This is the accepted manuscript made available via CHORUS. The article has been published as:

Unified theory of optical absorption and luminescence including both direct and phonon-assisted processes

Sabyasachi Tiwari, Emmanouil Kioupakis, José Menendez, and Feliciano Giustino

Phys. Rev. B **109**, 195127 — Published 8 May 2024

DOI: 10.1103/PhysRevB.109.195127

Unified theory of optical absorption and luminescence including both direct and phonon-assisted processes

Sabyasachi Tiwari,[1,2] Emmanouil Kioupakis,[3] José Menendez,[4] and Feliciano Giustino[1,2]

[1] *Oden Institute for Computational Engineering and Sciences, The University of Texas at Austin, Austin, Texas 78712, USA*
[2] *Department of Physics, The University of Texas at Austin, Austin, Texas 78712, USA*
[3] *Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan, 48109, USA*
[4] *Department of Physics, Arizona State University, Tempe, Arizona 85287-1504, USA*

(Dated: April 11, 2024)

Most semiconductors and insulators exhibit indirect band gaps, but no theory is currently available to calculate light absorption and emission spectra of these systems over a wide spectral range with predictive accuracy. The standard textbook theory of indirect absorption becomes ill-defined and yields infinite absorption strength when a photon can promote both direct and phonon-assisted transitions. As a result, state-of-the-art *ab initio* methods for calculating optical spectra of solids are unable to describe direct and phonon-assisted transitions on the same footing. Here, we develop a rigorous first-principles approach that overcomes this limitation by including electron-phonon correlations via many-body quasi-degenerate perturbation theory. Our present formalism enables accurate calculations of the optical spectra of materials with direct, indirect, and quasi-direct band gaps, and reduces to the standard theories of direct-only absorption and indirect-only absorption in the appropriate limits. We demonstrate this methodology by investigating the optical absorption spectra of silicon, germanium, gallium arsenide, and diamond. In all cases we obtain spectra in excellent agreement with experiments. As a more ambitious test, we investigate the temperature-dependent photoluminescence of germanium, and we obtain for the first time quantitative agreement with experiments. This work supersedes the standard textbook theory of phonon-assisted optical processes.

I. INTRODUCTION

Light absorption and emission in solids are the prime manifestations of light-matter interactions, and form the basis for numerous equilibrium and non-equilibrium optical probes of excited states in condensed matter [1–3]. The ability to perform quantitatively accurate *ab initio* calculations of absorption and emission spectra is key to interpret experiments on excitations in bulk as well as low-dimensional materials and to shed light on the underlying many-body physics [4–9].

The central quantities in optical spectroscopy are the absorption coefficient $\alpha(\omega)$, whose inverse measures the characteristic attenuation length of an electromagnetic wave of frequency ω in a solid [10]; and the recombination rate $R(\omega)$, which quantifies the intensity of the electromagnetic field generated by an excited solid. At a microscopic level, light is absorbed when the electromagnetic wave creates an electron-hole pair or an exciton, and is emitted upon their recombination. Since the photon momentum of infrared, visible, and ultraviolet radiation is much smaller than the typical span of the Brillouin zone, the electron crystal momentum is conserved during the transition. This selection rule leads to two fundamental types of optical processes, namely direct and indirect transitions [11]. In a direct transition, the electron and the hole have the same crystal momentum; in an indirect transition, one phonon is also absorbed or emitted, so that the crystal momenta of the electron and the hole differ by the phonon momentum. These processes are of first and second order in perturbation theory, respectively, therefore the oscillator strength of direct processes is typically much higher than for indirect processes, except for symmetry-forbidden direct transitions [10]. Owing to the strong oscillator strength of direct transitions, direct-gap materials are sought-after for applications in solar technology, light-emitting devices, and lasers. However, the *majority* of existing non-metallic compounds exhibit an indirect band gap. In fact, 70% of the 81,865 inorganic materials with a band gap reported in the Materials Project database are classified as indirect [12].

Ab initio calculations of optical spectra including phonon-assisted processes have become possible during the past decade [13, 14]. In Ref. [13], the textbook perturbation theory approach to indirect absorption by Cheeseman [15] and by Hall, Bardeen, and Blatt [16] (CHBB henceforth) was used in combination with Wannier-Fourier interpolation of the electron-phonon matrix elements [17] to obtain the optical absorption spectrum of silicon (Si). In Ref. [14], the authors investigated the indirect absorption in Si by employing a non-perturbative approach based on a set of special atomic displacements in a large supercell [18, 19]. Additional calculations based on these methods have been reported more recently [20–22].

These approaches carry both advantages and limitations: (i) The perturbative approach yields good agreement with experiments for photon energies in between the indirect band gap and the direct band gap, and captures the phonon fine structure near the absorption edge. However, as we discuss in detail in Sec. II B, this approach yields a divergent absorption coefficient for photon energies above the direct gap. This limitation is especially problematic for materials where direct and indirect

gaps are close in energy, the so-called quasi-direct semiconductors [23]. This situation is fairly common, for example many transition metal dichalcogenides fall in this category [24, 25]. (ii) Unlike the CHBB theory, the special displacement method [19] yields good agreement with experiments throughout the entire spectral range. However, this method does not capture the phonon fine structure at the onset of absorption or emission, because it is based on the adiabatic approximation. In an attempt to overcome these challenges, recent work proposed to partition optical calculations into an indirect range (between the indirect gap and the the direct gap) and a direct range (above the direct gap) [26]. This approach mitigates the difficulties of the CHBB theory, but it is ill-defined near the direct gap, and misses the contributions to the spectrum from phonon-assisted processes around and above the direct gap. Moreover, such an approach cannot resolve the fine structure of phonon-assisted-processes near the direct gap, which is necessary to obtain the luminescence spectra of materials whose direct and indirect gaps are close in energy such as germanium [27–29]. Overall, a first-principles theory for calculating optical absorption and emission spectra including phonon-assisted processes across a wide spectral range, which can capture the phonon fine structure, is currently missing.

Here, we develop a unified *ab initio* theory of light absorption and emission in solids which overcomes the above difficulties. We show how the textbook CHBB approach becomes ill-defined in the presence of resonant electron-hole and electron-hole-phonon excitations, and we cure this divergence by means of many-body quasi-degenerate perturbation theory. We show how the present formulation provides a robust methodology for calculating optical spectra including electron-phonon interactions which are free of singularities, can be applied across the entire spectral range, and captures the phonon fine structure at the onset of absorption or emission. We also show how this more general method reduces to the CHBB theory in the appropriate limit. To demonstrate this methodology, we compute the optical absorption spectra of Si, germanium (Ge), gallium arsenide (GaAs), and diamond. In all cases we obtain excellent agreement with experiments. As a more ambitious test of the theory, we also compute the photoluminescence spectrum of Ge, and obtain quantitative agreement with high-resolution experiments. The methodology presented in this manuscript is implemented and available in the `EPW v5.8` release [30].

The manuscript is organized as follows. In Sec. II we outline our theoretical approach. Specifically, in Sec. II B we briefly review the textbook CHBB theory of indirect absorption. In Sec. II C we identify the problematic aspect of the CHBB theory by reformulating this approach via many-body perturbation theory. In Sec. II D we introduce our quasi-degenerate perturbation theory strategy to cure the divergences in the CHBB theory. Section III summarizes our computational setup. In this section we also provide extensive convergence tests. In Sec. IV we apply the present method to the optical absorption spectrum and near-edge fine structure of Si, and to the optical absorption and luminescence spectra of Ge. For compactness, we discuss the absorption spectra of diamond and GaAs in Appendices A and B, and the possible alternative approaches to our presented method in appendix C. In Sec. V we summarize our findings and discuss possible future generalizations of the theory.

II. FIRST-PRINCIPLES THEORY OF DIRECT AND PHONON-ASSISTED OPTICAL TRANSITIONS

A. Hamiltonian in Fock space

To obtain optical spectra, we aim to evaluate the imaginary part of the dielectric function of a crystal, $\varepsilon_2(\omega)$, where ω is the photon frequency. Once obtained the imaginary part of the dielectric function $\varepsilon_2(\omega)$, we can compute the absorption coefficient as [31]:

$$\alpha(\omega) = \frac{\omega\varepsilon_2(\omega)}{cn(\omega)}, \tag{1}$$

where c is the speed of light and $n(\omega)$ is the real part of the refractive index. Similarly, we can compute the total rate of radiative recombination per unit volume via the van Roosbroek-Shockley relation [10, 32]:

$$R(\omega) = \frac{2n}{\pi c^3}\frac{\omega^3\varepsilon_2(\omega)}{\exp(\hbar\omega/k_{\mathrm{B}}T)-1}, \tag{2}$$

where k_{B} is the Boltzmann constant and T is the temperature.

The imaginary part of the dielectric function is related to the number of photons absorbed per unit time and per unit volume, N_{p}, by [31]:

$$\frac{dN_{\mathrm{p}}}{dt} = \omega\varepsilon_2(\omega)N_{\mathrm{p}} \ . \tag{3}$$

The rate of photon absorption on the left-hand side is equal to the number of optical transitions per unit time, and can be evaluated via Fermi's golden rule. In the following we consider cubic crystals for notational simplicity.

To write Fermi's golden rule for the transition rates, we need to specify the Hamiltonian of the system. Using many-body operator notation, we have:

$$\hat{H} = \hat{H}_0 + \hat{V}_{\mathrm{er}} + \hat{V}_{\mathrm{ep}} \ , \tag{4}$$

where the terms on the right-hand side denote the electronic Hamiltonian in the absence of phonons and radiation, the electron-radiation interaction term, and the electron-phonon interaction term, respectively. The unperturbed Hamiltonian is given by:

$$\hat{H}_0 = \sum_{n\mathbf{k}} \varepsilon_{n\mathbf{k}}\, \hat{c}^{\dagger}_{n\mathbf{k}}\hat{c}_{n\mathbf{k}} + \sum_{\mathbf{q}\nu} \hbar\omega_{\mathbf{q}\nu}\, \hat{a}^{\dagger}_{\mathbf{q}\nu}\hat{a}_{\mathbf{q}\nu} \ . \tag{5}$$

In this expression, $\varepsilon_{n\mathbf{k}}$ denote electron quasi-particle energies (for example, Kohn-Sham or GW energies), $\omega_{\mathbf{q}\nu}$ are phonon frequencies in the harmonic approximation, and $\hat{c}_{n\mathbf{k}}$, $\hat{a}_{\mathbf{q}\nu}$ are fermion and boson operators, respectively; n denotes the combined band and spin index, and $\mathbf{k}$ is the electron wavevector; similarly, ν and $\mathbf{q}$ are phonon branch index and wavevector, respectively. The summations run over all electron band and phonon modes, and over uniform Brillouin-zone grids with N points. The Hamiltonian $\hat{H}_0$ describes the energy of a Born-von Kármán (BvK) supercell consisting of N primitive unit cells.

For an electromagnetic field consisting of a monochromatic plane wave of frequency ω, linearly polarized along the direction $\mathbf{e}$, the vector potential is $\mathbf{A}(\mathbf{r},t) = A_0\mathbf{e}\cos(\omega t)$, where A_0 denotes the amplitude, $\mathbf{r}$ is the position, and t is the time. The corresponding electron-radiation Hamiltonian in the electric-dipole approximation is:

$$\hat{V}_{\rm er} = eA_0\,\mathbf{e}\cdot\sum_{mn\mathbf{k}}\mathbf{v}_{mn\mathbf{k}}\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}\cos(\omega t)\ , \tag{6}$$

where e and m are the electron charge and mass, respectively, and $\mathbf{v}_{mn\mathbf{k}}$ denote velocity matrix elements among Kohn-Sham states. These matrix elements can be obtained using the approach of Ref. [33].

The electron-phonon interaction term in Eq. (4) is given by [34]:

$$\hat{V}_{\rm ep} = N^{-\frac{1}{2}}\sum_{mn\nu,\mathbf{k},\mathbf{q}} g_{mn\nu}(\mathbf{k},\mathbf{q})\,\hat{c}^\dagger_{m\mathbf{k}+\mathbf{q}}\hat{c}_{n\mathbf{k}}(\hat{a}_{\mathbf{q}\nu}+\hat{a}^\dagger_{-\mathbf{q}\nu})\ , \tag{7}$$

where the electron-phonon matrix elements $g_{mn\nu}(\mathbf{k},\mathbf{q})$ denote the probability amplitude for the transition from the initial electronic state $n\mathbf{k}$ to the final state $m\mathbf{k}+\mathbf{q}$ via the phonon $\mathbf{q}\nu$. Unlike in standard textbook approaches [11], here the electron-phonon phonon interaction Hamiltonian $\hat{V}_{\rm ep}$ is time-independent. This choice relates to the fact that we start from the many-body Hamiltonian of electrons and phonons in the absence of radiation, which does not contain the time variable; see for example Eqs. (1) and (62) of Ref. 34. The present choice is more convenient than the semiclassical theory of Ref. 11 when constructing correlated electron-hole-phonon excitations which are needed in Sec. II D.

B. Textbook CHBB approach to direct and phonon-assisted optical absorption

In the textbook theory of direct and phonon-assisted absorption, one evaluates optical transition rates using Fermi's golden rule, within a single-particle picture. The rate of direct transitions in the BvK supercell is given by:

$$\frac{dN_{\rm p}}{dt} = \frac{2\pi}{\hbar}(eA_0/2)^2\sum_{cv\mathbf{k}}|\mathbf{e}\cdot\mathbf{v}_{cv\mathbf{k}}|^2\delta(\varepsilon_{c\mathbf{k}}-\varepsilon_{v\mathbf{k}}-\hbar\omega)\ , \tag{8}$$

where v and c refer to valence and conduction states, respectively. Furthermore, the number of photons N_p present in the BvK supercell can be obtained from the time-average of the energy density of the electromagnetic field:

$$N_{\rm p} = N\Omega\frac{\epsilon_0\omega A_0^2}{2\hbar}, \tag{9}$$

where Ω is the volume of the primitive unit cell. By combining Eqs. (8) and (9) one obtains the standard expression for the macroscopic dielectric function resulting from direct optical transitions [10, 31]:

$$\epsilon_2(\omega) = \frac{\pi e^2}{\epsilon_0\Omega}\frac{1}{\omega^2}\frac{1}{N}\sum_{cv\mathbf{k}}|\mathbf{e}\cdot\mathbf{v}_{cv\mathbf{k}}|^2\delta(\varepsilon_{c\mathbf{k}}-\varepsilon_{v\mathbf{k}}-\hbar\omega)\ . \tag{10}$$

The band summations in Eqs. (8) and (10) also run over the spin index, so that an extra factor of 2 is needed for spin-unpolarized calculations.

