Quadratic response theory for spin-orbit coupling in semiconductor heterostructures

Bradley A. Foreman*
Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
(Received 17 June 2005; revised manuscript received 31 August 2005; published 28 October 2005)

This paper examines the properties of the self-energy operator in lattice-matched semiconductor heterostructures, focusing on nonanalytic behavior at small values of the crystal momentum, which gives rise to long-range Coulomb potentials. A nonlinear response theory is developed for nonlocal spin-dependent perturbing potentials. The ionic pseudopotential of the heterostructure is treated as a perturbation of a bulk reference crystal, and the self-energy is derived to second order in the perturbation. If spin-orbit coupling is neglected outside the atomic cores, the problem can be analyzed as if the perturbation were a local spin scalar, since the nonlocal spin-dependent part of the pseudopotential merely renormalizes the results obtained from a local perturbation. The spin-dependent terms in the self-energy therefore fall into two classes: short-range potentials that are analytic in momentum space, and long-range nonanalytic terms that arise from the screened Coulomb potential multiplied by a spin-dependent vertex function. For an insulator at zero temperature, it is shown that the electronic charge induced by a given perturbation is exactly linearly proportional to the charge of the perturbing potential. These results are used in a subsequent paper to develop a first-principles effective-mass theory with generalized Rashba spin-orbit coupling.

DOI: 10.1103/PhysRevB.72.165344 PACS number(s): 73.21.−b, 73.61.Ey, 71.15.Ap

I. INTRODUCTION

The Rashba Hamiltonian[1] is the prototype of a class of effective-mass Hamiltonians describing spin-orbit coupling in semiconductors.[2] These models have been under intensive study in the past several years due to theoretical and experimental advances in spin-related phenomena such as the intrinsic spin Hall effect[3–17] and the spin galvanic and circular photogalvanic effects.[18–25] Such effects are generated by spin-orbit coupling terms in the conduction or valence bands of clean nonmagnetic semiconductors. Although a variety of different two- and three-dimensional semiconductor systems are under investigation, one of the most widely studied is a heterostructure between semiconductors with the zinc-blende structure, in which an external electric field can be used to tune the relative contributions from the Rashba and Dresselhaus spin-splitting terms.[25–27]

In a two-dimensional effective-mass model, the Rashba coupling has no coordinate dependence. But in a three-dimensional theory, it is usually separated into (1) a contribution proportional to the macroscopic electric field generated by gate voltages, dopants, and free carriers; and (2) δ functions representing the contribution from the rapid change in potential at a heterojunction.[2,28–39] However, the assumption that a heterojunction can be represented by a short-range δ potential has never been justified from first principles. In a self-consistent theory with electron-electron interactions, there are in general long-range Coulomb multipole potentials that are not well localized at the interface. These long-range potentials contribute spin-dependent terms to the Hamiltonian. Thus it is important to establish the conditions under which such terms will appear in the effective-mass Hamiltonian for a heterojunction.

This paper examines the long-range terms in the self-energy of a quasiparticle for the case of a lattice-matched semiconductor heterostructure. The potential energy of the ions is described using norm-conserving pseudopotentials.[40–42] The heterostructure pseudopotential is treated as a perturbation of a bulk reference crystal, with the self-energy calculated to second order in the perturbation using quadratic response theory.[43–48] This approach is well justified numerically, since the linear response alone has been shown to give excellent predictions for the valence band offset in a variety of material systems, including isovalent and heterovalent heterostructures.[49–57]

The approach used here follows closely earlier work by Sham[58] on the theory of shallow impurity states in bulk semiconductors. Sham's work is generalized to include nonlocal spin-dependent potentials and terms of higher order in the crystal momentum. Even for local spin-independent potentials, the present work includes terms neglected in Sham's analysis, such as dipole potentials in the quadratic response.

This paper is limited to a study of the electron self-energy in the limit of small crystal momentum. The derivation of an effective-mass Hamiltonian from these results is presented in the following paper.[59]

As a workable approximation scheme, the calculation of the linear response is carried out to terms two orders in q higher than the lowest nonvanishing term, where $\mathbf{q}$ is the crystal momentum transfer of the perturbing potential. The quadratic response is calculated to the same order in q as the lowest nonvanishing term in the linear response. (See the following paper[59] for further discussion of this approximation scheme.) Three classes of heterostructure perturbations are considered:

(I) Heterovalent perturbations with nonzero charge. In this case the perturbing potential includes a monopole of $O(q^{-2})$, and the analysis is performed to an accuracy of $O(q^0)$ in the linear response and $O(q^{-2})$ in the quadratic response.

(II) Isovalent perturbations for which the linear response has a nonzero dipole moment. In this case the leading term in the linear response is $O(q^{-1})$, so the linear response is evaluated to $O(q)$ and the quadratic response is evaluated to $O(q^{-1})$.

(III) Isovalent perturbations for which the linear response has no dipole moment. In this case the leading term in the linear response is $O(q^0)$, so the linear response is evaluated to $O(q^2)$ and the quadratic response is evaluated to $O(q^0)$.

The perturbations that make up a given heterostructure are generally a mixture of classes I, II, and III. The simplest situation is that of an isovalent heterostructure made up of semiconductors with the zinc-blende structure, such as GaAs/AlAs or InAs/GaSb. In this case, every ionic perturbation is an isovalent perturbation from class III.

Most theoretical and experimental studies of the Rashba spin-splitting Hamiltonian have dealt with this type of heterostructure. This case is therefore studied in greatest detail here, by working out the explicit form of the self-energy from crystal symmetry. The results show that in this case the Rashba Hamiltonian contains only short-range terms (to within the accuracy of the stated approximation scheme). However, there are long-range spin-dependent terms that are not of the Rashba form.

In a heterovalent system such as Ge/GaAs, the ionic perturbations are from class I. However, since macroscopic accumulations of charge are energetically unfavorable, real heterostructures tend to be macroscopically neutral.[51,60] Such a nominally heterovalent class I problem can therefore often be reduced to an isovalent class II or III problem by grouping the ions together in neutral clusters and treating these clusters as the basic unit. This approach is discussed further in Appendix A.

In wurtzite heterostructures such as GaN/AlN, the ionic perturbations are from class II, since the site symmetry of atoms in the wurtzite structure (space group C_{6v}^4) permits a dipole moment. These dipole terms produce spontaneous polarization along the hexagonal c axis in bulk wurtzite crystals, leading to macroscopic interface charge at heterojunctions.[61–68] Such charge produces macroscopic electric fields that generate different piezoelectric strain fields in different materials. The present theory, which is restricted to lattice-matched heterostructures, is therefore not generally applicable to wurtzite systems (except in the unrealistic special case[61] where the interface polarization charge is exactly cancelled by an external interface charge). However, the results derived here provide a first step towards a more general theory dealing with lattice-mismatched heterostructures.

The paper begins in Sec. II by establishing the basic definitions and notation for the Green function and self-energy that are used throughout the paper. The finite-temperature formalism is used (both for generality and because it facilitates the derivation of Ward identities), although the main interest of this paper is the limit of an insulator at zero temperature. In Sec. III, the self-energy is expanded in powers of the perturbing potential using vertex functions. A set of Ward identities is derived for the vertex functions at finite and zero temperature. Section IV presents general expressions for the nonlinear density response, including Ward identities for the static polarization. The bare ionic perturbations are screened in Sec. V, where the proper vertex functions and proper polarization are introduced.

A detailed analysis of the small wave vector properties of the linear screened potential is carried out in Sec. VI for the special case of a local spin-independent perturbation. The quadratic response for the same case is considered in Sec. VII, and the linear and quadratic contributions to the self-energy are derived in Sec. VIII. In Sec. IX it is shown that in the norm-conserving pseudopotential formalism, the contributions from the nonlocal spin-dependent part of the perturbing potential merely renormalize the contributions from the local part of the perturbation. The main results of the paper are discussed and summarized in Sec. X.

II. GREEN FUNCTION AND SELF-ENERGY

This section establishes the notation, basic definitions, and symmetry properties of the Green function and self-energy used in subsequent sections of the paper. The starting point is the definition of the one-particle thermal Green function[69–71]

$$G_{ss'}(\mathbf{x},\tau;\mathbf{x}',\tau') = -\langle T_\tau[\hat{\psi}_s(\mathbf{x},\tau)\hat{\psi}_{s'}^\dagger(\mathbf{x}',\tau')]\rangle, \tag{2.1}$$

where $s=\pm\frac{1}{2}$ labels the z component of the spin, τ is the imaginary time, T_τ is the time ordering operator, and $\hat{\psi}_s(\mathbf{x},\tau)=e^{\hat{K}\tau}\hat{\psi}_s(\mathbf{x})e^{-\hat{K}\tau}$ and $\hat{\psi}_s^\dagger(\mathbf{x},\tau)=e^{\hat{K}\tau}\hat{\psi}_s^\dagger(\mathbf{x})e^{-\hat{K}\tau}$ are field operators in the Heisenberg picture. The angular brackets denote a thermal average

$$\langle\hat{O}\rangle = e^{\beta\Omega}\mathrm{Tr}(e^{-\beta\hat{K}}\hat{O}), \tag{2.2}$$

where $\beta=1/k_BT$ is the inverse temperature, Tr denotes a trace over the many-particle Fock space, $\hat{K}=\hat{H}-\mu\hat{N}$ is the grand Hamiltonian (with μ the chemical potential and $\hat{N}$ the number operator), and $e^{-\beta\Omega}=\mathrm{Tr}(e^{-\beta\hat{K}})$. The many-particle Hamiltonian is defined by

$$\hat{H} = \sum_{s,s'}\int\int \hat{\psi}_s^\dagger(\mathbf{x})h_{ss'}(\mathbf{x},\mathbf{x}')\hat{\psi}_{s'}(\mathbf{x}')d^3x\,d^3x' + \frac{1}{2}\sum_{s,s'}\int\int\frac{\hat{\psi}_s^\dagger(\mathbf{x})\hat{\psi}_{s'}^\dagger(\mathbf{x}')\hat{\psi}_{s'}(\mathbf{x}')\hat{\psi}_s(\mathbf{x})}{|\mathbf{x}-\mathbf{x}'|}d^3x\,d^3x', \tag{2.3}$$

where $h=h^\dagger$ is the Hamiltonian of a single noninteracting particle, and Hartree atomic units are used.

Since $\hat{K}$ is time independent, G has the form $G_{ss'}(\mathbf{x},\tau;\mathbf{x}',\tau')=G_{ss'}(\mathbf{x},\mathbf{x}',\tau-\tau')$, with $G_{ss'}(\mathbf{x},\mathbf{x}',\tau-\beta)=-G_{ss'}(\mathbf{x},\mathbf{x}',\tau)$ for $0<\tau<\beta$. This permits the Fourier series representation (for $-\beta<\tau<\beta$)

$$G(\tau) = \frac{1}{\beta}\sum_{n=-\infty}^{\infty}G(\zeta_n)e^{-i\zeta_n\tau}, \tag{2.4a}$$

$$G(\zeta_n) = \int_0^\beta G(\tau)e^{i\zeta_n\tau}d\tau, \tag{2.4b}$$

where $\zeta_n=(2n+1)\pi/\beta$, and $G(\tau)$ denotes a single-particle operator whose matrix elements in the $|\mathbf{x},s\rangle$ basis are $G_{ss'}(\mathbf{x},\mathbf{x}',\tau)$. A continuous Green function $G(\omega)$ may then

be defined by analytic continuation of $G(\zeta_n)$ from the discrete frequencies $\omega=\mu+i\zeta_n$.

The Green function $G(\omega)$ satisfies Dyson's equation

$$[\omega - h - \Sigma(\omega)]G(\omega) = G(\omega)[\omega - h - \Sigma(\omega)] = 1, \quad (2.5)$$

which is an implicit definition for the self-energy operator

$$\Sigma(\omega) = \omega - h - G^{-1}(\omega). \quad (2.6)$$

A formal solution to Eq. (2.5) can be constructed by solving the nonhermitian eigenvalue equations

$$[h + \Sigma(\omega)]|\psi_n(\omega)\rangle = E_n(\omega)|\psi_n(\omega)\rangle, \quad (2.7a)$$

$$[h + \Sigma^\dagger(\omega)]|\chi_n(\omega)\rangle = E_n^*(\omega)|\chi_n(\omega)\rangle, \quad (2.7b)$$

which are also referred to as Dyson's equations. It is usually assumed that the solutions to (2.7) form a complete biorthonormal[72] set with the properties

$$\langle\chi_n(\omega)|\psi_{n'}(\omega)\rangle = \delta_{nn'}, \quad (2.8a)$$

$$\sum_n |\psi_n(\omega)\rangle\langle\chi_n(\omega)| = 1, \quad (2.8b)$$

although this is difficult to prove in general.[73] If Eqs. (2.8) are valid, then $G(\omega)$ is given by[72,74]

$$G(\omega) = \sum_n \frac{|\psi_n(\omega)\rangle\langle\chi_n(\omega)|}{\omega - E_n(\omega)}, \quad (2.9)$$

which satisfies the Dyson equation (2.5) by construction.

Symmetries of the many-particle Hamiltonian $\hat{H}$ under time reversal and space group operations imply corresponding symmetries of the one-particle operators G and Σ. Derivations of the most useful symmetry relations are presented in Appendix B.

III. VERTEX FUNCTIONS

In this section, a perturbative approach to the Dyson equation (2.7a) is developed by using vertex functions to expand the self-energy in powers of the perturbing potential.

A. Definitions

The single-particle Hamiltonian h is chosen here to have the form

$$h_{ss'}(\mathbf{x},\mathbf{x}') = -\tfrac{1}{2}\nabla^2\delta(\mathbf{x}-\mathbf{x}')\delta_{ss'} + v^{\text{ext}}_{ss'}(\mathbf{x},\mathbf{x}'), \quad (3.1)$$

where the fixed external potential v^{ext} is a norm-conserving ionic pseudopotential,[40–42] which accounts for both spin-orbit coupling[75–78] and scalar relativistic effects. The Dyson equation (2.7a) is therefore

$$-\frac{1}{2}\nabla^2\psi_s(\mathbf{x},\omega) + \sum_{s'}\int V_{ss'}(\mathbf{x},\mathbf{x}',\omega)\psi_{s'}(\mathbf{x}',\omega)d^3x' = E(\omega)\psi_s(\mathbf{x},\omega), \quad (3.2)$$

in which ψ is a spinor wave function, and the total potential energy V is

$$V_{ss'}(\mathbf{x},\mathbf{x}',\omega) = v^{\text{ext}}_{ss'}(\mathbf{x},\mathbf{x}') + \Sigma_{ss'}(\mathbf{x},\mathbf{x}',\omega). \quad (3.3)$$

The external pseudopotential can be separated as $v^{\text{ext}} = v^{(0)} + v$, where $v^{(0)}$ is the potential of some periodic reference crystal, and v is a nonperiodic perturbation associated with a heterostructure or an impurity. It is assumed that the total potential (3.3) can be represented as a power series in the perturbation v:

$$V_{s_1s_2}(\mathbf{x}_1,\tau_1;\mathbf{x}_2,\tau_2) \equiv V(12) = \sum_{\nu=0}^{\infty} V^{(\nu)}(12), \quad (3.4)$$

where $V^{(0)}$ is the potential (3.3) when $v^{\text{ext}}=v^{(0)}$, and the numerical arguments on the right-hand side are shorthand for the space, spin, and time coordinates $(1)=(\mathbf{x}_1,s_1,\tau_1)$. Although the upper limit of the formal expansion (3.4) is written as $\nu=\infty$, this may well be an asymptotic series, and in practice only a finite number of terms are retained.

