Linear and nonlinear optical response of crystals using length and velocity gauges: Effect of basis truncation

Alireza Taghizadeh,[1,*] F. Hipolito,[1] and T. G. Pedersen[1,2]

[1]*Department of Physics and Nanotechnology, Aalborg University, DK-9220 Aalborg Øst, Denmark*
[2]*Center for Nanostructured Graphene, DK-9220 Aalborg Øst, Denmark*

We study the effects of a truncated band structure on the linear and nonlinear optical response of crystals using four methods. These are constructed by (i) choosing either the length or velocity gauge for the perturbation and (ii) computing the current density either directly or via the time derivative of the polarization density. In the infinite-band limit, the results of all four methods are identical, but basis truncation breaks their equivalence. In particular, certain response functions vanish identically and unphysical low-frequency divergences are observed for few-band models in the velocity gauge. Using a hexagonal boron nitride (hBN) monolayer as a case study, we analyze the problems associated with all methods and identify the optimal one. Our results show that the length-gauge calculations provide the fastest convergence rates as well as the most accurate spectra for any basis size and, moreover, that low-frequency divergences are eliminated.

I. INTRODUCTION

The optical response of crystals is essential for countless technological applications of solids as well as for characterization of materials. In semiconductors and insulators, optics provide access to important features of the band structure, including band gaps and transitions at high-symmetry points [1,2]. The optical response can be characterized by the linear response as well as various nonlinear responses, e.g., second-/third-harmonic generation (SHG/THG), optical rectification (OR), sum-/difference-frequency generation, etc. [3]. Several nonlinear optical (NLO) phenomena have important scientific and technical applications at energies ranging from the terahertz to visible wavelengths such as in laser technology, optical communication, biomolecular detection, and surface characterization [3,4]. The interest in NLO processes has recently grown dramatically due to the large response and exotic phenomena observed in two-dimensional (2D) materials such as graphene [5–10], hexagonal boron nitride (hBN) [11–14], and transition-metal dichalcogenides [15–19].

From a theoretical point of view, a reliable method for the computation of linear and nonlinear optical response functions based on the material band structure is crucial. For the linear optical response, such calculations are now routinely performed, and excellent agreement with experiments is obtained (see, e.g., Ref. [2]). However, the calculation of the NLO response of crystals remains an open subject as various methods for calculation, e.g., the choice between length and velocity gauges, frequently generate different results [20–33]. An extreme example of these differences emerges when considering systems whose electronic properties can be captured accurately by a two-band Hamiltonian, e.g., hBN [12,34] or the low-energy properties of biased bilayer graphene [13,35,36]. In such systems, the evaluation of the second-order (in fact, all even-order) response in the velocity gauge is identically zero at all frequencies, whereas the equivalent

calculation using the length gauge leads to finite results [12,14]. Moreover, computing the NLO response in the length gauge is significantly more complex than in the velocity gauge due to the appearance of generalized derivatives (GDs) [12,23,24].

In spite of the above-mentioned complexity, the linear and second-order conductivity expressions in the length-gauge approach can be derived from the expressions in the velocity gauge by a procedure explained in Ref. [22]. However, the procedure is exact only for an infinite number of bands. Identical length-gauge expressions were subsequently derived using an elegant and simple commutator-based approach in Ref. [23]. More recently, it has been shown that gauge invariance is upheld for any order of perturbation if a complete basis set is used for both calculations [32]. However, for many practical reasons, both analytic and numerical approaches to the calculation of the optical response rely on truncated basis sets that break the gauge invariance. The influence of basis truncation on the optical response in length and velocity gauges was discussed recently in Ref. [32], where it was predicted qualitatively that at low frequencies, the contributions from the omitted bands in the velocity gauge could be considerable. The gauge freedom, however, is not the only source of differences between commonly used computational approaches. Hence, different choices exist for the observable providing the optical response. Moreover, the GD in length-gauge calculations can be evaluated by its definition but also circumvented using an approximate sum rule presented in Refs. [23,37].

In this paper, all of these alternatives will be examined, emphasizing the effects of basis truncation for a real material (hBN) and in a broad frequency range. Hence, we compare four computational approaches to the optical response, including the linear optical conductivity (OC), OR, SHG, and THG of periodic systems, and study the convergence of each approach as a function of basis size. The four methods consist of the combinations of two choices of gauge, i.e., length and velocity gauges, and two ways of computing the current-density response: direct evaluation of the current density and evaluation via the time derivative of the polarization density. In addition, we investigate the effects arising from evaluating
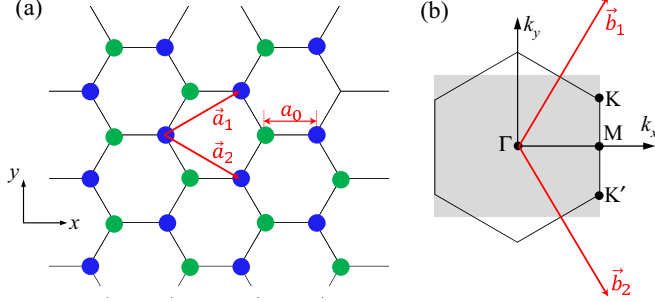
---

*ata@nano.aau.dk

(a)

(b)

FIG. 1. (a) Illustration of a 2D honeycomb lattice with a two-atom unit cell, i.e., a boron atom (green) and a nitrogen atom (blue), and the primitive vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. (b) The first Brillouin zone (BZ) with primitive reciprocal vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. The shaded region shows the integration area used in this work, which is equivalent to the first BZ.

the GDs by the above-mentioned sum rule for a truncated basis set. We choose monolayer hBN as a test case due to the simplicity of the band structure and the pronounced two-band character of the material. In order to have access to a variable-size basis set, we use an empirical pseudopotential Hamiltonian [38,39] that reproduces the low-energy properties of hBN monolayers [12,34]. We find that the length-gauge approach generates the most accurate results among the considered methods for any basis size. Moreover, we study the effect of basis-set truncation on the unphysical zero-frequency divergences plaguing velocity-gauge calculations [21–23]. Thus, our results provide guidelines for choosing the optimal computational method for the optical response based on a truncated band structure. This is essential in cases where the number of available bands is typically limited such as tight-binding (TB) models. Similarly, many-body calculations employing the Bethe-Salpeter equation frequently rely on a truncated band structure (see, e.g., Ref. [17]) due to the computational complexity. In both cases, an optimal combination of gauge and observable is crucial.

This paper is organized as follows. First, we present the pseudopotential approach for computing the electronic band structure of hBN using the energy dispersion of a two-band TB model for parametrization. Then, the dynamical equation of motion is reviewed in Sec. III, and its perturbative solution is derived up to the third order. Based on this solution, we numerically compute and compare the linear, SHG, OR, and THG conductivity spectra for the hBN monolayer using the four above-mentioned approaches and analyze the influence of basis truncation on the calculated spectra. Finally, a summary of results is presented in Sec. V.

II. PSEUDOPOTENTIAL HAMILTONIAN

The hBN monolayer is a 2D crystal with a honeycomb lattice consisting of two different atoms per unit cell as depicted in Fig. 1(a), with primitive lattice vectors $\mathbf{a}_1 = a(\sqrt{3}\mathbf{e}_x + \mathbf{e}_y)/2$ and $\mathbf{a}_2 = a(\sqrt{3}\mathbf{e}_x - \mathbf{e}_y)/2$, where a is the lattice constant related to the interatomic distance a_0 via $a = \sqrt{3}a_0$ ($a_0 = 1.45$ Å for the hBN [12,34]). The first Brillouin zone (BZ), associated high-symmetric points, and primitive reciprocal vectors $\mathbf{b}_1$ and $\mathbf{b}_2$ are illustrated in Fig. 1(b). It should be noted that, throughout the text, all vectors are

TABLE I. Pseudopotential form factors (in eV) for monolayer hBN obtained by fitting the band structure to a TB model [34]. The reciprocal vectors **G** are normalized by $2\pi/a$.

	$\|\mathbf{G}\|^2$		
	4/3	4	16/3
$\mathcal{V}_{\mathbf{G}}^S$	10.785	1.472	7.8
$\mathcal{V}_{\mathbf{G}}^{AS}$	8.007	0	3.697

indicated by bold letters, and Greek subscripts and superscripts denote the Cartesian components of vectors and tensors.

