förderung des Walter-Schottky-Inst. der Techn. Univ. München, 2011).

[27] Z. R. Wasilewski, M. M. Dion, D. J. Lockwood, P. Poole, R. W. Streater, and A. J. SpringThorpe, Journal of Applied Physics **81**, 1683 (1997), https://doi.org/10.1063/1.364012, URL `https://doi.org/10.1063/1.364012`.

[28] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, Journal of Applied Physics **89**, 5815 (2001), http://dx.doi.org/10.1063/1.1368156, URL `http://dx.doi.org/10.1063/1.1368156`.

[29] T. Unuma, M. Yoshita, T. Noda, H. Sakaki, and H. Akiyama, Journal of Applied Physics **93**, 1586 (2003), https://doi.org/10.1063/1.1535733, URL `https://doi.org/10.1063/1.1535733`.

[30] C. A. Ullrich and G. Vignale, Phys. Rev. Lett. **87**, 037402 (2001), URL `https://link.aps.org/doi/10.1103/PhysRevLett.87.037402`.

[31] A. P. Hibbins, J. R. Sambles, C. R. Lawrence, and J. R. Brown, Physical review letters **92**, 143904 (2004).

[32] Y. Todorov, L. Tosetto, J. Teissier, A. M. Andrews, P. Klang, R. Colombelli, I. Sagnes, G. Strasser, and C. Sirtori, Optics Express **18**, 13886 (2010), URL `http://www.opticsexpress.org/abstract.cfm?URI=oe-18-13-13886`.

[33] Y. Todorov, A. M. Andrews, R. Colombelli, S. De Liberato, C. Ciuti, P. Klang, G. Strasser, and C. Sirtori, Physical Review Letters **105**, 196402 (2010).

[34] P. Jouy, A. Vasanelli, Y. Todorov, A. Delteil, G. Biasiol, L. Sorba, and C. Sirtori, Applied Physics Letters **98**, 231114 (2011), ISSN 0003-6951.

[35] J. M. Manceau, S. Zanotto, I. Sagnes, G. Beaudoin, and R. Colombelli, Applied Physics Letters **103**, 091110 (2013).

[36] J.-M. Manceau, S. Zanotto, T. Ongarello, L. Sorba, A. Tredicucci, G. Biasiol, and R. Colombelli, Applied Physics Letters **105**, 081105 (2014), http://dx.doi.org/10.1063/1.4893730.

[37] S. Zanotto, F. P. Mezzapesa, F. Bianco, G. Biasiol, L. Baldacci, M. S. Vitiello, L. Sorba, R. Colombelli, and A. Tredicucci, Nat Phys **10**, 830 (2014), ISSN 1745-2473.

[38] A. Vasanelli, Y. Todorov, and C. Sirtori, Comptes Rendus Physique **17**, 861 (2016), ISSN 1631-0705, URL `http://www.sciencedirect.com/science/article/pii/S1631070516300299`.

[39] B. Askenazi, A. Vasanelli, Y. Todorov, E. Sakat, J.-J. Greffet, G. Beaudoin, I. Sagnes, and C. Sirtori, ACS Photonics **4**, 2550 (2017), http://dx.doi.org/10.1021/acsphotonics.7b00838, URL `http://dx.doi.org/10.1021/acsphotonics.7b00838`.

[40] C. Y.-P. Chao and S. L. Chuang, Phys. Rev. B **43**, 7027 (1991), URL `https://link.aps.org/doi/10.1103/PhysRevB.43.7027`.

[41] T. B. Bahder, Phys. Rev. B **41**, 11992 (1990), URL `https://link.aps.org/doi/10.1103/PhysRevB.41.11992`.

[42] A. C. Rodrigues Bittencourt, A. M. Cohen, and G. E. Marques, Phys. Rev. B **57**, 4525 (1998), URL `https://link.aps.org/doi/10.1103/PhysRevB.57.4525`.

[43] X. C. Soler, Ph.D. thesis, California Institute of Technology (2003), URL `http://resolver.caltech.edu/CaltechETD:etd-05232003-104331`.

[44] C. R. Pidgeon and R. N. Brown, Phys. Rev. **146**, 575 (1966), URL `https://link.aps.org/doi/10.1103/PhysRev.146.575`.

[45] Y. Varshni, Physica **34**, 149 (1967), ISSN 0031-8914, URL `http://www.sciencedirect.com/science/article/pii/0031891467900626`.

[46] M. Helm, *The basic physics of intersubband transitions* (Academic Press, 1999), book section 1, ISBN 0080864600.

[47] J. Nespolo and I. Carusotto, Phys. Rev. B **100**, 035305 (2019), URL `https://link.aps.org/doi/10.1103/PhysRevB.100.035305`.

[48] R. Cominotti, H. A. M. Leymann, J. Nespolo, J. M. Manceau, M. Jeannin, R. Colombelli, and I. Carusotto, *Theory of coherent optical nonlinearities of intersubband transitions in semiconductor quantum wells* (2021), URL `https://arxiv.org/abs/2109.00285`.

[49] R. Harrington, *Time-harmonic electromagnetic fields* (Wiley-IEEE press, New York, NY, 1961).

[50] S. Zanotto, G. Biasiol, R. Degl'Innocenti, L. Sorba, and A. Tredicucci, Applied Physics Letters **97**, 231123 (2010).

[51] E. R. Weber, R. K. Willardson, H. Liu, and F. Capasso, *Intersubband transitions in quantum wells: physics and device applications* (Academic Press, 1999), ISBN 0080864600.