The linear and quadratic terms of (3.4) are

$$V^{(1)}(12) = \Gamma^{(1)}(1243)v(34),$$

$$V^{(2)}(12) = \frac{1}{2}\Gamma^{(2)}(124365)v(34)v(56), \quad (3.5)$$

where $\Gamma^{(\nu)}$ is called the vertex function of order ν. Here a summation or integration of repeated coordinates is assumed, and the labels are ordered as the trace of a matrix product. The perturbation v is taken to be an instantaneous static potential of the form

$$v(34) = v_{s_3s_4}(\mathbf{x}_3,\mathbf{x}_4)\delta(\tau_3-\tau_4). \quad (3.6)$$

The vertex functions are by definition functional derivatives of V with respect to v:

$$\Gamma^{(1)}(1243) = \frac{\delta V(12)}{\delta v(34)},$$

$$\Gamma^{(2)}(124365) = \frac{\delta^2 V(12)}{\delta v(34)\delta v(56)} = \frac{\delta\Gamma^{(1)}(1243)}{\delta v(56)}, \quad (3.7)$$

which may also be expressed as

$$\Gamma^{(1)}(1243) = \delta(13)\delta(24) + \frac{\delta\Sigma(12)}{\delta v(34)},$$

$$\Gamma^{(2)}(124365) = \frac{\delta^2\Sigma(12)}{\delta v(34)\delta v(56)}, \quad (3.8)$$

in which $\delta(12)=\delta_{s_1s_2}\delta(\mathbf{x}_1-\mathbf{x}_2)\delta(\tau_1-\tau_2)$. Note that upon application of the Fourier transforms defined in Appendix C, the above equations hold equally well in momentum and frequency space.

It is convenient at this point to carry out the time integrals in Eq. (3.5). This eliminates the variables $\tau_3,\tau_4,\ldots$ from $\Gamma^{(\nu)}$ and v, and reduces the time dependence of the $\delta(13)\delta(24)$ term in Eq. (3.8) to $\delta(\tau_1-\tau_2)$. It is assumed below that this has been done.

B. Ward identities

The vertex functions satisfy various Ward identities[79–81] for certain limiting values of their arguments. One set of these can be derived by varying the chemical potential μ by a small amount $\delta\mu$. Since $\hat{K}=\hat{H}-\mu\hat{N}$, this is equivalent to varying v by $\delta v(12)=-\delta\mu\,\delta(12)$. For this special case, Eqs. (3.7) and (3.8) reduce to

$$\Gamma^{(1)}(1233)=\delta(12)-\frac{\delta\Sigma(12)}{\delta\mu},$$

$$\Gamma^{(2)}(124355)=-\frac{\delta\Gamma^{(1)}(1243)}{\delta\mu}, \tag{3.9}$$

which involve a trace over the input variables of Γ. If Σ is analytically continued as a function of $\omega=\mu+i\zeta_n$, the variation with respect to μ may be written as

$$\frac{\delta}{\delta\mu}\rightarrow\frac{\partial}{\partial\omega}+\frac{\partial}{\partial\mu}. \tag{3.10}$$

Now for the special case of an insulator at $T=0$, the chemical potential can have any value in the range $\mu_{N-1}<\mu<\mu_N$, where μ_N is the minimum energy needed to add one particle to the ground state of an N-particle system, and the energy gap is $E_g=\mu_N-\mu_{N-1}$. [For small temperatures μ approaches the well-defined limit $\frac{1}{2}(\mu_N+\mu_{N-1})$,[82] but at exactly $T=0$ it becomes ill defined.] Since μ can vary arbitrarily within the gap for a system with finite E_g, the Ward identities for the insulator reduce to

$$\Gamma^{(1)}(1233)=\delta(12)-\frac{\partial\Sigma(12)}{\partial\omega},$$

$$\Gamma^{(2)}(124355)=-\frac{\partial\Gamma^{(1)}(1243)}{\partial\omega},$$

$$\Gamma^{(2)}(123344)=\frac{\partial^2\Sigma(12)}{\partial\omega^2}, \tag{3.11}$$

which generalize and extend the results derived for spin-independent local potentials in Refs. 58 and 80.

IV. NONLINEAR DENSITY RESPONSE

A. Definitions

In this section, perturbation theory (see Appendix D) is used to evaluate the electron density of the system with Hamiltonian $\hat{H}=\hat{H}_0+\hat{H}_1$, in which $\hat{H}_0$ is the Hamiltonian of the reference crystal and $\hat{H}_1$ is the perturbation due to v:

$$\hat{H}_1=\mathrm{tr}(\hat{\rho}v)=\sum_{s,s'}\int\int\hat{\rho}_{s's}(\mathbf{x}',\mathbf{x})v_{ss'}(\mathbf{x},\mathbf{x}')d^3xd^3x', \tag{4.1}$$

where $\hat{\rho}$ is the density operator

$$\hat{\rho}_{s's}(\mathbf{x}',\mathbf{x})=\hat{\psi}_s^\dagger(\mathbf{x})\hat{\psi}_{s'}(\mathbf{x}'). \tag{4.2}$$

The mean nonlocal electron density in the perturbed system is defined as

$$n_{ss'}(\mathbf{x},\mathbf{x}')=\langle\hat{\rho}_{ss'}(\mathbf{x},\mathbf{x}',\tau)\rangle, \tag{4.3}$$

which is independent of τ. If n is evaluated using the perturbation theory formula (D2), one obtains a power series in v:

$$n_{ss'}(\mathbf{x},\mathbf{x}')=\sum_{\nu=0}^{\infty}n_{ss'}^{(\nu)}(\mathbf{x},\mathbf{x}'), \tag{4.4}$$

in which

$$n_{ss'}^{(0)}(\mathbf{x},\mathbf{x}')=\langle\hat{\rho}_{ss'}(\mathbf{x},\mathbf{x}',\tau)\rangle_0 \tag{4.5}$$

is the density of the reference crystal. (The notation $\langle\hat{O}\rangle_0$ refers to a thermal average with respect to the reference crystal; see Appendix D.) The terms of order $\nu>0$ are given by

$$n^{(\nu)}(00')=\frac{1}{\nu!}\Pi^{(\nu)}(00',1'1,\ldots,\nu'\nu)v(11')\cdots v(\nu\nu'), \tag{4.6}$$

where $\Pi^{(\nu)}$ is the νth-order static polarization (or density correlation function), which is defined in Eq. (E1). Here and below, the numerical arguments of $n^{(\nu)}$, $\Pi^{(\nu)}$, and v are time-independent quantities of the form $(0)=(\mathbf{x}_0,s_0)$.

B. Ward identities

The linear and quadratic density response are given explicitly by

$$n^{(1)}(12)=\Pi^{(1)}(1243)v(34),$$

$$n^{(2)}(12)=\frac{1}{2}\Pi^{(2)}(124365)v(34)v(56), \tag{4.7}$$

which shows that Π may be defined as a functional derivative of n with respect to v:

$$\Pi^{(1)}(1243)=\frac{\delta n(12)}{\delta v(34)},$$

$$\Pi^{(2)}(124365)=\frac{\delta^2 n(12)}{\delta v(34)\delta v(56)}=\frac{\delta\Pi^{(1)}(1243)}{\delta v(56)}. \tag{4.8}$$

For the special case $\delta v(12)=-\delta\mu\delta(12)$ representing a variation in chemical potential of $\delta\mu$ (see Sec. III B), these expressions give

$$\Pi^{(1)}(1233)=-\frac{\partial n(12)}{\partial\mu},$$

$$\Pi^{(2)}(124355)=-\frac{\partial\Pi^{(1)}(1243)}{\partial\mu}, \tag{4.9}$$

which are the Ward identities for the static polarization. For an insulator at $T=0$, μ is indefinite, and these reduce to

$$\Pi^{(1)}(1233)=0, \quad \Pi^{(2)}(124355)=0. \tag{4.10}$$

V. SCREENING

A. Potential

In this section, the concept of screening is used to extract the long-range Coulomb interaction terms from the polarization and vertex functions. The first-order screened potential φ is defined by adding the Coulomb potential generated by $n^{(1)}$ to v:

$$\varphi(12)=v(12)+u(1243)\Pi^{(1)}(3465)v(56). \tag{5.1}$$

Here u represents the Coulomb interaction, which is spin-independent and local in coordinate space at both the input and output:

$$u(1243)=\delta_{s_1s_2}\delta_{s_3s_4}\delta(\mathbf{x}_1-\mathbf{x}_2)\delta(\mathbf{x}_3-\mathbf{x}_4)u(\mathbf{x}_1-\mathbf{x}_3). \tag{5.2}$$

In momentum space (see Appendix C) this has the form

$$u(1243)=\delta_{s_1s_2}\delta_{s_3s_4}\delta_{\mathbf{k}_1-\mathbf{k}_2,\mathbf{k}_3-\mathbf{k}_4}u(\mathbf{k}_1-\mathbf{k}_2), \tag{5.3}$$

in which $u(\mathbf{k})=v_c(\mathbf{k})/\Omega$, where Ω is the crystal volume and

$$v_c(\mathbf{k})=\begin{cases}4\pi/k^2 & \text{if } k\neq 0,\\ 0 & \text{if } k=0.\end{cases} \tag{5.4}$$

Another way of writing the screened potential is

$$\varphi(12)=\epsilon^{-1}(1243)v(34), \tag{5.5}$$

in which the inverse static electronic dielectric matrix is

$$\epsilon^{-1}(1243)=\delta(13)\delta(24)+u(1265)\Pi^{(1)}(5643). \tag{5.6}$$

The dielectric matrix ϵ satisfies

$$\epsilon(1243)\epsilon^{-1}(3465)=\epsilon^{-1}(1243)\epsilon(3465)=\delta(15)\delta(26), \tag{5.7}$$

and is given explicitly below in Eq. (5.17). For an insulator, the Ward identity (4.10) yields

$$\epsilon^{-1}(1233)=\delta(12). \tag{5.8}$$

The second-order potential $\varphi^{(2)}$ is just the Coulomb potential generated by $n^{(2)}$:

$$\varphi^{(2)}(12)=u(1243)n^{(2)}(34). \tag{5.9}$$

By translation symmetry, $\Pi^{(1)}(1243)=0$ unless $\mathbf{k}_1-\mathbf{k}_2=\mathbf{k}_3-\mathbf{k}_4+\mathbf{G}$, where $\mathbf{G}$ is a reciprocal lattice vector of the reference crystal. Equation (5.1) may therefore be written as

$$\varphi_{ss'}(\mathbf{k},\mathbf{k}')=v_{ss'}(\mathbf{k},\mathbf{k}')+\delta_{ss'}v_c(\mathbf{q})\sum_{\mathbf{k}''}\sum_{\mathbf{G}}\Pi_{\lambda'\lambda}(\mathbf{q};\mathbf{k}'',\mathbf{k}''+\mathbf{q}+\mathbf{G})v_{\lambda\lambda'}(\mathbf{k}''+\mathbf{q}+\mathbf{G},\mathbf{k}''), \tag{5.10}$$

where $\mathbf{q}\equiv\mathbf{k}-\mathbf{k}'$ and

$$\Pi_{\lambda'\lambda}(\mathbf{q};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G})=\frac{1}{\Omega}\sum_{\mathbf{k}_2}\Pi^{(1)}_{\alpha\alpha,\lambda'\lambda}(\mathbf{q}+\mathbf{k}_2,\mathbf{k}_2;\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}). \tag{5.11}$$

This simplified form of Π is introduced because the Coulomb potential depends only on the local spin-independent density $n(\mathbf{x})\equiv n_{\alpha\alpha}(\mathbf{x},\mathbf{x})$.

B. Vertex functions

Given Eqs. (5.5) and (5.9), one can rewrite the total potentials (3.5) as

$$V^{(1)}(12)=\tilde{\Gamma}^{(1)}(1243)\varphi(34),$$

$$V^{(2)}(12)=\frac{1}{2}\tilde{\Gamma}^{(2)}(124365)\varphi(34)\varphi(56)+\tilde{\Gamma}^{(1)}(1243)\varphi^{(2)}(34), \tag{5.12}$$

in which the *proper* vertex function $\tilde{\Gamma}^{(\nu)}$ is defined in perturbation theory as the sum of all νth-order vertex diagrams that cannot be separated into two disconnected parts by cutting one Coulomb interaction line or one electron propagator. (An alternative definition would be as the νth functional derivative of V with respect to φ.[83,84]) From the above results, Γ and $\tilde{\Gamma}$ are related by

$$\Gamma^{(1)}(1243)=\tilde{\Gamma}^{(1)}(1265)\epsilon^{-1}(5643),$$

$$\Gamma^{(2)}(124365)=\tilde{\Gamma}^{(2)}(128709)\epsilon^{-1}(7843)\epsilon^{-1}(9065)+\tilde{\Gamma}^{(1)}(1287)u(8709)\Pi^{(2)}(904365). \tag{5.13}$$

In an insulator, the Ward identities (4.10) and (5.8) yield

$$\Gamma^{(1)}(1233)=\tilde{\Gamma}^{(1)}(1244),$$

$$\Gamma^{(2)}(124355)=\tilde{\Gamma}^{(2)}(128799)\epsilon^{-1}(7843),$$

$$\Gamma^{(2)}(123344)=\tilde{\Gamma}^{(2)}(125566). \tag{5.14}$$

Hence, for insulators, the Ward identities (3.11) are valid for both Γ and $\tilde{\Gamma}$.

C. Polarization

In a similar fashion, one can define the proper polarization $\tilde{\Pi}$ as the sum of all static polarization diagrams that cannot be split by cutting a Coulomb line. Thus

$$\Pi^{(1)}(1243)=\tilde{\Pi}^{(1)}(1265)\epsilon^{-1}(5643), \tag{5.15}$$

which has the form of a Dyson equation:[81]

$$\Pi^{(1)}(1243)=\tilde{\Pi}^{(1)}(1243)+\tilde{\Pi}^{(1)}(1265)u(5687)\Pi^{(1)}(7843)=\tilde{\Pi}^{(1)}(1243)+\Pi^{(1)}(1265)u(5687)\tilde{\Pi}^{(1)}(7843). \tag{5.16}$$

This can be used to verify that the dielectric matrix

$$\epsilon(1243) = \delta(13)\delta(24) - u(1265)\tilde{\Pi}^{(1)}(5643) \quad (5.17)$$

is indeed the inverse of Eq. (5.6). Equations (5.15) and (5.16) can also be written as

$$\Pi^{(1)}(1243) = \epsilon^{-1}(6512)\tilde{\Pi}^{(1)}(5643), \quad (5.18)$$

in which the symmetry property (E4) was used. The total and proper quadratic polarizations are likewise related by (see Fig. 5 of Ref. 58)

$$\Pi^{(2)}(124365) = \epsilon^{-1}(8712)\tilde{\Pi}^{(2)}(78092'1') \times \epsilon^{-1}(9043)\epsilon^{-1}(1'2'65). \quad (5.19)$$

In an insulator, Eqs. (5.8) and (5.15) give $\tilde{\Pi}^{(1)}(1233) = \Pi^{(1)}(1244) = 0$, which implies that $\epsilon(1233) = \delta(12)$. The inverses of Eqs. (5.15) and (5.19), i.e.,

$$\tilde{\Pi}^{(1)}(1243) = \Pi^{(1)}(1265)\epsilon(5643), \quad (5.20)$$

$$\tilde{\Pi}^{(2)}(124365) = \epsilon(8712)\Pi^{(2)}(78092'1')\epsilon(9043)\epsilon(1'2'65), \quad (5.21)$$

then show that the insulator Ward identities (4.10) are valid for both Π and $\tilde{\Pi}$.

