PHYSICAL REVIEW B **92**, 235432 (2015)

Intraband effects in excitonic second-harmonic generation

Thomas Garm Pedersen*
Department of Physics and Nanotechnology, Aalborg University, DK-9220 Aalborg Øst, Denmark
and Center for Nanostructured Graphene (CNG), DK-9220 Aalborg Øst, Denmark

(Received 23 September 2015; revised manuscript received 20 November 2015; published 17 December 2015)

A theory for the nonlinear excitonic optical response of semiconductors is developed. By adopting the length gauge, intraband effects are rigorously taken into account. We show that the second-order nonlinear response mixing intra- and interband transitions can be expressed in terms of generalized derivatives of the exciton Green's function. The theory is applied to hexagonal boron-nitride monolayers. For both the linear and nonlinear response, a dramatic influence of excitons is found. Hence, new discrete resonances appear as well as pronounced changes in the continuum spectrum.

DOI: 10.1103/PhysRevB.92.235432 PACS number(s): 42.65.−k, 71.35.Cc, 72.20.Ht

I. INTRODUCTION

It is by now very well documented that electron-hole interactions, i.e., excitonic effects, play an important role in the linear optical properties of semiconductors [1]. In particular, low-dimensional semiconductors are sensitive to excitons because of reduced screening and increased electron-hole overlap as a consequence of confinement. Prominent examples include carbon nanotubes [2] and transition-metal dichalcogenides [3], in which exciton binding energies can amount to a significant part of the band gap. In the linear optical response, the spectral modifications reveal themselves both as discrete exciton resonances in the band gap and as a renormalization of the continuous part of the spectrum. Much less is understood about the nonlinear optical properties, although several intriguing experimental indications of excitons have been found [4–6]. A few early theoretical publications have considered the influence of electron-hole interactions on the nonlinear response of bulk semiconductors such as GaAs [7–9]. Recently, the demonstration of efficient second-harmonic generation in two-dimensional materials has led to a renewed interest in this problem [10,11]. Intuitively, modifications similar to the linear case are expected for low-order nonlinear processes such as second-harmonic generation [9]. In fact, in Ref. [10], the second-order spectra were partly explained in terms of the linear response at the fundamental frequency ω and the replica at $\omega/2$. However, notable differences exist between linear and nonlinear response. Primarily, the selection rules of nonlinearities are highly sensitive to symmetry with all even order responses being forbidden in the dipole approximation for centrosymmetric structures [12]. Even more subtle differences are found, though. For a cold, clean (intrinsic) semiconductor the linear response is entirely due to interband transitions. At second and higher order, mixed intra- and interband terms will appear in addition to pure interband transitions [13–15]. The prototypical example of such mixed terms is found for two-band semiconductors. As the second-order response involves three consecutive transitions, at least one must be of intraband type when only two bands are available. More generally, excitons couple correlated intra- and interband transitions.

In the present work, a model of the second-order nonlinear excitonic optical response is developed and applied to the two-dimensional two-band semiconductor hexagonal boron nitride (h-BN). In the monolayer form, the two π bands of this material constitute the highest valence and lowest conduction band, respectively. The odd parity of the π states with respect to the crystal plane ensures complete decoupling from the even parity σ states. Hence, to a good approximation the in-plane optical response in the vicinity of the band gap is attributed to the two π bands. This greatly simplifies the computation of the linear excitonic results. We demonstrate here that both linear and nonlinear excitonic responses follow from a general density matrix approach. We have previously applied a similar model to the excitonic second-harmonic response of MoS_2 [11], however, taking only purely interband contributions into account. In the present work, we extend the model to mixed intra- and interband terms. We follow Aversa and Sipe [14] and work in the length gauge in order to manifestly eliminate unphysical low-frequency divergences. Applied to h-BN sheets, we demonstrate that dramatic excitonic effects are found in the nonlinear response. Thus, discrete excitons in the band gap lead to very distinct features in the spectra. Moreover, the continuum part is strongly affected as well.

II. MIXED INTRA- AND INTERBAND RESPONSE

Working in the length gauge entails handling intraband dipole matrix elements $\langle n\boldsymbol{k}|\boldsymbol{r}|n\boldsymbol{k}\rangle$, where $\boldsymbol{r}$ is the position operator, n is the band index, and $\boldsymbol{k}$ is the wave vector. By themselves, such terms are ill-defined in the infinite crystal limit but by formulating the response in terms of commutators between the ill-defined intraband dipole operator and simple operators, such as momenta, a systematic way of including intraband effects is established [14]. The dipole matrix elements are then handled according to the separation

$$\langle m\boldsymbol{k}|\boldsymbol{r}|n\boldsymbol{k}\rangle = \langle n\boldsymbol{k}|\boldsymbol{r}^i|n\boldsymbol{k}\rangle\delta_{nm} + \langle m\boldsymbol{k}|\boldsymbol{r}^e|n\boldsymbol{k}\rangle(1-\delta_{nm}), \quad (1)$$

where, formally, $\boldsymbol{r}^i$ and $\boldsymbol{r}^e$ are the intra- and interband dipole operators. Thus, $\boldsymbol{r}^i$ and $\boldsymbol{r}^e$ only have nonvanishing diagonal and off-diagonal matrix elements in the band basis, respectively. The perturbation due to the optical field $\boldsymbol{F}$ is then $H_1 = \sum_n h^{\text{int}}(\boldsymbol{r}_n)$, where the sum is over all electrons and $h^{\text{int}}(\boldsymbol{r}) = e\boldsymbol{F}\cdot(\boldsymbol{r}^i+\boldsymbol{r}^e)$ is the single-body interaction Hamiltonian. An appropriate starting point for the mixed intra- and interband

---

[*] tgp@nano.aau.dk

response is provided by the density matrix formalism [16]. The dynamics of the density matrix ρ is governed by $H_0 + H_1$ with H_0 the unperturbed Hamiltonian including Coulomb interactions among electrons. We use a mean-field approach [17] to describe excitons. Hence, correlations beyond two-particle effects such as biexcitons are ignored. In Appendix A, we derive the dynamical equations for a general multiband material. However, for clarity we restrict the analysis below to a two-band material with a single occupied valence (v) band and a single empty conduction (c) band. The diagonal elements $f_{c\boldsymbol{k}} \equiv \rho_{cc\boldsymbol{k}}$ and $f_{v\boldsymbol{k}} \equiv \rho_{vv\boldsymbol{k}}$ represent the band populations, whereas the off-diagonal elements $\rho_{cv\boldsymbol{k}} = \rho^*_{vc\boldsymbol{k}}$ provide the coherences. As demonstrated in Appendix A, coupling to the optical field is described by the commutator $\langle j\boldsymbol{k}|[h^{\text{int}},\rho]|i\boldsymbol{k}\rangle$. For general operators, we denote matrix elements by $O_{nm\boldsymbol{k}} \equiv \langle n\boldsymbol{k}|O|m\boldsymbol{k}\rangle$. The interband part of the interaction Hamiltonian is straightforward and simply leads to

$$[e\boldsymbol{F}\cdot\boldsymbol{r}^e,\rho]_{cv\boldsymbol{k}} = e\boldsymbol{F}\cdot\boldsymbol{r}^e_{cv\boldsymbol{k}}(f_{c\boldsymbol{k}} - f_{v\boldsymbol{k}}). \tag{2}$$