In the case of indirect materials, due to crystal momentum conservation, Eq. (10) misses optical transitions across the indirect gap. The standard textbook approach to incorporate these transitions is provided by the CHBB theory [15, 16]. The CHBB theory describes phonon-assisted processes using *second-order* perturbation theory. In this approach, both radiation and phonons are regarded as time-dependent external fields acting on the electronic subsystem, and electrons are described in the single-particle picture [11]. Within this formalism, and in the case of gapped systems where the band gap is larger than phonon energies, the total rate of photon absorption in the BvK supercell is given by [35]:

$$\begin{aligned}\frac{dN_{\rm p}}{dt} = \frac{2\pi}{\hbar}\frac{e^2A_0^2}{2^2}\frac{1}{N}\sum_{cv\nu,\mathbf{k},\mathbf{q}}^{\eta=\pm1}\left|\mathbf{e}\cdot\left[\mathbf{S}^{(1)}_{cv\nu\eta}(\mathbf{k},\mathbf{q})+\mathbf{S}^{(2)}_{cv\nu\eta}(\mathbf{k},\mathbf{q})\right]\right|^2 \\ \times\,[n_{\mathbf{q}\nu}+(1+\eta)/2]\delta(\varepsilon_{c\mathbf{k}+\mathbf{q}}-\varepsilon_{v\mathbf{k}}+\eta\hbar\omega_{\mathbf{q}\nu}-\hbar\omega)\ ,\end{aligned} \tag{11}$$

where the oscillator strengths are:

$$-\mathbf{S}^{(1)}_{cv\nu\eta}(\mathbf{k},\mathbf{q}) = \sum_n\frac{g_{cn\nu}(\mathbf{k},\mathbf{q})\,\mathbf{v}_{nv\mathbf{k}}}{\varepsilon_{c\mathbf{k}+\mathbf{q}}-\varepsilon_{n\mathbf{k}}+\eta\hbar\omega_{\mathbf{q}\nu}}\ , \tag{12}$$

$$\mathbf{S}^{(2)}_{cv\nu\eta}(\mathbf{k},\mathbf{q}) = \sum_n\frac{\mathbf{v}_{cn\mathbf{k}+\mathbf{q}}\,g_{nv\nu}(\mathbf{k},\mathbf{q})}{\varepsilon_{n\mathbf{k}+\mathbf{q}}-\varepsilon_{v\mathbf{k}}+\eta\hbar\omega_{\mathbf{q}\nu}}\ . \tag{13}$$

In both expressions, the summation runs over all states, occupied and empty. Equation (12) describes a process whereby an electron first absorbs or emits a photon and then absorbs or emits a phonon, while in Eq. (13) the electron first absorbs or emits a phonon, and then absorbs or emits a photon [10]. Also in this case, a factor of 2 needs to be added to Eq. (11) for spin-unpolarized calculations. In practical *ab initio* calculations, to accelerate the convergence of the Brillouin zone sums, one adds a smearing parameter $i\gamma$ in the denominators of Eqs. (12) and (13). After including this smearing, Eq. (11) coincides with the expression used in Ref. 13.

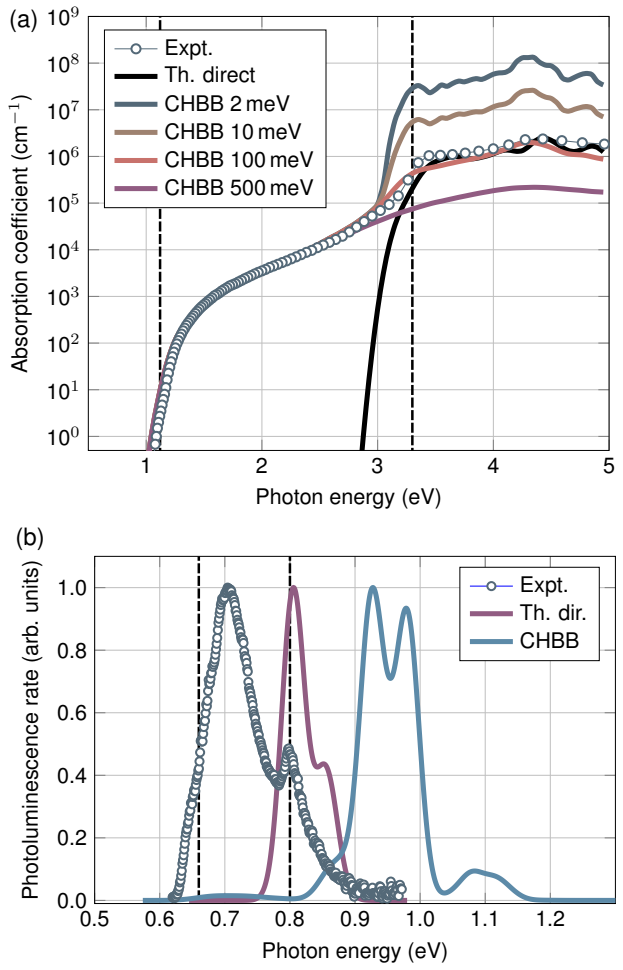

FIG. 1. **Inadequacy of the textbook CHBB theory.** (a) Absorption coefficient of Si at 300 K, calculated using the CHBB theory for various choices of the artificial broadening $i\gamma$ in the denominators of of Eqs. (12) and (13). As the photon energy exceeds the direct band gap, the absorption coefficient becomes very sensitive to γ and diverges in the limit $\gamma \to 0$. The red line shows the absorption coefficient computed by including only direct transitions. Disks are experimental data from Ref. 36. Details of computational parameters are provided in Sec. III. (b) Luminescence spectrum of Ge at 295 K, calculated by considering only direct transitions (purple), or only indirect transitions using the CHBB theory (blue), calculated with an artificial broadening of 2 meV. Disks are experimental data from Ref. 27. In all panels, the indirect and direct band gaps are indicated by vertical lines.

Calculations performed using Eqs. (11)-(13) carry two shortcomings: (i) They are based on single-particle perturbation theory, and as such they do not include Pauli blocking; (ii) The denominators of Eqs. (12) and (13) *diverge* whenever the energy of the intermediate state is $\varepsilon_{n\mathbf{k}} = \varepsilon_{c\mathbf{k}+\mathbf{q}} \pm \hbar\omega_{\mathbf{q}\nu}$ or $\varepsilon_{n\mathbf{k}} = \varepsilon_{v\mathbf{k}+\mathbf{q}} \pm \hbar\omega_{\mathbf{q}\nu}$. These conditions can be met when the photon energy exceeds the energy of the direct band gap. As a consequence of these singularities, the dielectric function becomes infinite at and above the direct gap. To mitigate this singularity, in practical calculations one incorporates an artificial broadening parameter $i\gamma$ in the denominators of Eqs. (12) and (13) [13]. However, the resulting absorption spectrum becomes extremely sensitive to the broadening γ, and predictive power is lost. This effect is illustrated in Fig. 1(a) for the case of Si: by varying the broadening parameter γ between 2 meV and 500 meV, the absorption coefficient for a photon energy of 3 eV changes by three orders of magnitude. The same behavior is observed for all other materials considered in this study.

In Fig. 1(b) we also show how the photoluminescence spectrum of Ge computed by using either Eqs. (10) or (11) is in *qualitative* disagreement with experiments, by exhibiting a single peak in the wrong location instead of a peak-shoulder structure as in the experiments. These two examples for prototypical semiconductors illustrate the limitations of the standard textbook approach.

C. Many-body Fock-space approach

In this section we derive Eqs. (10) and (11) of the CHBB theory, using an alternative approach with respect to standard textbooks [11]. This derivation serves two purposes: (i) To clearly identify the origin of the divergence in the CHBB dielectric function; (ii) To prepare for the derivation of the quasi-degenerate perturbation theory framework in Sec. II D.

In the CHBB theory [11], both radiation and phonons are considered as external, time-dependent fields. In contrast, we here employ first-order perturbation theory for the electron-radiation Hamiltonian of Eq. (6), starting from coupled many-body electron-phonon states associated with the Hamiltonian $\hat{H}_0+\hat{V}_{\rm ep}$ given by Eqs. (5) and (7). In this section we begin with non-degenerate perturbation theory, in Sec. II D we generalize this derivation to include degeneracies.

We use the notation $|i\rangle$ and $|f\rangle$ to denote initial and final many-body states that are solutions of the Hamiltonian $\hat{H}_0+\hat{V}_{\rm ep}$ with energies E_i and E_f, respectively. The electron-radiation term in Eq. (6) induces optical transitions between these states with a rate given by Fermi's golden rule:

$$\Gamma_{i\to f} = \frac{\pi e^2 A_0^2}{2\hbar} \left| \mathbf{e}\cdot\sum_{mn\mathbf{k}} \mathbf{v}_{mn\mathbf{k}} \langle f|\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}|i\rangle \right|^2 \delta(E_f - E_i \pm \hbar\omega). \tag{14}$$

In general, we do not have an explicit expression for the many-body eigenstates $|i\rangle$ and $|f\rangle$. Therefore, we construct approximate states by using the first-order correction to the eigenstates of $\hat{H}_0$ resulting from the phonon perturbation $\hat{V}_{\rm ep}$. We use $|i_0\rangle$ and $|f_0\rangle$ to indicate the unperturbed initial and final states involved in the optical transition, respectively. The eigenstates of $\hat{H}_0$ are sometimes referred to as Born-Oppenheimer product states.

The states $|i_0\rangle$ correspond to uncorrelated electron-phonon states where the electron subsystem is in the ground state, and the phonon subsystem is specified by a

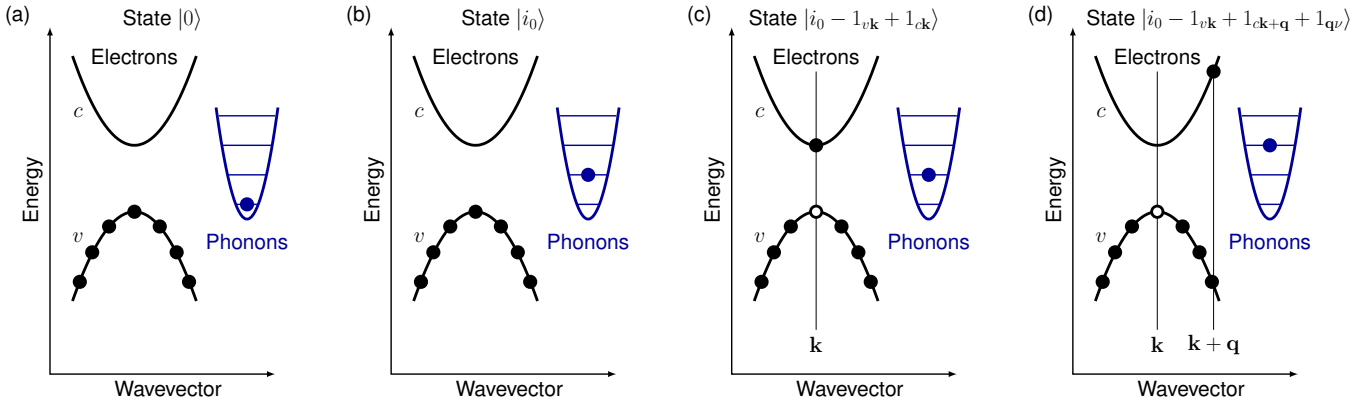

FIG. 2. **Many-body electron-phonon states in the occupation-number representation**. (a) Electron and phonon vacuum $|0\rangle$, corresponding to filled valence bands, empty conduction bands, and no phonons. v and c denote representative valence and conduction bands, respectively, and black disks indicate filled valence states. The blue parabola and ladder represent the eigenvalues of the quantum harmonic oscillator, with the blue disk indicating the phonon eigenvalue. In this case, the oscillator is in its ground state. (b) Born-Oppenheimer product state $|i_0\rangle$, corresponding to the electronic ground state and a finite number of phonons. (c) Electron-hole excitation $|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle$ created by removing an electron from $|i_0\rangle$ in the valence state $v\mathbf{k}$, and adding one electron to the conduction state $c\mathbf{k}$. The phonon configuration is the same as in $|i_0\rangle$. (d) Electron-hole-phonon excitation $|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} + 1_{\mathbf{q}\nu}\rangle$ obtained by removing an electron from $|i_0\rangle$ in the valence state $v\mathbf{k}$, adding an electron to the conduction state $c\mathbf{k}+\mathbf{q}$, and adding one phonon to the vibrational configuration of $|i_0\rangle$.

set of phonon occupation numbers [37, 38], see Fig. 2(b). Let us call $|0\rangle$ the electron-phonon vacuum, corresponding to an insulator with all valence bands filled, all conduction bands empty, and no phonons [Fig. 2(a)]. To fix ideas, we take the electron vacuum state to be a Slater determinant of Kohn-Sham states. Then the states $|i_0\rangle$ are of the form $\hat{a}^\dagger_{\mathbf{q}\nu}|0\rangle$, $\hat{a}^\dagger_{\mathbf{q}'\nu'}\hat{a}^\dagger_{\mathbf{q}\nu}|0\rangle$, and so on. Similarly, the states $|f_0\rangle$ correspond to uncorrelated electron-phonon states, where the electronic subsystem is in a many-body excited state, and the phonon subsystem is specified by a given set of phonon occupation numbers, see Fig. 2(c) and (d). At the end of the derivation we will perform a thermal average over phonon occupations. First-order non-degenerate perturbation theory gives:

$$|i\rangle = |i_0\rangle + \sum_{r_0\neq i_0} \frac{\langle r_0|\hat{V}_{\rm ep}|i_0\rangle}{E_{i_0} - E_{r_0}}|r_0\rangle \ , \tag{15}$$

$$|f\rangle = |f_0\rangle + \sum_{t_0\neq f_0} \frac{\langle t_0|\hat{V}_{\rm ep}|f_0\rangle}{E_{f_0} - E_{t_0}}|t_0\rangle \ . \tag{16}$$

where $|r_0\rangle$ and $|t_0\rangle$ denote intermediate Born-Oppenheimer product states. By replacing the last two relations in the matrix elements appearing in Eq. (14), to first order in perturbation theory we obtain:

$$\begin{aligned}\langle f|\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}|i\rangle = {}& \langle f_0|\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}|i_0\rangle \\ & + \sum_{t_0\neq f_0} \frac{\langle f_0|\hat{V}_{\rm ep}|t_0\rangle\langle t_0|\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}|i_0\rangle}{E_{f_0} - E_{t_0}} + \\ & + \sum_{r_0\neq i_0} \frac{\langle f_0|\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}|r_0\rangle\langle r_0|\hat{V}_{\rm ep}|i_0\rangle}{E_{i_0} - E_{r_0}} \ . \end{aligned} \tag{17}$$

The terms on each line of the r.h.s. of this equation correspond to direct optical transitions, indirect transitions with photon process first and phonon process second, and indirect transition with phonon process first and photon process second, respectively.

The term on the first line on the r.h.s. of Eq. (17) can be analyzed as follows. The index n must belong to the valence otherwise the braket vanishes, say $n = v$ where v indicate valence states. The state $\hat{c}_{v\mathbf{k}}|i_0\rangle$ has a hole in $v\mathbf{k}$, so there are only two options for m: either $m = v$, in which case $|f_0\rangle$ must coincide with $|i_0\rangle$ and the braket yields 1; or $m = c$, where c is a conduction state. We indicate this state using the notation $|f_0\rangle = |i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle$ from Kittel [38], see Fig. 2(c). In the former case, the summation $\sum_{v\mathbf{k}} \mathbf{v}_{vv\mathbf{k}}$ in Eq. (14) vanishes by time-reversal symmetry. In the latter case, $|f_0\rangle$ is a state with an electron-hole pair excited out of the ground state, and with the same phonon occupations as in the ground state. The energy of this state is $E_{f_0} = E_{i_0} - \varepsilon_{v\mathbf{k}} + \varepsilon_{c\mathbf{k}}$. The rate of optical transitions associated with this term can therefore be written from Eq. (14) as:

$$\sum_f \Gamma^{\rm (d)}_{i\to f} = \frac{\pi e^2 A_0^2}{2\hbar} \sum_{cv\mathbf{k}} |\mathbf{e}\cdot\mathbf{v}_{cv\mathbf{k}}|^2\, \delta(\varepsilon_{c\mathbf{k}} - \varepsilon_{v\mathbf{k}} - \hbar\omega) \ , \tag{18}$$

where the superscript "d" stands for "direct". This expression coincides with the conventional rate of Eq. (8) obtained using a single-particle formalism.