Since the Coulomb interaction depends only on the reduced polarization matrix (5.11), Eq. (5.16) can be reduced to

$$\Pi_{ss'}(\mathbf{q}+\mathbf{G};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}') = \tilde{\Pi}_{ss'}(\mathbf{q}+\mathbf{G};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}') + \sum_{\mathbf{G}''} \tilde{\Pi}(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}'') \times v_c(\mathbf{q}+\mathbf{G}'')\Pi_{ss'}(\mathbf{q}+\mathbf{G}'';\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}'), \quad (5.22)$$

in which a scalar version of Π is defined by

$$\Pi(\mathbf{q},\mathbf{q}+\mathbf{G}) = \sum_{\mathbf{k}} \Pi_{\lambda\lambda}(\mathbf{q};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}). \quad (5.23)$$

Now $v_c(\mathbf{q}+\mathbf{G}'')$ is nonsingular in the limit $q \to 0$ when $\mathbf{G}'' \neq \mathbf{0}$, so it is convenient to regroup the series expansion of Eq. (5.22) so as to isolate the terms $v_c(\mathbf{q})$:[85,86]

$$\Pi_{ss'}(\mathbf{q}+\mathbf{G};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}') = P_{ss'}(\mathbf{q}+\mathbf{G};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}') + P(\mathbf{q}+\mathbf{G},\mathbf{q})v_c(\mathbf{q})\Pi_{ss'} \times (\mathbf{q};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}'). \quad (5.24)$$

Here P is the sum of all polarization diagrams that cannot be separated by cutting a Coulomb line labeled with $\mathbf{q}$ (although they may be split by cutting lines labeled $\mathbf{q}+\mathbf{G}''$ with $\mathbf{G}'' \neq \mathbf{0}$). This will be called the *regular* polarization; it is related to the proper polarization $\tilde{\Pi}$ by Eq. (5.22) with $\Pi \to P$ and $\mathbf{G}'' \neq \mathbf{0}$.

Both $\tilde{\Pi}$ and P are well behaved in the limit $q \to 0$, but P is more convenient for analysis of the small-q behavior of Π because, unlike the case for $\tilde{\Pi}$, it does not require the inversion of matrices (see Ref. 87 for further discussion and an alternative approach). From the relationship between P and $\tilde{\Pi}$, it is apparent that the insulator Ward identities (4.10) for Π and $\tilde{\Pi}$ hold for P as well.

VI. LINEAR RESPONSE TO A LOCAL PERTURBATION

In this section the properties of the screened potential φ are examined in greater detail for the case of a local perturbing potential,[58] which by definition has the form

$$v(\mathbf{x},\mathbf{x}') = \delta(\mathbf{x}-\mathbf{x}')v(\mathbf{x}),$$

$$v(\mathbf{k},\mathbf{k}') = v(\mathbf{k}-\mathbf{k}'). \quad (6.1)$$

Here the spin indices were omitted because a hermitian, time-reversal invariant, local potential must be a spin scalar (see Sec. IX). Contributions from the nonlocal part of the perturbation are considered in Sec. IX.

A. Screened potential

With this simplification, all of the polarization matrices can be reduced to the scalar form (5.23), and the screened potential (5.10) simplifies to[58]

$$\varphi(\mathbf{q}) = v(\mathbf{q}) + v_c(\mathbf{q})\sum_{\mathbf{G}} \Pi(\mathbf{q},\mathbf{q}+\mathbf{G})v(\mathbf{q}+\mathbf{G}). \quad (6.2)$$

Likewise, the local version of Eq. (5.24) is

$$\Pi(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}') = P(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}') + P(\mathbf{q}+\mathbf{G},\mathbf{q})v_c(\mathbf{q})\Pi(\mathbf{q},\mathbf{q}+\mathbf{G}'). \quad (6.3)$$

It is convenient at this point to define a macroscopic static electronic dielectric function[88]

$$\epsilon(\mathbf{k}) = 1 - v_c(\mathbf{k})P(\mathbf{k},\mathbf{k}) = 1/[1 + v_c(\mathbf{k})\Pi(\mathbf{k},\mathbf{k})], \quad (6.4)$$

which may be used to express Π as a function of the regular polarization P:

$$\Pi(\mathbf{q},\mathbf{q}+\mathbf{G}) = \epsilon^{-1}(\mathbf{q})P(\mathbf{q},\mathbf{q}+\mathbf{G}), \quad (6.5a)$$

$$\Pi(\mathbf{q}+\mathbf{G},\mathbf{q}) = \epsilon^{-1}(\mathbf{q})P(\mathbf{q}+\mathbf{G},\mathbf{q}), \quad (6.5b)$$

$$\Pi(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}') = P(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}') + P(\mathbf{q}+\mathbf{G},\mathbf{q})\epsilon^{-1}(\mathbf{q})v_c(\mathbf{q})P(\mathbf{q},\mathbf{q}+\mathbf{G}'). \quad (6.5c)$$

Here all of the nonanalytic behavior at small q is contained in the factors $v_c(\mathbf{q})$ and $\epsilon^{-1}(\mathbf{q})$.

Upon substituting (6.5) into (6.2), one obtains the screened potentials

$$\varphi(\mathbf{q}) = \frac{v(\mathbf{q})}{\epsilon(\mathbf{q})} + \frac{v_c(\mathbf{q})}{\epsilon(\mathbf{q})}\sum_{\mathbf{G}\neq\mathbf{0}} P(\mathbf{q},\mathbf{q}+\mathbf{G})v(\mathbf{q}+\mathbf{G}), \quad (6.6)$$

$$\varphi(\mathbf{q}+\mathbf{G}) = v(\mathbf{q}+\mathbf{G}) + v_c(\mathbf{q}+\mathbf{G})P(\mathbf{q}+\mathbf{G},\mathbf{q})\varphi(\mathbf{q}) + v_c(\mathbf{q}+\mathbf{G})\sum_{\mathbf{G}'\neq\mathbf{0}} P(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}')v(\mathbf{q}+\mathbf{G}'), \quad (6.7)$$

where both expressions are valid for arbitrary $\mathbf{q}$ and $\mathbf{G}$, but

the latter is more useful for investigating the behavior of φ in the neighborhood of a nonzero reciprocal lattice vector. For small $\mathbf{q}$, the first term in (6.6) is the macroscopic screening that occurs even for slowly varying potentials [with $v(\mathbf{k})=0$ for $\mathbf{k}$ outside the first Brillouin zone], while the second term is a local-field correction[89,90] arising from the microscopic inhomogeneity of the reference crystal.

B. Power series expansions

The next step is to establish the small-q properties of P. Since the only singular Coulomb terms in $P(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}')$ are the factors $v_c(\mathbf{q}+\mathbf{G}'')$ with $\mathbf{G}''\neq\mathbf{0}$, the regular polarization $P(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}')$ is analytic for $q<G_{\min}$, where $G_{\min}$ is the magnitude of the smallest nonzero reciprocal lattice vector. In addition, the symmetry property (E4) gives $P(\mathbf{x},\mathbf{x}')=P(\mathbf{x}',\mathbf{x})$ or

$$P(\mathbf{k},\mathbf{k}')=P(-\mathbf{k}',-\mathbf{k}), \tag{6.8}$$

while the Ward identity (4.10) for an insulator implies that

$$\lim_{q\to 0}P(\mathbf{q},\mathbf{q}+\mathbf{G})=0. \tag{6.9}$$

From these results, we see that the matrix $P_{\mathbf{G}\mathbf{G}'}(\mathbf{q})\equiv P(\mathbf{q}+\mathbf{G},\mathbf{q}+\mathbf{G}')$ has the Taylor series expansion

$$P_{\mathbf{00}}(\mathbf{q})=P_{\mathbf{00}}^{[2]}+P_{\mathbf{00}}^{[4]}+P_{\mathbf{00}}^{[6]}+O(q^8), \tag{6.10a}$$

$$P_{\mathbf{0G}}(\mathbf{q})=P_{\mathbf{0G}}^{[1]}+P_{\mathbf{0G}}^{[2]}+P_{\mathbf{0G}}^{[3]}+P_{\mathbf{0G}}^{[4]}+O(q^5), \tag{6.10b}$$

$$P_{\mathbf{GG}'}(\mathbf{q})=P_{\mathbf{GG}'}^{[0]}+P_{\mathbf{GG}'}^{[1]}+P_{\mathbf{GG}'}^{[2]}+O(q^3), \tag{6.10c}$$

with $P_{\mathbf{G0}}(\mathbf{q})=P_{\mathbf{0},-\mathbf{G}}(-\mathbf{q})$. Here $P_{\mathbf{GG}'}^{[l]}$ denotes a general polynomial of order l in the Cartesian components of $\mathbf{q}$; for example,

$$P_{\mathbf{00}}^{[2]}=P_{\mathbf{00}}^{[2]}(\mathbf{q})=q_\alpha q_\beta P_{\mathbf{00}}^{\alpha\beta}, \tag{6.11}$$

in which $P_{\mathbf{00}}^{\alpha\beta}$ is a constant, and a sum over the Cartesian components α and β is implicit.

In the limit of small q, the dielectric function of an insulator therefore tends toward a finite but direction-dependent limit:

$$\lim_{q\to 0}\epsilon(\mathbf{q})\equiv\epsilon(\hat{\mathbf{q}})=1-4\pi P_{\mathbf{00}}^{\alpha\beta}\frac{q_\alpha q_\beta}{q^2}, \tag{6.12}$$

in which $\hat{\mathbf{q}}=\mathbf{q}/q$. This behavior contributes nonanalytic terms in the small-q expansion of $\epsilon^{-1}(\mathbf{q})$:

$$\frac{1}{\epsilon(\mathbf{q})}=\frac{1}{\epsilon(\hat{\mathbf{q}})}[1+w_c(\mathbf{q})(P_{\mathbf{00}}^{[4]}+P_{\mathbf{00}}^{[6]})+w_c^2(\mathbf{q})(P_{\mathbf{00}}^{[4]})^2]+O(q^6), \tag{6.13}$$

in which $w_c(\mathbf{q})=v_c(\mathbf{q})/\epsilon(\hat{\mathbf{q}})$.

C. Pseudopotential

To proceed further it is necessary to make some assumptions about the perturbing pseudopotential $v(\mathbf{k})$. This is taken to be a superposition of spherically symmetric ionic perturbations, each of which has the Gaussian form used in Refs. 40–42. The pseudopotential can therefore be written as

$$v(\mathbf{k})=v_{\mathrm{an}}(\mathbf{k})+v_c(\mathbf{k})\rho(\mathbf{k}), \tag{6.14}$$

in which $v_{\mathrm{an}}(\mathbf{k})$ and $\rho(\mathbf{k})$ are entire analytic functions of $\mathbf{k}$, and $\rho(\mathbf{k})$ represents a portion of the pseudocharge density of the perturbation [the other portion being given by $k^2v_{\mathrm{an}}(\mathbf{k})/4\pi$].

For a single ion, these functions have the form of a Gaussian times a polynomial in k^2:[41,42]

$$\rho(\mathbf{k})=-\frac{Z_v}{\Omega}\left[1+\tfrac{1}{2}(kr_0)^2\right]e^{-(kr_0)^2/2},$$

$$v_{\mathrm{an}}(\mathbf{k})=\frac{1}{\Omega}[2\pi Z_v r_0^2+g(k^2r_0^2)]e^{-(kr_0)^2/2}, \tag{6.15}$$

where Z_v is the charge of the ion, r_0 is a core radius parameter, and $g(x)$ is a cubic polynomial given in Refs. 41 and 42. Here $\rho(\mathbf{k})$ has been defined in such a way that its Taylor series contains no term proportional to k^2:

$$\rho(\mathbf{k})=\rho_0+\rho_4k^4+\rho_6k^6+\cdots. \tag{6.16}$$

Therefore, the only term in $v_c(\mathbf{k})\rho(\mathbf{k})$ that does not vanish in the limit $k\to 0$ is the divergent term $-4\pi Z_v/\Omega k^2$. This term has been eliminated[91] at $k=0$ [by the definition (5.4) of $v_c(\mathbf{k})$] because the ion is assumed to be accompanied by Z_v electrons, so that the crystal remains neutral after the perturbation.

If the perturbation contains more than one ion (e.g., the quasiatoms defined in Appendix A), $\rho(\mathbf{k})$ is just a general Taylor series

$$\rho(\mathbf{k})=\rho_0+\rho^{[1]}+\rho^{[2]}+\cdots, \tag{6.17}$$

although the linear response can always be treated as a superposition of individual ions.

For $\mathbf{G}\neq\mathbf{0}$, $v(\mathbf{q}+\mathbf{G})$ is analytic for $q<G$, with the Taylor series

$$v(\mathbf{q}+\mathbf{G})=v_{\mathbf{G}}^{[0]}+v_{\mathbf{G}}^{[1]}+v_{\mathbf{G}}^{[2]}+v_{\mathbf{G}}^{[3]}+O(q^4), \tag{6.18}$$

in which $v_{\mathbf{G}}^{[0]}=v(\mathbf{G})$. A similar expansion is valid for $v_c(\mathbf{q}+\mathbf{G})$.