In contrast, the intraband operator requires greater care. The analysis is simplified by making use of the intraband commutator relation [14]

$$\begin{aligned}[][\boldsymbol{r}^i,O]_{nm\boldsymbol{k}} &= i(O_{nm\boldsymbol{k}})_{;\boldsymbol{k}},\\ (O_{nm\boldsymbol{k}})_{;\boldsymbol{k}} &= \nabla_{\boldsymbol{k}}O_{nm\boldsymbol{k}} - iO_{nm\boldsymbol{k}}[\boldsymbol{\Omega}_{nn\boldsymbol{k}} - \boldsymbol{\Omega}_{mm\boldsymbol{k}}],\end{aligned} \tag{3}$$

where $(O_{nm\boldsymbol{k}})_{;\boldsymbol{k}}$ is the generalized derivative given in terms of the Berry connection $\boldsymbol{\Omega}_{nm\boldsymbol{k}} = iV_{\text{UC}}^{-1}\int_{\text{UC}} u^*_{n\boldsymbol{k}}\nabla_{\boldsymbol{k}}u_{m\boldsymbol{k}}d\boldsymbol{r}$ with $u_{m\boldsymbol{k}}$ the periodic Bloch part of the wave function and V_{UC} the unit cell volume. Note that while the ordinary derivative is sensitive to the choice of $\boldsymbol{k}$-dependent phase of the band states, the generalized derivative is invariant. Thus, physically meaningful quantities should be defined in terms of the generalized derivative. It is a vectorial quantity and, moreover, $\boldsymbol{r}^e_{cv\boldsymbol{k}} = \boldsymbol{\Omega}_{cv\boldsymbol{k}}$. It follows from Appendix A that

$$\begin{aligned}i\hbar\frac{d\rho_{cv\boldsymbol{k}}}{dt} &- E_{cv\boldsymbol{k}}\rho_{cv\boldsymbol{k}}\\ &= \left\{e\boldsymbol{F}\cdot\boldsymbol{\Omega}_{cv\boldsymbol{k}} - \sum_{\boldsymbol{k}'}v_{cv}(\boldsymbol{k},\boldsymbol{k}')\rho_{cv\boldsymbol{k}'}\right\}(f_{v\boldsymbol{k}} - f_{c\boldsymbol{k}})\\ &\quad+\sum_{\boldsymbol{k}'}\{v_{vv}(\boldsymbol{k},\boldsymbol{k}')(f_{v\boldsymbol{k}'} - 1) - v_{cc}(\boldsymbol{k},\boldsymbol{k}')f_{c\boldsymbol{k}'}\}\rho_{cv\boldsymbol{k}}\\ &\quad+ie\boldsymbol{F}\cdot(\rho_{cv\boldsymbol{k}})_{;\boldsymbol{k}},\end{aligned} \tag{4}$$

and

$$\begin{aligned}i\hbar\frac{df_{c\boldsymbol{k}}}{dt} &= e\boldsymbol{F}\cdot(\boldsymbol{\Omega}_{cv\boldsymbol{k}}\rho_{vc\boldsymbol{k}} - \boldsymbol{\Omega}_{vc\boldsymbol{k}}\rho_{cv\boldsymbol{k}})\\ &\quad+\sum_{\boldsymbol{k}'}[v_{vc}(\boldsymbol{k},\boldsymbol{k}')\rho_{cv\boldsymbol{k}}\rho_{vc\boldsymbol{k}'} - v_{cv}(\boldsymbol{k},\boldsymbol{k}')\rho_{vc\boldsymbol{k}}\rho_{cv\boldsymbol{k}'}]\\ &\quad+ie\boldsymbol{F}\cdot(f_{c\boldsymbol{k}})_{;\boldsymbol{k}}.\end{aligned} \tag{5}$$

In these expressions, $E_{cv\boldsymbol{k}} = E_{c\boldsymbol{k}} - E_{v\boldsymbol{k}}$ is the transition energy and sums over $\boldsymbol{k}$ should be understood as integrals over the Brillouin zone. Similar relations are found for $f_{v\boldsymbol{k}}$ and $\rho_{vc\boldsymbol{k}} = \rho^*_{cv\boldsymbol{k}}$. Here, excitonic effects are included via the $\boldsymbol{k}$-space Coulomb interaction $v(\boldsymbol{k}-\boldsymbol{k}')$ and we have introduced

$$v_{nm}(\boldsymbol{k},\boldsymbol{k}') = v(\boldsymbol{k}-\boldsymbol{k}')I_{n\boldsymbol{k},n\boldsymbol{k}'}I_{m\boldsymbol{k}',m\boldsymbol{k}} \tag{6}$$

with $I_{n\boldsymbol{k},m\boldsymbol{k}'}$ the overlap between Bloch parts of the band states (see Appendix A). In a mean-field approach, screening is not properly included [17]. Hence, in the numerical application below, screening will be introduced via a phenomenological dielectric constant.

The general equations of motion can be iterated to any given order starting from initial equilibrium conditions. However, we wish to focus on cold, clean semiconductors with a sizable band gap, for which one can ignore the field-induced changes in occupation, i.e., we take $f_{v\boldsymbol{k}} = 1$ and $f_{c\boldsymbol{k}} = 0$. In this case, the evolution of the coherence is governed by the simpler dynamical equation

$$i\hbar\frac{d\rho_{cv\boldsymbol{k}}}{dt} - H_{eh}[\rho_{cv}] = e\boldsymbol{F}\cdot\boldsymbol{\Omega}_{cv\boldsymbol{k}} + ie\boldsymbol{F}\cdot(\rho_{cv\boldsymbol{k}})_{;\boldsymbol{k}}, \tag{7}$$

where we have introduced the electron-hole Hamiltonian H_{eh} via its action on the coherence

$$H_{eh}[\rho_{cv}] = E_{cv\boldsymbol{k}}\rho_{cv\boldsymbol{k}} - \sum_{\boldsymbol{k}'}v_{cv}(\boldsymbol{k},\boldsymbol{k}')\rho_{cv\boldsymbol{k}'}. \tag{8}$$

The square bracket notation $[\cdots]$ indicates a functional dependence leading to a coupling of all $\boldsymbol{k}$ points. We solve this by iteration using the exciton Green's function defined in time domain by the relation

$$-i\hbar\frac{dG}{dt} + H_{eh}[G] = \delta_{\boldsymbol{k},\boldsymbol{k}'}\delta(t-t'). \tag{9}$$

By Fourier transforming, we find the equivalent frequency ω domain condition $-\hbar\omega G^{(\omega)}_{\boldsymbol{k},\boldsymbol{k}'} + H_{eh}[G^{(\omega)}] = \delta_{\boldsymbol{k},\boldsymbol{k}'}$, which has a simple solution in the Lehmann spectral representation

$$G^{(\omega)}_{\boldsymbol{k},\boldsymbol{k}'} = \sum_n\frac{\psi^{(n)}_{cv\boldsymbol{k}}\psi^{(n)*}_{cv\boldsymbol{k}'}}{E_n - \hbar\omega}. \tag{10}$$

Here, the eigenstates follow from the eigenvalue problem $H_{eh}[\psi^{(n)}] = E_n\psi^{(n)}$ and $\psi^{(n)}_{cv\boldsymbol{k}} = \langle v\boldsymbol{k}\to c\boldsymbol{k}\mid\psi^{(n)}\rangle$ is the projection onto a singlet band-to-band transition. This problem is essentially the Bethe-Salpeter equation for the exciton eigenstates [17]. Note, however, that exchange terms are ignored (see Appendix A) and only the direct Coulomb interaction is retained here.