The characterization of the electronic properties of the system is based on an empirical pseudopotential Hamiltonian [38–40]. In this method, the real potential that governs the motion of electrons in the system is replaced by a simple effective potential, the *pseudopotential*, which lumps together the effects of core electrons as well as crystal nuclei [2]. By writing the Hamiltonian eigenstates as $\phi(\mathbf{r}) = \exp(i\mathbf{k}\cdot\mathbf{r})u(\mathbf{r})$, with a crystal momentum $\mathbf{k}$ and lattice-periodic function $u(\mathbf{r})$, the Schrödinger equation is transformed into an eigenvalue problem in reciprocal space as [38]

$$\sum_{\mathbf{G}'}\left[\frac{\hbar^2}{2m}|\mathbf{k}+\mathbf{G}|^2\delta_{\mathbf{G}\mathbf{G}'} + \mathcal{V}_{\mathbf{G}-\mathbf{G}'}\right]u_{\mathbf{G}'} = E\,u_{\mathbf{G}}, \quad (1)$$

where $\mathbf{G}$ and $\delta_{\mathbf{G}\mathbf{G}'}$ represent the reciprocal lattice vectors and Kronecker delta, respectively. The Fourier coefficients of the pseudopotential in the reciprocal space read $\mathcal{V}_{\mathbf{G}}$, while $u_{\mathbf{G}}$ and E denote the eigenvectors (the Fourier coefficients of u) and eigenvalues, respectively. The Fourier coefficients of the pseudopotential $\mathcal{V}_{\mathbf{G}}$ can be decomposed into symmetric and antisymmetric parts, the so-called *form factors* $\mathcal{V}_{\mathbf{G}}^S$ and $\mathcal{V}_{\mathbf{G}}^{AS}$, as [38]

$$\mathcal{V}_{\mathbf{G}} \equiv \mathcal{V}_{\mathbf{G}}^S\cos(\mathbf{G}\cdot\boldsymbol{\tau}) + i\mathcal{V}_{\mathbf{G}}^{AS}\sin(\mathbf{G}\cdot\boldsymbol{\tau}), \quad (2)$$

where $2\boldsymbol{\tau} = a/\sqrt{3}\mathbf{e}_x$ is the vector connecting the two atoms in the unit cell.

We limit the calculation of the form factors to the first four smallest $|\mathbf{G}|$ with squared magnitudes of 0, 4/3, 4, and 16/3 [normalized by $(2\pi/a)^2$]. The value of $\mathcal{V}_0^S$ is not important since it only shifts the energies, while the antisymmetric form factors for $|\mathbf{G}|^2 = 0$ and $4(2\pi/a)^2$ are not important since $\mathbf{G}\cdot\boldsymbol{\tau}$ is zero or $\pm\pi$, respectively. In addition, we adopt the spherical approximation for the pseudopotential [38], which reduces the total number of unknown form factors to five: three symmetric and two antisymmetric form factors. The form factors used for the hBN monolayer are presented in Table I. These were determined by fitting the pseudopotential band structure to the low-energy part, i.e., the vicinity of the K and M k points, of the TB band structure for the hBN monolayer [12,34,41] employing a nearest-neighbor hopping integral of $\gamma_0 = 2.33$ eV and on-site energies of ±3.9 eV.

In Fig. 2, we compare the energy dispersions of the pseudopotential and TB Hamiltonians. The pseudopotential Hamiltonian reproduces accurately the energy dispersion in the vicinity of the K and M points but deviates from the TB model at the BZ center, i.e., the Γ point, similar to the *ab initio* calculations of Ref. [34]. In this paper, we have employed 43
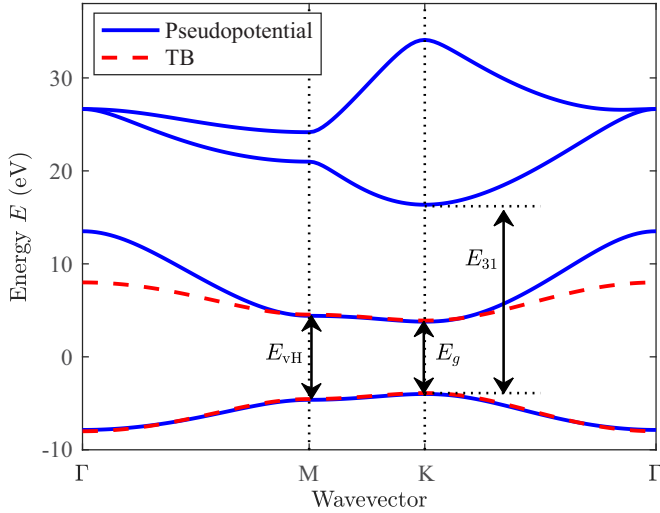
FIG. 2. Band structures of the hBN monolayer computed by the two-π-band TB model (red) and the pseudopotential Hamiltonian (blue). The TB parametrization follows Ref. [34], and the pseudopotential Hamiltonian is parametrized according to the form factors presented in Table I, using a total of 43 reciprocal lattice vectors for the eigenvalue problem [Eq. (1)]. Three important transition energies are defined: at the band gap $E_g \equiv E_2(K) - E_1(K)$, van Hove (vH) singularity $E_{\text{vH}} \equiv E_2(M) - E_1(M)$, and $E_{31} \equiv E_3(M) - E_1(M)$, where E_i denote the energy of the ith band.

reciprocal lattice vectors, which generates a total of 43 bands, 41 of which have dispersions above the TB conduction band. This large number of bands allows us to study the convergence of the optical response as a function of basis size, i.e., the number of bands used in the calculation. The band gap and van Hove transition energies are $E_g = 7.78$ eV and $E_{\text{vH}} = 9.04$ eV, respectively, while the transition to the second conduction band occurs at an energy of $E_{31} = 20.4$ eV (see Fig. 2).

III. CURRENT-DENSITY RESPONSE

Here, we briefly review the calculation of the optical response of a periodic system in equilibrium under the influence of an external electromagnetic field. The periodic system is characterized by an unperturbed Hamiltonian $\hat{H}_0$ and an external time-dependent perturbation $\hat{V}(t)$, such that the total Hamiltonian reads $\hat{H} = \hat{H}_0 + \hat{V}(t)$. The unperturbed Hamiltonian $\hat{H}_0$ leads to the pseudopotential eigenvalue problem (1). The form of the external potential depends on the gauge choice, and in the length gauge $\hat{V}_l(t) = e\hat{\mathbf{r}} \cdot \boldsymbol{\mathcal{E}}$, whereas in the velocity gauge, $V_v(t) = e(\hat{\mathbf{p}} \cdot \boldsymbol{\mathcal{A}} + e\boldsymbol{\mathcal{A}}^2/2)/m$. Here, $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{A}}$ are the electric field and vector potential, respectively. Both choices have their merits and shortcomings in the context of periodic systems. The latter benefits from the fact that the matrix elements of the momentum operator are easily computed but is plagued by spurious divergences since the electric field has to be mapped to the vector potential via $\boldsymbol{\mathcal{E}} = -\partial \boldsymbol{\mathcal{A}}/\partial t$. In contrast, the former requires a more elaborate calculation of the optical response but circumvents the unphysical divergences at zero frequency [23]. Throughout this work, the external electromagnetic field is defined by its

TABLE II. Four methods for computing the current-density response and their respective labels. The methods arise from the combination of gauge freedom and choice of physical observable under consideration, i.e. direct evaluation of the current or the time derivative of the polarization. Here, $\hat{\mathbf{r}}$, $\hat{\mathbf{p}}$, $\boldsymbol{\mathcal{E}}$, and $\boldsymbol{\mathcal{A}}$ represent the position operator, momentum operator, electric field, and vector potential, respectively.

label	$\hat{V}(t) \propto$	$\mathbf{J}(t) \propto$
$\mathcal{A}$	$\hat{\mathbf{r}} \cdot \boldsymbol{\mathcal{E}}$	$\text{tr}\{\hat{\mathbf{p}}\,\hat{\rho}(t)\}$
$\mathcal{B}$	$\hat{\mathbf{r}} \cdot \boldsymbol{\mathcal{E}}$	$\partial\text{tr}\{\hat{\mathbf{r}}\,\hat{\rho}(t)\}/\partial t$
$\mathcal{C}$	$\hat{\mathbf{p}} \cdot \boldsymbol{\mathcal{A}} + e\boldsymbol{\mathcal{A}}^2/2$	$\text{tr}\{(\hat{\mathbf{p}} + e\boldsymbol{\mathcal{A}})\,\hat{\rho}(t)\}$
$\mathcal{D}$	$\hat{\mathbf{p}} \cdot \boldsymbol{\mathcal{A}} + e\boldsymbol{\mathcal{A}}^2/2$	$\partial\text{tr}\{\hat{\mathbf{r}}\,\hat{\rho}(t)\}/\partial t$

decomposition into harmonic components,

$$\boldsymbol{\mathcal{E}}(t) = \frac{1}{2}\sum_p \sum_\alpha \mathbf{e}_\alpha \mathcal{E}_\alpha(\omega_p)e^{-i\omega_p t}, \tag{3}$$

where the p summation is performed over both positive and negative frequencies.

The optical response calculation relies on the evaluation of the time-dependent density operator, $\hat{\rho}(t) = \sum_{mn}\rho_{mn}(t)|m\rangle\langle n|$, governed by the quantum Liouville equation $i\hbar\partial\hat{\rho}(t)/\partial t = [\hat{H}, \hat{\rho}(t)]$. This equation is solved perturbatively (see the Appendix for details) to obtain the optical response either by evaluating directly the expectation value of the current-density operator, $\mathbf{J}(t) = \text{tr}\{\hat{\mathbf{J}}\,\hat{\rho}(t)\}$, or by computing the time derivative of the expectation value of the polarization-density operator, $\mathbf{J}(t) \equiv \partial\mathbf{P}(t)/\partial t = \partial\text{tr}\{\hat{\mathbf{P}}\,\hat{\rho}(t)\}/\partial t$ [12,42]. Here, the current- and polarization-density operators read $\hat{\mathbf{J}} = -eg\hat{\mathbf{p}}/(mA)$ and $\hat{\mathbf{P}} = -eg\hat{\mathbf{r}}/A$, respectively, where $g = 2$ accounts for the spin degeneracy and A is the crystal area. The combination of two gauges and two ways of evaluating the current-density response leads to a total of four approaches to compute the response, as summarized and labeled in Table II. Below, we briefly discuss the important details regarding the calculation of linear and nonlinear current-density responses in these four approaches using the perturbative expansion of the density matrix.