D. Effective macroscopic density

It is now convenient to rewrite Eq. (6.6) in a form modeled after the familiar expressions for screening in a homogeneous system:[92]

$$\varphi(\mathbf{q})=v_{\mathrm{an}}(\mathbf{q})+\frac{v_c(\mathbf{q})\bar{n}(\mathbf{q})}{\epsilon(\mathbf{q})}, \tag{6.19}$$

in which

$$\bar{n}(\mathbf{q}) = \rho(\mathbf{q}) + P_{\mathbf{00}}(\mathbf{q})v_{\mathrm{an}}(\mathbf{q}) + \sum_{\mathbf{G}\neq\mathbf{0}} P_{\mathbf{0G}}(\mathbf{q})v(\mathbf{q}+\mathbf{G}) \tag{6.20}$$

is an effective macroscopic electron density, which is analytic for $q<G_{\min}$. This includes the (partial) bare charge $\rho(\mathbf{q})$, the macroscopic charge induced by $v_{\mathrm{an}}(\mathbf{q})$, and the local-field corrections. Note that since $\varphi(\mathbf{q})=v(\mathbf{q})+v_c(\mathbf{q})n^{(1)}(\mathbf{q})$, one can also write $\bar{n}(\mathbf{q})=[\rho(\mathbf{q})+n^{(1)}(\mathbf{q})]\epsilon(\mathbf{q})$. Also note that $\bar{n}_0=\rho_0$, because the second term in (6.20) is $O(q^2)$ and the last term is $O(q)$. The leading contributions to $\varphi(\mathbf{q})$ for small q are therefore

$$\begin{aligned}\varphi(\mathbf{q}) = \rho_0[w_c(\mathbf{q}) + w_c^2(\mathbf{q})(P_{\mathbf{00}}^{[4]} + P_{\mathbf{00}}^{[6]}) + w_c^3(\mathbf{q})(P_{\mathbf{00}}^{[4]})^2] + w_c(\mathbf{q}) \\ \times(\bar{n}^{[1]} + \bar{n}^{[2]} + \bar{n}^{[3]} + \bar{n}^{[4]}) + w_c^2(\mathbf{q})P_{\mathbf{00}}^{[4]}(\bar{n}^{[1]} + \bar{n}^{[2]}) \\ + v_{\mathrm{an}}(\mathbf{q}) + O(q^3),\end{aligned} \tag{6.21}$$

in which $v_{\mathrm{an}}(\mathbf{q})$ is to be replaced by its Taylor series expansion.

The first set of terms in (6.21) is proportional to $\rho_0=-Z_v/\Omega$. These terms are just the power series expansion for $\rho_0 v_c(\mathbf{q})/\epsilon(\mathbf{q})$. Such terms are present in general, but they vanish for isovalent perturbations (e.g., Al substituting for Ga in GaAs).

The remaining nonanalytic terms depend on $\bar{n}^{[l]}$, where $l\geqslant 1$. The symmetry of $\bar{n}^{[l]}$ may differ from that of $\rho^{[l]}$. For example, the term $w_c(\mathbf{q})n^{[1]}(\mathbf{q})$ would contribute a dipole field if it were present, but $\rho^{[1]}$ vanishes for a spherically symmetric atom. However, since the symmetry of P is the same as that of the reference crystal, the symmetry of $\bar{n}$ is just the site symmetry at the position of the ionic perturbation (i.e., the maximal common subgroup of the reference crystal space group and the full rotation group at the given atomic site). Therefore, for crystals of sufficiently low symmetry (e.g., wurtzite), $\bar{n}^{[1]}$ may contribute a dipole field even though $\rho^{[1]}$ does not.

E. Special cases

The general expression (6.21) is quite cumbersome and is unlikely to be used in its entirety for any particular material system. In many cases one would only be interested in retaining terms that are two orders in q higher than the lowest nonvanishing term. Thus, for heterovalent perturbations with $\rho_0\neq 0$ (i.e., class I of the Introduction), Eq. (6.21) could be simplified to

$$\varphi(\mathbf{q}) = v_{\mathrm{an}}(\mathbf{0}) + w_c(\mathbf{q})(\rho_0 + \bar{n}^{[1]} + \bar{n}^{[2]}) + \rho_0 w_c^2(\mathbf{q})P_{\mathbf{00}}^{[4]} + O(q), \tag{6.22}$$

which contains monopole, dipole, and quadrupole terms, plus a correction to the monopole term describing the wave vector dependence of the dielectric function. For isovalent perturbations in crystals with atomic site symmetry that supports a dipole moment (class II), a suitable approximation would be

$$\varphi(\mathbf{q}) = v_{\mathrm{an}}(\mathbf{q}) + w_c(\mathbf{q})(\bar{n}^{[1]} + \bar{n}^{[2]} + \bar{n}^{[3]}) + w_c^2(\mathbf{q})P_{\mathbf{00}}^{[4]}\bar{n}^{[1]} + O(q^2), \tag{6.23}$$

which includes additional octopole terms. Finally, for isovalent perturbations in crystals with site symmetry that does not support a dipole moment (class III), one would use

$$\varphi(\mathbf{q}) = v_{\mathrm{an}}(\mathbf{q}) + w_c(\mathbf{q})(\bar{n}^{[2]} + \bar{n}^{[3]} + \bar{n}^{[4]}) + w_c^2(\mathbf{q})P_{\mathbf{00}}^{[4]}\bar{n}^{[2]} + O(q^3). \tag{6.24}$$

As an explicit example, consider the case of isovalent substitutions in a crystal with the zinc-blende or diamond structure (space group T_d^2 or O_h^7), both of which have site symmetry T_d at the atomic sites. In this case $\rho_0=0$, and the only invariants of order q^4 or lower are 1, q^2, $q_xq_yq_z$, q^4, and $q_x^4+q_y^4+q_z^4$. Hence, the quadratic terms are isotropic ($P_{\mathbf{00}}^{\alpha\beta}=P_2\delta_{\alpha\beta}, \bar{n}^{[2]}=\bar{n}_2q^2$) and the long-wavelength dielectric function reduces to a constant [$\epsilon(\hat{\mathbf{q}})=\epsilon=1-4\pi P_2$]. The leading contributions to $\varphi(\mathbf{q})$ may be written as

$$\varphi(\mathbf{q}) = \frac{4\pi\bar{n}_2}{\epsilon}(1-\delta_{\mathbf{q0}}) + C_1 + C_2q^2 + C_3\frac{q_xq_yq_z}{q^2} + C_4\frac{q_x^4+q_y^4+q_z^4}{q^2} + O(q^3), \tag{6.25}$$

where C_i is a constant. The terms C_3 and C_4 represent octopole and hexadecapole moments, respectively.

F. Nonzero reciprocal lattice vectors

Turning now to $\varphi(\mathbf{q}+\mathbf{G})$, Eq. (6.7) can be written in the condensed notation

$$\varphi(\mathbf{q}+\mathbf{G}) = R_{\mathbf{G0}}(\mathbf{q})\varphi(\mathbf{q}) + \xi_{\mathbf{G}}(\mathbf{q}), \tag{6.26}$$

in which $R_{\mathbf{G0}}(\mathbf{q})$ and $\xi_{\mathbf{G}}(\mathbf{q})$ are analytic for $q<G_{\min}$:

$$R_{\mathbf{G0}}(\mathbf{q}) = \begin{cases} 1 & \text{if } \mathbf{G}=\mathbf{0}, \\ v_c(\mathbf{q}+\mathbf{G})P_{\mathbf{G0}}(\mathbf{q}) & \text{if } \mathbf{G}\neq\mathbf{0} \end{cases} \tag{6.27}$$

$$\xi_{\mathbf{G}}(\mathbf{q}) = (1-\delta_{\mathbf{G0}})\left[v(\mathbf{q}+\mathbf{G}) + v_c(\mathbf{q}+\mathbf{G}) \times \sum_{\mathbf{G}'\neq\mathbf{0}} P_{\mathbf{GG}'}(\mathbf{q})v(\mathbf{q}+\mathbf{G}')\right]. \tag{6.28}$$

For $\mathbf{G}\neq\mathbf{0}$, Eqs. (6.10b) and (6.18) show that $R_{\mathbf{G0}}(\mathbf{q})$ is of order q or higher. The explicit form of the Taylor series for $R_{\mathbf{G0}}(\mathbf{q})$ is determined by finding the invariants of the group of the wave vector $\mathbf{G}$ in the reference crystal, where different $\mathbf{G}$ vectors are treated as inequivalent. Thus, for general $\mathbf{G}$, the linear term is nonvanishing. However, the leading term in the Taylor series for $\xi_{\mathbf{G}}(\mathbf{q})$ (with $\mathbf{G}\neq\mathbf{0}$) is a constant.

Hence, for $\mathbf{G}\neq\mathbf{0}$, the nonanalytic terms in $\varphi(\mathbf{q}+\mathbf{G})$ are at least one order in q higher than those in $\varphi(\mathbf{q})$. In the zinc-blende example discussed above, one has

$$\varphi(\mathbf{q}+\mathbf{G}) = C_3R_{\mathbf{G0}}^{\alpha}\frac{q_\alpha q_xq_yq_z}{q^2} + \chi_{\mathbf{G}}(\mathbf{q}) + O(q^3), \tag{6.29}$$

in which $\chi_{\mathbf{G}}(\mathbf{q})$ is analytic, and $R_{\mathbf{G0}}^{\alpha}$ is the linear coefficient in the Taylor series for $R_{\mathbf{G0}}(\mathbf{q})$. The nonanalytic term in

(6.29) is a hexadecapole moment that is invariant with respect to the group of **G**.

VII. QUADRATIC RESPONSE TO A LOCAL PERTURBATION

To calculate the quadratic density $n^{(2)}(\mathbf{k})$, it is helpful to begin by considering the following partial density obtained from the local version of Eqs. (4.7), (5.5), and (5.19):

$$\tilde{n}^{(2)}(\mathbf{k}) = \frac{1}{2}\sum_{\mathbf{k}_1}{}' \sum_{\mathbf{G}_1\mathbf{G}_2} \tilde{\Pi}^{(2)}(\mathbf{k},\mathbf{k}_1+\mathbf{G}_1,\mathbf{k}-\mathbf{k}_1+\mathbf{G}_2) \times \varphi(\mathbf{k}_1+\mathbf{G}_1)\varphi(\mathbf{k}-\mathbf{k}_1+\mathbf{G}_2), \tag{7.1}$$

where the summation on $\mathbf{k}_1$ is limited to the first Brillouin zone of the reference crystal. Here the proper polarization $\tilde{\Pi}^{(2)}(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2)$ vanishes unless $\mathbf{k}=\mathbf{k}_1+\mathbf{k}_2+\mathbf{G}$, where **G** is any reciprocal lattice vector. It satisfies the symmetry relations (E4), the reduced form of which is

$$\tilde{\Pi}^{(2)}(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = \tilde{\Pi}^{(2)}(\mathbf{k},\mathbf{k}_2,\mathbf{k}_1) = \tilde{\Pi}^{(2)}(-\mathbf{k}_1,-\mathbf{k},\mathbf{k}_2). \tag{7.2}$$

It also satisfies the Ward identity (4.10):

$$\lim_{k\to 0}\tilde{\Pi}^{(2)}(\mathbf{k},\mathbf{k}_1,\mathbf{k}-\mathbf{k}_1+\mathbf{G}_2) = 0. \tag{7.3}$$

With these constraints, the Taylor series expansion of the polarization matrix $\tilde{\Pi}^{(2)}_{\mathbf{G}\mathbf{G}_1\mathbf{G}_2}(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2)=\tilde{\Pi}^{(2)}(\mathbf{k}+\mathbf{G},\mathbf{k}_1+\mathbf{G}_1,\mathbf{k}_2+\mathbf{G}_2)$ has a form similar to that given for P in Eq. (6.10):

$$\tilde{\Pi}^{(2)}_{\mathbf{000}}(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = k_\alpha k_{1\beta} k_{2\gamma}\tilde{\Pi}^{\alpha\beta\gamma}_{\mathbf{000}} + O(k^4),$$

$$\tilde{\Pi}^{(2)}_{\mathbf{G}\mathbf{00}}(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = k_{1\beta} k_{2\gamma}\tilde{\Pi}^{\beta\gamma}_{\mathbf{G}\mathbf{00}} + O(k^3),$$

$$\tilde{\Pi}^{(2)}_{\mathbf{0}\mathbf{G}_1\mathbf{G}_2}(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = k_\alpha\tilde{\Pi}^{\alpha}_{\mathbf{0}\mathbf{G}_1\mathbf{G}_2} + O(k^2),$$

$$\tilde{\Pi}^{(2)}_{\mathbf{G}\mathbf{G}_1\mathbf{G}_2}(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = \tilde{\Pi}_{\mathbf{G}\mathbf{G}_1\mathbf{G}_2} + O(k), \tag{7.4}$$

where the order of the leading term is equal to the number of **G** vectors that are zero. Here $O(k^n)$ denotes a term of order $k^p k_1^q k_2^r$, where $n=p+q+r$.

The partial quadratic density (7.1) generates a Coulomb potential $v^{(2)}(\mathbf{q})=v_c(\mathbf{q})\tilde{n}^{(2)}(\mathbf{q})$, which is then screened to produce $\varphi^{(2)}(\mathbf{q})$ of Eq. (5.9). This potential is calculated by replacing $v(\mathbf{q})$ with $v^{(2)}(\mathbf{q})$ in Eqs. (6.6) and (6.7). The result may be written as

$$\varphi^{(2)}(\mathbf{q}) = \frac{v_c(\mathbf{q})\bar{n}^{(2)}(\mathbf{q})}{\epsilon(\mathbf{q})}, \tag{7.5}$$

where $\bar{n}^{(2)}(\mathbf{q})=n^{(2)}(\mathbf{q})\epsilon(\mathbf{q})$ is an effective "external" density

$$\bar{n}^{(2)}(\mathbf{q}) = \sum_{\mathbf{G}} R_{\mathbf{0}\mathbf{G}}(\mathbf{q})\tilde{n}^{(2)}(\mathbf{q}+\mathbf{G}), \tag{7.6}$$

in which

$$R_{\mathbf{0}\mathbf{G}}(\mathbf{q}) = \begin{cases} 1 & \text{if } \mathbf{G}=\mathbf{0}, \\ P_{\mathbf{0}\mathbf{G}}(\mathbf{q})v_c(\mathbf{q}+\mathbf{G}) & \text{if } \mathbf{G}\neq\mathbf{0} \end{cases} \tag{7.7}$$

In Eq. (7.6), $\tilde{n}^{(2)}(\mathbf{q}+\mathbf{G})$ is given by (7.1), where $\varphi(\mathbf{k}+\mathbf{G})$ can be expressed in terms of $\varphi(\mathbf{k})$ using Eq. (6.26). The resulting expression for (7.6) can be written as $\bar{n}^{(2)}(\mathbf{q})=\bar{n}^{(2)}_A(\mathbf{q})+\bar{n}^{(2)}_B(\mathbf{q})+\bar{n}^{(2)}_C(\mathbf{q})$, in which

$$\bar{n}^{(2)}_A(\mathbf{q}) = \tfrac{1}{2}\sum_{\mathbf{k}}{}' A(\mathbf{q},\mathbf{k},\mathbf{q}-\mathbf{k})\varphi(\mathbf{k})\varphi(\mathbf{q}-\mathbf{k}),$$

$$\bar{n}^{(2)}_B(\mathbf{q}) = \tfrac{1}{2}\sum_{\mathbf{k}}{}' [B(\mathbf{q},\mathbf{k},\mathbf{q}-\mathbf{k})\varphi(\mathbf{q}-\mathbf{k}) + B(\mathbf{q},\mathbf{q}-\mathbf{k},\mathbf{k})\varphi(\mathbf{k})],$$

$$\bar{n}^{(2)}_C(\mathbf{q}) = \tfrac{1}{2}\sum_{\mathbf{k}}{}' C(\mathbf{q},\mathbf{k},\mathbf{q}-\mathbf{k}). \tag{7.8}$$

Here the functions A, B, and C, which are defined in Appendix F, have the Taylor series expansions

$$A(\mathbf{k},\mathbf{k_1},\mathbf{k_2}) = k_\alpha k_{1\beta} k_{2\gamma} A_{\alpha\beta\gamma} + O(k^4),$$

$$B(\mathbf{k},\mathbf{k_1},\mathbf{k_2}) = k_\alpha k_{2\gamma} B_{\alpha\gamma} + O(k^3),$$

$$C(\mathbf{k},\mathbf{k_1},\mathbf{k_2}) = k_\alpha C_\alpha + O(k^2). \tag{7.9}$$

For **q** values inside the first Brillouin zone, the functions A, B, and C in (7.8) are analytic for all **k** values included in the summation, but $\varphi(\mathbf{k})$ is nonanalytic at $\mathbf{k}=\mathbf{0}$. Therefore it is possible that the φ terms in (7.8) may produce nonanalytic behavior in $\bar{n}^{(2)}(\mathbf{q})$. $\bar{n}^{(2)}_C(\mathbf{q})$ is obviously analytic in **q**, as is the second term in $\bar{n}^{(2)}_B(\mathbf{q})$. The first term in $\bar{n}^{(2)}_B(\mathbf{q})$ is as well, since a small variation $\delta\mathbf{q}$ can be eliminated from $\varphi(\mathbf{q}-\mathbf{k})$ with an equal variation $\delta\mathbf{k}=\delta\mathbf{q}$. This slightly shifts the zone boundary in the summation, but $\varphi(\mathbf{q}-\mathbf{k})$ is analytic at the zone boundary, so $\bar{n}^{(2)}_B(\mathbf{q})$ is analytic for small **q**.