Throughout, we consider perturbation by a monochromatic field $\boldsymbol{F} = \frac{1}{2}(\boldsymbol{F}_\omega e^{-i\omega t} + \boldsymbol{F}^*_\omega e^{i\omega t})$. Now, to zeroth order $\rho^{(0)}_{cv\boldsymbol{k}} = 0$ and we write the first-order solution as $\rho^{(1)}_{cv\boldsymbol{k}} = \frac{1}{2}(\rho^{(\omega)}_{cv\boldsymbol{k}}e^{-i\omega t} + \rho^{(-\omega)}_{cv\boldsymbol{k}}e^{i\omega t})$. Hence, from Eq. (7),

$$\begin{aligned}\rho^{(\omega)}_{cv\boldsymbol{k}} &= -e\boldsymbol{F}_\omega\cdot\sum_{\boldsymbol{k}'}G^{(\omega)}_{\boldsymbol{k},\boldsymbol{k}'}\boldsymbol{\Omega}_{cv\boldsymbol{k}'},\\ \rho^{(-\omega)}_{cv\boldsymbol{k}} &= -e\boldsymbol{F}^*_\omega\cdot\sum_{\boldsymbol{k}'}G^{(-\omega)}_{\boldsymbol{k},\boldsymbol{k}'}\boldsymbol{\Omega}_{cv\boldsymbol{k}'}.\end{aligned} \tag{11}$$

Next, we consider the second-order equation

$$-i\hbar\frac{d\rho^{(2)}_{cv\boldsymbol{k}}}{dt} + H_{eh}[\rho^{(2)}_{cv}] = -ie\boldsymbol{F}\cdot(\rho^{(1)}_{cv\boldsymbol{k}})_{;\boldsymbol{k}}. \tag{12}$$

In analogy with the first-order problem, we separate into frequency components. However, at second order we have a time-independent (dc) term in addition to 2ω terms. Consequently, we write $\rho^{(2)}_{cv\boldsymbol{k}} = \frac{1}{4}(\rho^{(2\omega)}_{cv\boldsymbol{k}}e^{-2i\omega t} + \rho^{(-2\omega)}_{cv\boldsymbol{k}}e^{2i\omega t} + \rho^{(\text{dc})}_{cv\boldsymbol{k}})$

and find

$$\rho^{(2\omega)}_{cv\boldsymbol{k}} = -ie\boldsymbol{F}_\omega \cdot \sum_{\boldsymbol{k}'} G^{(2\omega)}_{\boldsymbol{k},\boldsymbol{k}'}\big(\rho^{(\omega)}_{cv\boldsymbol{k}'}\big)_{;\boldsymbol{k}'}. \tag{13}$$

Analogous expressions for $\rho^{(-2\omega)}_{cv\boldsymbol{k}}$ and $\rho^{(\mathrm{dc})}_{cv\boldsymbol{k}}$ are readily obtained. This procedure can be continued to arbitrary order. Our aim is the second-order nonlinear optical response and, so, higher-order terms are not needed. To complete the calculation, we apply the density matrix to compute the macroscopic induced current. As only contributions from the coherence contribute, we find generally

$$\boldsymbol{j}(t) = -\frac{e}{8\pi^3 m}\sum_{\boldsymbol{k}}\{\boldsymbol{p}_{vc\boldsymbol{k}}\rho_{cv\boldsymbol{k}} + \boldsymbol{p}_{cv\boldsymbol{k}}\rho_{vc\boldsymbol{k}}\}. \tag{14}$$

Inserting the various orders of the coherence, the current is found to contain a first-order term $\boldsymbol{j}^{(1)}(t) = \frac{1}{2}(\boldsymbol{j}_\omega e^{-i\omega t} + \boldsymbol{j}^*_\omega e^{i\omega t})$ as well as a second-order term $\boldsymbol{j}^{(2)}(t) = \frac{1}{4}(\boldsymbol{j}_{2\omega}e^{-2i\omega t} + \boldsymbol{j}^*_{2\omega}e^{2i\omega t} + \boldsymbol{j}_0)$ and higher. The current oscillating at 2ω describes the second-harmonic generation process, whereas the dc term $\boldsymbol{j}_0$ describes optical rectification, also known as the photogalvanic effect [18]. Applying the above results we then find $\boldsymbol{j}_\omega = \boldsymbol{\sigma}^{(1)} \cdot \boldsymbol{F}_\omega$ and $\boldsymbol{j}_{2\omega} = \boldsymbol{\sigma}^{(2)} : \boldsymbol{F}_\omega\boldsymbol{F}_\omega$ as well as $\boldsymbol{j}_0 = \boldsymbol{\sigma}^{(0)} : \boldsymbol{F}_\omega\boldsymbol{F}^*_\omega$ with

$$\boldsymbol{\sigma}^{(1)}(\omega) = \frac{e^2}{8\pi^3 m}\sum_{\boldsymbol{k},\boldsymbol{k}'}\boldsymbol{p}_{vc\boldsymbol{k}}G^{(\omega)}_{\boldsymbol{k},\boldsymbol{k}'}\boldsymbol{\Omega}_{cv\boldsymbol{k}'} + (\omega \to -\omega)^*, \tag{15}$$

$$\boldsymbol{\sigma}^{(2)}(\omega) = -\frac{ie^3}{8\pi^3 m}\sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{k}''}\boldsymbol{p}_{vc\boldsymbol{k}}G^{(2\omega)}_{\boldsymbol{k},\boldsymbol{k}'}\big(G^{(\omega)}_{\boldsymbol{k}',\boldsymbol{k}''}\big)_{;\boldsymbol{k}'}\boldsymbol{\Omega}_{cv\boldsymbol{k}''} + (\omega \to -\omega)^*, \tag{16}$$

$$\boldsymbol{\sigma}^{(0)}(\omega) = -\frac{ie^3}{8\pi^3 m}\sum_{\boldsymbol{k},\boldsymbol{k}',\boldsymbol{k}''}\boldsymbol{p}_{vc\boldsymbol{k}}G^{(0)}_{\boldsymbol{k},\boldsymbol{k}'}\big(G^{(\omega)}_{\boldsymbol{k}',\boldsymbol{k}''}\big)_{;\boldsymbol{k}'}\boldsymbol{\Omega}_{cv\boldsymbol{k}''} + (\omega \to -\omega)^*. \tag{17}$$

These expressions are the excitonic generalizations of the independent-particle (free-carrier) expressions in Refs. [13–15]. It follows that intraband effects manifest themselves via generalized derivatives of the exciton Green's function. Note that because the Hamiltonian is defined in terms of its action on ρ_{cv}, this derivative is to be taken for $nm = cv$ [c.f. Eq. (3)]. We stress that, in the expressions above, only well-defined matrix elements, i.e., momenta and Berry connections, appear. To express these results in terms of exciton matrix elements, we utilize Eq. (10) and introduce

$$\boldsymbol{P}_n = \sum_{\boldsymbol{k}}\psi^{(n)}_{cv\boldsymbol{k}}\boldsymbol{p}_{vc\boldsymbol{k}}, \quad \boldsymbol{\Omega}_n = \sum_{\boldsymbol{k}}\psi^{(n)}_{cv\boldsymbol{k}}\boldsymbol{\Omega}_{vc\boldsymbol{k}},$$

$$\boldsymbol{Q}_{mn} = i\sum_{\boldsymbol{k}}\psi^{(m)*}_{cv\boldsymbol{k}}\big(\psi^{(n)}_{cv\boldsymbol{k}}\big)_{;\boldsymbol{k}}. \tag{18}$$

This allows us to write

$$\boldsymbol{\sigma}^{(1)}(\omega) = \frac{e^2}{8\pi^3 m}\sum_n\frac{\boldsymbol{P}_n\boldsymbol{\Omega}^*_n}{E_n - \hbar\omega} + (\omega \to -\omega)^*, \tag{19}$$

$$\boldsymbol{\sigma}^{(2)}(\omega) = -\frac{e^3}{8\pi^3 m}\sum_{m,n}\frac{\boldsymbol{P}_m\boldsymbol{Q}_{mn}\boldsymbol{\Omega}^*_n}{(E_m - 2\hbar\omega)(E_n - \hbar\omega)} + (\omega \to -\omega)^*, \tag{20}$$

$$\boldsymbol{\sigma}^{(0)}(\omega) = -\frac{e^3}{8\pi^3 m}\sum_{m,n}\frac{\boldsymbol{P}_m\boldsymbol{Q}_{mn}\boldsymbol{\Omega}^*_n}{E_m(E_n - \hbar\omega)} + (\omega \to -\omega)^*. \tag{21}$$