The term on the second line on the r.h.s. of Eq. (17) is analyzed similarly: n must belong to the valence to have a nonvanishing braket ($n = v$); the case $m = v$ can be discarded because it corresponds to real phonon-induced transitions across the gap, which can be ignored for systems with band gaps larger than the maximum phonon energy; therefore we must have $|t_0\rangle = |i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle$

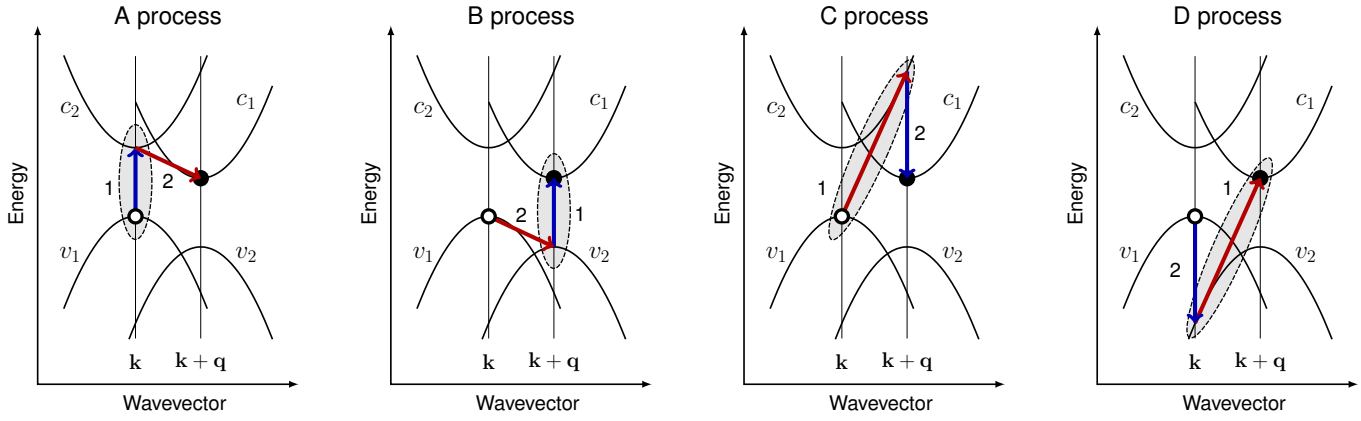

FIG. 3. **Phonon-assisted optical transitions in the occupation-number representation**. Schematic illustration of the indirect phonon-assisted optical processes A, B, C, and D described by Eq. (19). Blue arrows indicate photon-mediated transitions, which conserve crystal momentum; red arrows indicate phonon-assisted transitions, which do not conserve crystal momentum. Shaded ovals represent the intermediate excited state in each case. The open circle is the hole, the disk is the electron. The labels $v_{1,2}$ and $c_{1,2}$ denote representative valence and conduction bands; "1" and "2" indicate the order in which each process occurs. In processes B and D, the order of transitions is dictated by Pauli blocking.

and $E_{t_0} = E_{i_0} - \varepsilon_{v\mathbf{k}} + \varepsilon_{c\mathbf{k}}$. The coupling $\hat{V}_{\rm ep}$ can now lead to the following processes: (i) Generate a second electron-hole pair, which we ignore since it represents a higher-order process; (ii) Cause the transition of the electron in $c\mathbf{k}$ to another conduction state $c'\mathbf{k}+\mathbf{q}$, which is illustrated in Fig. 3 as process A; (iii) Cause the transition of the hole in $v\mathbf{k}$ to another valence state $v'\mathbf{k}+\mathbf{q}$, which is illustrated as process B in Fig. 3; (iv) Remove the electron from the state $c\mathbf{k}$ and create it again in the same state; this process corresponds to $\mathbf{q}=0$ phonons and its weight is negligible in the thermodynamic limit of an extended solid (i.e. dense Brillouin-zone sampling); (v) Remove the hole from $v\mathbf{k}$ and create it again, which is also negligible as per the previous point. In practice, only processes A and B contribute to the transition rates.

Since $\hat{V}_{\rm ep}$ in (7) contains phonon creation and annihilation operators, the final state $|f_0\rangle$ will also differ from $|t_0\rangle$ by having one phonon added or removed, that is $|f_0\rangle = \sqrt{n_{-\mathbf{q}\nu}+1}|t_0 - 1_{c\mathbf{k}} + 1_{c'\mathbf{k}+\mathbf{q}} + 1_{-\mathbf{q}\nu}\rangle$ or $|f_0\rangle = \sqrt{n_{\mathbf{q}\nu}}|t_0 - 1_{c\mathbf{k}} + 1_{c'\mathbf{k}+\mathbf{q}} - 1_{\mathbf{q}\nu}\rangle$ for A-processes, and $|f_0\rangle = \sqrt{n_{-\mathbf{q}\nu}+1}|t_0 + 1_{v\mathbf{k}} - 1_{v'\mathbf{k}+\mathbf{q}} + 1_{-\mathbf{q}\nu}\rangle$ or $|f_0\rangle = \sqrt{n_{\mathbf{q}\nu}}|t_0 + 1_{v\mathbf{k}} - 1_{v'\mathbf{k}+\mathbf{q}} - 1_{\mathbf{q}\nu}\rangle$ for B-processes, see Fig. 2(d). In these expressions, $n_{\mathbf{q}\nu}$ denotes the number of phonons of the intermediate state $|t_0\rangle$ in the mode $\mathbf{q}\nu$, which coincides with the number of phonons in the initial state $|i_0\rangle$. The energy of the final state is therefore $E_{f_0} = E_{t_0} - \varepsilon_{c\mathbf{k}} + \varepsilon_{c'\mathbf{k}+\mathbf{q}} + \hbar\omega_{-\mathbf{q}\nu}$ or $E_{f_0} = E_{t_0} - \varepsilon_{c\mathbf{k}} + \varepsilon_{c'\mathbf{k}+\mathbf{q}} - \hbar\omega_{\mathbf{q}\nu}$. The energies E_{f_0} and E_{t_0} thus determined are used to evaluate the denominators in the second line of Eq. (17).

A similar reasoning can be performed for the third line of Eq. (17). In this case the non-negligible contributions arise from the processes denoted by C and D in Fig. 3. In process C, $\hat{V}_{\rm ep}$ creates an electron-hole pair, and $\hat{V}_{\rm er}$ transfers an electron somewhere else in the conduction bands, or a hole somewhere else in the valence bands. We ignore processes that create two electron-hole pairs, and processes where $\hat{V}_{\rm ep}$ annihilates and creates the same electron, which require $\mathbf{q}=0$ and have negligible weight in extended systems.

Taking the above considerations into account, the rate for transitions associated with the phonon-mediated processes A, B, C, and D in Fig. 3 can be written starting from Eq. (14) as follows:

$$\begin{aligned}\sum_f \Gamma^{(\rm i)}_{i\to f} = \frac{\pi e^2 A_0^2}{2\hbar}\frac{1}{N}\sum_{\substack{cv\mathbf{k}\\ \mathbf{q}\nu,\eta}}[n_{\mathbf{q}\nu}+(1+\eta)/2] \\ \times\Bigg|\mathbf{e}\cdot\sum_{c'}\frac{g_{cc'\nu}(\mathbf{k},\mathbf{q})\mathbf{v}_{c'v\mathbf{k}}}{\varepsilon_{c\mathbf{k}+\mathbf{q}}-\varepsilon_{c'\mathbf{k}}+\eta\hbar\omega_{\mathbf{q}\nu}} \qquad \text{(A)} \\ +\ \mathbf{e}\cdot\sum_{v'}\frac{\mathbf{v}_{cv'\mathbf{k}+\mathbf{q}}g_{v'v\nu}(\mathbf{k},\mathbf{q})}{\varepsilon_{v'\mathbf{k}+\mathbf{q}}-\varepsilon_{v\mathbf{k}}+\eta\hbar\omega_{\mathbf{q}\nu}} \qquad \text{(B)} \\ +\ \mathbf{e}\cdot\sum_{c'}\frac{\mathbf{v}_{cc'\mathbf{k}+\mathbf{q}}g_{c'v\nu}(\mathbf{k},\mathbf{q})}{\varepsilon_{v\mathbf{k}}-\varepsilon_{c'\mathbf{k}+\mathbf{q}}-\eta\hbar\omega_{\mathbf{q}\nu}} \qquad \text{(C)} \\ +\ \mathbf{e}\cdot\sum_{v'}\frac{g_{cv'\nu}(\mathbf{k},\mathbf{q})\mathbf{v}_{v'v\mathbf{k}}}{\varepsilon_{v'\mathbf{k}}-\varepsilon_{c\mathbf{k}+\mathbf{q}}-\eta\hbar\omega_{\mathbf{q}\nu}}\Bigg|^2 \qquad \text{(D)} \\ \times\,\delta(\varepsilon_{c\mathbf{k}+\mathbf{q}}-\varepsilon_{v\mathbf{k}}+\eta\hbar\omega_{\mathbf{q}\nu}-\hbar\omega). \qquad (19)\end{aligned}$$

In this expression, the superscript "i" stands for "indirect", and we have added the labels A, B, C, D to indicate the contributions from the corresponding terms. The quantities $n_{\mathbf{q}\nu}$ denote the number of phonons in the ground state $|i_0\rangle$, and η takes the values $+1$ for phonon emission and -1 for phonon absorption. To reach Eq. (19), we have used $\omega_{\mathbf{q}\nu} = \omega_{-\mathbf{q}\nu}$ and $n_{-\mathbf{q}\nu} = n_{\mathbf{q}\nu}$. As for Eq. (10), band summations also run over the spin index, hence a factor of 2 needs to be added for spin-unpolarized calculations.

The numbers of phonons $n_{\mathbf{q}\nu}$ in Eq. (19) pertain to

the state $|i_0\rangle$. In order to obtain the transition rate in thermal equilibrium at the temperature T, we perform a canonical average of $\Gamma^{(\mathrm{i})}_{i\to f}$ over all possible phonon occupations, $\Gamma^{(\mathrm{i})}_{i\to f} \to \langle\Gamma^{(\mathrm{i})}_{i\to f}\rangle_T$. Since the rate is linear in $n_{\mathbf{q}\nu}$, this operation corresponds to performing the canonical average of the phonon number, which is a standard textbook exercise in statistical mechanics and yields the Bose-Einstein function:

$$n_{\mathbf{q}\nu} \;\longrightarrow\; n_{\mathbf{q}\nu}(T) = [\exp(\hbar\omega_{\mathbf{q}\nu}/k_{\mathrm{B}}T) - 1]^{-1}. \qquad (20)$$

Therefore the transition rate in thermal equilibrium is simply obtained by replacing $n_{\mathbf{q}\nu}(T)$ for $n_{\mathbf{q}\nu}$ in Eq. (19). Then, the total photon absorption rate is obtained by adding up Eqs. (18) and (19), $dN_{\mathrm{p}}/dt = \sum_f \langle\Gamma^{(\mathrm{d})}_{i\to f}\rangle_T + \sum_f \langle\Gamma^{(\mathrm{i})}_{i\to f}\rangle_T$, and the resulting dielectric function is found via Eqs. (3) and (9).

The indirect absorption rate derived here from Fock-space perturbation theory, Eq. (19), is almost identical to the textbook CHBB result in Eq. (11). The only difference is that the signs of the terms corresponding to process C and process D are inverted. This curious difference comes from the fact that the standard approach is a single-particle theory and does not take into account Pauli blocking. To better understand this difference, let us consider process D in Fig. 3. In CHBB one considers single-particle virtual transitions from v_1 to v_2, and then from v_2 to c_1; in this case, the order of processes is photon first, phonon second. However, in the many-body approach this sequence is forbidden by Pauli blocking: since the state v_2 is occupied, the many-body matrix element associated with the virtual transition v_1 to v_2 vanishes identically. The only nonzero matrix elements occur when considering first the phonon process (v_2 to c_1), which is not forbidden by Pauli blocking because c_1 is empty; and then the photon process (v_1 to v_2), which is now allowed because v_2 has been vacated in the previous step. In practice, this slight difference between the two approaches is almost inconsequential because processes C and D involve a phonon-assisted transition across the gap, and carry energy denominators that are at least as large as the gap; therefore these terms are usually much smaller than those from processes A and B which involve phonon-assisted transitions within the valence bands or within the conduction bands.

Equation (17) offers a clue to understand the problem of divergences with the CHBB theory: When $E_{t_0} = E_{f_0}$ in the denominators on the second line of Eq. (17), intermediate states in phonon-assisted processes are in resonance with direct processes. This situation can occur only for processes A and B shown in Fig. 3, and only when the photon energy exceeds the energy of the direct band gap. Conversely, the denominators on the third line of Eq. (17) cannot vanish because $E_{r_0} > E_{i_0}$ by construction. From a mathematical standpoint, these resonances correspond to vanishing energy denominators for A and B processes in Eq. (19), which involve energy differences within the valence bands or within the conduction bands; instead, the denominators in Eq. (19) corresponding to processes C and D cannot vanish because the difference between band energies is at least as large as the band gap.

In order to address the resonance between intermediate states $|t_0\rangle$ and final states $|f_0\rangle$ in Eq. (19), Eq. (16) must be upgraded to the case of *degenerate* perturbation theory. The next section is devoted to this generalization.

D. Many-body Fock-space quasi-degenerate perturbation theory approach

Equation (16) is no longer valid when $|f_0\rangle$ and $|t_0\rangle$ are degenerate. Furthermore, when $|f_0\rangle$ and $|t_0\rangle$ are quasi-degenerate, i.e. their energy difference is small but nonzero, the evaluation of Eq. (16) becomes numerically unstable due to the small-denominator problem. In this section we lift degeneracies by diagonalizing the electron-phonon interaction Hamiltonian $\hat{V}_{\mathrm{ep}}$ in the subspace of quasi-degenerate states, and we use the states thus determined to evaluate optical transition rates.

1. *Perturbation of quasi-degenerate states*

To handle the perturbation of quasi-degenerate states, we partition the photon energy axis into a series of contiguous intervals bounded by the energies $E_1, E_2, \cdots$, with $E_{n+1} = E_n + \Delta E$. The width ΔE of each interval is an externally-defined parameter, and the final results must be checked for convergence in the limit $\Delta E \to \infty$. For each interval, we treat all states $|t_0\rangle$ with energy E_{t_0} inside the interval as quasi-degenerate, and all other states outside this window as non-degenerate. This partitioning is shown in Fig. 4. In the following we concentrate on the states that belong to the lowest energy interval, $[E_1, E_2)$. After completing the process for one interval, we repeat the procedure for the next interval, until we span the entire range of photon energies of interest.

Let us rename the states $|t_0\rangle$ with $E_1 \le E_{t_0} < E_2$ as $|d_0; p\rangle$, with $p = 1, 2, \cdots N_{\mathrm{qd}}$ to distinguish them from all other $|t_0\rangle$ states with energies outside of this window. Barring accidental degeneracies, the states $|d_0; p\rangle$ have slightly different energies and cannot be treated using degenerate perturbation theory. To use perturbation theory, we proceed as follows:

(i) We define the midpoint energy of the interval $[E_1, E_2)$ as $\bar{E} = (E_1 + E_2)/2$.