However, for $\bar{n}^{(2)}_A(\mathbf{q})$ this argument is no longer valid. The singularities in $\varphi(\mathbf{k})$ and $\varphi(\mathbf{q}-\mathbf{k})$ merge when $\mathbf{q}=\mathbf{0}$, producing nonanalytic behavior in $\bar{n}^{(2)}_A(\mathbf{q})$ at this point. The contribution from the nonanalytic part of $\bar{n}^{(2)}_A(\mathbf{q})$ is examined in Appendix G, where it is shown to be negligible under all three approximation schemes defined in the Introduction. Therefore, only the analytic part of $\bar{n}^{(2)}(\mathbf{q})$ is retained here.

The leading contributions to the quadratic screened potential are therefore

$$\varphi^{(2)}(\mathbf{q}) = w_c(\mathbf{q})(q_\alpha\bar{n}^{(2)}_\alpha + q_\alpha q_\beta\bar{n}^{(2)}_{\alpha\beta}) + O(q), \tag{7.10}$$

where $\bar{n}^{(2)}_\alpha$ and $\bar{n}^{(2)}_{\alpha\beta}$ are Taylor series coefficients for the analytic part of $\bar{n}^{(2)}(\mathbf{q})$. The absence of a constant term in the power series for $\bar{n}^{(2)}(\mathbf{q})$ is a consequence of the Ward identity (7.3). Equation (7.10) is used in its full form only for isovalent class III perturbations. For class II, the $\bar{n}^{(2)}_{\alpha\beta}$ term is negligible, while for class I, the entire contribution from $\varphi^{(2)}(\mathbf{q})$ is negligible.[93]

In the vicinity of a nonzero reciprocal lattice vector, $\varphi^{(2)}(\mathbf{q}+\mathbf{G})$ can be written in a form similar to (6.26):

$$\varphi^{(2)}(\mathbf{q}+\mathbf{G}) = R_{\mathbf{G0}}(\mathbf{q})\varphi^{(2)}(\mathbf{q}) + \xi^{(2)}_{\mathbf{G}}(\mathbf{q}), \tag{7.11}$$

in which $\xi^{(2)}_{\mathbf{G}}(\mathbf{q})$ is given by Eq. (6.28) with $v(\mathbf{k})$ replaced by $v^{(2)}(\mathbf{k})=v_c(\mathbf{k})\bar{n}^{(2)}(\mathbf{k})$. Using the same type of analysis as before, one finds that the nonanalytic part of $\xi^{(2)}_{\mathbf{G}}(\mathbf{q})$ is $O(q)$ for class I, $O(q^3)$ for class II, and $O(q^5)$ for class III. Therefore, the limit $\xi^{(2)}_{\mathbf{G}}(\mathbf{0})$ is well defined, and the leading terms in $\varphi^{(2)}(\mathbf{q}+\mathbf{G})$ are given by

$$\varphi^{(2)}(\mathbf{q}+\mathbf{G}) = w_c(\mathbf{q})(q_\alpha q_\beta \bar{n}^{(2)}_\alpha R^\beta_{\mathbf{G0}}) + \xi^{(2)}_{\mathbf{G}}(\mathbf{0}) + O(q). \tag{7.12}$$

Since the leading terms here are $O(q^0)$, this contribution is negligible for classes I and II.

Note that the quadratic response for a heterostructure cannot be written as a superposition of spherically symmetric atomic perturbations; one must also include diatomic perturbations with axial symmetry $C_{\infty v}$ (for a heteronuclear diatomic molecule) or $D_{\infty h}$ (for a homonuclear diatomic molecule).[94] The symmetry of $\bar{n}^{(2)}(\mathbf{q})$ is determined by the maximal common subgroup of the reference crystal space group and the molecular point group. For example, for a perturbation at neighboring atomic sites in zinc-blende, the symmetry of $\bar{n}^{(2)}(\mathbf{q})$ is C_{3v}, which supports a nonvanishing dipole moment $\bar{n}^{(2)}_\alpha$.

In general, $\bar{n}^{(2)}_\alpha$ is nonvanishing for any heteronuclear perturbation, because $C_{\infty v}$ itself permits the existence of a dipole. Such dipoles therefore always appear in heterostructures involving more than one type of atomic perturbation (e.g., InAs/GaSb). (This property of the nonlinear response was deduced from numerical calculations of band offsets in Ref. 95.) Furthermore, the quadrupole term $\bar{n}^{(2)}_{\alpha\beta}-\frac{1}{3}\bar{n}^{(2)}_{\lambda\lambda}\delta_{\alpha\beta}$ is nonvanishing for any diatomic perturbation, since isotropy requires cubic symmetry.

VIII. SELF-ENERGY FOR A LOCAL PERTURBATION

A. Linear terms

The above results may now be used to calculate the self-energy Σ and the total potential V defined in Eqs. (3.3) and (5.12). The total linear potential (5.12) is given for the case of a local perturbation by

$$V^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) = \sum_{\mathbf{G}''} \tilde{\Gamma}^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\mathbf{q}+\mathbf{G}'';\omega)\varphi(\mathbf{q}+\mathbf{G}''), \tag{8.1}$$

where $\mathbf{q}=\mathbf{k}-\mathbf{k}'$. Here $\varphi(\mathbf{q}+\mathbf{G}'')$ can be expressed in terms of $\varphi(\mathbf{q})$ using Eq. (6.26); this yields

$$V^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) = \Lambda_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega)\varphi(\mathbf{q}) + W^{(1)}_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega), \tag{8.2}$$

in which

$$W^{(1)}_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega) = \sum_{\mathbf{G}''\neq\mathbf{0}} \tilde{\Gamma}^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\mathbf{q}+\mathbf{G}'';\omega)\xi_{\mathbf{G}''}(\mathbf{q}) \tag{8.3}$$

is an analytic function of $\mathbf{k}$ and $\mathbf{k}'$ (and therefore also of $\mathbf{q}$). The nonanalytic terms are all contained in the screened potential $\varphi(\mathbf{q})=v_{\mathrm{an}}(\mathbf{q})+v_c(\mathbf{q})\bar{n}(\mathbf{q})/\epsilon(\mathbf{q})$, which is multiplied by the effective macroscopic vertex function

$$\Lambda_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega) = \sum_{\mathbf{G}''} \tilde{\Gamma}^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+(\mathbf{G}';\mathbf{q}+\mathbf{G}'';\omega)R_{\mathbf{G}''\mathbf{0}}(\mathbf{q}). \tag{8.4}$$

Since $\Lambda_{ss'}$ is an analytic function of $\mathbf{k}$ and $\mathbf{k}'$, it can be expanded in a Taylor series [treating $\mathbf{q}=\mathbf{k}-\mathbf{k}'$ and $\mathbf{Q}=\frac{1}{2}(\mathbf{k}+\mathbf{k}')$ as the independent variables], with the result

$$\begin{aligned}\Lambda_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega) = {}& [\delta_{ss'}\delta_{\mathbf{GG}'} - \partial\Sigma^{(0)}_{ss'}(\mathbf{G},\mathbf{G}';\omega)/\partial\omega] \\ & + q_\alpha\Lambda^{(\alpha|\cdot)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + Q_\alpha\Lambda^{(\cdot|\alpha)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) \\ & + q_\alpha q_\beta\Lambda^{(\alpha\beta|\cdot)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + Q_\alpha Q_\beta\Lambda^{(\cdot|\alpha\beta)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) \\ & + q_\alpha Q_\beta\Lambda^{(\alpha|\beta)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + O(q^3).\end{aligned} \tag{8.5}$$

Here the lowest-order term is determined by the Ward identity (3.11), and the Taylor series coefficients such as $\Lambda^{(\alpha|\beta)}_{ss'}$ are given in Appendix H. The analytic potential (8.3) can be expanded in the same way:

$$\begin{aligned}W^{(1)}_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega) = {}& W^{(1)}_{ss'}(\mathbf{0},\mathbf{0};\mathbf{G},\mathbf{G}';\omega) \\ & + q_\alpha W^{(\alpha|\cdot)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + Q_\alpha W^{(\cdot|\alpha)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) \\ & + q_\alpha q_\beta W^{(\alpha\beta|\cdot)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + Q_\alpha Q_\beta W^{(\cdot|\alpha\beta)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) \\ & + q_\alpha Q_\beta W^{(\alpha|\beta)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + O(q^3).\end{aligned} \tag{8.6}$$

In the expansion (8.5), a term such as q_α appears in the total potential (8.2) as a multiplicative factor in front of the screened potential $\varphi(\mathbf{q})$. In coordinate space, this term therefore takes the gradient of $\varphi(\mathbf{x})$, generating the α component of the macroscopic electric field produced by the perturbation $v(\mathbf{x})$. Likewise, the term Q_α has the form of a (symmetrized) crystal momentum operator that acts upon the envelope functions in an effective-mass theory. The Taylor series (8.6) for the analytic potential (8.3) is interpreted in the same way, except that these terms produce only short-range localized potentials because they are analytic functions of $\mathbf{q}$.

The various terms in Eq. (8.5) therefore give rise to various long-range spin-dependent potentials whose particular form is determined by the symmetry of the coefficients $\Lambda^{(\alpha|\cdot)}_{ss'}$, etc. The specific term that generates the long-range Rashba effect is $\Lambda^{(\alpha|\beta)}_{ss'}$, since this term is linear in the electric field $q_\alpha\varphi(\mathbf{q})$ and linear in the crystal momentum Q_β. The usual short-range part of the Rashba coupling is generated by the analogous term $W^{(\alpha|\beta)}_{ss'}$ in Eq. (8.6).

The complete expression for $V^{(1)}$ is obtained by inserting the expansion (8.5) for Λ and one of the three expansions

(6.22), (6.23), and (6.24) for $\varphi(\mathbf{q})$ into Eq. (8.2). For the specific example of isovalent perturbations in zinc-blende materials treated in Eq. (6.25), one finds

$$\begin{aligned} V^{(1)}_{ss'}&(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) \\ &= \frac{1-\delta_{\mathbf{q}\mathbf{0}}}{q^2}\{[\delta_{ss'}\delta_{\mathbf{G}\mathbf{G}'} - \partial\Sigma^{(0)}_{ss'}(\mathbf{G},\mathbf{G}';\omega)/\partial\omega] \\ &\times [4\pi\bar{n}_2 q^2/\epsilon + C_3 q_x q_y q_z + C_4(q_x^4+q_y^4+q_z^4)] \\ &+ C_3 q_x q_y q_z [q_\alpha \Lambda^{(\alpha|\cdot)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + Q_\alpha \Lambda^{(\cdot|\alpha)}_{ss'}(\mathbf{G},\mathbf{G}';\omega)]\} \\ &+ \text{analytic terms} + O(q^3), \end{aligned} \tag{8.7}$$

where the analytic terms include $W^{(1)}$ and contributions from the analytic part of $\varphi(\mathbf{q})$. From this result it can be seen that the Rashba effect in isovalent zinc-blende materials does not include any long-range terms (to within the accuracy of the present approximation scheme), since $\Lambda^{(\alpha|\beta)}_{ss'}$ contributes only to $O(q^3)$. However, there are other long-range spin-splitting terms of $O(q^2)$ or lower, and the Rashba effect does contribute nonnegligible long-range terms for perturbations in classes I and II.

B. Quadratic terms

Turning now to the quadratic response, the two contributions to $V^{(2)}$ in Eq. (5.12) will be denoted $V^{(2a)}$ and $V^{(2b)}$, respectively. The first of these is given by

$$\begin{aligned} V^{(2a)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) = \tfrac{1}{2}\sum_{\mathbf{k}_1}{}' \sum_{\mathbf{G}_1}\sum_{\mathbf{G}_2} \tilde{\Gamma}^{(2)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\mathbf{k}_1 \\ +\mathbf{G}_1,\mathbf{q}-\mathbf{k}_1+\mathbf{G}_2;\omega)\varphi(\mathbf{k}_1+\mathbf{G}_1) \\ \times\varphi(\mathbf{q}-\mathbf{k}_1+\mathbf{G}_2). \end{aligned} \tag{8.8}$$

Upon inserting Eq. (6.26) for $\varphi(\mathbf{k}+\mathbf{G})$ into the right-hand side, one obtains an expression for $V^{(2a)}$ very similar to that found in Eqs. (7.8) and (F1) for the effective quadratic density $\bar{n}^{(2)}(\mathbf{q})$. Just as before, there are both analytic and nonanalytic contributions. The nonanalytic contributions can be evaluated using the method outlined in Appendix G; the results show that the nonanalytic terms in $V^{(2a)}$ are $O(q^{-1})$ for class I, $O(q)$ for class II, and $O(q^3)$ for class III. (An explicit expression for the leading nonanalytic term in class I was given previously by Sham.[58,96]) Therefore, the nonanalytic contributions are negligible in all three cases, and the leading term in $V^{(2a)}$ is just a constant:

$$V^{(2a)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) = V^{(2a)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + O(q). \tag{8.9}$$

This term is negligible under the approximation schemes for classes I and II.