Thus, in this formulation, all response functions are given in terms of exciton matrix elements such as momentum $\boldsymbol{P}_n$, etc. For the unperturbed independent-particle Hamiltonian $H_{IP} = \sum_{n\boldsymbol{k}} E_{n\boldsymbol{k}}|n\boldsymbol{k}\rangle\langle n\boldsymbol{k}|$, the commutator $[H_{IP},\boldsymbol{r}^e] = \hbar\boldsymbol{p}/(im)$ leads to the rule $\boldsymbol{\Omega}_{cv\boldsymbol{k}} = \hbar\boldsymbol{p}_{cv\boldsymbol{k}}/(imE_{cv\boldsymbol{k}})$. Similarly, introducing the many-body interband position and momentum operators $\boldsymbol{R}^e$ and $\boldsymbol{P}$, the commutator with H_0 yields $\boldsymbol{\Omega}^*_n = \hbar\boldsymbol{P}^*_n/(imE_n)$. Finally, the expressions above can be converted into formulas for the associated susceptibilities via $\boldsymbol{\chi}^{(1)} = \frac{i}{\omega\varepsilon_0}\boldsymbol{\sigma}^{(1)}$ and $\boldsymbol{\chi}^{(2)} = \frac{i}{2\omega\varepsilon_0}\boldsymbol{\sigma}^{(2)}$:

$$\boldsymbol{\chi}^{(1)}(\omega) = \frac{e^2\hbar}{8\pi^3\varepsilon_0 m^2\omega}\sum_n\frac{\boldsymbol{P}_n\boldsymbol{P}^*_n}{E_n(E_n - \hbar\omega)} + (\omega \to -\omega)^*, \tag{22}$$

$$\boldsymbol{\chi}^{(2)}(\omega) = -\frac{e^3\hbar}{16\pi^3\varepsilon_0 m^2\omega}\sum_{m,n}\frac{\boldsymbol{P}_m\boldsymbol{Q}_{mn}\boldsymbol{P}^*_n}{E_n(E_m - 2\hbar\omega)(E_n - \hbar\omega)} + (\omega \to -\omega)^*. \tag{23}$$

We note that although the derivation above was restricted to a single pair of bands, it is straightforward to extend to multiband systems. Thus, in Appendix B, expressions that apply to multiband semiconductors are derived. In fact, the expressions (19)–(21) are still valid provided the exciton matrix elements are generalized slightly by summing over pairs of bands [see Eq. (B7)]. Moreover, a purely interband second-order response, Eq. (B8), is found.

We proceed by deriving approximate expressions for the free-carrier limit, in which electron-hole interaction is ignored. This is easily accomplished as, in the independent-particle limit, $G^{(\omega)}_{\boldsymbol{k},\boldsymbol{k}'} = \delta_{\boldsymbol{k},\boldsymbol{k}'}/(E_{cv\boldsymbol{k}} - \hbar\omega)$ and, so,

$$\boldsymbol{\chi}^{(1)}(\omega) = \frac{e^2\hbar^2}{4\pi^3\varepsilon_0 m^2}\sum_{\boldsymbol{k}}\frac{\boldsymbol{p}_{vc\boldsymbol{k}}\boldsymbol{p}_{cv\boldsymbol{k}}}{E_{cv\boldsymbol{k}}\big(E^2_{cv\boldsymbol{k}} - \hbar^2\omega^2\big)}, \tag{24}$$

$$\boldsymbol{\chi}^{(2)}(\omega) = \frac{3ie^3\hbar^2}{8\pi^3 m^2\varepsilon_0}\sum_{\boldsymbol{k}}\frac{\boldsymbol{p}_{vc\boldsymbol{k}}(\boldsymbol{p}_{cv\boldsymbol{k}})_{;\boldsymbol{k}}}{\big(E^2_{cv\boldsymbol{k}} - 4\hbar^2\omega^2\big)\big(E^2_{cv\boldsymbol{k}} - \hbar^2\omega^2\big)}. \tag{25}$$

The expression for $\boldsymbol{\chi}^{(2)}$ can be shown to be equivalent to the ones applied in Refs. [15,19]. Note that the integrands in Eqs. (24) and (25) are manifestly finite in the dc limit $\omega \to 0$. The excitonic expressions (19) and (20) can be written

$$\boldsymbol{\sigma}^{(1)}(\omega) = \frac{e^2\hbar}{8\pi^3 im^2}\sum_n\left\{\frac{\boldsymbol{P}_n\boldsymbol{P}^*_n}{E_n - \hbar\omega} - \frac{\boldsymbol{P}^*_n\boldsymbol{P}_n}{E_n + \hbar\omega}\right\}, \tag{26}$$

$$\boldsymbol{\sigma}^{(2)}(\omega) = \frac{ie^3\hbar}{8\pi^3 m^2}\sum_{m,n}\left\{\frac{\boldsymbol{P}_m\boldsymbol{Q}_{mn}\boldsymbol{P}^*_n}{E_n(E_m - 2\hbar\omega)(E_n - \hbar\omega)} - \frac{\boldsymbol{P}^*_m\boldsymbol{Q}^*_{mn}\boldsymbol{P}_n}{E_n(E_m + 2\hbar\omega)(E_n + \hbar\omega)}\right\}. \tag{27}$$

Now, for systems with negligible spin-orbit interaction, time reversal symmetry means that one can choose phases such that $\psi^{(n)}_{cv,-\boldsymbol{k}} = \psi^{(n)*}_{cv\boldsymbol{k}}$ and $\boldsymbol{p}_{vc,-\boldsymbol{k}} = -\boldsymbol{p}_{cv\boldsymbol{k}}$. Hence, by shifting the sums in Eq. (18) from $\boldsymbol{k}$ to $-\boldsymbol{k}$ it is readily shown that, with this phase choice, $\boldsymbol{P}_n^*\boldsymbol{P}_n = \boldsymbol{P}_n\boldsymbol{P}_n^*$ and $\boldsymbol{P}_m^*\boldsymbol{Q}_{mn}^*\boldsymbol{P}_n = \boldsymbol{P}_m\boldsymbol{Q}_{mn}\boldsymbol{P}_n^*$ and, so, the conductivities vanish in the dc limit. In turn, this implies that the susceptibilities remain finite. Since all expressions should be invariant under the choice of phase, it follows that this conclusion is generally valid. Numerically, we indeed find finite dc susceptibilities in the example studied below.

Our results have been derived using perturbations formulated in the length $\boldsymbol{r}\cdot\boldsymbol{F}$ gauge rather than velocity $\boldsymbol{p}\cdot\boldsymbol{A}$ gauge with $\boldsymbol{A}$ the vector potential. However, an intimately related issue is concerned with invariance of the responses, i.e., current $\boldsymbol{j}(t)$ and polarization $\boldsymbol{P}(t)$. Physically, these are related by $\boldsymbol{j}(t) = d\boldsymbol{P}(t)/dt$, which leads to the relation between conductivity and susceptibility applied above. Our approach is focused on the calculation of the current and then deriving the associated polarization. However, one could have chosen a different strategy and computed the susceptibilities directly rather than through the conductivities. To this end, the interband polarization could have been found in analogy with Eq. (14) by replacing velocity with dipole matrix elements, i.e., $\boldsymbol{P}(t) = -e/(8\pi^3)\sum_{\boldsymbol{k}}\{\boldsymbol{\Omega}_{vc\boldsymbol{k}}\rho_{cv\boldsymbol{k}} + \boldsymbol{\Omega}_{cv\boldsymbol{k}}\rho_{vc\boldsymbol{k}}\}$. Focusing on the 2ω response and applying the relation between dipole and momentum matrix elements, this would imply a nonlinear dipole susceptibility given by

$$\chi^{(2)}_{\mathrm{dipole}}(\omega) = -\frac{e^3\hbar^2}{8\pi^3\varepsilon_0 m^2}\sum_{m,n}\frac{\boldsymbol{P}_m\boldsymbol{Q}_{mn}\boldsymbol{P}_n^*}{E_mE_n(E_m-2\hbar\omega)(E_n-\hbar\omega)} + (\omega\to-\omega)^*. \tag{28}$$