The evaluation of the density-matrix elements and direct calculation of the current-density response in the velocity gauge are a rather straightforward problem since the momentum operator is a well-defined operator in periodic systems. In contrast, both the perturbation evaluation in the length gauge and the calculation of the polarization density involve the ill-defined (in periodic systems) position operator $\hat{\mathbf{r}}$. In spite of the *prima facie* problems associated with the position operator, it has been shown in Ref. [23] (and references therein) that the optical response can be computed in this gauge by separating formally the interband and intraband parts of the position operator as $\hat{\mathbf{r}} = \hat{\mathbf{r}}^{(e)} + \hat{\mathbf{r}}^{(i)}$,

$$\mathbf{r}_{nm}^{(e)} \equiv \langle\phi_{n\mathbf{k}}|\hat{\mathbf{r}}^{(e)}|\phi_{m\mathbf{k}'}\rangle = (1 - \delta_{nm})\delta_{\mathbf{k}\mathbf{k}'}\boldsymbol{\Omega}_{nm}, \tag{4a}$$

$$\mathbf{r}_{nm}^{(i)} \equiv \langle\phi_{n\mathbf{k}}|\hat{\mathbf{r}}^{(i)}|\phi_{m\mathbf{k}'}\rangle = \delta_{nm}(\boldsymbol{\Omega}_{nn} + i\nabla_\mathbf{k})\delta_{\mathbf{k}\mathbf{k}'}, \tag{4b}$$

where the generalized Berry connections are defined as

$$\boldsymbol{\Omega}_{nm} \equiv \langle u_{n\mathbf{k}}|i\nabla_\mathbf{k}|u_{m\mathbf{k}}\rangle = \frac{i}{A_{\text{uc}}}\int_{\text{uc}} u_{n\mathbf{k}}^* \nabla_\mathbf{k} u_{m\mathbf{k}}\,d^2\mathbf{r}, \tag{5}$$

with the cell-periodic functions $u_{n\mathbf{k}}$ and unit-cell area A_{uc}. To simplify the notation, we frequently suppress the explicit dependence of quantities on the wave vector. The interband matrix elements of position and momentum operators are related by $im\mathbf{\Omega}_{nm} = \hbar\mathbf{p}_{nm}/E_{nm}$ [43] for $n \neq m$, where $E_{nm} \equiv E_n - E_m$. In addition, the intraband part of the position operator leads to the appearance of the GD, denoted typically by $(\cdot)_{;\mathbf{k}}$ [23,24]. For any simple operator (diagonal in $\mathbf{k}$) $\hat{O}$ the following expressions are then derived [23]:

$$\langle\phi_{n\mathbf{k}}|[\hat{\mathbf{r}}^{(i)}, \hat{O}]|\phi_{m\mathbf{k}'}\rangle = i\delta_{\mathbf{k}\mathbf{k}'}(O_{nm})_{;\mathbf{k}}, \tag{6a}$$

$$(O_{nm})_{;\mathbf{k}} \equiv \nabla_{\mathbf{k}}O_{nm} - i[\mathbf{\Omega}_{nn} - \mathbf{\Omega}_{mm}]O_{nm}. \tag{6b}$$

In addition, by virtue of the canonical commutation relation, i.e., $[\hat{r}^\alpha, \hat{p}^\beta] = i\hbar\delta_{\alpha\beta}$, and by separating the interband and intraband parts of the position operator, a sum rule is derived for the GD:

$$\left(p^\beta_{mn}\right)_{;k^\alpha} = \hbar\delta_{\alpha\beta}\delta_{mn} + \frac{\hbar}{m}\sum_l\left[\frac{\bar{\delta}_{ml}p^\alpha_{ml}p^\beta_{ln}}{E_{ml}} - \frac{\bar{\delta}_{ln}p^\beta_{ml}p^\alpha_{ln}}{E_{ln}}\right], \tag{7}$$

where we introduce $\bar{\delta}_{ml} \equiv 1 - \delta_{ml}$. An equivalent procedure was used for the GD evaluation in Refs. [23,24,37]. The basis truncation breaks this sum rule, thus opening a door for an additional convergence problem, as discussed in Sec. IV. Making use of the perturbative solution of the density matrix, Eq. (A2), and the above-mentioned results, we evaluate the current density up to the third order in the electric field strength.

We begin by addressing in detail the four possible methods to compute the linear response. Without loss of generality, the first-order current density $J^{(1)}_\lambda(t)$ reads

$$J^{(1)}_\lambda(t) = \sum_p\sum_\alpha \sigma^{(1)}_{\lambda\alpha}(\omega_p)\mathcal{E}_\alpha(\omega_p)e^{-i\omega_p t}. \tag{8}$$

The different methods of calculation lead to four conductivity tensors $\sigma^{(1)}_{\lambda\alpha}$, defined as

$$\sigma^{\mathcal{A}(1)}_{\lambda\alpha} \equiv C_e\sum_{\substack{\mathbf{k},n,m\\n\neq m}}\frac{p^\lambda_{nm}g^\alpha_{mn}}{E_{mn}} - C_i\sum_{\mathbf{k},n}\frac{p^\lambda_{nn}}{\hbar\omega_p}\frac{\partial f_n}{\partial k^\alpha}, \tag{9a}$$

$$\sigma^{\mathcal{B}(1)}_{\lambda\alpha} \equiv C_e\sum_{\substack{\mathbf{k},n,m\\n\neq m}}\frac{p^\lambda_{nm}g^\alpha_{mn}\hbar\omega_p}{E^2_{mn}} - C_i\sum_{\mathbf{k},n}\frac{m}{\hbar^2\omega_p}\frac{\partial E_n}{\partial k^\lambda}\frac{\partial f_n}{\partial k^\alpha}, \tag{9b}$$

$$\sigma^{\mathcal{C}(1)}_{\lambda\alpha} \equiv C_e\sum_{\mathbf{k},n,m}\frac{p^\lambda_{nm}g^\alpha_{mn}}{\hbar\omega_p} + \frac{ige^2}{2mA_{\text{uc}}\omega_p}\delta_{\lambda\alpha}, \tag{9c}$$

$$\sigma^{\mathcal{D}(1)}_{\lambda\alpha} \equiv \sigma^{\mathcal{A}(1)}_{\alpha\lambda}. \tag{9d}$$

Here, $f_{nm} \equiv f_n - f_m$, with $f_n \equiv f(E_n)$ being the Fermi-Dirac distribution, and the summation over $\mathbf{k}$ implies an integral over the BZ, i.e., $(2\pi)^D\sum_{\mathbf{k}} \to A\int_{\text{BZ}}d^D\mathbf{k}$, with D being the dimension ($D = 2$ for hBN). Also, the indices $m,n,l \in \{1,2,\dots\}$ run over all the bands, and the constant coefficients C_e and C_i and variable g^α_{mn} are defined as

$$C_e \equiv C_i\frac{\hbar}{m} \equiv \frac{ige^2\hbar}{2m^2A}, \quad g^\alpha_{mn} \equiv \frac{f_{nm}p^\alpha_{mn}}{\hbar\omega_p - E_{mn}}. \tag{10}$$

Similarly, the second- and third-order current-density responses are determined by

$$J^{(2)}_\lambda(t) = \sum_{p,q}\sum_{\alpha,\beta}\sigma^{(2)}_{\lambda\alpha\beta}(\omega_p,\omega_q)\mathcal{E}_\alpha(\omega_p)\mathcal{E}_\beta(\omega_q)e^{-i(\omega_p+\omega_q)t}, \tag{11}$$

$$J^{(3)}_\lambda(t) = \sum_{p,q,s}\sum_{\alpha\beta\gamma}\sigma^{(3)}_{\lambda\alpha\beta\gamma}(\omega_p,\omega_q,\omega_s)$$
$$\times \mathcal{E}_\alpha(\omega_p)\mathcal{E}_\beta(\omega_q)\mathcal{E}_\gamma(\omega_s)e^{-i(\omega_p+\omega_q+\omega_s)t}. \tag{12}$$

Given their complicated form, the expressions for the quadratic and cubic conductivity tensors can be found in Eqs. (A3), (A6), and (A7) in the Appendix.

To characterize the dependence of the optical response on the number of bands (including both valence and conduction bands) N_b, we define a convergence measure that quantifies the difference with respect to the evaluation with a large number of bands. In our calculations, the reference number was set to $N_{\text{ref}} = 20$. For the quantification, a *truncation inaccuracy* $\Delta(N_b)$ is defined as

$$\Delta(N_b) \equiv \frac{\langle|\sigma(\omega,N_b) - \sigma(\omega,N_{\text{ref}})|^2\rangle}{\langle|\sigma(\omega,N_{\text{ref}})|^2\rangle}, \tag{13}$$

where $\langle|\sigma(\omega)|^2\rangle \equiv \int_{\omega_i}^{\omega_f}|\sigma(\omega)|^2 d\omega/(\omega_f - \omega_i)$, with ω_i and ω_f being the integration bounds.