(ii) We add a scissor operator to the non-interacting Hamiltonian $\hat{H}_0$ in Eq. (5). This operator is chosen so as to make the states $|d_0; p\rangle$ exactly degenerate by shifting their individual energies to $\bar{E}$:

$$\hat{H}'_0 = \hat{H}_0 + \sum_p (\bar{E} - E_{d_0;p})|d_0; p\rangle\langle d_0; p| \; . \qquad (21)$$

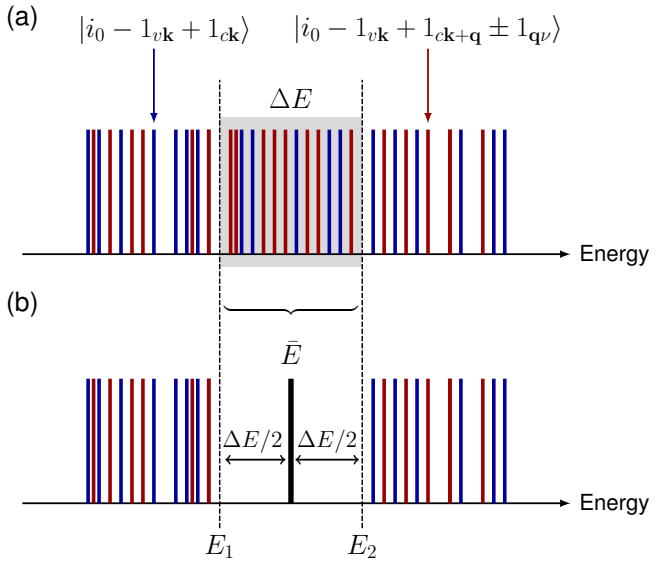

FIG. 4. **Handling of quasi-degenerate states in quasi-degenerate perturbation theory**. (a) Schematic illustration of the energy window employed to identify quasi-degenerate electron-hole and electron-hole-phonon excitations needed in Eq. (23). Vertical bars indicate the energy of an excited electron-hole pair $|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle$ (blue) or an electron-hole-phonon excitation $|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} \pm 1_{\mathbf{q}\nu}\rangle$ (red). States with energy between E_1 and E_2 are considered quasi-degenerate. (b) We make quasi-degenerate states exactly degenerate by using a scissor operator. After this operation, we have a degenerate manifold of states at the energy $\bar{E}$, separated by all other states by the energy $\Delta E/2$.

We then subtract the same operator from the electron-phonon Hamiltonian $\hat{V}_{\rm ep}$, in order to keep the total Hamiltonian $\hat{H}$ unchanged, i.e. $\hat{H}_0 + \hat{V}_{\rm ep} = \hat{H}'_0 + \hat{V}'_{\rm ep}$:

$$\hat{V}'_{\rm ep} = \hat{V}_{\rm ep} - \sum_p (\bar{E} - E_{d_0;p}) |d_0;p\rangle\langle d_0;p| \ . \tag{22}$$

(iii) As a result of steps (i) and (ii), we obtain a manifold of degenerate states $|d_0;p\rangle$ with energy $\bar{E}$, which is separated by all other states $|t_0\rangle$ by an energy of at least $\Delta E/2$, see Fig. 4. The first-order change of these states can be obtained using the standard prescription of degenerate perturbation theory [39, 40]. This procedure requires us to first lift the degeneracy by diagonalizing $\hat{V}'_{\rm ep}$ in the degenerate manifold, and then to apply non-degenerate perturbation theory to the zeroth-order states just obtained. Let us call U_{mp} the diagonalizer of this perturbation, and λ_m is the associated eigenvalues, so that the excited-state energies including electron-phonon interactions are $E_p = \bar{E} + \lambda_p$.

$$\langle d_0;s|\hat{V}'_{\rm ep}|d_0;p\rangle = \sum_m U_{sm}\lambda_m U^{-1}_{mp} \ , \tag{23}$$

and the states with lifted degeneracy at zeroth order are given by:

$$|\tilde{d}_0;m\rangle = \sum_p U_{mp}|d_0;p\rangle \ . \tag{24}$$

(iv) Using the broken-degeneracy states $|\tilde{d}_0;m\rangle$ from (iii), we build the perturbed final states from first-order non-degenerate perturbation theory:

$$|f;m\rangle = |\tilde{d}_0;m\rangle + {\sum_{t_0}}' \frac{\langle t_0|\hat{V}_{\rm ep}|\tilde{d}_0;m\rangle}{\bar{E} - E_{t_0}}|t_0\rangle + \sum_{n\neq m}\sum_{t_0}{}' \frac{\langle \tilde{d}_0;n|\hat{V}_{\rm ep}|t_0\rangle\langle t_0|\hat{V}_{\rm ep}|\tilde{d}_0;m\rangle}{\bar{E} - E_{t_0}} \frac{|\tilde{d}_0;n\rangle}{\lambda_m - \lambda_n}. \tag{25}$$

The prime in the summations indicates that only states t_0 such that $E_{t_0} < E_1$ or $E_{t_0} \geq E_2$ are included. In Eq. (25) we have added an index m to $|f\rangle$ to emphasize that this procedure leads to $m = 1, 2, \cdots N_{\rm qd}$ final states.

The sum in the second line of Eq. (25) can be ignored on the grounds that it leads to a two-phonon process. In fact, after replacing this expression inside Eq. (14), we see that the corresponding process consists of a photon-assisted transition from the ground state to an intermediate state, followed by a phonon-assisted transition to another intermediate state, and a second phonon-assisted transition to the final state. Upon ignoring these higher-order processes, Eq. (25) can be rewritten more conveniently by making the original states and the unitary matrices explicit:

$$|f;m\rangle = \sum_p U_{mp}\left[|d_0;p\rangle + {\sum_{t_0}}' \frac{\langle t_0|\hat{V}_{\rm ep}|d_0;p\rangle}{\bar{E} - E_{t_0}}|t_0\rangle\right] \ . \tag{26}$$

This result is analogous to the non-degenerate case in Eq. (16), except for the unitary matrix and for the removal of quasi-degenerate states in the primed summation.

If we reduce the size ΔE of the energy intervals to the point of having only either one or no states within each interval, then in the case of a single state we have $N_{\rm qd} = 1$, U_{mp} is of size 1×1, and Eq. (26) reduces to the simpler non-degenerate prescription of Eq. (16).

2. *Explicit expression for the perturbation matrix*

In order to employ Eq. (26), we need to diagonalize the perturbation matrix $\langle d_0;s|\hat{V}'_{\rm ep}|d_0;p\rangle$ appearing in Eq. (23). To this end, we express the quasi-degenerate manifold of states $|d_0;p\rangle$ in the language of single-particle excitations and phonon excitations.

The simplest states $|t_0\rangle$ are of the type $|t_0\rangle = |i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle$ with energy $E_{t_0} = E_{i_0} - \varepsilon_{v\mathbf{k}} + \varepsilon_{c\mathbf{k}}$ [Sec. II C and Fig. 2(b)]. These states are connected to other quasi-degenerate states in processes A and B of Fig. 3 by $\hat{V}'_{\rm ep}$. States $|d_0; s\rangle$ with nonvanishing matrix elements $\langle d_0; s|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle$ are of the type:

$$|i_0 - 1_{v\mathbf{k}} + 1_{c'\mathbf{k}+\mathbf{q}} - 1_{\mathbf{q}\nu}\rangle \ , \tag{27}$$
$$|i_0 - 1_{v\mathbf{k}} + 1_{c'\mathbf{k}+\mathbf{q}} + 1_{-\mathbf{q}\nu}\rangle \ , \tag{28}$$
$$|i_0 - 1_{v'\mathbf{k}+\mathbf{q}} + 1_{c\mathbf{k}} + 1_{\mathbf{q}\nu}\rangle \ , \tag{29}$$
$$|i_0 - 1_{v'\mathbf{k}+\mathbf{q}} + 1_{c\mathbf{k}} - 1_{-\mathbf{q}\nu}\rangle \ , \tag{30}$$

and the corresponding energies are:

$$E_{i_0} - \varepsilon_{v\mathbf{k}} + \varepsilon_{c'\mathbf{k}+\mathbf{q}} - \hbar\omega_{\mathbf{q}\nu} \ , \tag{31}$$
$$E_{i_0} - \varepsilon_{v\mathbf{k}} + \varepsilon_{c'\mathbf{k}+\mathbf{q}} + \hbar\omega_{-\mathbf{q}\nu} \ , \tag{32}$$
$$E_{i_0} - \varepsilon_{v'\mathbf{k}+\mathbf{q}} + \varepsilon_{c\mathbf{k}} + \hbar\omega_{\mathbf{q}\nu} \ , \tag{33}$$
$$E_{i_0} - \varepsilon_{v'\mathbf{k}+\mathbf{q}} + \varepsilon_{c\mathbf{k}} - \hbar\omega_{-\mathbf{q}\nu} \ . \tag{34}$$

To build the perturbation matrix, we search for all such states with energies within a given energy window, say $[E_1, E_2)$. From the above relations, we see that all relevant states can be expressed in one of the following forms:

$$\text{(i)} \quad |i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle \ , \tag{35}$$
$$\text{(ii)} \quad |i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} - 1_{\mathbf{q}\nu}\rangle \ , \tag{36}$$
$$\text{(iii)} \quad |i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} + 1_{-\mathbf{q}\nu}\rangle \ , \tag{37}$$

with $v, c, \nu, \mathbf{k}, \mathbf{q}$ spanning the entire electron and phonon band manifold. The dimensionality of the corresponding Hilbert space is $N_v \times N_c \times N_{\rm ph} \times N^2$, where N_v, N_c, and $N_{\rm ph}$ are the number of valence bands, conduction bands, and phonon branches, respectively (we recall that N is the number of points in the Brillouin zone). In practice, since we only consider states within a small energy window, the size of the Hilbert space is drastically reduced.

The matrix elements of $\hat{V}'_{\rm ep}$ between the states in Eqs. (35)-(37) can be evaluated directly using Eqs. (7) and (22). The results are:

$$\langle i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle = (E_{i_0} + \varepsilon_{c\mathbf{k}} - \varepsilon_{v\mathbf{k}} - \bar{E}) \ , \tag{38}$$
$$\langle i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} - 1_{\mathbf{q}\nu}|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} - 1_{\mathbf{q}\nu}\rangle = (E_{i_0} + \varepsilon_{c\mathbf{k}+\mathbf{q}} - \varepsilon_{v\mathbf{k}} - \hbar\omega_{\mathbf{q}\nu} - \bar{E}) \ , \tag{39}$$
$$\langle i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} + 1_{-\mathbf{q}\nu}|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}+\mathbf{q}} + 1_{-\mathbf{q}\nu}\rangle = (E_{i_0} + \varepsilon_{c\mathbf{k}+\mathbf{q}} - \varepsilon_{v\mathbf{k}} + \hbar\omega_{-\mathbf{q}\nu} - \bar{E}) \ , \tag{40}$$
$$\langle i_0 - 1_{v\mathbf{k}} + 1_{c'\mathbf{k}+\mathbf{q}} - 1_{\mathbf{q}\nu}|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle = N^{-\frac{1}{2}}\sqrt{n_{\mathbf{q}\nu}}\, g_{c'c\nu}(\mathbf{k}, \mathbf{q}) \ , \tag{41}$$
$$\langle i_0 - 1_{v\mathbf{k}} + 1_{c'\mathbf{k}+\mathbf{q}} + 1_{-\mathbf{q}\nu}|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle = N^{-\frac{1}{2}}\sqrt{n_{-\mathbf{q}\nu} + 1}\, g_{c'c\nu}(\mathbf{k}, \mathbf{q}) \ , \tag{42}$$
$$\langle i_0 - 1_{v'\mathbf{k}+\mathbf{q}} + 1_{c\mathbf{k}} - 1_{-\mathbf{q}\nu}|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle = N^{-\frac{1}{2}}\sqrt{n_{-\mathbf{q}\nu}}\, g_{vv'\nu}(\mathbf{k}+\mathbf{q}, -\mathbf{q}) \ , \tag{43}$$
$$\langle i_0 - 1_{v'\mathbf{k}+\mathbf{q}} + 1_{c\mathbf{k}} + 1_{\mathbf{q}\nu}|\hat{V}'_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle = N^{-\frac{1}{2}}\sqrt{n_{\mathbf{q}\nu} + 1}\, g_{vv'\nu}(\mathbf{k}+\mathbf{q}, -\mathbf{q}) \ . \tag{44}$$

All the other matrix elements that can be formed with the states of the type given in Eqs. (35)-(37) vanish.

By diagonalizing the matrix with elements given by Eqs. (38)-(44), we obtain the unitary transformation U_{mp} needed in Eq. (24), as well as the final-state energies to be used in Eq. (14). Using this matrix, we can write the broken-degeneracy states in Eq. (26) as:

$$\begin{aligned}
|f; m\rangle = &\sum_{cv\mathbf{k}} U_{m,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}}}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle \\
&+ \sum_{cv\mathbf{k},\mathbf{q}\nu} U_{m,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}-1_{\mathbf{q}\nu}}|i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}-1_{\mathbf{q}\nu}\rangle \\
&+ \sum_{cv\mathbf{k},\mathbf{q}\nu} U_{m,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+1_{-\mathbf{q}\nu}}|i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+1_{-\mathbf{q}\nu}\rangle \\
&+ \sum_{cv\mathbf{k}} U_{m,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}}} {\sum_{t_0}}' \frac{\langle t_0|\hat{V}_{\rm ep}|i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\rangle}{\bar{E} - E_{t_0}}|t_0\rangle \\
&+ \sum_{cv\mathbf{k},\mathbf{q}\nu} U_{m,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}-1_{\mathbf{q}\nu}} \\
&\qquad \times {\sum_{t_0}}' \frac{\langle t_0|\hat{V}_{\rm ep}|i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}-1_{\mathbf{q}\nu}\rangle}{\bar{E} - E_{t_0}}|t_0\rangle \\
&+ \sum_{cv\mathbf{k},\mathbf{q}\nu} U_{m,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+1_{-\mathbf{q}\nu}} \\
&\qquad \times {\sum_{t_0}}' \frac{\langle t_0|\hat{V}_{\rm ep}|i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+1_{-\mathbf{q}\nu}\rangle}{\bar{E} - E_{t_0}}|t_0\rangle
\end{aligned} \tag{45}$$

In this equation, we used an explicit notation for the index p in U_{mp}, e.g. $p = \{i_0 - 1_{v\mathbf{k}} + 1_{c\mathbf{k}}\}$, to show the relation to single-particle excitations in band and phonon indices.

We note that the matrix elements given by Eqs. (41)-(44) depend on the phonon occupations in the initial state $|i_0\rangle$. Since the diagonalization of this matrix is a nonlinear operation, the procedure employed in Eq. (20) to obtain the thermal average of the transition rates is no longer valid. Efficient strategies to overcome this obstacle are presented in Sec. III B 2.