This expression differs from Eq. (23) in that the energy E_m replaces the two-photon energy $2\hbar\omega$. Hence, the two are not identical. In fact, equivalence of the two expressions (in the independent-particle limit) can be demonstrated [20] but implicitly relies on the assumption that all summations are over a complete set of states. Obviously, this assumption is invalid in the present two-band model. Expanding out the denominator of Eq. (28) in partial fractions, we find

$$\frac{1}{E_mE_n(E_m-2\hbar\omega)(E_n-\hbar\omega)} = \frac{1}{2\hbar^2\omega^2}\left(\frac{1}{E_n-\hbar\omega}-\frac{1}{E_n}\right)\left(\frac{1}{E_m-2\hbar\omega}-\frac{1}{E_m}\right). \tag{29}$$

For Eq. (23), the result is

$$\frac{1}{2\hbar\omega E_n(E_m-2\hbar\omega)(E_n-\hbar\omega)} = \frac{1}{2\hbar^2\omega^2}\left(\frac{1}{E_n-\hbar\omega}-\frac{1}{E_n}\right)\left(\frac{1}{E_m-2\hbar\omega}\right). \tag{30}$$

Thus, the difference between the two is contained in the last term in the second set of brackets in Eq. (29). Near the 2ω resonance, this extra term is negligible but it cannot be ignored generally. Below, we will provide an example illustrating the difference between the two expressions. Note that the partial fractions are useful for obtaining formulas that are amenable to a Lanczos Green's function evaluation [11]. Thus, Eq. (30) leads to the convenient form $(2\hbar^2\omega^2)^{-1}(G^{(\omega)}-G^{(0)})G^{(2\omega)}$.

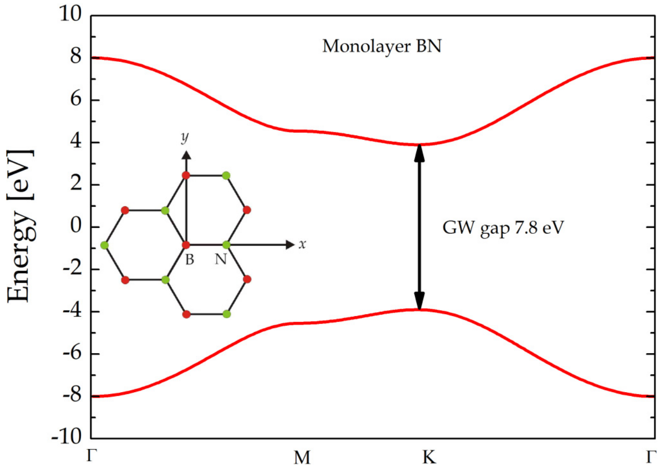

FIG. 1. (Color online) Tight-binding band structure of monolayer BN. The inset shows the atomic arrangement.

III. APPLICATION TO HEXAGONAL BORON NITRIDE

Several well-converged *ab initio* studies of the linear response of h-BN monolayers have been published [21,22]. The local density approximation density functional theory band gap of approximately 4.4 eV opens to much larger values of 7.8 eV [21] or 7.37 eV [23] due to GW quasiparticle corrections. However, electron-hole effects included via the Bethe-Salpeter equation reduce the position of the lowest optical absorption peak to around 6.2 eV [21], in good agreement with experiments on bulk h-BN [24] and exfoliated nanosheets [25]. We demonstrate here that these features are nicely captured in a minimal tight-binding plus Bethe-Salpeter model. Due to the minimal basis and accompanying low computational cost, this type of model is ideally suited for spectra requiring very dense k-point sampling. In a nearest-neighbor tight-binding approach, the only adjustable parameters are the hopping integral γ and the difference between B and N on-site energies. Taking the middle of the band gap as the energy zero point, the on-site energies are $\pm\alpha$ and the band gap is 2α. Hence, $\alpha = 3.9$ eV is adopted in order to match a GW gap of 7.8 eV. Assuming a hopping integral of $\gamma = 2.33$ eV obtained in Ref. [26] by fitting to *ab initio* results we find the band structure shown in Fig. 1.

The Coulomb interaction responsible for excitonic effects is incorporated similarly to the case of MoS_2 [11]. Hence, we take

$$v_{cv}(\boldsymbol{k},\boldsymbol{k}') = -\frac{e^2 I_{v\boldsymbol{k}',v\boldsymbol{k}} I_{c\boldsymbol{k},c\boldsymbol{k}'}}{8\pi^2\varepsilon\varepsilon_0|\boldsymbol{k}-\boldsymbol{k}'|}\exp(-l|\boldsymbol{k}-\boldsymbol{k}'|). \tag{31}$$

Here, ε is the screening dielectric constant and l is a length on the order of the sheet thickness representing charge smearing. Taking the layer thickness to be $l = 2$ Å we find that the correct exciton binding energy is obtained assuming screening of $\varepsilon = 1.5$. Finally, a lattice constant of 2.49 Å is assumed and line shape broadening is accounted for by adding a phenomenological imaginary part $\hbar\Gamma = 0.03$ eV to the optical frequency.

As h-BN is a two-dimensional material, it is convenient to replace the bulk response formulas derived in the previous section by sheet quantities. These differ by (i) restricting the

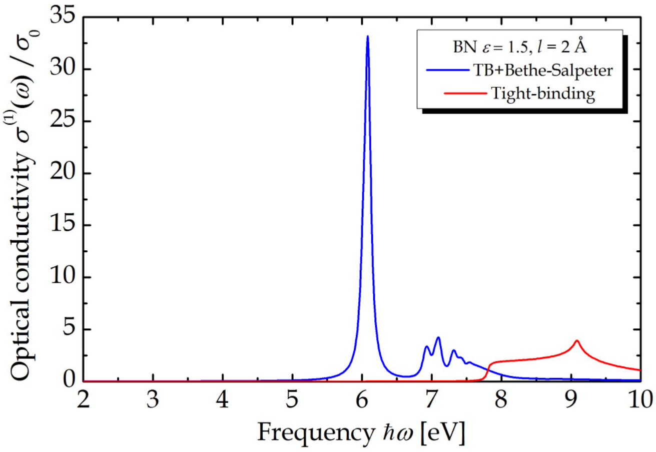

FIG. 2. (Color online) Real part of the linear optical conductivity with (blue curve) and without (red curve) excitonic effects.

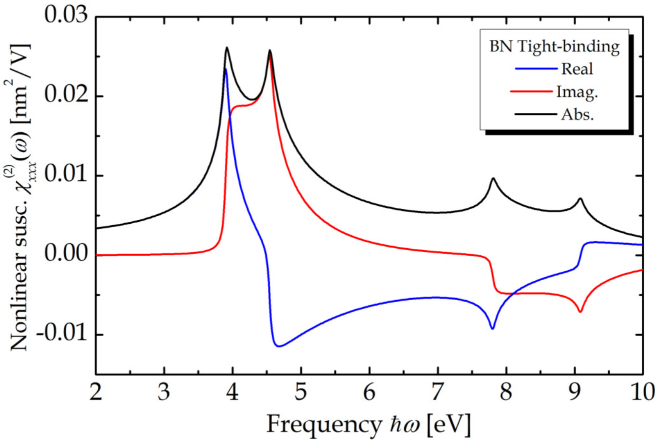

FIG. 3. (Color online) Second-harmonic nonlinear optical susceptibility in the independent-particle approximation.

k-space integrals to a two-dimensional Brillouin zone and (ii) a factor of 2π reflecting the reduced dimensionality. Since spin-orbit coupling is neglected, all response functions should be multiplied by a factor of 2 to account for spin summation. We use a dense k-vector grid of 100×100 points and adopt a Lanczos approach [11] to the evaluation of response functions. The linear sheet conductivity $\sigma^{(1)}(\omega)$ is conveniently expressed in units of the conductance quantum $\sigma_0 = e^2/4\hbar$. The result for the real (absorptive) part is shown in Fig. 2. The fundamental exciton leads to a distinct absorption peak around 6.1 eV. In addition, several higher excitons are visible. Finally, the continuum absorption above the quasiparticle band gap is found to be greatly reduced as a consequence of oscillator strength transferred into bound excitons. Thus, the excitonic modifications of the linear response are seen to be very dramatic. These results are in good quantitative agreement with first-principles calculations in Refs. [10,21–22], in particular, regarding the location of fundamental and higher exciton resonances. Also, the position and symmetric line shape of the fundamental resonance is in excellent agreement with recent absorption measurements on exfoliated nanosheets [25].