IV. RESULTS AND DISCUSSION

In this section, we address the dependence of the optical conductivity and several nonlinear processes on the basis truncation. Given the symmetry properties of the honeycomb lattice for hBN and restricting the external field to the in-plane directions (the crystal plane), it is sufficient to consider the diagonal components of the first-, second- [14], and third-order [44] conductivity tensors, namely. $\sigma^{(1)}_{xx}$, $\sigma^{(2)}_{xxx}$, and $\sigma^{(3)}_{xxxx}$. It should be noted that to ensure an adiabatic turn-on of the field, a positive infinitesimal value, $\eta = 0^+$, should be added to the frequency, i.e., $\omega_p \to \omega_p + i\eta$. Throughout the paper, we set $\hbar\eta = 0.03$ eV to account for line broadening. Regarding the integration over the BZ, we discretize the rectangular area of Fig. 1(b), which is equivalent to the first BZ, by at least 140 000 k points. We start by presenting results for the linear response (optical conductivity), then proceed to the second-order interactions SHG and OR and, finally, the third-order response. At the end, we compare quantitatively the truncation inaccuracy of the computed linear and nonlinear spectra.

A. Linear response

From the onset, Eqs. (9a)–(9d) show that in the presence of time-reversal symmetry the nondiagonal components vanish [45]. The results obtained from methods $\mathcal{D}$ and $\mathcal{A}$ are equivalent, and neither introduces unphysical divergences in the evaluation of the current-density response of cold insulators [23]. In addition, $\sigma^{\mathcal{A}(1)}$ can be shown to be formally equivalent to $\sigma^{\mathcal{B}(1)}$ regardless of the basis size. The intraband parts of $\sigma^{\mathcal{A}(1)}$ and $\sigma^{\mathcal{B}(1)}$ are identical simply due to the well-known result $m\partial E_n/\partial k^\lambda = \hbar p^\lambda_{nn}$ [46]. The interband part of $\sigma^{\mathcal{A}(1)}_{\alpha\alpha}$
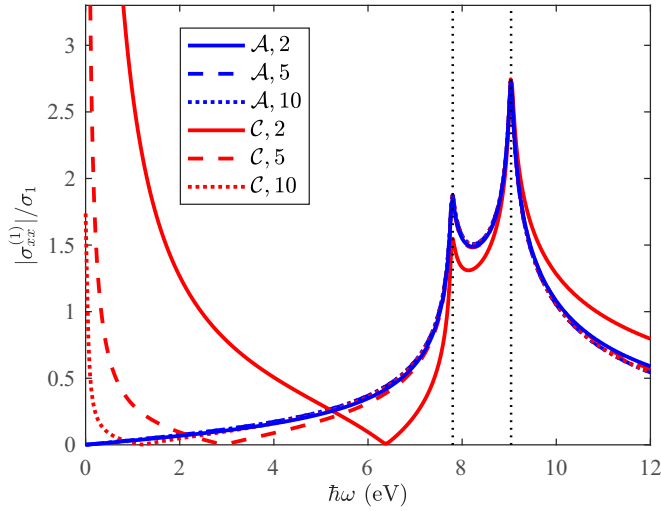
FIG. 3. OC spectrum of the hBN monolayer obtained from methods $\mathcal{A}$ (blue) and $\mathcal{C}$ (red) for $N_b = 2$ (solid lines), $N_b = 5$ (dashed lines), and $N_b = 10$ (dotted lines). The values are normalized to $\sigma_1 \equiv e^2/4\hbar = 6.0853 \times 10^{-5}$ S. The black dotted lines indicate $\hbar\omega = \{E_g, E_{vH}\}$.
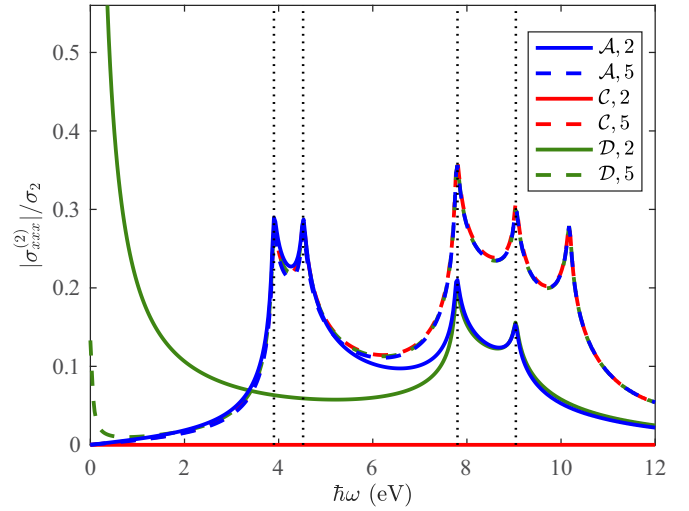

FIG. 4. SHG spectrum of the hBN monolayer obtained from methods $\mathcal{A}$ (blue), $\mathcal{C}$ (red), and $\mathcal{D}$ (green) for $N_b = 2$ (solid lines) and $N_b = 5$ (dashed lines). The values are normalized to $\sigma_2 \equiv e^3 a/8\gamma_0\hbar = 3.2797 \times 10^{-15}$ S m V^{-1}. The black dotted lines from left to right mark $2\hbar\omega = E_g$, $2\hbar\omega = E_{vH}$, $\hbar\omega = E_g$, and $\hbar\omega = E_{vH}$, respectively.

can be rewritten as

$$
C_e \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{p_{nm}^\alpha g_{mn}^\alpha}{E_{mn}} = \frac{C_e}{2} \left[\sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{p_{nm}^\alpha g_{mn}^\alpha}{E_{mn}} + (n \leftrightarrow m) \right]
$$

$$
= C_e \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{f_{nm} |p_{nm}^\alpha|^2}{E_{mn}} \frac{\hbar\omega_p}{(\hbar\omega_p)^2 - E_{mn}^2}, \quad (14)
$$

where in the first line $(n \leftrightarrow m)$ indicates that the preceding term should be written with exchanged dummy indices n and m. By the same token, the interband part of $\sigma_{\alpha\alpha}^{B(1)}$ can be rewritten as Eq. (14). With respect to the result derived with method $\mathcal{C}$, it has been shown in Ref. [22] that the expression of method $\mathcal{A}$, Eq. (9a), can be derived from the expression for method $\mathcal{C}$, Eq. (9c), by means of a sum rule. This sum rule is exact if and only if a complete basis set is used [32].

The truncation of the basis set breaks the equivalence between $\mathcal{A}$ and $\mathcal{C}$, leading to deviations between the optical response computed in these two methods. In Fig. 3, we compare the frequency dependence of the OC magnitude computed with methods $\mathcal{A}$ and $\mathcal{C}$ for three basis sets, $N_b = \{2, 5, 10\}$. The variation of the length-gauge results is quite small and not visible on the scale of Fig. 3. In contrast, the velocity-gauge response $\sigma^{\mathcal{C}(1)}$ is strongly dependent on the number of bands. In particular, the zero-frequency divergence is strongly suppressed with increasing N_b. Notwithstanding this strong suppression, the zero-frequency divergence remains present for any finite basis set. In addition, the features associated with the band gap and van Hove (vH) singularity also converge to the results computed using the length gauge.