3. Evaluating the transition rates and dielectric function

We can now derive the optical transition rates and the dielectric function by taking into account quasi-degeneracies as discussed in Sec. II D 2. To this end, we first rewrite the matrix elements $\langle f; p|\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}|i\rangle$ in Eq. (17) by taking into account Eqs. (26) and (45) instead of their non-degenerate counterpart Eq. (16). Explicit evaluation of the fermionic and bosonic operators gives:

$$\langle f; p|\hat{c}^\dagger_{m\mathbf{k}}\,\hat{c}_{n\mathbf{k}}|i\rangle = f_{n\mathbf{k}}(1-f_{m\mathbf{k}})U^*_{p,i_0-1_{n\mathbf{k}}+1_{m\mathbf{k}}} + N^{-1/2}\sum_{\mathbf{q}\nu,\eta}\sqrt{n_{-\eta\mathbf{q}\nu}+(\eta+1)/2}$$

$$\times\Bigg\{\sum_{c} U^*_{p,i_0-1_{n\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+\eta 1_{-\eta\mathbf{q}\nu}} g_{cm\nu}(\mathbf{k},\mathbf{q})$$
$$\times\left[\frac{f_{n\mathbf{k}}(1-f_{m\mathbf{k}})\theta_{mn\mathbf{k}}}{(\bar{E}-E_{i_0})-(\varepsilon_{m\mathbf{k}}-\varepsilon_{n\mathbf{k}})}+\frac{f_{m\mathbf{k}}f_{n\mathbf{k}}}{\varepsilon_{n\mathbf{k}}-\varepsilon_{c\mathbf{k}+\mathbf{q}}-\eta\hbar\omega_{-\eta\mathbf{q}\nu}}\right]$$
$$+\quad\sum_{v} U^*_{p,i_0-1_{v\mathbf{k}-\mathbf{q}}+1_{m\mathbf{k}}+\eta 1_{-\eta\mathbf{q}\nu}} g_{nv\nu}(\mathbf{k}-\mathbf{q},\mathbf{q})$$
$$\times\left[\frac{f_{n\mathbf{k}}(1-f_{m\mathbf{k}})\theta_{mn\mathbf{k}}}{(\bar{E}-E_{i_0})-(\varepsilon_{m\mathbf{k}}-\varepsilon_{n\mathbf{k}})}+\frac{(1-f_{n\mathbf{k}})(1-f_{m\mathbf{k}})}{\varepsilon_{v\mathbf{k}-\mathbf{q}}-\varepsilon_{n\mathbf{k}}-\eta\hbar\omega_{-\eta\mathbf{q}\nu}}\right]\Bigg\}\,. \tag{46}$$

In this expression, $f_{n\mathbf{k}}=1$ for valence bands, and 0 for conduction bands; $\theta_{mn\mathbf{k}}=1$ when the denominator $|(\bar{E}-E_{i_0})-(\varepsilon_{m\mathbf{k}}-\varepsilon_{n\mathbf{k}})|$ is larger than $\Delta E/2$, and 0 otherwise. These factors *eliminate* the small-denominator problem from all terms in Eq. (46). In fact, the factor $\theta_{mn\mathbf{k}}$ removes denominators smaller than $\Delta E/2$ from the fractions on the left, and the occupations $f_{n\mathbf{k}}$ ensure that the denominators of the fractions on the right are at least as large as the band gap.

The optical matrix elements obtained in Eq. (46) can be used inside Eq. (14) to obtain the transition rates from the initial state $|i\rangle$ to one of the quasi-degenerate final states $|f;p\rangle$ within the energy window $[E_1,E_2)$:

$$\Gamma_{i\to(f;p)}=\frac{\pi e^2A_0^2}{2\hbar}\Bigg|\mathbf{e}\cdot\sum_{cv\mathbf{k}}\Bigg\{U^*_{p,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}}}\mathbf{v}_{cv\mathbf{k}}$$
$$+\quad N^{-1/2}\sum_{\mathbf{q}\nu\eta}\sqrt{n_{\mathbf{q}\nu}+\frac{1+\eta}{2}}\;U^*_{p,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+\eta 1_{-\eta\mathbf{q}\nu}}$$
$$\times\Bigg[\sum_{c'}\frac{g_{cc'\nu}(\mathbf{k},\mathbf{q})\mathbf{v}_{c'v\mathbf{k}}}{(\bar{E}-E_{i_0})-(\varepsilon_{c'\mathbf{k}}-\varepsilon_{v\mathbf{k}})}\,\theta_{c'v\mathbf{k}} \qquad \text{(A)}$$
$$+\quad\sum_{v'}\frac{\mathbf{v}_{cv'\mathbf{k}+\mathbf{q}}\,g_{v'v\nu}(\mathbf{k},\mathbf{q})}{(\bar{E}-E_{i_0})-(\varepsilon_{c\mathbf{k}+\mathbf{q}}-\varepsilon_{v'\mathbf{k}+\mathbf{q}})}\,\theta_{cv'\mathbf{k}+\mathbf{q}} \qquad \text{(B)}$$
$$+\quad\sum_{c'}\frac{\mathbf{v}_{cc'\mathbf{k}+\mathbf{q}}\,g_{c'v\nu}(\mathbf{k},\mathbf{q})}{\varepsilon_{v\mathbf{k}}-\varepsilon_{c'\mathbf{k}+\mathbf{q}}-\eta\hbar\omega_{-\eta\mathbf{q}\nu}} \qquad \text{(C)}$$
$$+\quad\sum_{v'}\frac{g_{cv'\nu}(\mathbf{k},\mathbf{q})\mathbf{v}_{v'v\mathbf{k}}}{\varepsilon_{v'\mathbf{k}}-\varepsilon_{c\mathbf{k}+\mathbf{q}}-\eta\hbar\omega_{-\eta\mathbf{q}\nu}}\Bigg]\Bigg\}\Bigg|^2 \qquad \text{(D)}$$
$$\times\;\delta(E_p-E_{i_0}-\hbar\omega)\,. \tag{47}$$

In order to obtain the last expression, we used $\omega_{-\mathbf{q}\nu}=\omega_{\mathbf{q}\nu}$ and $n_{-\mathbf{q}\nu}=n_{\mathbf{q}\nu}$, and we expressed $g_{mn\nu}(\mathbf{k}-\mathbf{q},\mathbf{q})$ in terms of $g_{mn\nu}(\mathbf{k},\mathbf{q})$ by a change of summation variables.

Equation (47) is the generalization of Eqs. (18) and (19) to the case of quasi-degenerate perturbation theory. We see that this equations now contains contributions from both direct transitions (first line) and phonon-assisted transitions (second to fifth line). In this case, all transitions are correlated via the unitary matrices U_{mp}, and as a result the summation $\sum_{cv\mathbf{k}}$ takes place *inside* the absolute value unlike in Eqs. (18) and (19). Accordingly, the Dirac delta in Eq. (47) selects a single quasi-degenerate final state $E_p=\bar{E}+\lambda_p$, which consists of a superposition of electron-hole pair excitations and phonons, as shown by Eq. (45). This formulation describes excited states as linear combinations of Born-Oppenheimer product states, therefore it includes electron-hole-phonon *correlations* from the start.

We note that the terms in the third to sixth line of Eq. (47) correspond precisely to the processes A, B, C, D identified in Eq. (19) and illustrated in Fig. 3. The main difference with respect to Eq. (19) lies in the fact that, by construction, the denominators of processes A and B cannot vanish, therefore the small-denominator problem of the CHBB theory is completely eliminated.

To obtain the dielectric function from Eq. (47) we proceed as follows: (i) We sum the transition rates over all quasi-degenerate final states; (ii) We carry out a canonical average of the transition rates over the phonon occupations of the initial states $|i_0\rangle$ using the Gibbs factor $Z^{-1}\exp(-\beta E_{i_0})$, where $Z=\sum_{i_0}\exp(-\beta E_{i_0})$ is the partition function and $\beta=(k_{\mathrm{B}}T)^{-1}$; (iii) We make use of Eqs. (3) and (9). The result is:

$$\varepsilon_2(\omega)=\frac{\pi e^2}{\epsilon_0\Omega}\frac{1}{\omega^2}\frac{1}{N}\sum_{i_0,p}Z^{-1}\exp(-\beta E_{i_0})$$
$$\times\Bigg|\mathbf{e}\cdot\sum_{cv\mathbf{k}}\Bigg\{U^*_{p,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}}}\mathbf{v}_{cv\mathbf{k}}$$
$$+\quad N^{-1/2}\sum_{\mathbf{q}\nu\eta}\sqrt{n_{\mathbf{q}\nu}+\frac{1+\eta}{2}}\;U^*_{p,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+\eta 1_{-\eta\mathbf{q}\nu}}$$
$$\times\Bigg[\sum_{c'}\frac{g_{cc'\nu}(\mathbf{k},\mathbf{q})\mathbf{v}_{c'v\mathbf{k}}}{(\bar{E}-E_{i_0})-(\varepsilon_{c'\mathbf{k}}-\varepsilon_{v\mathbf{k}})}\,\theta_{c'v\mathbf{k}}$$
$$+\quad\sum_{v'}\frac{\mathbf{v}_{cv'\mathbf{k}+\mathbf{q}}\,g_{v'v\nu}(\mathbf{k},\mathbf{q})}{(\bar{E}-E_{i_0})-(\varepsilon_{c\mathbf{k}+\mathbf{q}}-\varepsilon_{v'\mathbf{k}+\mathbf{q}})}\,\theta_{cv'\mathbf{k}+\mathbf{q}}$$
$$+\quad\sum_{c'}\frac{\mathbf{v}_{cc'\mathbf{k}+\mathbf{q}}\,g_{c'v\nu}(\mathbf{k},\mathbf{q})}{\varepsilon_{v\mathbf{k}}-\varepsilon_{c'\mathbf{k}+\mathbf{q}}-\eta\hbar\omega_{-\eta\mathbf{q}\nu}}$$
$$+\quad\sum_{v'}\frac{g_{cv'\nu}(\mathbf{k},\mathbf{q})\mathbf{v}_{v'v\mathbf{k}}}{\varepsilon_{v'\mathbf{k}}-\varepsilon_{c\mathbf{k}+\mathbf{q}}-\eta\hbar\omega_{-\eta\mathbf{q}\nu}}\Bigg]\Bigg\}\Bigg|^2$$
$$\times\;\delta(E_p-E_{i_0}-\hbar\omega)\,. \tag{48}$$

In this expression, the energy of the initial state $|i_0\rangle$ is given by $E_{i_0}=E_0+\sum_{\mathbf{q}\nu}\hbar\omega_{\mathbf{q}\nu}/2+\sum_{\mathbf{q}\nu}n_{\mathbf{q}\nu}\hbar\omega_{\mathbf{q}\nu}$, where E_0 is the energy of the electronic ground state. The canonical average must be performed over all possible occupations $n_{\mathbf{q}\nu}=0,1,2,\cdots$. Unlike in Eq. (19), in this case the transition rates are not simple linear functions of $n_{\mathbf{q}\nu}$, therefore the replacement of $n_{\mathbf{q}\nu}$ by the Bose-Einstein occupation as in Eq. (20) is not allowed in principle. In particular, the occupations $n_{\mathbf{q}\nu}$ are implicitly contained in the energies E_{i_0}, in the unitary matrices U_{mp}, and in the eigenvalues E_p. In Sec. III B 2 we discuss efficient strategies to address this point.

Equation (48) is the main result of this work, and constitutes the *unification* of the theories of direct and phonon-assisted absorption and emission into a single formalism. This equation reduces to the standard results in

Eqs. (18) and (19) when direct and phonon-assisted transitions are well separated in energy and correlation effects between electron-hole pair excitations and phonons are neglected. Furthermore, Eq. (48) resolves the small denominator problem of the CHBB theory and provides the correct dielectric function throughout the complete photon energy range, spanning both direct and phonon-assisted transitions. One appealing feature of Eq. (48) is that it maintains a close formal similarity with previous theories and has a transparent physical interpretation, as we discuss in the following sections.

III. COMPUTATIONAL SETUP

A. Calculation parameters

In this section we provide details of the computational setup used to validate the methodology presented in Sec. II D and to compute the optical spectra of Si, Ge, GaAs, and diamond. Readers primarily interested in the applications of this methodology can skip directly to Sec. IV for Si and Ge, and to Appendices A and B for GaAs and diamond, respectively.

Single particle wavefunctions, band structures, and phonon dispersions and eigenmodes are calculated using density functional theory (DFT) and density functional perturbation theory (DFPT) [41, 42], respectively, as implemented in the `Quantum ESPRESSO` suite [43]. Electron-phonon matrix elements are computed using the `EPW` code [17, 30, 44], which calss the `Wannier90` code in library mode [45]. We employ the local density approximation (LDA) to density functional theory [46, 47], and norm-conserving pseudopotentials [48, 49]. The planewaves kinetic energy cutoff is set to 60 Ry, 140 Ry, 100 Ry, and 120 Ry for Si, Ge, GaAs, and diamond, respectively, and a common energy convergence threshold of 10^{-13} Ry is used throughout.

For the Wannier interpolation of electron bands, phonon dispersions, and electron-phonon matrix elements, we use 6^3 coarse Brillouin-zone grids in all cases. The fine grids for Si, GaAs, diamond, and Ge are 32^3 **k**-points and 8^3 **q**-points, 84^3 **k**-points and 6^3 **q**-points, 48^3 **k**-points and 12^3 **q**-points, 64^3 **k**-points and 8^3 **q**-points, respectively. To resolve the phonon fine structure near the optical absorption or emission onset, in the case of Si and Ge we also employ denser fine grids with 60^3 **k**- and 30^3 **q**-points, and 140^3 **k**- and 20^3 **q**-points, respectively. These grids ensure convergence of all spectra reported in this manuscript.

For all four materials, we use their experimental lattice constants at 300 K. To correct for the DFT bandgap underestimation, we use scissor shifts $\Delta = 0.60$ eV, 0.30 eV, 0.54 eV, and 1.40 eV for Si, Ge, GaAs, and diamond, respectively. Accordingly, we renormalize the velocity matrix elements using the factor $(\varepsilon_{c\mathbf{k}} - \varepsilon_{v\mathbf{k}})/(\varepsilon_{c\mathbf{k}} - \varepsilon_{v\mathbf{k}} - \Delta)$, where the energies $\varepsilon_{c\mathbf{k}}$ include the scissor shift [50].

The energy interval ΔE required for evaluating Eq. (48) is set to 160 meV for Si, Ge, and GaAs, and to 400 meV for diamond. Convergence tests with respect to this parameter are discussed in Sec. III B 1.

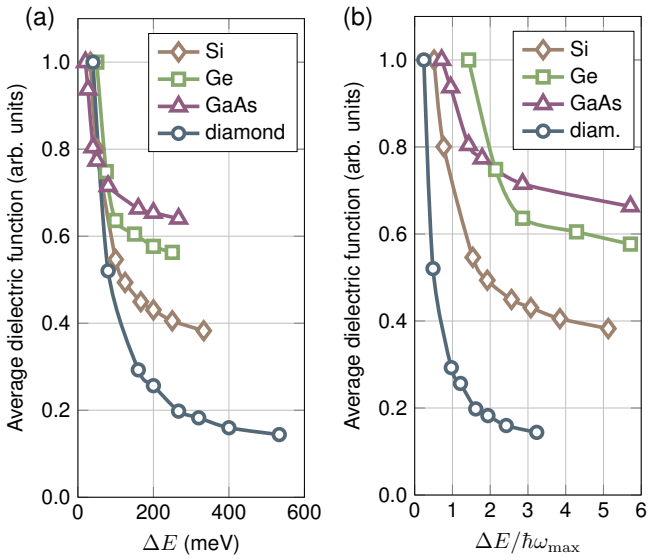

FIG. 5. **Convergence tests: Energy window**. Convergence of quasi-degenerate perturbation theory calculations with respect to the energy window ΔE. (a) Imaginary part of the dielectric function of Si, Ge, GaAs, and diamond calculated using Eq. (48) by varying the energy window ΔE. We provide the average $(\omega_2 - \omega_1)^{-1} \int_{\omega_1}^{\omega_2} d\omega\, \varepsilon_2(\omega)$ in the energy intervals $[\hbar\omega_1 : \hbar\omega_2] = 3.4{:}4.5$ eV, $0.8{:}1.5$ eV, $1.5{:}2.5$ eV, $6.7{:}7.5$ eV, for Si, Ge, GaAs, and diamond, respectively. (b) Same dataset as in (a), but this time we scale the horizontal axis by the largest phonon energy of each compound, $\hbar\omega_{\max}$. It is seen that convergence is achieved when the energy window is a few multiples of the characteristic phonon energy.

In calculations of the absorption coefficient via Eq. (1) and of the radiative emission rate via Eq. (2), we obtain the real part of the refractive index from $n^2 = (\sqrt{\varepsilon_1^2 + \varepsilon_2^2} + \varepsilon_1)/2$, where ε_1 is the real part of the dielectric function. $\varepsilon_1(\omega)$ is obtained as the Kramers-Kronig transform of $\varepsilon_2(\omega)$, and the unknown constant shift is set by using the DFPT value for $\varepsilon_1(\omega = 0)$.