The second-order nonlinear response is described by the tensors Eqs. (20) and (21). For the hexagonal symmetry with a mirror symmetry plane containing the x axis the four nonvanishing tensor elements are $\chi^{(2)}_{xxx} = -\chi^{(2)}_{xyy} = -\chi^{(2)}_{yyx} = -\chi^{(2)}_{yxy}$ and similarly for $\boldsymbol{\sigma}^{(0)}$. Ignoring for the moment excitonic effects, we find the independent-electron second-harmonic generation spectrum in Fig. 3. This can be seen to resemble the result in Ref. [19]. Note, however, that the Dirac approximation for the band structure as well as a different tight-binding parametrization was adopted in Ref. [19], which leads to noticeable differences. Focusing on the imaginary part, the main features of the spectrum resemble the linear spectrum to a large degree. Thus, starting at half the band gap, a step and a van Hove singularity are found. A similar but inverted set of features is found at the band gap. In both cases, the similarity to the linear independent-particle result is clear.

Adding excitons to the nonlinear response leads to modifications similar to the linear case, as evidenced in Fig. 4. Again, pronounced features produced by the fundamental exciton are found, i.e., features at $\hbar\omega \approx 6$ eV and $2\hbar\omega \approx 6$ eV. Moreover, higher excitons are visible as weaker peaks. A significant difference with respect to the independent-particle result in Fig. 3 is the relative weight of ω and 2ω resonances. Thus, in Fig. 3, the ω response in the range 8–10 eV is markedly weaker than the 2ω response around 4–5 eV. In contrast, the response due to the fundamental exciton is of similar strength in the ω and 2ω ranges when excitons are taken into account. Furthermore, the magnitude of the resonances is greatly enhanced due to the Coulomb coupling, as demonstrated by the different scales in Figs. 3 and 4. The spectrum for the absolute value of the response is seen to agree reasonably well with the time-dependent Hartree-Fock result in Ref. [10]. Overall, excitonic effects are clearly found to be indispensable for the nonlinear response. In the inset, we compare current-based and dipole-based susceptibilities, i.e., Eqs. (23) and (28). As expected, the 2ω ranges are basically identical, whereas clear differences are found at higher energies. These differences only pertain to the intensity of the peaks, however. The discrepancy is a consequence of the two-band approximation that breaks gauge invariance. Thus, the spectra are expected to converge as more bands are included.

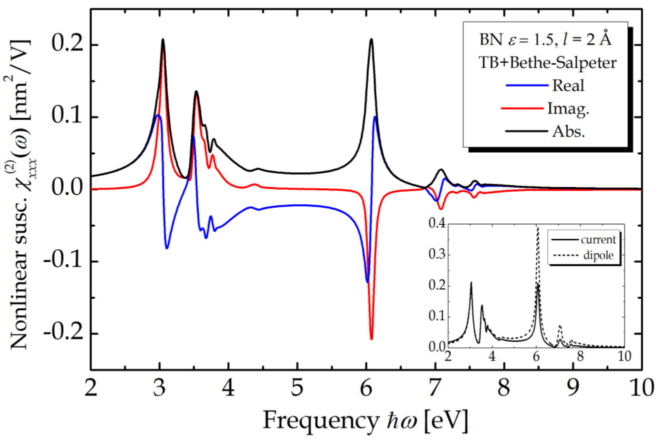

FIG. 4. (Color online) Second-harmonic optical susceptibility including excitonic effects. Note the different scale compared to Fig. 3. The inset shows a comparison between absolute values based on current and dipole approaches.

IV. SUMMARY

In summary, we have formulated a density matrix approach to the optical response of two-band semiconductors taking intraband effects into account rigorously. By iteration, the first- and second-order responses were derived but arbitrarily high orders are accessible, in principle. The results, formulated in terms of exciton Green's functions, are manifestly free from unphysical zero-frequency divergences. When applied to h-BN monolayers it is found that excitons dominate both linear and nonlinear optical properties. Fingerprints of both fundamental and higher bound excitons are seen. However, even above the band gap, Coulomb interactions lead to significant modifications. The present formalism is readily extended to describe various nonlinear effects in multiband materials.

ACKNOWLEDGMENTS

Useful discussions with Mads Lund Trolle are gratefully acknowledged. This work is financially supported by the Center for Nanostructured Graphene and the QUSCOPE center. CNG is sponsored by the Danish National Research Foundation, project DNRF58 and QUSCOPE is sponsored by the Villum Foundation.

APPENDIX A: DYNAMICAL EQUATION

The governing equations for the density matrix in the single-particle basis are conveniently obtained from the Hamiltonian in second quantization with $c,c^\dagger$ fermionic annihilation and creation operators

$$H=\sum_{kl}h_{kl}c_l^\dagger c_k+\frac{1}{2}\sum_{klmn}v_{klmn}c_k^\dagger c_l^\dagger c_n c_m, \tag{A1}$$

in which h is the single-electron part (including dipole coupling to the electromagnetic field) and v is the Coulomb potential. This leads to the usual equation of motion [17] for the density matrix $\rho_{ji}=\langle c_i^\dagger c_j\rangle$,

$$\begin{aligned} i\hbar\frac{d\rho_{ji}}{dt}&+\sum_l\{h_{li}\rho_{jl}-h_{jl}\rho_{li}\}\\ &=\sum_{lmn}\{v_{lmni}\langle c_l^\dagger c_m^\dagger c_n c_j\rangle-v_{jlmn}\langle c_i^\dagger c_l^\dagger c_m c_n\rangle\}. \end{aligned}\tag{A2}$$

We then apply the mean-field approximation $\langle c_l^\dagger c_m^\dagger c_n c_j\rangle\approx\rho_{jl}\rho_{nm}-\rho_{nl}\rho_{jm}$. The single-particle Hamiltonian has a zero order and an interaction part $h=h^0+h^{\text{int}}$. The first part only has diagonal elements, whereas the second (time-dependent) part has both diagonal and off-diagonal parts. Also, we introduce the quasiparticle energies $E_n=h^0_{nn}+\sum_l\{v_{nlnl}-v_{nlln}\}\delta_{lv}$. Here, the delta function serves to count occupied states only. Hence,

$$\begin{aligned} i\hbar\frac{d\rho_{ji}}{dt}&-E_{ji}\rho_{ji}-\langle j|[h^{\text{int}},\rho]|i\rangle\\ &=\sum_{lmn}\{v_{mlni}-v_{lmni}\}(\rho_{nl}-\delta_{mi}\delta_{nl}\delta_{lv})\rho_{jm}\\ &\quad+\sum_{lmn}\{v_{jlmn}-v_{jlnm}\}(\rho_{nl}-\delta_{mj}\delta_{nl}\delta_{lv})\rho_{mi}. \end{aligned}\tag{A3}$$

We now specialize to the case of Bloch states of the form $|n\boldsymbol{k}_n\rangle=V^{-1/2}u_{n\boldsymbol{k}_n}(\boldsymbol{r})e^{i\boldsymbol{k}_n\cdot\boldsymbol{r}}$. In terms of the Fourier transform of the Coulomb interaction $v(\boldsymbol{q})$, the matrix elements become

$$v_{abcd}=\sum_{\boldsymbol{q}}v(\boldsymbol{q})\langle a\boldsymbol{k}_a|e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}|c\boldsymbol{k}_c\rangle\langle b\boldsymbol{k}_b|e^{i\boldsymbol{q}\cdot\boldsymbol{r}}|d\boldsymbol{k}_d\rangle. \tag{A4}$$