B. Second-order response

The conductivity tensors for the four methods are shown in Eq. (A3). Based on these expressions, we numerically demonstrate the equivalence of Eqs. (A3a)–(A3d) for a large basis set. It should be mentioned that one can derive the expression of method $\mathcal{A}$ from method $\mathcal{C}$ by a procedure that includes separating the interband and intraband terms as demonstrated in Ref. [22]. However, the procedure is exact only if an infinite number of bands is employed. Figure 4 illustrates the SHG conductivities $\sigma_{xxx}^{(2)}(\omega, \omega)$ for two representative sizes of the basis set, namely, $N_b = \{2, 5\}$. The results obtained by method $\mathcal{B}$ are numerically identical to those of method $\mathcal{A}$ and hence are omitted. Considering the response in the vicinity of the lower-energy features, i.e., $2\hbar\omega \lesssim E_{vH}$, our results show that the calculation based on method $\mathcal{A}$ exhibits a small dependence on the basis size. In striking contrast, both methods $\mathcal{C}$ and $\mathcal{D}$ present highly different results. The former is identically zero for all frequencies since, for a two-band model, the integrand of Eq. (A3c) is an odd function of $\mathbf{k}$ due to time-reversal symmetry. It should be emphasized that the zero response of method $\mathcal{C}$ for the two-band model is a consequence of an oversimplification rather than a physical symmetry of the problem. Regarding the response in method $\mathcal{D}$, it is nonzero but exhibits a zero-frequency divergence and does not reproduce the SHG features associated with $2\hbar\omega \sim \{E_g, E_{vH}\}$. At higher energies, $\hbar\omega \sim \{E_g, E_{vH}\}$, the responses computed with the four methods show strong variations with the increase of N_b. It is important to note that this variation arises from interactions between the valence band and the second conduction band, i.e., $2\hbar\omega \sim E_{31}$ (see Fig. 2). Hence, the deviation of the spectra in this frequency range should not be considered a convergence issue. Rather, it is a consequence of the limited frequency range, for which the two-band model is applicable.
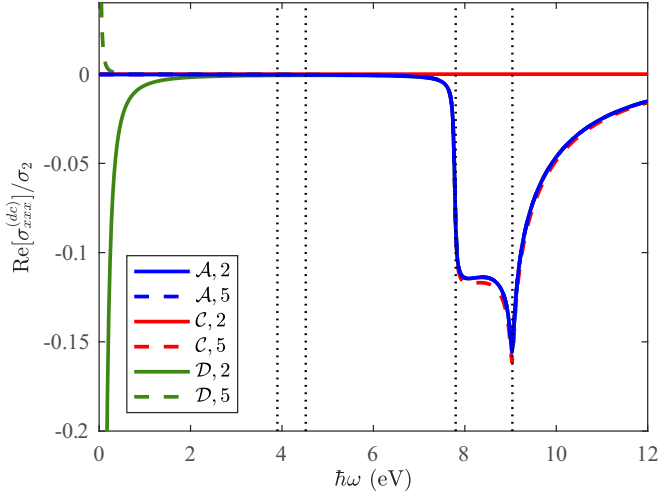
FIG. 5. OR spectrum of the hBN monolayer obtained from methods $\mathcal{A}$ (blue), $\mathcal{C}$ (red), and $\mathcal{D}$ (green) for $N_b = 2$ (solid lines) and $N_b = 5$ (dashed lines). The values are normalized to σ_2 (see the Fig. 4 caption). The dotted lines from left to right indicate $2\hbar\omega = E_g$, $2\hbar\omega = E_{\mathrm{vH}}$, $\hbar\omega = E_g$, and $\hbar\omega = E_{\mathrm{vH}}$, respectively.

The basis truncation also affects the calculation of other second-order processes such as OR, $\sigma_{xxx}^{(2)}(\omega, -\omega)$, as illustrated in Fig. 5. The OR results are similar to those obtained for SHG, but given the fact that the OR conductivity does not contain a $2\hbar\omega - E_{cv}$ term in the denominator, the response starts at $\hbar\omega \sim E_g$. In this case, the most significant differences appear in method $\mathcal{C}$, where, similar to the SHG process, the two-band calculation yields a zero response regardless of the external photon frequency. Moreover, the results generated by method $\mathcal{D}$ feature once more a spurious zero-frequency divergence, which is suppressed gradually by including more bands in the calculations.

For the results in Figs. 4 and 5, we evaluate the GDs appearing in Eqs. (A3) by employing the definition, Eq. (6). However, as pointed out in Sec. III, one may employ the sum rule, Eq. (7), for evaluating the GDs present in the intraband parts of $\sigma^{\mathcal{A}(2)}$, $\sigma^{\mathcal{B}(2)}$, and $\sigma^{\mathcal{D}(2)}$. Following this substitution, up to machine precision, all three methods generate SHG and OR spectra identical to those of method $\mathcal{C}$ for any size of basis set. This means that, for a two-band model, the length-gauge approach produces identically zero response similar to the velocity gauge if one uses the sum rule for evaluating the GD. Thus, the choice between exact and approximate implementations of the GD is of considerable importance for a truncated basis set.

C. Third-order response

Here, we limit our analysis to the effects of truncation in the calculation of THG, $\sigma_{xxxx}^{(3)}(\omega, \omega, \omega)$, via methods $\mathcal{A}$ and $\mathcal{C}$. In Fig. 6(a), we compare the THG conductivity computed by both methods for three values of N_b. The THG results follow the trends observed in the linear and quadratic responses. First, the results computed via method $\mathcal{A}$ display only a weak dependence on the size of the truncated basis. Second, the velocity-gauge results exhibit strong zero-frequency divergences that decrease, although slowly, with the
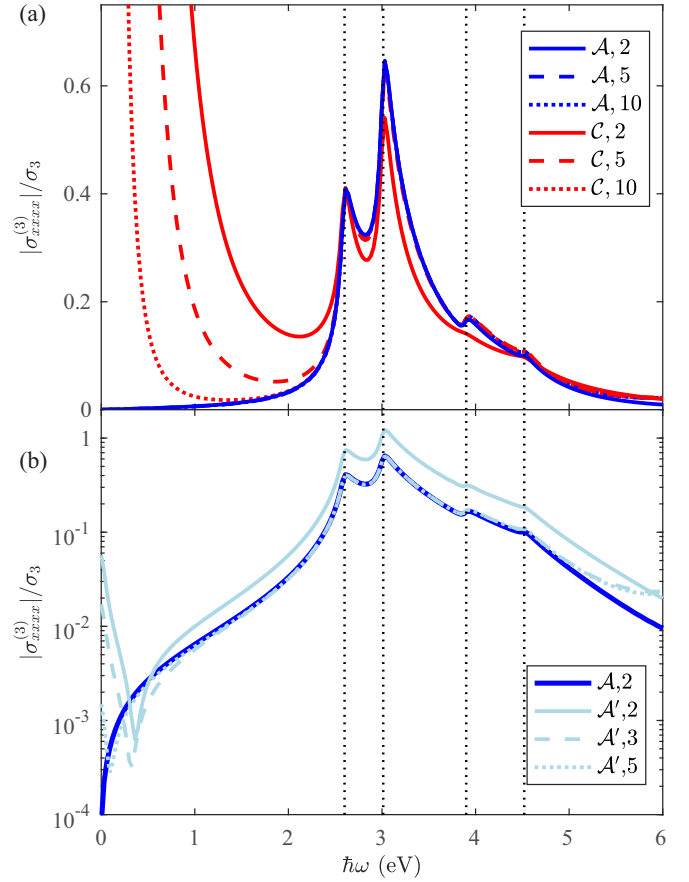

FIG. 6. (a) THG spectrum of the hBN monolayer obtained from methods $\mathcal{A}$ (blue) and $\mathcal{C}$ (red) for basis sets with a size of $N_b = 2$ (solid lines), $N_b = 5$ (dashed lines), and $N_b = 10$ (dotted lines). The values are normalized to $\sigma_3 \equiv e^4 a^2 / 16\gamma_0^2 \hbar = 1.7675 \times 10^{-25}$ Sm^2 V^{-2}. The black dotted lines from left to right mark $3\hbar\omega = E_g$, $3\hbar\omega = E_{\mathrm{vH}}$, $2\hbar\omega = E_g$, and $2\hbar\omega = E_{\mathrm{vH}}$, respectively. (b) THG spectrum of the hBN monolayer (on a log scale) obtained from method $\mathcal{A}$ by the direct evaluation of GD [Eq. (6); dark blue solid line] for $N_b = 2$ or by employing the sum rule [Eq. (7); light blue lines], labeled as $\mathcal{A}'$, for $N_b = 2$ (solid line), $N_b = 3$ (dashed line), and $N_b = 5$ (dotted line).

increasing number of bands. Compared to the linear response, the divergence at zero frequency in method $\mathcal{C}$ in the third-order spectrum is stronger, i.e., as ω^{-3} vs ω^{-1}.

In Fig. 6(b), we compare the THG spectra obtained with method $\mathcal{A}$ either by calculating the GDs directly using Eq. (6) or by employing the sum rule (7), labeling the latter approach as $\mathcal{A}'$. The spectra are plotted on a log scale to be more illustrative. For $N_b = 2$, the results show that the approximate GD from Eq. (7) overestimates the response in the frequency range of $3\hbar\omega \sim \{E_g, E_{\mathrm{vH}}\}$ and $2\hbar\omega \sim \{E_g, E_{\mathrm{vH}}\}$. By increasing N_b to 3 and then 5, however, the 2–5 eV features of method $\mathcal{A}'$ converge to those of method $\mathcal{A}$. Even so, the low-frequency response in $\mathcal{A}'$ still deviates considerably from method $\mathcal{A}$, which demonstrates the need for a large basis set. As in the SHG case, it should be noted that the high-frequency ($\hbar\omega > 5$ eV) deviation can be attributed to transitions involving higher conduction bands and hence is of less importance.
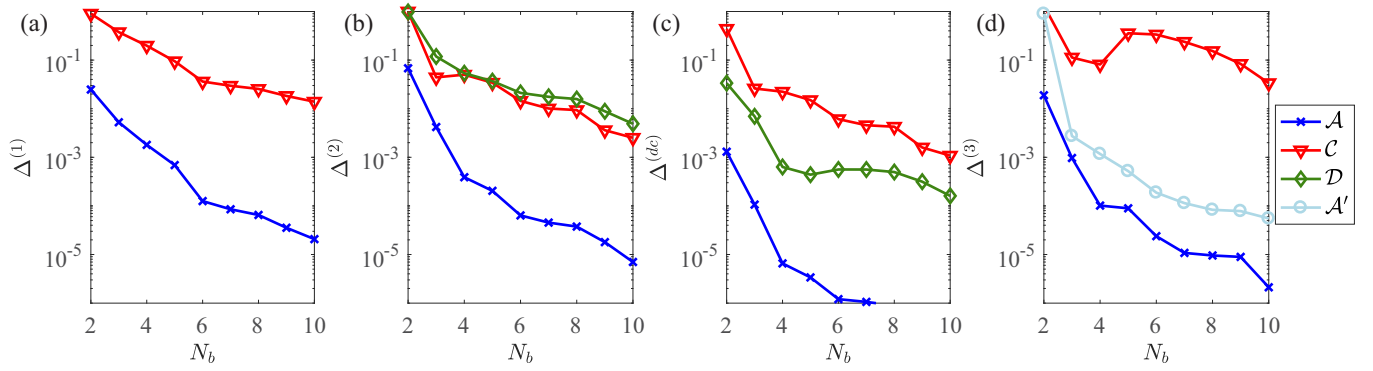
FIG. 7. Truncation inaccuracy $\Delta(N_b)$, defined in Eq. (13), for the four processes under consideration, namely, (a) OC, (b) SHG, (c) OR, and (d) THG, obtained by employing methods $\mathcal{A}$ (blue), $\mathcal{C}$ (red), and $\mathcal{D}$ (green). Method $\mathcal{A}'$ (light blue) indicates the modified $\mathcal{A}$, for which the GDs are approximated by Eq. (7). Reference results are computed with $N_{\mathrm{ref}} = 20$. The lower spectral limit for the analysis is set at $\hbar\omega_i = 0.1E_g$, and the upper limits read (a) $\hbar\omega_f = 1.5E_g$, (b) $\hbar\omega_f = 0.7E_g$, (c) $\hbar\omega_f = 1.5E_g$, and (d) $\hbar\omega_f = 0.7E_g$.