Finally, the second contribution to $V^{(2)}$ in Eq. (5.12) is given by

$$\begin{aligned} V^{(2b)}_{ss'}&(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) \\ &= \sum_{\mathbf{G}''} \tilde{\Gamma}^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\mathbf{q}+\mathbf{G}'';\omega)\varphi^{(2)}(\mathbf{q}+\mathbf{G}''), \end{aligned} \tag{8.10}$$

where $\varphi^{(2)}(\mathbf{q}+\mathbf{G}'')$ was given previously in Eq. (7.11). Inserting this expression into Eq. (8.10), one obtains

$$\begin{aligned} V^{(2b)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) &= \Lambda_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega)\varphi^{(2)}(\mathbf{q}) \\ &+ W^{(2b)}_{ss'}(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega), \end{aligned} \tag{8.11}$$

in which Λ was defined in Eq. (8.4), and

$$\begin{aligned} W^{(2b)}_{ss'}&(\mathbf{k},\mathbf{k}';\mathbf{G},\mathbf{G}';\omega) \\ &= \sum_{\mathbf{G}''\neq\mathbf{0}} \tilde{\Gamma}^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\mathbf{q}+\mathbf{G}'';\omega)\xi^{(2)}_{\mathbf{G}''}(\mathbf{q}). \end{aligned} \tag{8.12}$$

Unlike the case for Eq. (8.3), this is not an analytic function of $\mathbf{q}$. However, as discussed below Eq. (7.11), the nonanalytic portion is $O(q)$ and therefore vanishes at $q=0$. An explicit expression for $V^{(2b)}$ can now be obtained by inserting the expansion (7.10) for $\varphi^{(2)}(\mathbf{q})$ into Eq. (8.11):

$$\begin{aligned} V^{(2b)}_{ss'}&(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\omega) \\ &= q_\alpha w_c(\mathbf{q})\{\bar{n}^{(2)}_\alpha[\delta_{ss'}\delta_{\mathbf{G}\mathbf{G}'} - \partial\Sigma^{(0)}_{ss'}(\mathbf{G},\mathbf{G}';\omega)/\partial\omega \\ &\quad + q_\beta\Lambda^{(\beta|\cdot)}_{ss'}(\mathbf{G},\mathbf{G}';\omega) + Q_\beta\Lambda^{(\cdot|\beta)}_{ss'}(\mathbf{G},\mathbf{G}';\omega)]\} \\ &\quad + q_\alpha q_\beta w_c(\mathbf{q})\{\bar{n}^{(2)}_{\alpha\beta}[\delta_{ss'}\delta_{\mathbf{G}\mathbf{G}'} - \partial\Sigma^{(0)}_{ss'}(\mathbf{G},\mathbf{G}';\omega)/\partial\omega]\} \\ &\quad + W^{(2b)}_{ss'}(\mathbf{0},\mathbf{0};\mathbf{G},\mathbf{G}';\omega) + O(q). \end{aligned} \tag{8.13}$$

The first and second terms are dipole and quadrupole potentials, respectively, while the last term is just a constant. In class II, only the leading $O(q^{-1})$ dipole term is retained; in class I, the entire expression (8.13) is neglected.

IX. NONLOCAL PERTURBATIONS

An arbitrary nonlocal potential can be separated into local and nonlocal parts (although this separation is not unique):

$$v_{ss'}(\mathbf{x},\mathbf{x}') = v^{\text{loc}}_{ss'}(\mathbf{x},\mathbf{x}') + v^{\text{nl}}_{ss'}(\mathbf{x},\mathbf{x}'). \tag{9.1}$$

Here the local part $v^{\text{loc}}_{ss'}(\mathbf{x},\mathbf{x}')$ has the form of Eq. (6.1) and is treated according to the methods developed above. This section considers the changes in the preceding expressions that may be necessary in the case of nonlocal perturbations, particularly those involving spin-orbit coupling.

A. Analytic form

A general nonlocal potential can be written as

$$v_{ss'}(\mathbf{x},\mathbf{x}') = \delta_{ss'}v_0(\mathbf{x},\mathbf{x}') + \boldsymbol{\sigma}_{ss'}\cdot\mathbf{v}(\mathbf{x},\mathbf{x}'), \tag{9.2}$$

where v_0 is a scalar relativistic potential, $\boldsymbol{\sigma}$ is the Pauli matrix, and $\mathbf{v}$ is a pseudovector (similar to orbital angular momentum) that accounts for spin-orbit coupling. If v is Hermitian and time-reversal invariant, then v_0 is real and

symmetric, while $\mathbf{v}(\mathbf{x},\mathbf{x}')=-\mathbf{v}(\mathbf{x}',\mathbf{x})$ is imaginary and antisymmetric. Thus $\mathbf{v}$ can have no local component, and a local time-reversal-invariant potential must be a spin scalar:

$$v^{\mathrm{loc}}_{ss'}(\mathbf{x},\mathbf{x}') = \delta_{ss'}\delta(\mathbf{x}-\mathbf{x}')v_{\mathrm{loc}}(\mathbf{x}). \tag{9.3}$$

In the norm-conserving pseudopotential formalism, the nonlocal part of the ionic pseudopotential $v^{\mathrm{nl}}_{ss'}(\mathbf{x},\mathbf{x}')$ is confined to a small region near the nucleus, typically either having the form of a polynomial times a Gaussian[40–42] or vanishing absolutely outside a core region of radius r_c.[97,98] As a result, $v^{\mathrm{nl}}_{ss'}(\mathbf{k},\mathbf{k}')$ is an entire analytic function of $\mathbf{k}$ and $\mathbf{k}'$.

It is important to note that this analytic form relies upon a physical approximation. In an all-electron calculation where the pseudopotential approximation is not used, the spin-orbit coupling does in general include a contribution from the long-range Coulomb part of the ionic potential. The choice of an analytic pseudopotential $v^{\mathrm{nl}}_{ss'}(\mathbf{k},\mathbf{k}')$ is therefore an approximation, in which the spin-orbit coupling is assumed to be dominated by the contribution from the ionic core. Conventional norm-conserving pseudopotentials incorporate all relativistic corrections of order $Z^2\alpha^2$ (where Z is the atomic number and α is the fine-structure constant), but neglect various terms of order α^2,[99,100] including the spin-orbit coupling from the long-range (but slowly varying) Coulomb potential outside the core region. This approximation is used in all that follows. It greatly simplifies the analysis of spin-dependent perturbations, as shown below.

B. Screening

Consider now the description of screening for a spin-dependent perturbation. The relationship between the total polarization Π and the regular polarization P was given above in Eq. (5.24) for a general nonlocal potential. Setting $\mathbf{G}=\mathbf{0}$ in this equation gives

$$\Pi_{ss'}(\mathbf{q};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}') = \epsilon^{-1}(\mathbf{q})P_{ss'}(\mathbf{q};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}'), \tag{9.4}$$

where $\epsilon(\mathbf{q})$ is the same scalar dielectric function defined above in Eq. (6.4). Substituting this result into Eq. (5.24) then yields

$$\Pi_{ss'}(\mathbf{q}+\mathbf{G};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}') = P_{ss'}(\mathbf{q}+\mathbf{G};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}') + P(\mathbf{q}+\mathbf{G},\mathbf{q})\frac{v_c(\mathbf{q})}{\epsilon(\mathbf{q})}P_{ss'}(\mathbf{q};\mathbf{k},\mathbf{k}+\mathbf{q}+\mathbf{G}'). \tag{9.5}$$

Equations (9.4) and (9.5) replace the scalar equations (6.5) derived previously.

If the perturbation is now separated into local and nonlocal parts, the linear screened potential (5.10) can be written in a form similar to (6.19):

$$\varphi_{ss'}(\mathbf{k},\mathbf{k}') = v^{\mathrm{an}}_{ss'}(\mathbf{k},\mathbf{k}') + \delta_{ss'}\frac{v_c(\mathbf{q})[\bar{n}(\mathbf{q})+n_{\mathrm{nl}}(\mathbf{q})]}{\epsilon(\mathbf{q})}, \tag{9.6}$$

in which $v^{\mathrm{an}}_{ss'}(\mathbf{k},\mathbf{k}')=v^{\mathrm{nl}}_{ss'}(\mathbf{k},\mathbf{k}')+\delta_{ss'}v_{\mathrm{an}}(\mathbf{q}),\mathbf{q}=\mathbf{k}-\mathbf{k}',\bar{n}(\mathbf{q})$ is the effective density (6.20) for the local potential, and

$$n_{\mathrm{nl}}(\mathbf{q}) = \sum_{\mathbf{k}''}\sum_{\mathbf{G}} P_{\lambda'\lambda}(\mathbf{q};\mathbf{k}'',\mathbf{k}''+\mathbf{q}+\mathbf{G})v^{\mathrm{nl}}_{\lambda\lambda'}(\mathbf{k}''+\mathbf{q}+\mathbf{G},\mathbf{k}'') \tag{9.7}$$

is a correction to $\bar{n}(\mathbf{q})$ from the nonlocal part of the perturbation. Since $n_{\mathrm{nl}}(\mathbf{q})=O(q)$ is an analytic function with the same site symmetry as $\bar{n}(\mathbf{q})$, this term does not produce any qualitative changes in φ; it merely renormalizes $\bar{n}(\mathbf{q})$. The only qualitatively new contribution to φ is the spin-dependent term $v^{\mathrm{nl}}_{ss'}(\mathbf{k},\mathbf{k}')$ itself, which is analytic.

For wave vectors in the vicinity of a nonzero reciprocal lattice vector, it is convenient to write Eq. (5.10) in the following alternative form:

$$\varphi_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}') = R_{\mathbf{G0}}(\mathbf{q})\varphi_{ss'}(\mathbf{k},\mathbf{k}') + \xi^{\mathbf{G}}_{ss'}(\mathbf{k},\mathbf{k}'). \tag{9.8}$$

Here $R_{\mathbf{G0}}(\mathbf{q})$ was defined in Eq. (6.27), and

$$\xi^{\mathbf{G}}_{ss'}(\mathbf{k},\mathbf{k}') = \delta_{ss'}[\xi_{\mathbf{G}}(\mathbf{q}) + (1-\delta_{\mathbf{G0}})v_c(\mathbf{q}+\mathbf{G})n_{\mathrm{nl}}(\mathbf{q}+\mathbf{G})] + v^{\mathrm{nl}}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}') - R_{\mathbf{G0}}(\mathbf{q})v^{\mathrm{nl}}_{ss'}(\mathbf{k},\mathbf{k}') \tag{9.9}$$

is a generalization of the function $\xi_{\mathbf{G}}(\mathbf{q})$ defined in Eq. (6.28). This is an analytic function of $\mathbf{k}$ and $\mathbf{k}'$ for $q<G_{\min}$.

If Eqs. (9.6) and (9.8) are inserted into the nonlocal version of Eq. (8.1) [i.e., Eq. (5.12)], it is apparent that the nonlocal part of the perturbation produces no qualitative change in the total linear potential $V^{(1)}$. The only change is a simple renormalization of the analytic and nonanalytic terms in $V^{(1)}$.

The same conclusion also holds for the quadratic potential $V^{(2)}$. Thus, the correct qualitative form of $V^{(1)}$ and $V^{(2)}$ can be derived by ignoring the nonlocal (and spin-dependent) part of the perturbing potential, and including spin only in the vertex function Λ and the analytic parts of $V^{(1)}$ and $V^{(2)}$. This is precisely the approach used in Secs. VI–VIII.

The key to obtaining this simple result is the fact that $v^{\mathrm{nl}}_{ss'}(\mathbf{k},\mathbf{k}')$ is analytic. As shown above, this relies upon the approximation of neglecting spin-orbit coupling outside the atomic cores. Such an approximation would also be possible (and even desirable for its simplicity) in an all-electron calculation where the core electrons are treated explicitly.

X. SUMMARY AND CONCLUSIONS

This paper has presented an analysis of the self-energy of an electron in a lattice-matched semiconductor heterostructure for small values of the crystal momentum. A general theory of nonlinear response for nonlocal spin-dependent perturbations was developed in terms of vertex functions and the static polarization, and applied to the case of quadratic response in a periodic insulator at zero temperature. A set of Ward identities was established for nonlocal spin-dependent potentials. The heterostructure perturbation was separated into a local spin-independent part and a nonlocal spin-dependent part, and the contributions from these were analyzed separately. Due to the neglect of spin-orbit coupling outside the atomic cores, the nonlocal part of the potential is analytic in momentum space. As a result, the nonlocal part of

the perturbation merely renormalizes the contributions from the local part.

The main results of the paper are presented in Eqs. (8.2), (8.9), and (8.11). The total linear potential (8.2) has the form $V^{(1)}=\Lambda\varphi+W^{(1)}$, in which all of the nonanalytic contributions come from the screened scalar potential φ. This has a form $\varphi(\mathbf{q})=v_{\mathrm{an}}(\mathbf{q})+v_c(\mathbf{q})\bar{n}(\mathbf{q})/\epsilon(\mathbf{q})$ similar to that for screening in a homogeneous system, except that the effective density $\bar{n}(\mathbf{q})$ has the site symmetry of the perturbation and the macroscopic dielectric function $\epsilon(\mathbf{q})$ has the symmetry of the reference crystal.

Spin-dependent contributions come from the analytic part $W^{(1)}$ and the vertex function Λ. The vertex function can be expanded in a Taylor series (8.5), in which q_α takes the gradient (in coordinate space) of $\varphi(\mathbf{q})$, while Q_α is a crystal momentum operator in effective-mass theory. The generalized Rashba effect comes from the term linear in q_α and Q_β, but there are other spin-splitting contributions from the lower-order terms as well. A more detailed analysis of the various terms is given in the following paper on effective-mass theory.[59]

The total quadratic potential of Eqs. (8.9) and (8.11) has a similar form $V^{(2)}=\Lambda\varphi^{(2)}+W^{(2)}$, in which $W^{(2)}$ is analytic to within the accuracy of the approximation scheme defined in the Introduction. The quadratic screened potential is $\varphi^{(2)}(\mathbf{q})=v_c(\mathbf{q})\bar{n}^{(2)}(\mathbf{q})/\epsilon(\mathbf{q})$, where the effective external density $\bar{n}^{(2)}(\mathbf{q})$ has the site symmetry of a diatomic perturbation in the reference crystal. Due to the Ward identities for an insulator, the leading term in the power series expansion of $\bar{n}^{(2)}(\mathbf{q})$ is a dipole term. The Rashba term in the quadratic potential is always negligible under the approximation scheme used here.

The results derived in this paper are used to develop a first-principles effective-mass theory in the following paper.[59] The present results are of crucial importance in establishing clearly defined limitations on the validity of this theory. Most previous formulations of effective-mass theory have been based on non-self-consistent empirical pseudopotentials, for which the possibility of long-range Coulomb interactions is not even considered. However, as shown here, long-range potentials arising from nonanalytic terms in the screening potential—and even the charge density itself—must be considered in general.

The omission of such terms is partially justified (to a certain order of approximation) in isovalent zinc-blende systems, where high crystal symmetry eliminates the contributions from dipole and quadrupole terms in the linear response.[51] However, it is not fully justified even in zinc-blende, since the leading octopole terms are of a lower order than the position dependence of the effective mass, which is often included in heterostructure effective-mass calculations. The following paper[59] accounts for all terms of the same order as the position dependence of the effective mass, including the octopole and hexadecapole potentials derived here in Eqs. (6.25) and (8.7).

A pioneering paper by Sham on effective-mass theory for shallow impurity states[58] dealt with many of the same issues (for local spin-independent potentials), but at a lower order of approximation. In particular, Sham considered only the lowest-order terms in cubic crystals. At this level of approximation, the total polarization can be treated as analytic [see Eq. (4.8) of Ref. 58], whereas the present Eq. (6.5) shows that this is no longer true for terms of higher order (such as those needed for the analysis in Ref. 59) or crystals of lower symmetry. The present work provides a systematic framework for extending Sham's analysis to crystals of general symmetry and terms of arbitrary order.