B. Convergence tests

1. *Convergence with the quasi-degeneracy energy window*

The calculation of the dielectric function using Eq. (48) requires us to specify the energy window ΔE to identify quasi-degenerate states, see Fig. 4. For small ΔE, only a few states are treated as quasi-degenerate; conversely, as $\Delta E \to \infty$, the energy window encompasses the entire electron-hole-phonon Hilbert space. The latter scenario leads to the most accurate result but is computationally prohibitive due to the fine Brillouin zone grids required for **k**- and **q**-points. In practice, we need to choose the

smallest ΔE that leads to accurate values of $\varepsilon_2(\omega)$ by performing a convergence test with respect to this parameter.

In Fig. 5(a) we show the imaginary part of the dielectric function of Si, Ge, GaAs, and diamond, as calculated using Eq. (48) by varying the energy window ΔE. For conciseness, we provide the average values of $\varepsilon_2(\omega)$ over the photon energy ranges 3.4-4.5 eV, 0.8-1.5 eV, 1.5-2.5 eV, and 6.7-7.5 eV for Si, Ge, GaAs, and diamond, respectively. These ranges have been chosen to be slightly above the onset for direct transitions, where quasi-degeneracy effects are significant. The sensitivity of the results to ΔE is highest in these energy ranges, and becomes less important away from these intervals. Therefore, Fig. 5 illustrates a worst-case scenario. Calculations can be considered as converged for $\Delta E = 280$ meV, 200 meV, 200 meV, and 400 meV for Si, Ge, GaAs, and diamond, respectively.

The size of the energy window is expected to scale with the magnitude of the characteristic electron-phonon matrix elements $g_{mn\nu}(\mathbf{k},\mathbf{q})$, which enter the perturbation $\hat{V}'_{\rm ep}$ discussed in Sec. II D 2. As a simple proxy for the average matrix elements, we consider the energy of the highest optical phonon, $\hbar\omega_{\rm max}$. In the case of Si, Ge, GaAs, and diamond, we have $\hbar\omega_{\rm max} = 62$ meV, 38 meV, 35 meV, and 162 meV, respectively. Figure 4(b) shows the same data as in Fig. 4(a), except that on the horizontal axis we use $\Delta E/\hbar\omega_{\rm max}$ instead of ΔE. We can see that, with this scaling of the energy axis, convergence is achieved when the energy window corresponds to a few multiples of the characteristic phonon energy.

2. *Canonical average over the phonon population*

The canonical average of Eq. (48) over the phonon population is computationally challenging because the number of initial states $|i_0\rangle$ grows combinatorially with the number of $\mathbf{q}$-points and phonon occupations at each wavevector. Specifically, if we have N $\mathbf{q}$-points and $N_{\rm ph}$ phonon branches, and we consider the phonon occupations $n_{\mathbf{q}\nu} = 0, 1, \cdots, n_{\rm max} - 1$ for each mode $\mathbf{q}\nu$, there are $n_{\rm max}^{N_{\rm ph}N}$ distinct $|i_0\rangle$ states to be included in the canonical average. For the compounds and Brillouin-zone grids considered in this work, the number of initial states is of the order of $n_{\rm max}^{60,000}$.

To address this challenge, we propose to *replace* the phonon occupations in Eq. (48) and in the perturbation matrix elements in Eqs. (41)-(44) by the corresponding Bose-Einstein occupation, as in Eq. (20), and we validate this choice using a Monte Carlo calculation. The motivation for this choice is twofold: (i) In the case of vanishing temperature, $T \to 0$, this replacement becomes exact; (ii) In the case where phonon-assisted transitions are not in resonance with direct optical transitions, this choice ensures that Eq. (48) reduces to the CHBB theory.

To assess the validity of the above replacement, we perform an explicit canonical average over the initial states

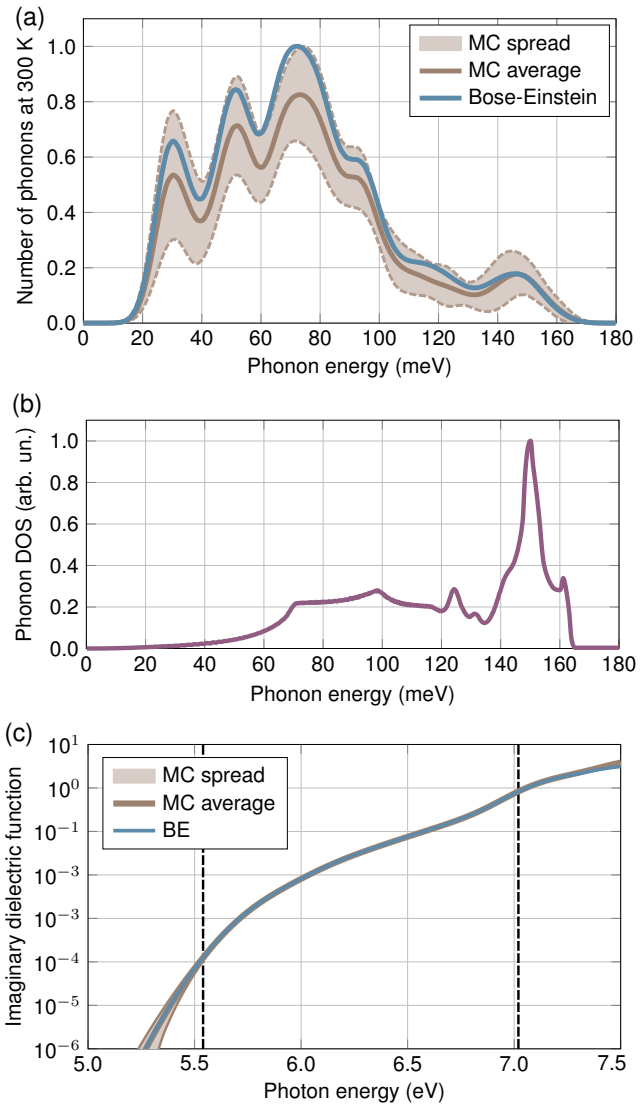

FIG. 6. **Convergence tests: Phonon occupations**. Test of the "average phonon approximation" in the evaluation of Eq. (48). (a) Number of phonons per unit energy in diamond at 300 K, as obtained from rejection Monte Carlo sampling of 10,000 configurations: average (brown line), and standard deviation (shaded brown area bounded by dashed line). For comparison, we also show the product of the phonon DOS and the thermal Bose-Einstein occupations (blue). It is seen that the two curves match closely throughout the entire phonon energy range. The peaks in these functions correspond to the peaks in the phonon DOS, as shown in panel (b). (c) Imaginary part of the dielectric function of diamond at 300 K, as obtained from Eq. (48) using explicit Monte Carlo sampling (20 configurations, brown line and shading), and the average-phonon approximation (blue). The Monte Carlo average and the average-phonon approximation are indistinguishable. The direct and indirect gaps are indicated by vertical lines.

$|i_0\rangle$ for diamond as a test case, by employing a rejection Monte Carlo sampling algorithm [51, 52]. To this end, we proceed as follows:

(i) We generate random non-negative integers $n_{\mathbf{q}\nu}$ for all phonon modes $\mathbf{q}\nu$;

(ii) We evaluate the probability of this configuration using the Gibbs distribution $\exp(-\beta E)$, where $E = \sum_{\mathbf{q}\nu} n_{\mathbf{q}\nu}\hbar\omega_{\mathbf{q}\nu}$ is the thermal energy;

(iii) We draw a random number r from a uniform distribution in the interval $[0, 1)$;

(iv) If $r \leq \exp(-\beta\, E)$, we accept the configuration and evaluate the dielectric function $\varepsilon_2(\omega)$ using Eq. (48); otherwise we go back to step (i);

Using this method, we generate multiple initial many-body states $|i_0\rangle$ which automatically satisfy the Gibbs distribution, and calculate the imaginary dielectric function for each configuration. At the end of this process, we take the average of ε_2 over all accepted configurations.

Figure 6 shows the result of this procedure for the dielectric function $\varepsilon_2(\omega)$ of diamond. Panel (a) shows the spectral distribution of phonon occupations $n_{\mathbf{q}\nu}$ at $T = 300$ K, as obtained by generating 10,000 configurations using the above Monte Carlo algorithm. In the limit of dense sampling, this distribution should converge to the product of the DOS and the Bose-Einstein distribution, $\mathrm{DOS}(\hbar\omega) \times n(\hbar\omega)$, where $n(\hbar\omega) = [\exp(\hbar\omega/k_{\mathrm{B}}T) - 1]^{-1}$. Here we see that the Monte Carlo sampling leads to a distribution that follows very closely the expected behavior. The peak structure of this distribution is inherited from the phonon DOS, which is shown in panel (b) for comparison. In Fig. 6(c), we show the imaginary part of the dielectric function of diamond at $T = 300$ K, as obtained using 20 Monte Carlo samples. The black like indicates the average of the distribution, and the shaded area indicates the standard deviation. For comparison, we show in the same panel the dielectric function obtained via a single calculation with the phonon occupations set to the Bose-Einstein distribution, Eq. (20). Clearly, the two results are in very good agreement, thereby validating our choice.

All calculations discussed in the following are based on the "average phonon approximation" to Eqs. (41)-(44) and Eq. (48), as given by Eq. (20). Whenever very high accuracy is necessary, we recommend using the rejection sampling Monte Carlo method described above or other similar stochastic methods.

IV. APPLICATIONS

In this section, we discuss the optical absorption spectrum and the near-edge absorption fine structure of Si (Sec. IV A), and the absorption and photoluminescence spectra of Ge (Sec. IV B). To keep the presentation compact, we present the corresponding discussions for GaAs and diamond in Appendices A and B, respectively.

A. Optical absorption spectrum of Si

In Fig. 7 we report our results for the optical absorption spectrum and the imaginary part of the dielectric function of Si as the prototypical indirect-gap semiconductor.

Figure 7(a) shows the standard Si band structure, for reference. The maximum of the valence band is at the Γ point, and the minimum of the conduction band is near the X point. Upon applying the scissor correction discussed in Sec. III A, we have a fundamental indirect gap $E_{\mathrm{g}}^{\mathrm{ind}} = 1.12$ eV [36] and a direct gap $E_{\mathrm{g}}^{\mathrm{dir}} = 3.3$ eV.

Figure 7(b) shows the absorption coefficient computed using Eq. (48). The calculations are seen to be in close agreement with experimental data [36] (circles) over a wide energy range spanning several electronvolts, and capture both the regime of indirect phonon-assisted transitions (between $E_{\mathrm{g}}^{\mathrm{ind}}$ and $E_{\mathrm{g}}^{\mathrm{dir}}$) and that of direct processes (above $E_{\mathrm{g}}^{\mathrm{dir}}$). This behavior should be contrasted with the textbook CHBB theory, which diverges at the onset of direct transitions, as shown in Fig. 1(a). Importantly, the present calculations do *not* include any artificial broadening of the energy denominators in Eq. (48).

In Fig. 7(c), we quantify the importance of phonon-assisted transitions by separating the contributions of direct and indirect processes to the dielectric function ε_2. To identify these contributions in Eq. (48), we eliminate in turn terms proportional to $U^*_{p,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}}}$ and $U^*_{p,i_0-1_{v\mathbf{k}}+1_{c\mathbf{k}+\mathbf{q}}+\eta 1_{-\eta\mathbf{q}\nu}}$, respectively. We emphasize that this decomposition is only approximate insofar it neglects interference terms when taking the square modulus in Eq. (48), and should only be regarded as a semi-quantitative analysis tool. As expected, indirect transitions dominate in the energy range between $E_{\mathrm{g}}^{\mathrm{ind}}$ and $E_{\mathrm{g}}^{\mathrm{dir}}$, and direct processes dominate for energies above $E_{\mathrm{g}}^{\mathrm{dir}}$. In contrast, what we were not expecting is that indirect processes remain important at high photon energies. For example, at photon energies near $\hbar\omega = 4$ eV, the contribution to the dielectric function from indirect processes represents 10% of the total.

In Fig. 7(d) we zoom near the optical absorption onset, and compare with high-resolution experimental data [53]. To facilitate comparison at various temperatures, we use additional temperature-dependent scissor shifts as indicated in the figure caption. The same shifts could be obtained from first principles using the Allen-Heine theory [54–59], but we leave this additional refinement to future work. Near the absorption onset, we expect the function $\alpha^{1/2}$ to exhibit a piece-wise linear behavior as a function of the photon energy, with two absorption edges [10, 11]. The first edge corresponds to the onset of phonon absorption processes at the energy $E_{\mathrm{g}}^{\mathrm{ind}} - \hbar\omega_{\mathrm{ph}}$, where $\hbar\omega_{\mathrm{ph}} \simeq 62$ meV is the characteristic phonon frequency; the second edge is at the onset of phonon emission processes at the energy $E_{\mathrm{g}}^{\mathrm{ind}} + \hbar\omega_{\mathrm{ph}}$. As seen in panel (d), our calculations correctly reproduce this piecewise linear behavior, and are in extremely good quanti-

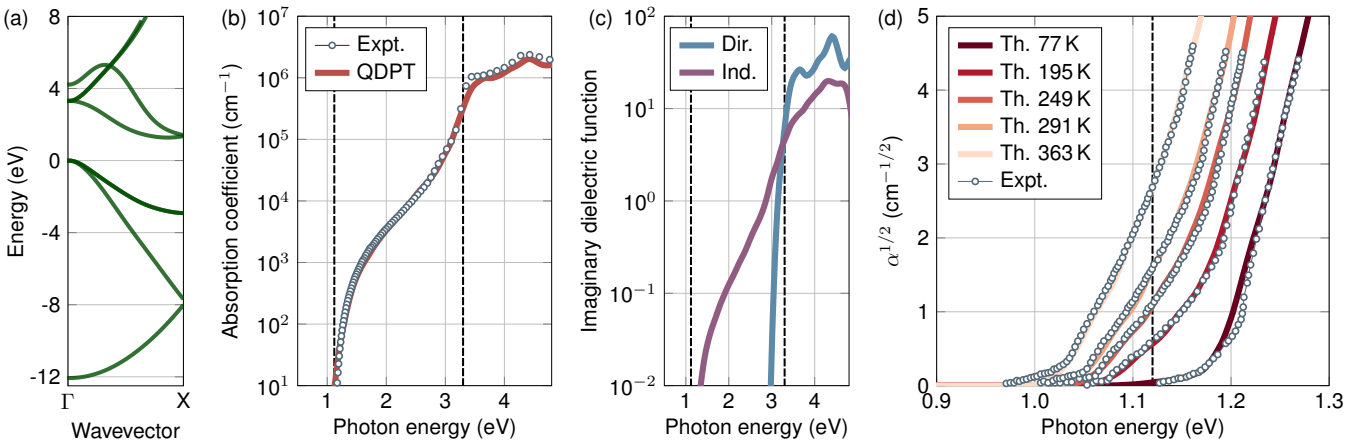

FIG. 7. **Full-range absorption spectrum and low-energy absorption fine structure in Si**. (a) Band structure of Si, with the energy referred to the valence band top. (b) Optical absorption coefficient of Si at 300 K including both direct and phonon-assisted transitions, as calculated using Eq. (48) (red). Black circles are experimental data from Ref. [36]. The gaps are indicated by vertical lines. (c) Decomposition of the imaginary part of the dielectric function of Si into contributions from direct (cyan) and phonon-assisted (purple) processes. (d) Near-edge fine structure of the optical absorption spectrum of Si, for the temperatures 77 K, 195 K, 249 K, 291 K, and 363 K (shades of red). Experimental values from Ref. 53 are shown as black circles. In these curves, we employ temperature-dependent scissor shifts of 72 meV, 52 meV, 35 meV, 29 meV, and 5 meV at the temperatures 77 K, 195 K, 249 K, 291 K, and 363 K, respectively.

tative agreement with experimental data. We also note that the present generalized theory and implementation delivers improved agreement with experiments in the near-edge fine structure, as compared to earlier *ab initio* calculations based on the CHBB theory [35].