D. Convergence analysis

Figures 7(a) to 7(d) show the truncation inaccuracy $\Delta(N_b)$ defined in Eq. (13) as a function of basis size N_b for linear, SHG, OR, and THG responses of monolayer hBN, respectively. Since the two length-gauge approaches $\mathcal{A}$ and $\mathcal{B}$ generate numerically identical spectra for all linear and nonlinear processes, the truncation inaccuracy of method $\mathcal{B}$ is omitted. Note, however, that method $\mathcal{B}$ typically requires additional efforts compared to $\mathcal{A}$ in its numerical implementation due to the additional position operator. The convergence behavior of all four investigated linear and nonlinear processes is qualitatively similar. For instance, the length-gauge approaches with the direct evaluation of GDs converge faster with respect to the basis set size than the velocity-gauge methods, i.e., $\mathcal{C}$ and $\mathcal{D}$, in all cases. If the sum rule of Eq. (7) is employed to evaluate the GD appearing in methods $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{D}$, the truncation inaccuracies will be identical to that of method $\mathcal{C}$ for the second-order responses. Moreover, the truncation inaccuracy computed in the velocity gauge, method $\mathcal{C}$, increases significantly for the third-order response due to the strong zero-frequency divergence. Also, for the third-order responses, the modified $\mathcal{A}$ method, $\mathcal{A}'$, generates more accurate spectra than method $\mathcal{C}$ when at least three bands are used in the calculation. Nevertheless, it underperforms when compared with the original method $\mathcal{A}$ for all basis sizes considered.

V. SUMMARY

In summary, we have investigated the effects of basis truncation on several linear and nonlinear optical response functions, including the OC, SHG, OR, and THG. The conductivity tensors were derived and compared using four computationally different approaches. These result from combining two choices of gauges and two ways of evaluating the current density, i.e., directly or via the polarization. For the OC, the equivalence of all four methods has been demonstrated analytically provided a complete basis set is used, whereas for the NLO response, we have demonstrated it numerically by employing a large basis set in calculations. The length-gauge approaches, i.e., tensors labeled with $\mathcal{A}$

or $\mathcal{B}$, generate the most accurate spectra, compared to the velocity-gauge approaches, particularly for small basis sets. In addition, it has been shown that the choice of method to compute the GD is crucial as the evaluation based on the sum rule, Eq. (7), may result in degrading convergence. Finally, although the well-known zero-frequency divergences in the velocity-gauge responses vanish by increasing the size of basis set, the calculated spectra are far less accurate than the ones generated by the length-gauge methods. Our results shed light on the source of the differences arising from several commonly used computational approaches to the linear and nonlinear optical responses.

APPENDIX: NONLINEAR CONDUCTIVITY TENSORS

1. Perturbative density matrix

The equation of motion can be solved by employing a perturbative approach and expanding the solution as a power series of the perturbation: $\rho_{mn}(t) = \sum_{N=0}^{\infty} \rho_{mn}^{(N)}(t)$, where $\rho_{mn}^{(N)}(t)$ at order N is determined iteratively from the previous order via

$$\rho_{mn}^{(N)}(t) = \frac{1}{i\hbar} \int_{-\infty}^{t} [\hat{V}(t'), \hat{\rho}^{(N-1)}(t')]_{mn} e^{iE_{mn}(t'-t)/\hbar} dt'. \quad \text{(A1)}$$

In the absence of any perturbation, the system is at equilibrium, and its density-matrix elements are determined by $\rho_{mn}^{(0)} \equiv \delta_{mn} f_n$. Hence, by integrating Eq. (A1) for a set of time-harmonic perturbations, $\hat{V}(t) = 1/2 \sum_p \hat{V}(\omega_p) \exp(-i\omega_p t)$,

the first three terms read [3]

$$\rho_{mn}^{(1)}(t) = \frac{1}{2} \sum_p f_{nm} \frac{V_{mn}(\omega_p)}{\hbar\omega_p - E_{mn}} e^{-i\omega_p t}, \tag{A2a}$$

$$\rho_{mn}^{(2)}(t) = \frac{1}{4} \sum_{p,q} \sum_l \frac{e^{-i(\omega_p+\omega_q)t}}{\hbar(\omega_p+\omega_q) - E_{mn}} \left[\frac{f_{nl} V_{ml}(\omega_q) V_{ln}(\omega_p)}{\hbar\omega_p - E_{ln}} - \frac{f_{lm} V_{ml}(\omega_p) V_{ln}(\omega_q)}{\hbar\omega_p - E_{ml}} \right], \tag{A2b}$$

$$\rho_{mn}^{(3)}(t) = \frac{1}{8} \sum_{p,q,s} \sum_{j,l} \frac{e^{-i(\omega_p+\omega_q+\omega_s)t}}{\hbar(\omega_p+\omega_q+\omega_s) - E_{mn}} \left\{ \frac{V_{mj}(\omega_p)}{\hbar(\omega_q+\omega_s) - E_{jn}} \left[\frac{f_{nl} V_{jl}(\omega_q) V_{ln}(\omega_s)}{\hbar\omega_s - E_{ln}} - \frac{f_{lj} V_{jl}(\omega_s) V_{ln}(\omega_q)}{\hbar\omega_s - E_{jl}} \right] \right.$$

$$\left. + \frac{V_{ln}(\omega_p)}{\hbar(\omega_q+\omega_s) - E_{ml}} \left[\frac{f_{jm} V_{mj}(\omega_s) V_{jl}(\omega_q)}{\hbar\omega_s - E_{mj}} - \frac{f_{lj} V_{mj}(\omega_q) V_{jl}(\omega_s)}{\hbar\omega_s - E_{jl}} \right] \right\}, \tag{A2c}$$

where V_{mn} are the matrix elements of the perturbation, i.e., $V_{mn}(\omega) \equiv \langle m|\hat{V}(\omega)|n\rangle$. Two choices of gauge are employed for the perturbative Hamiltonian $\hat{V}$, i.e., $\hat{V}_l$ and $\hat{V}_v$, as defined in Sec. III.

2. Second-order tensors

Here, we show the tensor expressions derived for the four methods in Table II using Eq. (A2b). The expression for method $\mathcal{C}$ is obtained straightforwardly since it contains only the matrix elements of the well-defined momentum operator. For methods $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{D}$, the position operator has to be separated into its interband and intraband parts, and it should be treated carefully, as outlined briefly in Sec. III and in detail in Ref. [23].