We restrict $\boldsymbol{q}$ to the Brillouin zone, which amounts to neglecting local-field effects. We introduce the Bloch overlaps $I_{n\boldsymbol{k}',m\boldsymbol{k}}=\langle n\boldsymbol{k}'|e^{i(\boldsymbol{k}'-\boldsymbol{k})\cdot\boldsymbol{r}}|m\boldsymbol{k}\rangle=V_{\text{UC}}^{-1}\int_{\text{UC}}u^*_{n\boldsymbol{k}'}u_{m\boldsymbol{k}}d\boldsymbol{r}$ and so

$$v_{abcd}=\sum_{\boldsymbol{q}}v(\boldsymbol{q})I_{a\boldsymbol{k}_a,c\boldsymbol{k}_a+\boldsymbol{q}}I_{b\boldsymbol{k}_b,d\boldsymbol{k}_b-\boldsymbol{q}}\delta_{\boldsymbol{k}_c,\boldsymbol{k}_a+\boldsymbol{q}}\delta_{\boldsymbol{k}_d,\boldsymbol{k}_b-\boldsymbol{q}}. \tag{A5}$$

Normalization is such that $I_{n\boldsymbol{k},m\boldsymbol{k}}=\delta_{nm}$. The density matrix generally is not diagonal in wave vector index. However, in the dipole approximation, only the diagonal elements matter. Having $\boldsymbol{k}_n=\boldsymbol{k}_l$ is only possible for the first and fourth terms on the right-hand side of Eq. (A3). The remaining terms are exchange terms that we neglect. Hence,

$$\begin{aligned} i\hbar\frac{d\rho_{ji\boldsymbol{k}}}{dt}&-E_{ji\boldsymbol{k}}\rho_{ji\boldsymbol{k}}-\langle j\boldsymbol{k}|[h^{\text{int}},\rho]|i\boldsymbol{k}\rangle\\ &=\sum_{lmn\boldsymbol{k}'}v(\boldsymbol{k}-\boldsymbol{k}')I_{m\boldsymbol{k},n\boldsymbol{k}'}I_{l\boldsymbol{k}',i\boldsymbol{k}}(\rho_{nl\boldsymbol{k}'}-\delta_{mi}\delta_{nl}\delta_{lv})\rho_{jm\boldsymbol{k}}\\ &\quad-\sum_{lmn\boldsymbol{k}'}v(\boldsymbol{k}-\boldsymbol{k}')I_{j\boldsymbol{k},n\boldsymbol{k}'}I_{l\boldsymbol{k}',m\boldsymbol{k}}(\rho_{nl\boldsymbol{k}'}-\delta_{mj}\delta_{nl}\delta_{lv})\rho_{mi\boldsymbol{k}}. \end{aligned}\tag{A6}$$

Under the assumption that coupling of nearby $\boldsymbol{k}$ points $\boldsymbol{k}$ and $\boldsymbol{k}'$ dominate, the fact that $I_{n\boldsymbol{k},m\boldsymbol{k}}=\delta_{nm}$ means that $|I_{v\boldsymbol{k},v\boldsymbol{k}'}|,|I_{c\boldsymbol{k},c\boldsymbol{k}'}|\gg|I_{v\boldsymbol{k},c\boldsymbol{k}'}|$, i.e., intraband overlap dominates over interband overlap. In this case, taking $j=c$ and $i=v$ means that for a two-band material

$$\begin{aligned} i\hbar\frac{d\rho_{cv\boldsymbol{k}}}{dt}&-E_{cv\boldsymbol{k}}\rho_{cv\boldsymbol{k}}-\langle c\boldsymbol{k}|[h^{\text{int}},\rho]|v\boldsymbol{k}\rangle\\ &=\sum_{\boldsymbol{k}'}v(\boldsymbol{k}-\boldsymbol{k}')\{|I_{v\boldsymbol{k},v\boldsymbol{k}'}|^2(\rho_{vv\boldsymbol{k}'}-1)-|I_{c\boldsymbol{k},c\boldsymbol{k}'}|^2\rho_{cc\boldsymbol{k}'}\}\rho_{cv\boldsymbol{k}}\\ &\quad+\sum_{\boldsymbol{k}'}v(\boldsymbol{k}-\boldsymbol{k}')I_{c\boldsymbol{k},c\boldsymbol{k}'}I_{v\boldsymbol{k}',v\boldsymbol{k}}(\rho_{cc\boldsymbol{k}}-\rho_{vv\boldsymbol{k}})\rho_{cv\boldsymbol{k}'}. \end{aligned}\tag{A7}$$

Similarly, for $i=j=c$,

$$\begin{aligned} i\hbar\frac{d\rho_{cc\boldsymbol{k}}}{dt}&-\langle c\boldsymbol{k}|[h^{\text{int}},\rho]|c\boldsymbol{k}\rangle\\ &=\sum_{\boldsymbol{k}'}v(\boldsymbol{k}-\boldsymbol{k}')\{I_{c\boldsymbol{k}',c\boldsymbol{k}}I_{v\boldsymbol{k},v\boldsymbol{k}'}\rho_{cv\boldsymbol{k}}\rho_{vc\boldsymbol{k}'}\\ &\quad-I_{c\boldsymbol{k},c\boldsymbol{k}'}I_{v\boldsymbol{k}',v\boldsymbol{k}}\rho_{vc\boldsymbol{k}}\rho_{cv\boldsymbol{k}'}\}. \end{aligned}\tag{A8}$$

APPENDIX B: MULTIBAND EXTENSION

If the restriction to a single pair of valence and conduction bands (v,c) is lifted, the dynamical equations are clearly more complicated but still tractable. For a cold, clean multiband semiconductor, the dominant extension is that the simplified electron-hole Hamiltonian, Eq. (8), is replaced by the multiband expression

$$H_{eh}[\rho]=E_{cv\boldsymbol{k}}\rho_{cv\boldsymbol{k}}-\sum_{v'c'\boldsymbol{k}'}v(\boldsymbol{k}-\boldsymbol{k}')I_{c\boldsymbol{k},c'\boldsymbol{k}'}I_{v'\boldsymbol{k}',v\boldsymbol{k}}\rho_{c'v'\boldsymbol{k}'}. \tag{B1}$$

In turn, the Green's function is a larger construction coupling a given pair of bands (v,c) to any other (v',c') and is defined by $-\hbar\omega G^{(\omega)}_{cv\boldsymbol{k},c'v'\boldsymbol{k}'} + H_{eh}[G^{(\omega)}] = \delta_{cv\boldsymbol{k},c'v'\boldsymbol{k}'}$, which in the Lehmann representation means that

$$G^{(\omega)}_{cv\boldsymbol{k},c'v'\boldsymbol{k}'} = \sum_n \frac{\psi^{(n)}_{cv\boldsymbol{k}}\psi^{(n)*}_{c'v'\boldsymbol{k}'}}{E_n - \hbar\omega}. \tag{B2}$$

The eigenstate problem $H_{eh}[\psi^{(n)}] = E_n\psi^{(n)}$ with H_{eh} defined by Eq. (B1) is now the full coupled Bethe-Salpeter equation. Hence, the results for the two-band case carry over apart from added summations over bands. As a consequence, Eqs. (15) and (16) now read

$$\boldsymbol{\sigma}^{(1)}(\omega) = \frac{e^2}{8\pi^3 m}\sum_{cv\boldsymbol{k},c'v'\boldsymbol{k}'} \boldsymbol{p}_{vc\boldsymbol{k}} G^{(\omega)}_{cv\boldsymbol{k},c'v'\boldsymbol{k}'}\boldsymbol{\Omega}_{c'v'\boldsymbol{k}'} + (\omega \to -\omega)^*, \tag{B3}$$

$$\boldsymbol{\sigma}^{(2)}(\omega) = -\frac{ie^3}{8\pi^3 m}\sum_{cv\boldsymbol{k},c'v'\boldsymbol{k}',c''v''\boldsymbol{k}''} \boldsymbol{p}_{vc\boldsymbol{k}} G^{(2\omega)}_{cv\boldsymbol{k},c'v'\boldsymbol{k}'}\left(G^{(\omega)}_{c'v'\boldsymbol{k}',c''v''\boldsymbol{k}''}\right)_{;\boldsymbol{k}'}\boldsymbol{\Omega}_{c''v''\boldsymbol{k}''} + (\omega \to -\omega)^*. \tag{B4}$$