B. Optical absorption and luminescence spectra of Ge

Ge is the prototypical quasi-direct semiconductor, since the energy separation between the direct and indirect band gaps is comparable to phonon energy scales. Figure 8(a) shows that Ge exhibits an indirect fundamental gap $E_{\rm g}^{\rm ind} = 0.66$ eV at room temperature, between the conduction band minimum at L and the valence band maximum at Γ. The minimum direct gap is $E_{\rm g}^{\rm dir} = 0.80$ eV at the Γ point. The difference between direct and indirect gaps is of only 141 meV [61]. For comparison, the highest-frequency phonon in Ge is the optical zone-center mode with energy $\hbar\omega = 35$ meV.

Figure 8(b) compares our calculated absorption coefficient to the experimental data from Ref. [60] at 300 K, from the absorption edge up to a photon energy of 3 eV. We see that the agreement between our calculations and experiment is excellent throughout the entire spectral range. We emphasize that our calculations do not employ any artificial broadening, and are well-behaved at the onset of direct transitions unlike the CHBB approach. In contrast, calculations based on the CHBB theory diverge in this range, similarly to the case of Si shown in Fig. 1(a). In Figs. 8(c),(d) we separate the contributions to the absorption coefficient resulting from from direct or phonon-assisted transitions; these processes dominate near the indirect gap [panel (c)], and remain sizable throughout the entire spectral range [panel (d)].

In Fig. 8(e)-(h) we compare the luminescence recombination rate of Ge, as calculated using Eqs. (2) and (48), and the experimental data from Ref. 27, for a range of temperatures. The impressive agreement between our calculations and experiments should be contrasted with the failure of prior theory shown in Fig. 1(b). Our present approach reproduces very closely the relative intensity of the measured peaks, as well as their temperature dependence.

Also in Fig. 8(e)-(h), we indicate the contributions to the luminescence intensity from direct band-to-band recombination (gray line and shading). We see that direct processes are responsible for the peak marked as P_2 in panels (e) and (f). The most significant contribution to the luminescence intensity results from phonon-assisted light emission, which is responsible for the large peak marked as P_1; this process occurs through an indirect transition from L to Γ mediated by zone edge optical phonons, followed by photon-assisted recombination at Γ, as schematically shown in Fig. 8(a).

At 295 K, both peaks P_1 and P_2 are clearly visible in both experimental and theoretical data, while at the lowest temperature of 75 K peak P_2 is completely suppressed. This temperature dependence can be rationalized using the van Roosbroek-Shockley relation in Eq. (2). In fact, the photon distribution function $\omega^3/[\exp(\hbar\omega/k_{\rm B}T) - 1]$ decreases to its half-maximum at $\hbar\omega \simeq 6k_{\rm B}T$, which corresponds to 153 meV at 295 K and 40 meV at 75 K, respectively. Therefore, at 295 K the cutoff energy set by the photon distribution is compa-

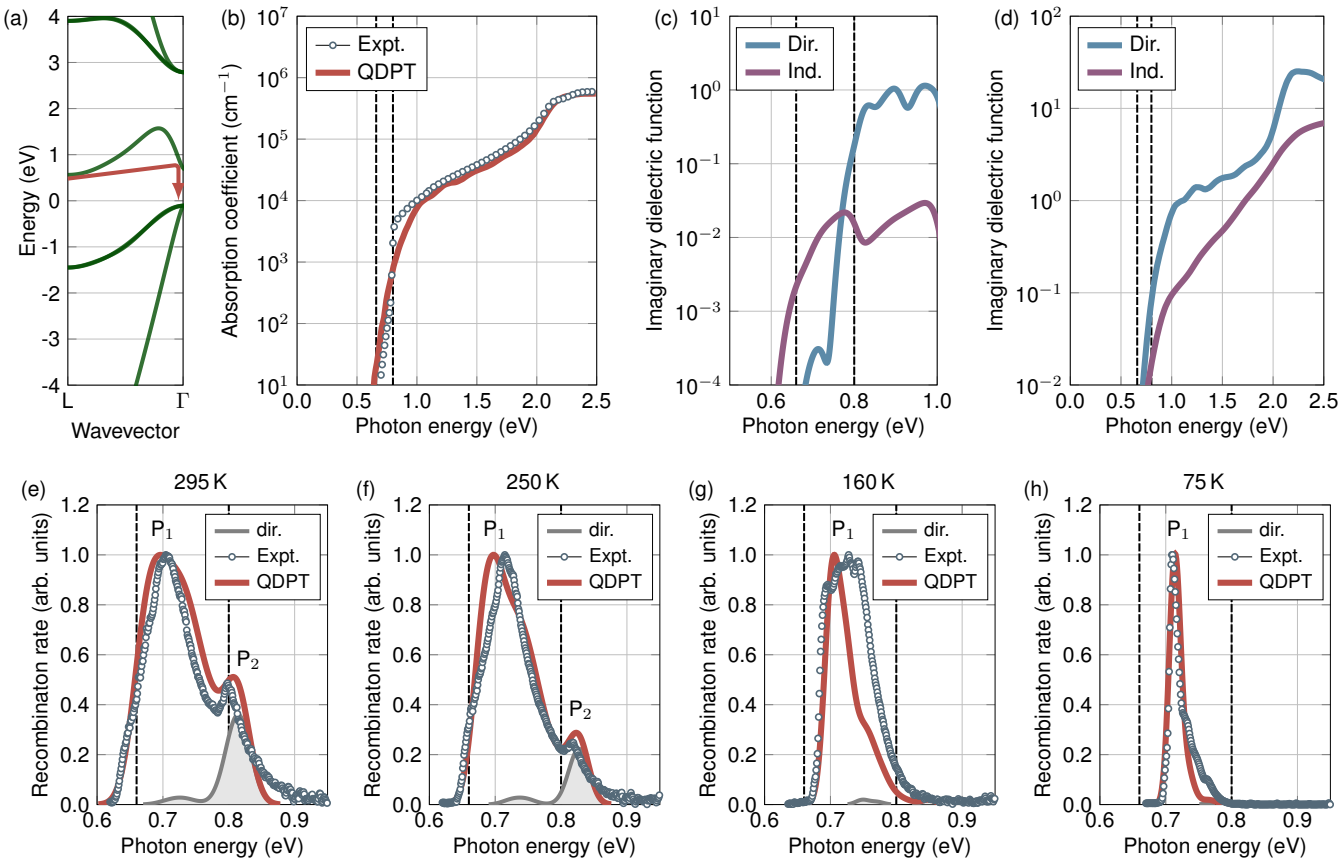

FIG. 8. **Phonon-assisted optical absorption and luminescence in Ge**. (a) Band structure of Ge, with the energy referred to the valence band top. The arrow indicates the phonon-assisted transition responsible for Peak P_1 in panels (e)-(h). (b) Optical absorption coefficient of Ge calculated using Eq. (48) (red), compared to experimental data from Ref. 60 (black circles). The indirect and direct gaps are indicated by vertical lines. (c) Decomposition of the imaginary part of the dielectric function of Ge into contributions from direct (cyan) and phonon-assisted (purple) processes, near the absorption onset. (d) Same as (c), but for an extended photon energy range. (e)-(h) Photoluminescence recombination rate of Ge, for the temperatures 295 K, 250 K, 160 K, and 75 K, respectively. Red lines are calculations using Eq. (48), experimental values from Ref. 27 are shown as black circles. We also show the contributions from direct recombination processes as gray lines and shaded areas. In panels (e)-(h) we employ temperature-dependent scissor shifts of 330 meV, 340 meV, 350 meV, and 325 meV, respectively.

rable to the separation between direct and indirect gaps (141 meV), and both peaks are visible; conversely, at low temperature the cutoff energy is much smaller than the separation between the gaps, hence peak P_2 is suppressed by the Bose-Einstein function.

V. CONCLUSIONS AND OUTLOOK

In this manuscript, we develop an *ab initio* theory of optical absorption and luminescence in semiconductors and insulators that unifies calculations of direct and indirect phonon-assisted optical transitions within a single conceptual framework.

The present work addresses a long-standing difficulty in the textbook CHBB theory of phonon-assisted optical processes, whereby the oscillator strength diverges at the onset of direct transitions. This fundamental limitation of the standard theory is a significant obstacle toward systematic and reliable *ab initio* calculations of the optical spectra of materials. For example, 70% of compounds in the Materials Project database are classified as indirect and would not be described correctly by the textbook CHBB framework. Here, we recognize that, for photon energies beyond the onset of direct transitions, zero-momentum electron-hole excitations and finite-momentum electron-hole-phonon excitations are resonant, and their coupling must be taken into account before applying perturbation theory. To this end, we develop a quasi-degenerate many-body perturbation theory framework in the occupation-number representation, and we obtain an expression for the imaginary part of the dielectric function that is free of the small-denominator problem, does not depend on arbitrary broadening parameters, and reduces to previous approaches in the limits where direct and phonon-assisted processes are not in resonance.

The main outcome of this development is that it is now

possible to compute complete optical absorption spectra by avoiding the artificial partitioning of the photon energy range into "direct" and an "indirect" regions, by removing the sensitivity of the results to the artificial broadening parameters, and by avoiding supercell calculations. We demonstrate this approach by performing *ab initio* calculations of the optical absorption spectra of prototypical semiconductors, such as Si, Ge, GaAs, and diamond, as well as the luminescence spectrum of Ge. In all cases we find excellent agreement with experiments, demonstrating the predictive power of the present theory.

It is natural to ask whether alternative approaches could deliver similar accuracy as the present method. An obvious choice would be to set up the problem in the language of many-body Green's functions [62] and use Feynman-Dyson perturbation theory. This approach was pursued in Ref. 63, 64 and requires expressing the electronic polarization propagator in terms of phonon-renormalized Green's functions; the renormalization is then achieved via the standard Fan-Migdal self-energy [34]. While this approach is formally exact, its usefulness is limited by the fact that in practical applications the fully interacting Green's function must be approximated by the non-interacting propagator. Under this approximation, the formalism reduces to the textbook CHHB theory [63] and suffers from the same limitations.

There are several opportunities for generalizing the present theory to even more accurate calculations. First, the spectra presented in this work could be computed using GW quasiparticle band structures [62, 65] and even GW perturbation theory for the electron-phonon matrix elements [66]. Second, the temperature-dependent scissor shifts employed here for the absorption and emission fine structures of Si and Ge could be computed directly from first principles using the Allen-Heine theory [54–59]. Third, the broadening of quasiparticle levels could be incorporated in our expressions via calculations of the Fan-Migdal self-energy [34]. Fourth, the equations could be generalized to incorporate excitonic effects within the Bethe-Salpeter framework [67–69]. All of these improvements are possible and within reach, and will be pursued in the future.

To ensure reproducibility and enable further work in this direction, we have made the present methodology publicly available in the `EPW v5.8` release [30].

We hope that this development will serve as a starting point for more systematic and more predictive calculations of the optical properties of solids including electron-phonon interactions.

ACKNOWLEDGMENTS

This work was supported by the U.S. National Science Foundation, DMREF Grant No. 2119555 (theory, implementation, calculations). Additional NSF support was provided by CSSI Grant No. 2103991 (Jupyter notebooks), and CSA Grant No. 2139536 (preparation for LCCF). Computational resources were provided by the Texas Advanced Computing Center (TACC) at The University of Texas at Austin, and the Argonne Leadership Computing Facility (a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357). This research was also supported by the Computational Materials Sciences Program funded by the US Department of Energy, Office of Science, Basic Energy Sciences, under award no. DE-SC0020129 (code refactoring for exascale). This research used resources of the National Energy Research Scientific Computing Center and the Argonne Leadership Computing Facility, which are DOE Office of Science User Facilities supported by the Office of Science of the U.S. Department of Energy, under Contracts No. DE-AC02-05CH11231 and DE-AC02-06CH11357, respectively.

Appendix A: Optical absorption spectrum of GaAs

To test the validity of the present approach in the case of direct-gap semiconductors, we consider GaAs as a prototypical example. In Fig. 9 we summarize our main results for the optical absorption spectrum and the imaginary part of the dielectric function of this compound.

Figure 9(a) shows the standard band structure of GaAs, with both the valence band maximum and conduction band minimum at the Γ point. The (scissor-corrected) direct gap is $E_{\rm g}^{\rm dir}$ =1.42 eV [71]. In Fig 9(b) we compare the optical absorption spectrum calculated using Eq. (48) and experiments [70]. It is seen that the present results, which *include* phonon-assisted transitions, are in close agreement with experimental data throughout the entire photon energy range. As expected, our formalism does not exhibit any pathological behavior when direct and phonon-assisted processes are in resonance.

In Fig. 9(c) we decompose the imaginary part of the dielectric function into contributions from direct transitions and phonon-assisted transitions. Near the absorption onset, phonon-assisted processes provide a negligible contribution to the oscillator strength. Conversely, at high photon energy, indirect processes provide a sizable enhancement of the total dielectric function; for example, for photons of energy $\hbar\omega = 2$ eV, the contribution from indirect processes is $\approx 8\%$ of the total. This finding indicates that, even in direct-gap semiconductors, inclusion of electron-phonon couplings may be important to achieve high accuracy. In the same panel we also see that the indirect contribution tends to increase with photon energy; we ascribe this process to the increase of phase space for phonon-assisted transitions as the joint electronic density of states increases away from the band edges.

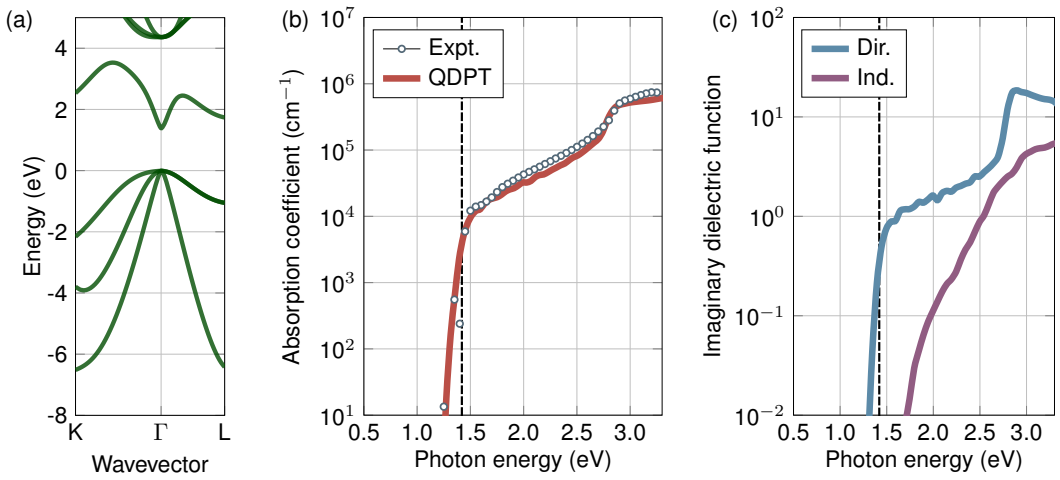

FIG. 9. **Contribution of phonon-assisted transitions to the absorption spectrum of GaAs**. (a) Band structure of GaAs. (b) Optical absorption spectrum of GaAs, including both direct and phonon-assisted transitions, as calculated using Eq. (48) (red). Black circles are experimental data from Ref. 70. The fundamental direct gap is indicated by a vertical dashed line. (c) Decomposition of the imaginary part of the dielectric function of GaAs into contributions from direct (cyan) and phonon-assisted (purple) processes.