The value of establishing such a framework is demonstrated by the result (7.10) derived here for the leading dipole term in the quadratic density response. Sham has stated that the quadratic density response contains no dipole terms,[101] but the justification for this statement is not clear because no details of his calculation were given. However, numerical evidence to the contrary was subsequently provided by Dandrea, Duke, and Zunger in a first-principles study of band offsets in InAs/GaSb superlattices.[95] They deduced that the calculated difference between the macroscopic interface dipoles for InSb and GaAs interfaces must be a nonlinear effect (because such differences do not exist in linear response theory[51] in cubic crystals), but did not inquire further into its origin.

To the author's knowledge, the present derivation provides the first direct explanation for their result, and the first demonstration that dipole terms are a general feature of the quadratic density response. The magnitude of such dipoles is small—contributing 50 to 100 meV to the band offset of typical no-common-atom heterojunctions[95,102]—but they play an important role in explaining the experimentally observed asymmetry of band offsets in such systems.[102]

As a final note, it is worth drawing attention to a fundamental property of the nonlinear response of insulators that apparently is not widely known. For example, in Refs. 51 and 52, Baroni *et al.* have pointed out that "within linear response theory, the electronic charge induced by a given perturbation is proportional to the charge of the perturbing potential,"[52] which implies that "within linear response theory, isovalent substitutional impurities carry no net charge."[51] Although the restriction to linear response theory is necessary in general, the results derived here (in Sec. VII) demonstrate that the quadratic density response of an *insulator* to a charged perturbation also carries no net charge (i.e., it vanishes in the limit of small wave vectors). Indeed, upon replacing Eqs. (4.10) and (5.19) with their higher-order generalizations, one finds that this statement remains true for the nonlinear density response (4.6) of arbitrary order.

This result stems from the Ward identities (4.10) for the total static polarization and proper polarization (see Sec. V C) of an insulator at zero temperature. As a consequence of these identities, one can therefore state that in an insulator, the total electronic charge induced by a given perturbation is exactly linearly proportional to the charge of the perturbing potential. Of course, this statement assumes that the system remains insulating over the full range of the perturbation (from zero to full strength); otherwise, the perturbation theory used here is no longer valid.

ACKNOWLEDGMENTS

This work was supported by Hong Kong UGC Grant HIA03/04.SC02.

APPENDIX A: REDUCING HETEROVALENT PERTURBATIONS TO ISOVALENT PERTURBATIONS

In a heterovalent system such as Ge/GaAs,[60] the ionic perturbations are from class I. However, it is often possible to reduce such problems to an equivalent class II or III problem, because accumulations of macroscopic charge are energetically unfavorable; therefore, the interfaces grown in real heterojunctions tend to be macroscopically neutral.

For Ge/GaAs, an ideal (110) heterojunction is already neutral, but an ideal (001) or (111) interface would have a large macroscopic interface charge, leading to a large compensating interface free-carrier density.[60] Since this is not observed experimentally, the atoms in a real interface are believed to be arranged in one or more layers of mixed composition, such that the net macroscopic interface charge is zero.[51,60] (This is similar to the concept of surface reconstruction, but the interface layers differ from the bulk only in chemical composition, not structure.) In such a system it is possible to replace the heterovalent ionic perturbations with a set of equivalent isovalent perturbations, simply by grouping the atoms together in clusters.

The first step is to define quasiatomic building blocks using a modified version of Evjen's technique.[103] Let $\Omega_0(\mathbf{r})$ be a Wigner-Seitz cell (of the reference crystal) that is centered on position $\mathbf{r}$, and let $\mathcal{N}$ be the number of atoms in any primitive cell of the reference crystal. For a given atom a at position $\mathbf{r}_a$ in the heterostructure, the quasiatomic potential $\bar{v}_a$ is defined in terms of the ionic potentials $v_{a'}$ for all atoms a' via

$$\bar{v}_a = \sum_{a'} w_{a'}(a) v_{a'}. \tag{A1}$$

Here $w_{a'}(a)$ is a weight factor, defined as $w_{a'}(a)=1/\mathcal{N}$ if atom a' lies inside $\Omega_0(\mathbf{r}_a)$, $w_{a'}(a)=0$ if a' lies outside $\Omega_0(\mathbf{r}_a)$, and $w_{a'}(a)=1/m\mathcal{N}$ if a' lies on the surface of $\Omega_0(\mathbf{r}_a)$ [where m is the number of cells $\Omega_0(\mathbf{r}_a+\mathbf{R})$ that share atom a', with $\mathbf{R}$ any Bravais lattice vector of the reference crystal]. In a bulk crystal, these quasiatoms are neutral objects with the site symmetry of atom a in the reference crystal. Therefore, in a heterostructure, the quasiatoms carry a charge only near the heterojunctions.

For Ge/GaAs, each quasiatomic building block is constructed from $\frac{1}{2}$ of the potential for a given ion plus $\frac{1}{8}$ of the potential for each of its four nearest neighbors. In a bulk zinc-blende crystal, these quasiatoms have T_d symmetry and possess no charge, no dipole moment, and no quadrupole moment. Therefore, in a heterostructure, the quasiatoms have monopole, dipole, and quadrupole moments only near the heterojunctions.

It is assumed here that the ions in the mixed-composition interface layers form a periodic array, so that a two-dimensional superlattice translation symmetry exists in the directions parallel to the junction plane. In this case, one can define a three-dimensional "slab-adapted"[104,105] unit cell of *quasiatoms* that is large enough to contain 100% of the ions in the mixed-composition layers. This unit cell has no net charge (since the interface is assumed to be macroscopically neutral), and for some choices of compositional mixing, it may also have no dipole moment.[60]

Thus, if one treats these slab-adapted unit cells as the fundamental perturbations, this type of heterovalent class I perturbation can be replaced by an equivalent neutral perturbation from class II or class III. In general the interface cells do have a dipole moment, so the interface perturbations are class II, while the bulk perturbations are class III. However, an interface dipole term of $O(q^{-1})$ is physically equivalent to a bulk quadrupole term of $O(q^0)$. Therefore, the approximation scheme defined in the Introduction yields results of the same overall accuracy for both the class II interface and class III bulk perturbations in Ge/GaAs.

APPENDIX B: SYMMETRY PROPERTIES

This appendix considers some symmetry properties of G and Σ. Time-reversal symmetry is developed from the properties of zero- and one-particle states. The vacuum state $|0\rangle$ is defined to be time-reversal invariant:

$$\hat{\Theta}|0\rangle = |0\rangle, \tag{B1}$$

where $\hat{\Theta}$ is the antiunitary time-reversal operator. The phase of $\hat{\Theta}$ may be partially defined by letting the single-particle basis states $|\mathbf{x},s\rangle = \hat{\psi}_s^\dagger(\mathbf{x})|0\rangle$ satisfy

$$\hat{\Theta}|\mathbf{x},s\rangle = (-1)^{s-1/2}|\mathbf{x},-s\rangle. \tag{B2}$$

This relation is consistent with the operator equation[106]

$$\hat{\Theta}\hat{\psi}_s^\dagger(\mathbf{x})\hat{\Theta}^\dagger = (-1)^{s-1/2}\hat{\psi}_{-s}^\dagger(\mathbf{x}). \tag{B3}$$

One may therefore define $\hat{\Theta}$ over the entire many-particle Fock space by Eqs. (B1) and (B3).

The next step is to use the identity[107]

$$\langle\beta|\hat{A}|\alpha\rangle = \langle\tilde{\beta}|\hat{\Theta}\hat{A}\hat{\Theta}^\dagger|\tilde{\alpha}\rangle^* = \langle\tilde{\alpha}|\hat{\Theta}\hat{A}^\dagger\hat{\Theta}^\dagger|\tilde{\beta}\rangle, \tag{B4}$$

in which $\hat{A}$ is a linear operator and $|\tilde{\alpha}\rangle = \hat{\Theta}|\alpha\rangle$. If the many-particle Hamiltonian $\hat{H}$ is time-reversal invariant (i.e., $[\hat{\Theta},\hat{H}]=0$), one has

$$\hat{\Theta}[\hat{\psi}_s(\mathbf{x},\tau)]^\dagger\hat{\Theta}^\dagger = (-1)^{s-1/2}\hat{\psi}_{-s}^\dagger(\mathbf{x},-\tau), \tag{B5}$$

which holds for complex τ. From this and Eq. (B4) one immediately obtains

$$G_{ss'}(\mathbf{x},\tau;\mathbf{x}',\tau') = (-1)^{s-s'}G_{-s',-s}(\mathbf{x}',-\tau';\mathbf{x},-\tau), \tag{B6}$$

which is the generalization of an ordinary Green-function "reciprocity" relation[108] to the interacting-particle case. (A similar expression was given in Ref. 79, but with the sign term omitted.) Since the change of time variables in (B6) does not alter $\tau-\tau'$, the Fourier transform of (B6) is just

$$G_{ss'}(\mathbf{x},\mathbf{x}',\omega) = (-1)^{s-s'}G_{-s',-s}(\mathbf{x}',\mathbf{x},\omega). \tag{B7}$$

Now $\hat{H}$ is time-reversal invariant if and only if h is, which from (B2) and (B4) implies that

$$h_{ss'}(\mathbf{x},\mathbf{x}') = (-1)^{s-s'} h_{-s',-s}(\mathbf{x}',\mathbf{x}). \quad \text{(B8)}$$

Equation (2.6) then shows that Σ has the same time-reversal properties as G:

$$\Sigma_{ss'}(\mathbf{x},\mathbf{x}',\omega) = (-1)^{s-s'} \Sigma_{-s',-s}(\mathbf{x}',\mathbf{x},\omega). \quad \text{(B9)}$$

G and Σ may also satisfy other conditions derived from linear symmetries of $\hat{H}$. Consider the linear many-particle operator $\hat{Q}$ defined by

$$\hat{Q} = \sum_{s,s'} \int\int \hat{\psi}_s^\dagger(\mathbf{x}) q_{ss'}(\mathbf{x},\mathbf{x}') \hat{\psi}_{s'}(\mathbf{x}') d^3x\, d^3x', \quad \text{(B10)}$$

in which q is a linear single-particle operator. $\hat{Q}$ obeys the commutation relations

$$[\hat{\psi}_s(\mathbf{x}), \hat{Q}] = \sum_{s'} \int q_{ss'}(\mathbf{x},\mathbf{x}') \hat{\psi}_{s'}(\mathbf{x}') d^3x', \quad \text{(B11a)}$$

$$[\hat{\psi}_s^\dagger(\mathbf{x}), \hat{Q}] = -\sum_{s'} \int \hat{\psi}_{s'}^\dagger(\mathbf{x}') q_{s's}(\mathbf{x}',\mathbf{x}) d^3x'. \quad \text{(B11b)}$$

The commutator of any two such operators is another operator with the same form:

$$[\hat{Q}_1, \hat{Q}_2] = \hat{Q}_3, \quad q_3 \equiv [q_1, q_2]. \quad \text{(B12)}$$

Now suppose that the Hamiltonian has the symmetry $[\hat{K},\hat{Q}]=[\hat{H},\hat{Q}]=0$. From (B12), this is possible only if $[h,q]=0$. One can then use Eqs. (2.1) and (B11) and the cyclic property of the trace to show that

$$[G,q] = 0, \quad \text{(B13)}$$

which further implies that $[\Sigma,q]=0$.

This result is applied to lattice translations throughout this paper, and to other space group operations in the following paper.[59] Also of interest in the present paper is the spin operator $\mathbf{s}_{ss'}(\mathbf{x},\mathbf{x}')=\frac{1}{2}\boldsymbol{\sigma}_{ss'}\delta(\mathbf{x}-\mathbf{x}')$, where $\boldsymbol{\sigma}$ is the Pauli spin matrix vector. If h is independent of spin (i.e., $[h,\mathbf{s}]=0$), then $[G,\mathbf{s}]=0$, and G and Σ have the scalar form

$$\Sigma_{ss'}(\mathbf{x},\mathbf{x}',\omega) = \delta_{ss'}\Sigma(\mathbf{x},\mathbf{x}',\omega). \quad \text{(B14)}$$

However, if h includes spin-orbit coupling (which is the case studied here), Σ is nondiagonal.

APPENDIX C: FOURIER TRANSFORMS

The Fourier transforms of the potential with respect to momentum and frequency are defined by

$$V(\mathbf{k},\mathbf{k}') = \frac{1}{\Omega}\int_\Omega \int_\Omega e^{-i\mathbf{k}\cdot\mathbf{x}} V(\mathbf{x},\mathbf{x}') e^{i\mathbf{k}'\cdot\mathbf{x}'} d^3x\, d^3x', \quad \text{(C1)}$$

$$V(\zeta_n,\zeta_{n'}) = \frac{1}{\beta}\int_0^\beta \int_0^\beta e^{i\zeta_n\tau} V(\tau,\tau') e^{-i\zeta_{n'}\tau'} d\tau\, d\tau'. \quad \text{(C2)}$$

Since $V(\tau,\tau')=V(\tau-\tau')$, the latter integral is always diagonal in n:

$$V(\zeta_n,\zeta_{n'}) = V(\zeta_n)\delta_{nn'}. \quad \text{(C3)}$$

For many-variable quantities such as the vertex function $\Gamma^{(\nu)}$, the Fourier integrals for $\Gamma^{(\nu)}(\mathbf{k},\mathbf{k}';\mathbf{q},\mathbf{q}';\ldots)$ have the same form as (C1) for each pair of $(\mathbf{k},\mathbf{k}')$ variables.