In the multiband case, however, the commutator expression involving the interaction Hamiltonian contains additional terms and, in general, reads

$$\langle j\boldsymbol{k}|[h^{\text{int}},\rho]|i\boldsymbol{k}\rangle = e\boldsymbol{F}\cdot\sum_{l\neq j}\boldsymbol{\Omega}_{jl\boldsymbol{k}}\rho_{li\boldsymbol{k}} - e\boldsymbol{F}\cdot\sum_{l\neq i}\rho_{jl\boldsymbol{k}}\boldsymbol{\Omega}_{li\boldsymbol{k}} + ie\boldsymbol{F}\cdot(\rho_{ji\boldsymbol{k}})_{;\boldsymbol{k}}. \tag{B5}$$

The additional terms mean that we now find a purely interband contribution to the second-order response, which can be written

$$\boldsymbol{\sigma}^{(2)}_{\text{inter}}(\omega) = -\frac{e^3}{8\pi^3 m}\sum_{cv\boldsymbol{k},c'v'\boldsymbol{k}',c''v''\boldsymbol{k}''} \boldsymbol{p}_{vc\boldsymbol{k}} G^{(2\omega)}_{cv\boldsymbol{k},c'v'\boldsymbol{k}'}\left\{\sum_{l\neq c'}\boldsymbol{\Omega}_{c'l\boldsymbol{k}'}G^{(\omega)}_{c'l\boldsymbol{k}',c''v''\boldsymbol{k}''} - \sum_{l\neq v'}\boldsymbol{\Omega}_{lv'\boldsymbol{k}'}G^{(\omega)}_{lv'\boldsymbol{k}',c''v''\boldsymbol{k}''}\right\}\boldsymbol{\Omega}_{c''v''\boldsymbol{k}''} + (\omega \to -\omega)^*. \tag{B6}$$

Finally, it may be noted that the compact expressions (19)–(21) still hold if one simply adds a sum over band pairs, i.e.,

$$\boldsymbol{P}_n = \sum_{cv\boldsymbol{k}}\psi^{(n)}_{cv\boldsymbol{k}}\boldsymbol{p}_{vc\boldsymbol{k}}, \qquad \boldsymbol{\Omega}_n = \sum_{cv\boldsymbol{k}}\psi^{(n)}_{cv\boldsymbol{k}}\boldsymbol{\Omega}_{vc\boldsymbol{k}}, \qquad \boldsymbol{Q}_{mn} = i\sum_{cv\boldsymbol{k}}\psi^{(m)*}_{cv\boldsymbol{k}}\left(\psi^{(n)}_{cv\boldsymbol{k}}\right)_{;\boldsymbol{k}}. \tag{B7}$$

Similarly, the purely interband response becomes

$$\boldsymbol{\sigma}^{(2)}_{\text{inter}}(\omega) = -\frac{e^3}{8\pi^3 m}\sum_{m,n}\frac{\boldsymbol{P}_m\boldsymbol{R}_{mn}\boldsymbol{\Omega}^*_n}{(E_m - 2\hbar\omega)(E_n - \hbar\omega)} + (\omega \to -\omega)^*, \tag{B8}$$

with

$$\boldsymbol{R}_{mn} = \sum_{cv\boldsymbol{k}}\psi^{(m)*}_{cv\boldsymbol{k}}\left\{\sum_{l\neq c}\boldsymbol{\Omega}_{cl\boldsymbol{k}}\psi^{(n)}_{lv\boldsymbol{k}} - \sum_{l\neq v}\boldsymbol{\Omega}_{lv\boldsymbol{k}}\psi^{(n)}_{cl\boldsymbol{k}}\right\}. \tag{B9}$$

[1] H. Haug and S.W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (World Scientific, Singapore, 1993).

[2] F. Wang, G. Dukovic, L. E. Brus, and T. F. Heinz, Science **308**, 838 (2005).

[3] K. He, N. Kumar, L. Zhao, Z. Wang, K. F. Mak, H. Zhao, and J. Shan, Phys. Rev. Lett. **113**, 026803 (2014).

[4] F. Minami, K. Inoue, Y. Kato, K. Yoshida, and K. Era, Phys. Rev. Lett. **67**, 3708 (1991).

[5] H. M. Su, J. T. Ye, Z. K. Tang, and K. S. Wong, Phys. Rev. B **77**, 125428 (2008).

[6] L. M. Malard, T. V. Alencar, Anna Paula M. Barboza, K. F. Mak, and A. M. de Paula, Phys. Rev. B. **87**, 201401 (2013).

[7] E. K. Chang, E. L. Shirley, and Z. H. Levine, Phys. Rev. B **65**, 035205 (2001).

[8] R. Leitsmann, W. G. Schmidt, P. H. Hahn, and F. Bechstedt, Phys. Rev. B **71**, 195209 (2005).

[9] T. G. Pedersen and H. Cornean, Europhys. Lett. **78**, 27005 (2007).

[10] M. Grüning and C. Attaccalite, Phys. Rev. B **89**, 081102 (2014).

[11] M. L. Trolle, G. Seifert, and T. G. Pedersen, Phys. Rev. B **89**, 235410 (2014).

[12] R. W. Boyd, *Nonlinear Optics* (Academic, San Diego, 1992).

[13] V. M. Genkin and P. M. Mednis, Sov. Phys. JETP **27**, 609 (1968).

[14] C. Aversa and J. E. Sipe, Phys. Rev. B **52**, 14636 (1995).

[15] B. Kirtman, F. L. Gu, and D. M. Bishop, J. Chem. Phys. **113**, 1294 (2000).

[16] I. Al-Naib, J. E. Sipe, and M. M. Dignam, Phys. Rev. B **90**, 245423 (2014).

[17] W. Schäfer and M. Wegener, *Semiconductor Optics and Transport Phenomena* (Springer, Berlin, 2002).

[18] H. Yuan, X. Wang, B. Lian, H. Zhang, X. Fang, B. Shen, G. Xu, Y. Xu, S.-C. Zhang, H. Y. Hwang, and Y. Cui, Nat. Nanotechnol. **9**, 851 (2014).

[19] V. A. Margulis, E. E. Muryumin, and E. A. Gaiduk, J. Phys.: Condens. Matter **25**, 195302 (2013).

[20] J. E. Sipe and E. Ghahramani, Phys. Rev. B. **48**, 11705 (1993).

[21] L. Wirtz, A. Marini, and A. Rubio, Phys. Rev. Lett. **96**, 126104 (2006).

[22] C. Attaccalite, M. Bockstedte, A. Marini, A. Rubio, and L. Wirtz, Phys. Rev. B. **83**, 144115 (2011).

[23] F. Hüser, T. Olsen, and K. S. Thygesen, Phys. Rev. B **87**, 235132 (2013).

[24] K. Watanabe, T. Taniguchi, and H. Kanda, Nat. Mater. **3**, 404 (2004).

[25] J. Zhu, J. Kang, J. Kang, D. Jariwala, J. D. Wood, J.-W. T. Seo, K.-S. Chen, T. J. Marks, and M. C. Hersam, Nano Lett. **15**, 7029 (2015).

[26] R. M. Ribeiro and N. M. R. Peres, Phys. Rev. B **83**, 235312 (2011).