$$\sigma_{\lambda\alpha\beta}^{\mathcal{A}(2)}(\omega_p,\omega_q) \equiv C_{ee} \sum_{\substack{\mathbf{k},n,m,l \\ n \neq l \neq m}} \frac{p_{nm}^\lambda (g_{ln}^\alpha p_{ml}^\beta - g_{ml}^\alpha p_{ln}^\beta)}{E_{ml} E_{ln}[\hbar(\omega_p+\omega_q) - E_{mn}]} + C_{ie} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{-p_{nm}^\lambda}{\hbar(\omega_p+\omega_q) - E_{mn}} \left(\frac{g_{mn}^\alpha}{E_{mn}} \right)_{;k^\beta}$$

$$+ C_{ie} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{-p_{nm}^\lambda}{E_{mn}[\hbar(\omega_p+\omega_q) - E_{mn}]} \frac{p_{mn}^\beta}{\hbar\omega_p} \frac{\partial f_{nm}}{\partial k^\alpha} + C_{ii} \sum_{\mathbf{k},n} \frac{p_{nn}^\lambda}{\hbar(\omega_p+\omega_q)(\hbar\omega_p)} \frac{\partial^2 f_n}{\partial k^\beta \partial k^\alpha}, \tag{A3a}$$

$$\sigma_{\lambda\alpha\beta}^{\mathcal{B}(2)}(\omega_p,\omega_q) \equiv C_{ee} \sum_{\substack{\mathbf{k},n,m,l \\ n \neq m \neq l \neq n}} \frac{\hbar(\omega_p+\omega_q) p_{nm}^\lambda (g_{ln}^\alpha p_{ml}^\beta - g_{ml}^\alpha p_{ln}^\beta)}{E_{mn} E_{ml} E_{ln}[\hbar(\omega_p+\omega_q) - E_{mn}]} + C_{ie} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{-\hbar(\omega_p+\omega_q) p_{nm}^\lambda}{E_{mn}[\hbar(\omega_p+\omega_q) - E_{mn}]} \left(\frac{g_{mn}^\alpha}{E_{mn}} \right)_{;k^\beta}$$

$$+ C_{ie} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{-\hbar(\omega_p+\omega_q) g_{mn}^\alpha (p_{nm}^\beta)_{;k^\lambda}}{2E_{mn}^2} \frac{1}{\hbar\omega_q + E_{mn}} + C_{ie} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{-\hbar(\omega_p+\omega_q) p_{nm}^\lambda}{E_{mn}^2[\hbar(\omega_p+\omega_q) - E_{mn}]} \frac{p_{mn}^\beta}{\hbar\omega_p} \frac{\partial f_{nm}}{\partial k^\alpha}, \tag{A3b}$$

$$\sigma_{\lambda\alpha\beta}^{\mathcal{C}(2)}(\omega_p,\omega_q) \equiv C_{ee} \frac{1}{(\hbar\omega_p)(\hbar\omega_q)} \sum_{\mathbf{k},n,m,l} \frac{p_{nm}^\lambda}{\hbar(\omega_p+\omega_q) - E_{mn}} (g_{ln}^\alpha p_{ml}^\beta - g_{ml}^\alpha p_{ln}^\beta), \tag{A3c}$$

$$\sigma_{\lambda\alpha\beta}^{\mathcal{D}(2)}(\omega_p,\omega_q) \equiv C_{ee} \frac{\hbar(\omega_p+\omega_q)}{(\hbar\omega_p)(\hbar\omega_q)} \sum_{\substack{\mathbf{k},n,m,l \\ n \neq m}} \frac{p_{nm}^\lambda (g_{ln}^\alpha p_{ml}^\beta - g_{ml}^\alpha p_{ln}^\beta)}{\hbar(\omega_p+\omega_q) - E_{mn}} + C_{ie} \frac{\hbar(\omega_p+\omega_q)}{2(\hbar\omega_p)(\hbar\omega_q)} \sum_{\mathbf{k},n,m} \frac{g_{mn}^\alpha (p_{nm}^\beta)_{;k^\lambda}}{\hbar\omega_q + E_{mn}}, \tag{A3d}$$

where g_{mn}^α has been defined in Eq. (10) and the constants C_{ee}, C_{ie}, and C_{ii} read

$$C_{ee} \equiv C_{ie} \frac{\hbar}{m} \equiv C_{ii} \frac{\hbar^2}{m^2} \equiv \frac{ge^3\hbar^2}{4m^3 A}. \tag{A4}$$

The expression for tensor $\sigma_{\lambda\alpha\beta}^{\mathcal{A}(2)}$ consists of four terms: one purely interband contribution, two mixed interband-intraband contributions, and one purely intraband contribution. By the same token, similar interband and intraband contributions in $\sigma_{\lambda\alpha\beta}^{\mathcal{B}(2)}$ can be identified. In contrast, the interband and intraband contributions are not separated in the expression for $\sigma_{\lambda\alpha\beta}^{\mathcal{C}(2)}$, and they are only partly divided in $\sigma_{\lambda\alpha\beta}^{\mathcal{D}(2)}$ due to the presence of $\hat{\mathbf{r}}$ in the current-density operator. For a cold, intrinsic semiconductor, the terms including derivatives of the band population f_n vanish, e.g., the last two terms of $\sigma_{\lambda\alpha\beta}^{\mathcal{A}(2)}$ and the last part of $\sigma_{\lambda\alpha\beta}^{\mathcal{B}(2)}$. The conductivity expressions in Eqs. (A3) can be symmetrized with respect to the frequencies and indices by performing a

permutation [24]. In deriving the conductivity tensors, the following useful expressions that can be derived from Eq. (6) have been used:

$$\lim_{\mathbf{k} \to \mathbf{k}'} (f_{n\mathbf{k}} - f_{m\mathbf{k}'}) \mathbf{r}_{mn}^{(i)} = i \delta_{nm} \nabla_{\mathbf{k}} f_n, \tag{A5a}$$

$$\left[\mathbf{r}_{nn}^{(i)} - \mathbf{r}_{mm}^{(i)} \right] p_{nm}^{\alpha} = i \delta_{\mathbf{k}\mathbf{k}'} \left(p_{nm}^{\alpha} \right)_{;\mathbf{k}}. \tag{A5b}$$

Finally, the GDs appearing in Eqs. (A3a), (A3b), and (A3d) can be computed either directly from the definition (6) or by using the sum rule (7). For the latter, the chain-rule property of the GD and an additional expression, $(E_{mn})_{;k^{\alpha}} = \hbar (p_{mm}^{\alpha} - p_{nn}^{\alpha})/m$, are utilized [23].

3. Third-order tensors

The third-order conductivity tensor components $\sigma_{\lambda\alpha\beta\gamma}^{(3)}$ are derived by inserting the length- and velocity-gauge perturbative Hamiltonian into the density-matrix elements of Eq. (A2c). Here, we present only the conductivity tensors obtained in methods $\mathcal{A}$ and $\mathcal{C}$. As in the case of the second-order tensor, $\sigma_{\lambda\alpha\beta\gamma}^{\mathcal{C}(3)}$ is obtained straightforwardly as

$$\sigma_{\lambda\alpha\beta\gamma}^{\mathcal{C}(3)}(\omega_p, \omega_q, \omega_s) \equiv C_{eee} \frac{1}{(\hbar\omega_p)(\hbar\omega_q)(\hbar\omega_s)} \sum_{\mathbf{k},n,m,j,l} \frac{p_{nm}^{\lambda}}{\hbar\omega_3 - E_{mn}} \left[\frac{p_{mj}^{\alpha}(g_{ln}^{\gamma} p_{jl}^{\beta} - g_{jl}^{\gamma} p_{ln}^{\beta})}{\hbar\omega_2 - E_{jn}} + \frac{p_{ln}^{\alpha}(g_{mj}^{\gamma} p_{jl}^{\beta} - g_{jl}^{\gamma} p_{mj}^{\beta})}{\hbar\omega_2 - E_{ml}} \right], \tag{A6}$$

where we introduce the coefficient $C_{eee} \equiv g e^4 \hbar^3/(i 8 m^4 A)$ and auxiliary variables $\omega_2 \equiv \omega_q + \omega_s$, $\omega_3 \equiv \omega_p + \omega_q + \omega_s$, $g_{mn}^{\gamma} \equiv f_{nm} p_{mn}^{\gamma}/(\hbar\omega_s - E_{mn})$. On the other hand, the calculation of $\sigma_{\lambda\alpha\beta\gamma}^{\mathcal{A}(3)}$ requires dividing the ill-defined position operator into interband and intraband parts. The resulting eight different combinations of interband and intraband terms, denoted by eee, eei, eie, eii, iee, iei, iie, and iii, are given by