Appendix B: Optical absorption spectrum of diamond

As a further test of the methodology developed in this work, we examine diamond as the prototypical example of indirect-gap semiconductor with strong electron-phonon couplings [73, 74].

Figure 10(a) shows the band structure of diamond for reference. The top of the valence band is at Γ, and the bottom of the conduction band is near the X point. The scissor-corrected fundamental indirect gap and direct gap are $E_{\rm g}^{\rm ind} = 5.54$ eV and $E_{\rm g}^{\rm dir} = 7.02$ eV, respectively.

In Fig. 10(b) we compare our calculations of the optical absorption coefficient of diamond, based on Eq. (48), to experiments [72]. We see that the agreement is very good both at the indirect onset and at the direct onset, which are marked by vertical dashed lines. It is also seen that our calculations overestimate the measured absorption coefficient around a photon energy of 6.5 eV. This effect was already reported in Ref. 18 and relates to the lack of GW corrections in out calculated band structures.

Figure 10(c) shows a decomposition of the imaginary part of the dielectric function into direct and phonon-assisted processes. We see that phonon-assisted processes tend to provide a contribution that becomes comparable to that of direct transitions at high photon energy, as already discussed above for Si and GaAs. For example, at the photon energy $\hbar\omega = 7.5\,$eV, the contribution from phonon-assisted processes to the oscillator strength represents $\approx 10\%$ of the total.

Appendix C: Alternative approaches

In Sec. II D we introduced a quasidegenerate many-body perturbation theory approach to deal with the resonance between intermediate and final states in indirect absorption processes. It is natural to ask whether this is the only way to avoid singularities such as those in Fig. 1(a), or there are alternative avenues worth investigating.

Generally speaking, it is expected that that account of higher-order processes leading to the formation of new quasiparticles will contribute to reducing the degree of degeneracy in the excited-state manifold. This is precisely what we achieve in Eq. (26) by taking into account the coupling of electron-hole pairs with phonons to produce coherent superpositions of electron-hole-phonon triples with renormalized excitation energies.

Similar to coupled electron-hole-phonon excitations, consideration of exciton-polaritons could be important to capture the fine structure of absorption and luminescence spectra of semiconductors [10]. However, these effects are still beyond the reach of *ab initio* methods, because the energy scales at play are in the 1-5 meV range (see, e.g., Refs. 10, 75), which is much smaller than the energy scales considered in this work and would require specialized interpolation techniques for the Bethe-Salpeter equation to obtain accurate exciton dispersion relations near critical points. Progress in this direction was made very recently [76], raising the hope that it will be possible to also incorporate exciton-polariton effects in the near future.

Whether or not consideration of more complex quasiparticles lifts degeneracies, the problem of near-degeneracies remains. Quasidegenerate states pose a significant numerical challenge since they require a pro-

,

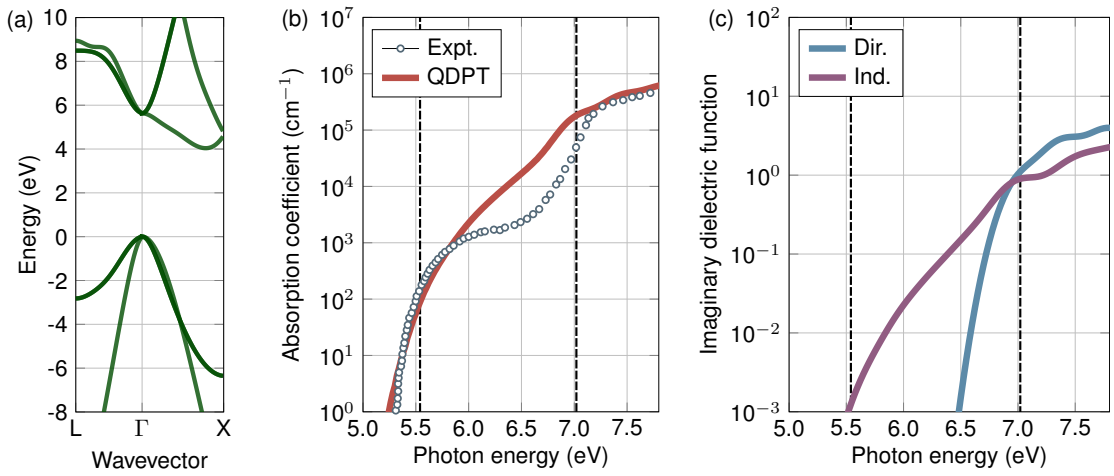

FIG. 10. **Full-range optical absorption spectrum of diamond**. (a) Band structure of diamond. (b) Optical absorption spectrum of diamond, including both direct and phonon-assisted transitions, as calculated using Eq. (48) (red). Black circles are experimental data from Ref. 72. The fundamental indirect gap and the direct gap are indicated by vertical dashed lines. (c) Decomposition of the imaginary part of the dielectric function of diamond into contributions from direct (cyan) and phonon-assisted (purple) processes.

hibitively fine sampling of the Brillouin zone to evaluate the integrals in principal value. Our present approach resolves this issue from the outset.

[1] A. Frisk Kockum, A. Miranowicz, S. De Liberato, S. Savasta, and F. Nori, Nat. Rev. Phys. **1**, 19 (2019).
[2] N. Rivera and I. Kaminer, Nat. Rev. Phys. **2**, 538 (2020).
[3] R. Gutzler, M. Garg, C. R. Ast, K. Kuhnke, and K. Kern, Nat. Rev. Phys. **3**, 441 (2021).
[4] M. H. Naik, E. C. Regan, Z. Zhang, Y.-H. Chan, Z. Li, D. Wang, Y. Yoon, C. S. Ong, W. Zhao, S. Zhao, M. I. B. Utama, B. Gao, X. Wei, M. Sayyad, K. Yumigeta, K. Watanabe, T. Taniguchi, S. Tongay, F. H. da Jornada, F. Wang, and S. G. Louie, Nature **609**, 52 (2022).
[5] D. Y. Qiu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **111**, 216805 (2013).
[6] L. Guo, M. Wu, T. Cao, D. M. Monahan, Y.-H. Lee, S. G. Louie, and G. R. Fleming, Nat. Phys. **15**, 228 (2019).
[7] Z. Ye, T. Cao, K. O'Brien, H. Zhu, X. Yin, Y. Wang, S. G. Louie, and X. Zhang, Nature **513**, 214 (2014).
[8] M. M. Ugeda, A. J. Bradley, S.-F. Shi, F. H. da Jornada, Y. Zhang, D. Y. Qiu, W. Ruan, S.-K. Mo, Z. Hussain, Z.-X. Shen, F. Wang, S. G. Louie, and M. F. Crommie, Nat. Mater. **13**, 1091 (2014).
[9] X.-X. Zhang, T. Cao, Z. Lu, Y.-C. Lin, F. Zhang, Y. Wang, Z. Li, J. C. Hone, J. A. Robinson, D. Smirnov, S. G. Louie, and T. F. Heinz, Nat. Nanotechnol. **12**, 883 (2017).
[10] P. Y. Yu and M. Cardona, *Fundamentals of semiconductors: physics and materials properties* (Springer Science & Business Media, 2010).
[11] G. Grosso and G. P. Parravicini, *Solid state physics* (Academic press, 2013).
[12] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 10.1063/1.4812323 (2013), 011002.
[13] J. Noffsinger, E. Kioupakis, C. G. Van de Walle, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **108**, 167402 (2012).
[14] M. Zacharias, C. E. Patrick, and F. Giustino, Phys. Rev. Lett. **115**, 177401 (2015).
[15] I. C. Cheeseman, Proc. Phys. Soc. A **65**, 25 (1952).
[16] L. H. Hall, J. Bardeen, and F. J. Blatt, Phys. Rev. **95**, 559 (1954).
[17] F. Giustino, M. L. Cohen, and S. G. Louie, Phys. Rev. B **76**, 165108 (2007).
[18] M. Zacharias and F. Giustino, Phys. Rev. B **94**, 075125 (2016).
[19] M. Zacharias and F. Giustino, Phys. Rev. Res. **2**, 013357 (2020).
[20] B. Monserrat, C. E. Dreyer, and K. M. Rabe, Phys. Rev. B **97**, 104310 (2018).
[21] T. A. Huang, M. Zacharias, D. K. Lewis, F. Giustino, and S. Sharifzadeh, J. Phys. Chem. Lett. **12**, 3802 (2021).
[22] X. Zhang and E. Kioupakis, Phys. Rev. B **107**, 115207 (2023).
[23] Y. Kang, Y. Youn, S. Han, J. Park, and C.-S. Oh, Chem. Mater. **31**, 4072 (2019).
[24] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.-Y. Chim, G. Galli, and F. Wang, Nano Lett. **10**, 1271 (2010).
[25] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz, Phys. Rev. Lett. **105**, 136805 (2010).
[26] J. Kangsabanik, M. K. Svendsen, A. Taghizadeh, A. Crovetto, and K. S. Thygesen, J. Am. Chem. Soc.

144, 19872 (2022).
[27] J. Menéndez, C. D. Poweleit, and S. E. Tilton, Phys. Rev. B **101**, 195204 (2020).
[28] J. Menéndez, M. Noël, J. C. Zwinkels, and D. J. Lockwood, Phys. Rev. B **96**, 121201 (2017).
[29] J. Menéndez, D. J. Lockwood, J. C. Zwinkels, and M. Noël, Phys. Rev. B **98**, 165207 (2018).
[30] H. Lee, S. Poncé, K. Bushick, S. Hajinazar, J. Lafuente-Bartolome, J. Leveillee, C. Lian, J.-M. Lihm, F. Macheda, H. Mori, H. Paudyal, W. H. Sio, S. Tiwari, M. Zacharias, X. Zhang, N. Bonini, E. Kioupakis, E. R. Margine, and F. Giustino, npj Comput. Mater. **9**, 156 (2023).
[31] F. Giustino, *Materials modelling using density functional theory: properties and predictions* (Oxford University Press, 2014).
[32] W. van Roosbroeck and W. Shockley, Phys. Rev. **94**, 1558 (1954).
[33] X. Wang, J. R. Yates, I. Souza, and D. Vanderbilt, Phys. Rev. B **74**, 195118 (2006).
[34] F. Giustino, Rev. Mod. Phys. **89**, 015003 (2017).
[35] J. Noffsinger, F. Giustino, B. D. Malone, C.-H. Park, S. G. Louie, and M. L. Cohen, Comput. Phys. Comm. **181**, 2140 (2010).
[36] R. Braunstein, A. R. Moore, and F. Herman, Phys. Rev. **109**, 695 (1958).
[37] M. Schlipf and F. Giustino, Phys. Rev. Lett. **127**, 237601 (2021).
[38] C. Kittel and C.-y. Fong, *Quantum theory of solids* (Wiley, 1987).
[39] I. Shavitt and R. J. Bartlett, *Many-body methods in chemistry and physics: MBPT and coupled-cluster theory* (Cambridge university press, 2009).
[40] I. Lindgren, J. Phys. B **7**, 2441 (1974).
[41] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001).
[42] X. Gonze and C. Lee, Phys. Rev. B **55**, 10355 (1997).
[43] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. D. Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio, A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. O. de-la Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni, J. Phys.: Condens. Matter **29**, 465901 (2017).
[44] S. Poncé, E. Margine, C. Verdi, and F. Giustino, Comput. Phys. Commun. **209**, 116 (2016).
[45] G. Pizzi, V. Vitale, R. Arita, S. Blügel, F. Freimuth, G. Géranton, M. Gibertini, D. Gresch, C. Johnson, T. Koretsune, J. Ibañez-Azpiroz, H. Lee, J.-M. Lihm, D. Marchand, A. Marrazzo, Y. Mokrousov, J. I. Mustafa, Y. Nohara, Y. Nomura, L. Paulatto, S. Poncé, T. Ponweiser, J. Qiao, F. Thöle, S. S. Tsirkin, M. Wierzbowska, N. Marzari, D. Vanderbilt, I. Souza, A. A. Mostofi, and J. R. Yates, J. Phys.: Condens. Matter **32**, 165902 (2020).
[46] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[47] D. M. Ceperley and B. J. Alder, Phys. Rev. Lett. **45**, 566 (1980).
[48] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
[49] X. Gonze, R. Stumpf, and M. Scheffler, Phys. Rev. B **44**, 8503 (1991).
[50] A. F. Starace, Phys. Rev. A **3**, 1242 (1971).
[51] R. E. Caflisch, Acta Numer **7**, 1–49 (1998).
[52] D. J. C. Mackay, NATO Adv Sci I D-BEH , 175 (1998).
[53] G. G. Macfarlane, T. P. McLean, J. E. Quarrington, and V. Roberts, Phys. Rev. **111**, 1245 (1958).
[54] P. B. Allen and V. Heine, J. Phys. C **9**, 2305 (1976).
[55] A. Marini, Phys. Rev. Lett. **101**, 106405 (2008).
[56] F. Giustino, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **105**, 265501 (2010).
[57] G. Antonius, S. Poncé, P. Boulanger, M. Côté, and X. Gonze, Phys. Rev. Lett. **112**, 215501 (2014).
[58] S. Poncé, Y. Gillet, J. Laflamme Janssen, A. Marini, M. Verstraete, and X. Gonze, J. Chem. Phys. **143**, 102813 (2015).
[59] J.-M. Lihm and C.-H. Park, Phys. Rev. X **11**, 041053 (2021).
[60] T. N. Nunley, N. S. Fernando, N. Samarasingha, J. M. Moya, C. M. Nelson, A. A. Medina, and S. Zollner, J. Vac. Sci. Technol. B **34** (2016).
[61] O. Madelung, W. Von der Osten, and U. Rössler, *Intrinsic Properties of Group IV Elements and III-V, II-VI and I-VII Compounds/Intrinsische Eigenschaften Von Elementen Der IV. Gruppe und Von III-V-, II-VI-und I-VII-Verbindungen*, Vol. 22 (Springer Science & Business Media, 1986).
[62] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[63] W. S. Chow, Phys. Rev. **185**, 1056 (1969).
[64] W. S. Chow, Phys. Rev. **185**, 1062 (1969).
[65] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[66] Z. Li, G. Antonius, M. Wu, F. H. da Jornada, and S. G. Louie, Phys. Rev. Lett. **122**, 186402 (2019).
[67] G. Strinati, Phys. Rev. B **29**, 5718 (1984).
[68] M. Rohlfing and S. G. Louie, Phys. Rev. Lett. **80**, 3320 (1998).
[69] M. Rohlfing and S. G. Louie, Phys. Rev. B **62**, 4927 (2000).
[70] K. Papatryfonos, T. Angelova, A. Brimont, B. Reid, S. Guldin, P. R. Smith, M. Tang, K. Li, A. J. Seeds, H. Liu, and D. R. Selviah, AIP Adv. **11**, 025327 (2021).
[71] J. S. Blakemore, J. Appl. Phys. **53**, R123 (1982).
[72] H. R. Phillip and E. A. Taft, Phys. Rev. **136**, A1445 (1964).
[73] B. Monserrat and R. J. Needs, Phys. Rev. B **89**, 214304 (2014).
[74] M. Cardona and M. L. W. Thewalt, Rev. Mod. Phys. **77**, 1173 (2005).
[75] H. M. Gibbs, G. Khitrova, and S. W. Koch, Nat. Photon. **5**, 273 (2011).
[76] J. B. Haber, D. Y. Qiu, F. H. da Jornada, and J. B. Neaton, Phys. Rev. B **108**, 125118 (2023).