$$\sigma_{\lambda\alpha\beta\gamma}^{\mathcal{A}(3)}(\omega_p, \omega_q, \omega_s) \equiv \sigma_{\lambda\alpha\beta\gamma}^{(3,eee)} + \sigma_{\lambda\alpha\beta\gamma}^{(3,eei)} + \sigma_{\lambda\alpha\beta\gamma}^{(3,eie)} + \sigma_{\lambda\alpha\beta\gamma}^{(3,eii)} + \sigma_{\lambda\alpha\beta\gamma}^{(3,iee)} + \sigma_{\lambda\alpha\beta\gamma}^{(3,iei)} + \sigma_{\lambda\alpha\beta\gamma}^{(3,iie)} + \sigma_{\lambda\alpha\beta\gamma}^{(3,iii)}, \tag{A7a}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,eee)} \equiv C_{eee} \sum_{\substack{\mathbf{k},n,m,j,l \\ n \neq l \neq j \neq m}} \frac{p_{nm}^{\lambda}}{\hbar\omega_3 - E_{mn}} \left[\frac{p_{mj}^{\alpha}(g_{ln}^{\gamma} p_{jl}^{\beta} - g_{jl}^{\gamma} p_{ln}^{\beta})}{E_{mj} E_{jl} E_{ln}(\hbar\omega_2 - E_{jn})} + \frac{p_{ln}^{\alpha}(g_{mj}^{\gamma} p_{jl}^{\beta} - g_{jl}^{\gamma} p_{mj}^{\beta})}{E_{mj} E_{jl} E_{ln}(\hbar\omega_2 - E_{ml})} \right], \tag{A7b}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,eei)} \equiv C_{iee} \frac{1}{\hbar\omega_s} \sum_{\substack{\mathbf{k},n,m,l \\ n \neq l \neq m}} \frac{-p_{nm}^{\lambda}}{E_{ln} E_{ml}(\hbar\omega_3 - E_{mn})} \left[\frac{p_{ml}^{\alpha} p_{ln}^{\beta}}{\hbar\omega_2 - E_{ln}} \frac{\partial f_{nl}}{\partial k^{\gamma}} + \frac{p_{ln}^{\alpha} p_{ml}^{\beta}}{\hbar\omega_2 - E_{ml}} \frac{\partial f_{ml}}{\partial k^{\gamma}} \right], \tag{A7c}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,eie)} \equiv C_{iee} \sum_{\substack{\mathbf{k},n,m,l \\ n \neq l \neq m}} \frac{p_{nm}^{\lambda}}{\hbar\omega_3 - E_{mn}} \left[\frac{p_{ml}^{\alpha}/E_{ml}}{\hbar\omega_2 - E_{ln}} \left(\frac{g_{ln}^{\gamma}}{E_{ln}} \right)_{;k^{\beta}} - \frac{p_{ln}^{\alpha}/E_{ln}}{\hbar\omega_2 - E_{ml}} \left(\frac{g_{ml}^{\gamma}}{E_{ml}} \right)_{;k^{\beta}} \right], \tag{A7d}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,eii)} \equiv C_{iie} \frac{1}{(\hbar\omega_s)(\hbar\omega_2)} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{p_{nm}^{\lambda} p_{mn}^{\alpha}/E_{mn}}{(\hbar\omega_3 - E_{mn})} \frac{\partial^2 f_{nm}}{\partial k^{\beta} \partial k^{\gamma}}, \tag{A7e}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,iee)} \equiv C_{iee} \sum_{\substack{\mathbf{k},n,m,l \\ n \neq l \neq m}} \left[\frac{p_{nm}^{\lambda}}{\hbar\omega_3 - E_{mn}} \right]_{;k^{\alpha}} \frac{g_{ln}^{\gamma} p_{ml}^{\beta} - g_{ml}^{\gamma} p_{ln}^{\beta}}{E_{ml} E_{nl}(\hbar\omega_2 - E_{mn})}, \tag{A7f}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,iei)} \equiv C_{iie} \frac{1}{\hbar\omega_s} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \frac{-p_{nm}^{\lambda}}{\hbar\omega_3 - E_{mn}} \left[\frac{p_{mn}^{\beta}/E_{mn}}{\hbar\omega_2 - E_{mn}} \frac{\partial f_{mn}}{\partial k^{\gamma}} \right]_{;k^{\alpha}}, \tag{A7g}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,iie)} \equiv C_{iie} \sum_{\substack{\mathbf{k},n,m \\ n \neq m}} \left[\frac{p_{nm}^{\lambda}}{\hbar\omega_3 - E_{mn}} \right]_{;k^{\alpha}} \frac{-1}{\hbar\omega_2 - E_{mn}} \left(\frac{g_{mn}^{\gamma}}{E_{mn}} \right)_{;k^{\beta}}, \tag{A7h}$$

$$\sigma_{\lambda\alpha\beta\gamma}^{(3,iii)} \equiv C_{iii} \frac{1}{(\hbar\omega_s)(\hbar\omega_2)(\hbar\omega_3)} \sum_{\mathbf{k},n} \left(-p_{nn}^{\lambda} \right) \frac{\partial^3 f_n}{\partial k^{\alpha} \partial k^{\beta} \partial k^{\gamma}}, \tag{A7i}$$

where $C_{iee} \hbar/m \equiv C_{iie} \hbar^2/m^2 \equiv C_{iii} \hbar^3/m^3 \equiv C_{eee}$ and integration by parts has been used in the k summation when deriving Eqs. (A7f) and (A7h). It should be noted that Eqs. (A6) and (A7) have not been symmetrized with respect to the frequencies and

indices, which can be performed by permutation of frequencies and component indices [47]. For a clean, cold semiconductor only the eee, eie, iee, and iie terms are nonzero.

[1] P. Basu, *Theory of Optical Processes in Semiconductors: Bulk and Microstructures*, Oxford Science Publications (Clarendon Press, Oxford, 2002).

[2] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors: Physics and Materials Properties* (Springer, Berlin, 2010).

[3] R. W. Boyd, *Nonlinear Optics*, 3rd ed. (Elsevier, Amsterdam, 2008).

[4] Y. R. Shen, *The Principles of Nonlinear Optics* (Wiley-Interscience, Hoboken, NJ, 2002).

[5] E. Hendry, P. J. Hale, J. Moger, A. K. Savchenko, and S. A. Mikhailov, Phys. Rev. Lett. **105**, 097401 (2010).

[6] S.-Y. Hong, J. I. Dadap, N. Petrone, P.-C. Yeh, J. Hone, and R. M. Osgood, Phys. Rev. X **3**, 021014 (2013).

[7] Y. Wang, M. Tokman, and A. Belyanin, Phys. Rev. B **94**, 195442 (2016).

[8] S. A. Mikhailov, Phys. Rev. B **93**, 085403 (2016).

[9] B. Semnani, A. H. Majedi, and S. Safavi-Naeini, J. Opt. **18**, 035402 (2016).

[10] Z. Sun, A. Martinez, and F. Wang, Nat. Photon. **10**, 227 (2016).

[11] Y. Li, Y. Rao, K. F. Mak, Y. You, S. Wang, C. R. Dean, and T. F. Heinz, Nano Lett. **13**, 3329 (2013).

[12] T. G. Pedersen, Phys. Rev. B **92**, 235432 (2015).

[13] S. J. Brun and T. G. Pedersen, Phys. Rev. B **91**, 205405 (2015).

[14] F. Hipolito, T. G. Pedersen, and V. M. Pereira, Phys. Rev. B **94**, 045434 (2016).

[15] N. Kumar, S. Najmaei, Q. Cui, F. Ceballos, P. M. Ajayan, J. Lou, and H. Zhao, Phys. Rev. B **87**, 161403 (2013).

[16] R. Wang, H.-C. Chien, J. Kumar, N. Kumar, H.-Y. Chiu, and H. Zhao, ACS Appl. Mater. Interfaces **6**, 314 (2013).

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

[18] K. F. Mak, K. L. McGill, J. Park, and P. L. McEuen, Science **344**, 1489 (2014).

[19] K. F. Mak and J. Shan, Nat. Photon. **10**, 216 (2016).

[20] D. H. Kobe, Phys. Rev. A **19**, 205 (1979).

[21] D. J. Moss, E. Ghahramani, J. E. Sipe, and H. M. van Driel, Phys. Rev. B **41**, 1542 (1990).

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

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

[24] J. E. Sipe and A. I. Shkrebtii, Phys. Rev. B **61**, 5337 (2000).

[25] L. B. Madsen, Phys. Rev. A **65**, 053417 (2002).

[26] K. Rzążewski and R. W. Boyd, J. Mod. Opt. **51**, 1137 (2004).

[27] K. S. Virk and J. E. Sipe, Phys. Rev. B **76**, 035213 (2007).

[28] Y. J. Chen and B. Hu, Phys. Rev. A **80**, 033408 (2009).

[29] H. M. Dong, K. Han, and W. Xu, J. Appl. Phys. **115**, 063503 (2014).

[30] M. Wu, S. Ghimire, D. A. Reis, K. J. Schafer, and M. B. Gaarde, Phys. Rev. A **91**, 043839 (2015).

[31] P. Földi, Phys. Rev. B **96**, 035112 (2017).

[32] G. B. Ventura, D. J. Passos, J. M. B. Lopes dos Santos, J. M. Viana Parente Lopes, and N. M. R. Peres, Phys. Rev. B **96**, 035431 (2017).

[33] C. Wang, X. Liu, L. Kang, B.-L. Gu, Y. Xu, and W. Duan, Phys. Rev. B **96**, 115147 (2017).

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

[35] E. V. Castro, K. S. Novoselov, S. V. Morozov, N. M. R. Peres, J. M. B. Lopes dos Santos, J. Nilsson, F. Guinea, A. K. Geim, and A. H. Castro Neto, Phys. Rev. Lett. **99**, 216802 (2007).

[36] A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, and A. K. Geim, Rev. Mod. Phys. **81**, 109 (2009).

[37] S. N. Rashkeev and W. R. L. Lambrecht, Phys. Rev. B **63**, 165212 (2001).

[38] M. L. Cohen and T. K. Bergstresser, Phys. Rev. **141**, 789 (1966).

[39] J. R. Chelikowsky and M. L. Cohen, Phys. Rev. B **14**, 556 (1976).

[40] J. C. Phillips, Phys. Rev. **112**, 685 (1958).

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

[42] S. Sharma and C. Ambrosch-Draxl, Phys. Scr. **2004**, 128 (2004).

[43] B. Gu, N. H. Kwong, and R. Binder, Phys. Rev. B **87**, 125301 (2013).

[44] X.-l. Yang and S.-w. Xie, Appl. Opt. **34**, 6130 (1995).

[45] J. L. P. Hughes and J. E. Sipe, Phys. Rev. B **53**, 10751 (1996).

[46] N. Ashcroft and N. Mermin, *Solid State Physics* (Saunders College, Philadelphia, 1976).

[47] F. Hipolito and T. G. Pedersen, arXiv:1708.06553.