For a function of the form $f(\mathbf{x})=f(r)Y_l^m(\hat{\mathbf{x}})$, where Y_l^m is a spherical harmonic, the Fourier transform is of the form $f(\mathbf{k})=f(k)Y_l^m(\hat{\mathbf{k}})$, where for $k>0$,

$$f(k) = \frac{4\pi}{\Omega}(-i)^l \int_0^\infty r^2 f(r) j_l(kr) dr, \quad \text{(C4)}$$

in which $j_l(kr)$ is a spherical Bessel function. For the special case $f(r)=r^{-n}$, Eq. (6.561.14) of Ref. 109 gives

$$f(k) = \frac{4\pi}{\Omega}(-i)^l \frac{\sqrt{\pi}\Gamma\left[\frac{1}{2}(l-n+3)\right]}{2^{n-1}\Gamma\left[\frac{1}{2}(l+n)\right]} k^{n-3}, \quad \text{(C5)}$$

in which $\Gamma(z)=\int_0^\infty e^{-t}t^{z-1}dt$. Equation (C5) is valid for $k>0, n>1$, and $l>n-3$.[109]

APPENDIX D: PERTURBATION THEORY

The starting point for the perturbation theory used in Sec. VI is the standard formula[69,70]

$$\langle \hat{A}_H(\tau)\rangle = \frac{\langle T_\tau[\hat{A}_I(\tau)\hat{\mathcal{U}}]\rangle_0}{\langle \hat{\mathcal{U}}\rangle_0}, \quad \text{(D1)}$$

where $\hat{A}_H(\tau)$ is a Heisenberg picture operator, $\hat{A}_I(\tau)$ is the same operator in the interaction picture, $\langle\hat{O}\rangle$ denotes the thermal average (2.2) with respect to $\hat{K}$, $\langle\hat{O}\rangle_0$ is a thermal average with respect to $\hat{K}_0=\hat{H}_0-\mu\hat{N}$, and $\hat{\mathcal{U}}=T_\tau\{\exp[-\int_0^\beta \hat{H}_1(\tau)d\tau]\}$. If $\hat{\mathcal{U}}$ is expanded in a power series, terms of equal order in the numerator and denominator can be grouped together as

$$\langle \hat{A}_H(\tau)\rangle = \langle \hat{A}_I(\tau)\rangle_0 + \sum_{\nu=1}^\infty \frac{(-1)^\nu}{\nu!}\int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_\nu \times \langle T_\tau[\Delta\hat{H}_1(\tau_1)\cdots\Delta\hat{H}_1(\tau_\nu)\Delta\hat{A}_I(\tau)]\rangle_0, \quad \text{(D2)}$$

where $\Delta\hat{A}_{I,H}(\tau)=\hat{A}_{I,H}(\tau)-\langle\hat{A}_I(\tau)\rangle_0$. A more compact expression for (D2) is

$$\langle \Delta\hat{A}_H(\tau)\rangle = \langle T_\tau[\Delta\hat{A}_I(\tau)\hat{\mathcal{V}}]\rangle_0, \quad \text{(D3)}$$

where $\hat{\mathcal{V}}=T_\tau\{\exp[-\int_0^\beta \Delta\hat{H}_1(\tau)d\tau]\}$.

APPENDIX E: POLARIZATION

The static polarization (4.6) is defined by

$$\Pi^{(\nu)}(00',11',\ldots,\nu\nu') = \int_0^\beta d\tau_1 \cdots \int_0^\beta d\tau_\nu D^{(\nu)}(00',11',\ldots,\nu\nu'), \tag{E1}$$

where D is the dynamic polarization

$$D^{(\nu)}(00',11',\ldots,\nu\nu') = (-1)^\nu \langle T_\tau[\Delta\hat{\rho}(00')\Delta\hat{\rho}(11')\cdots\Delta\hat{\rho}(\nu\nu')]\rangle_0. \tag{E2}$$

Here a superfluous second time variable has been added (for notational convenience) to the interaction picture operators according to the definition

$$\hat{\rho}_{ss'}(\mathbf{x},\tau;\mathbf{x}',\tau') \equiv \hat{\rho}_{ss'}(\mathbf{x},\mathbf{x}',\tau-\tau'), \tag{E3}$$

where $\tau' \equiv 0$. Since $D^{(\nu)}$ is periodic (with period β) in all time variables τ_λ, but depends on time only via the intervals $\tau_\lambda - \tau_0$ (for $\lambda = 1,2,\ldots,\nu$), it follows that $\Pi^{(\nu)}$ is independent of time.

By definition, D is symmetric with respect to interchange of any pair of operators $\Delta\hat{\rho}$; thus

$$D^{(\nu)}(\ldots,ii',\ldots,jj',\ldots) = D^{(\nu)}(\ldots,jj',\ldots,ii',\ldots). \tag{E4}$$

Another constraint on D can be derived from time-reversal symmetry using the methods of Appendix B:

$$D^{(\nu)}(00',11',\ldots,\nu\nu') = (-1)^s D^{(\nu)}(\bar{0}'\bar{0},\bar{1}'\bar{1},\ldots,\bar{\nu}'\bar{\nu}). \tag{E5}$$

Here $(\bar{\lambda}) = (\mathbf{x}_\lambda, -s_\lambda, \tau_\lambda)$ and

$$s = \sum_{\lambda=0}^{\nu} (s_\lambda - s'_\lambda). \tag{E6}$$

These symmetries are valid for the static polarization Π as well (with the time variables omitted).

APPENDIX F: FUNCTIONS A, B, AND C

The functions A, B, and C introduced in Eq. (7.8) are defined by

$$A(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = \sum_{\mathbf{G}}\sum_{\mathbf{G}_1}\sum_{\mathbf{G}_2} R_{\mathbf{0G}}(\mathbf{k})\tilde{\Pi}^{(2)} \times(\mathbf{k}+\mathbf{G},\mathbf{k}_1+\mathbf{G}_1,\mathbf{k}_2+\mathbf{G}_2)R_{\mathbf{G}_1\mathbf{0}}(\mathbf{k}_1)R_{\mathbf{G}_2\mathbf{0}}(\mathbf{k}_2),$$

$$B(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = \sum_{\mathbf{G}}\sum_{\mathbf{G}_1}\sum_{\mathbf{G}_2} R_{\mathbf{0G}}(\mathbf{k})\tilde{\Pi}^{(2)} \times(\mathbf{k}+\mathbf{G},\mathbf{k}_1+\mathbf{G}_1,\mathbf{k}_2+\mathbf{G}_2)\xi_{\mathbf{G}_1}(\mathbf{k}_1)R_{\mathbf{G}_2\mathbf{0}}(\mathbf{k}_2),$$

$$C(\mathbf{k},\mathbf{k}_1,\mathbf{k}_2) = \sum_{\mathbf{G}}\sum_{\mathbf{G}_1}\sum_{\mathbf{G}_2} R_{\mathbf{0G}}(\mathbf{k})\tilde{\Pi}^{(2)} \times(\mathbf{k}+\mathbf{G},\mathbf{k}_1+\mathbf{G}_1,\mathbf{k}_2+\mathbf{G}_2)\xi_{\mathbf{G}_1}(\mathbf{k}_1)\xi_{\mathbf{G}_2}(\mathbf{k}_2). \tag{F1}$$

APPENDIX G: NONANALYTIC TERMS IN $\bar{n}_A^{(2)}(\mathbf{q})$

To leading order, the term $\bar{n}_A^{(2)}(\mathbf{q})$ in Eq. (7.8) is

$$\bar{n}_A^{(2)}(\mathbf{q}) = \frac{1}{2}A_{\alpha\beta\gamma}q_\alpha {\sum_{\mathbf{k}}}' k_\beta(q_\gamma - k_\gamma)\varphi(\mathbf{k})\varphi(\mathbf{q}-\mathbf{k}) + O(q^3), \tag{G1}$$

in which $A_{\alpha\beta\gamma}$ is the Taylor series coefficient (7.9). The only contribution to Eq. (G1) that is of order q^2 comes from the monopole terms in $\varphi(\mathbf{q})$. For any perturbation comprising a finite number of atoms, the expansion (6.21) begins as $\varphi(\mathbf{q}) = -Z_v w_c(\mathbf{q})/\Omega + O(q^{-1})$, where Z_v is the net valence charge of the ionic perturbation.

If this lowest-order term is considered, the value of Eq. (G1) can be estimated by extending the summation to all values of $\mathbf{k}$. For a cubic crystal with scalar ϵ, the convolution can be performed using the Fourier transforms in Appendix C; the result is

$$\bar{n}_A^{(2)}(\mathbf{q}) = \frac{\pi^2 Z_v^2}{8\Omega\epsilon^2}A_{\alpha\beta\gamma}\frac{q_\alpha q_\beta q_\gamma}{q} + O(q^3). \tag{G2}$$

This shows explicitly that the leading nonanalytic term in $\bar{n}_A^{(2)}(\mathbf{q})$ is $O(q^2)$.

However, this term exists only for the heterovalent perturbations of class I in the Introduction. For the isovalent perturbations ($Z_v = 0$) of classes II and III, there is no $O(q^2)$ term in $\bar{n}_A^{(2)}(\mathbf{q})$. A similar analysis shows that for class II, $\bar{n}_A^{(2)}(\mathbf{q}) = O(q^4)$, while for class III, $\bar{n}_A^{(2)}(\mathbf{q}) = O(q^6)$. Therefore, according to the approximation scheme adopted in this paper, $\bar{n}_A^{(2)}(\mathbf{q})$ is negligible in all three cases.

APPENDIX H: VERTEX FUNCTION TAYLOR SERIES

Since the proper vertex function $\tilde{\Gamma}^{(1)}$ is an analytic function of $\mathbf{k}$ and $\mathbf{k}'$, it can be expanded in a Taylor series in the variables $\mathbf{q} = \mathbf{k} - \mathbf{k}'$ and $\mathbf{Q} = \frac{1}{2}(\mathbf{k}+\mathbf{k}')$. This yields an expression similar to Eq. (8.5):

$$\tilde{\Gamma}^{(1)}_{ss'}(\mathbf{k}+\mathbf{G},\mathbf{k}'+\mathbf{G}';\mathbf{q}+\mathbf{G}'';\omega) = \tilde{\Gamma}^{(1)}_{ss'}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega) + q_\alpha\tilde{\Gamma}^{(\alpha|\cdot)}_{ss'}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega) + Q_\alpha\tilde{\Gamma}^{(\cdot|\alpha)}_{ss'}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega) + q_\alpha q_\beta\tilde{\Gamma}^{(\alpha\beta|\cdot)}_{ss'} \times(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega) + Q_\alpha Q_\beta\tilde{\Gamma}^{(\cdot|\alpha\beta)}_{ss'}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega) + q_\alpha Q_\beta\tilde{\Gamma}^{(\alpha|\beta)}_{ss'}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega) + O(q^3). \tag{H1}$$

The Taylor series coefficients in Eq. (8.5) for the effective vertex function Λ are therefore given by

$$\Lambda_{ss'}^{(\alpha|\cdot)}(\mathbf{G},\mathbf{G}';\omega) = \tilde{\Gamma}_{ss'}^{(\alpha|\cdot)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega) + \sum_{\mathbf{G}''\neq\mathbf{0}} \tilde{\Gamma}_{ss'}^{(1)}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega)R_{\mathbf{G}''\mathbf{0}}^{\alpha},$$

$$\Lambda_{ss'}^{(\cdot|\alpha)}(\mathbf{G},\mathbf{G}';\omega) = \tilde{\Gamma}_{ss'}^{(\cdot|\alpha)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega),$$

$$\Lambda_{ss'}^{(\alpha\beta|\cdot)}(\mathbf{G},\mathbf{G}';\omega) = \tilde{\Gamma}_{ss'}^{(\alpha\beta|\cdot)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega) + \sum_{\mathbf{G}''\neq\mathbf{0}} [\tilde{\Gamma}_{ss'}^{(\alpha|\cdot)}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega)R_{\mathbf{G}''\mathbf{0}}^{\beta} + \tilde{\Gamma}_{ss'}^{(1)}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega)R_{\mathbf{G}''\mathbf{0}}^{\alpha\beta}],$$

$$\Lambda_{ss'}^{(\cdot|\alpha\beta)}(\mathbf{G},\mathbf{G}';\omega) = \tilde{\Gamma}_{ss'}^{(\cdot|\alpha\beta)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega),$$

$$\Lambda_{ss'}^{(\alpha|\beta)}(\mathbf{G},\mathbf{G}';\omega) = \tilde{\Gamma}_{ss'}^{(\alpha|\beta)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega) + \sum_{\mathbf{G}''\neq\mathbf{0}} \tilde{\Gamma}_{ss'}^{(\cdot|\beta)}(\mathbf{G},\mathbf{G}';\mathbf{G}'';\omega)R_{\mathbf{G}''\mathbf{0}}^{\alpha}, \tag{H2}$$

in which $R_{\mathbf{G}\mathbf{0}}^{\alpha}$ and $R_{\mathbf{G}\mathbf{0}}^{\alpha\beta}$ are the Taylor series coefficients for $R_{\mathbf{G}\mathbf{0}}(\mathbf{q})$. In these expressions, some special values of the coefficients for the case $\mathbf{G}''=\mathbf{0}$ are given by the Ward identity (3.11):

$$\tilde{\Gamma}_{ss'}^{(1)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega) = \delta_{ss'}\delta_{\mathbf{G}\mathbf{G}'} - \frac{\partial\Sigma_{ss'}^{(0)}(\mathbf{G},\mathbf{G}';\omega)}{\partial\omega},$$

$$\tilde{\Gamma}_{ss'}^{(\cdot|\alpha)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega) = -\frac{\partial}{\partial k_\alpha} \left.\frac{\partial\Sigma_{ss'}^{(0)}(\mathbf{k}+\mathbf{G},\mathbf{k}+\mathbf{G}';\omega)}{\partial\omega}\right|_{\mathbf{k}=\mathbf{0}},$$

$$\tilde{\Gamma}_{ss'}^{(\cdot|\alpha\beta)}(\mathbf{G},\mathbf{G}';\mathbf{0};\omega) = -\frac{1}{2}\frac{\partial^2}{\partial k_\alpha \partial k_\beta} \left.\frac{\partial\Sigma_{ss'}^{(0)}(\mathbf{k}+\mathbf{G},\mathbf{k}+\mathbf{G}';\omega)}{\partial\omega}\right|_{\mathbf{k}=\mathbf{0}}. \tag{H3}$$

*Electronic address: phbaf@ust.hk

[1] É. I. Rashba, Sov. Phys. Solid State **2**, 1109 (1960).

[2] R. Winkler, *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*, Springer Tracts in Modern Physics Vol. 191 (Springer, Berlin, 2003).

[3] S. Murakami, N. Nagaosa, and S.-C. Zhang, Science **301**, 1348 (2003).

[4] S. Murakami, N. Nagaosa, and S.-C. Zhang, Phys. Rev. B **69**, 235206 (2004).

[5] J. Sinova, D. Culcer, Q. Niu, N. A. Sinitsyn, T. Jungwirth, and A. H. MacDonald, Phys. Rev. Lett. **92**, 126603 (2004).

[6] D. Culcer, J. Sinova, N. A. Sinitsyn, T. Jungwirth, A. H. MacDonald, and Q. Niu, Phys. Rev. Lett. **93**, 046602 (2004).

[7] S.-Q. Shen, M. Ma, X. C. Xie, and F. C. Zhang, Phys. Rev. Lett. **92**, 256603 (2004).

[8] S.-Q. Shen, Y.-J. Bao, M. Ma, X. C. Xie, and F. C. Zhang, Phys. Rev. B **71**, 155316 (2005).

[9] X. Ma, L. Hu, R. Tao, and S.-Q. Shen, Phys. Rev. B **70**, 195343 (2004).

[10] S. Murakami, N. Nagaosa, and S.-C. Zhang, Phys. Rev. Lett. **93**, 156804 (2004).

[11] J. Schliemann and D. Loss, Phys. Rev. B **71**, 085308 (2005).

[12] B. A. Bernevig and S.-C. Zhang, Phys. Rev. Lett. **95**, 016801 (2005).

[13] B. A. Bernevig and S.-C. Zhang, cond-mat/0412550 (unpublished).

[14] S. Zhang and Z. Yang, Phys. Rev. Lett. **94**, 066602 (2005).

[15] J. Hu, cond-mat/0503149 (unpublished).

[16] Y. K. Kato, R. C. Meyers, A. C. Gossard, and D. D. Awschalom, Science **306**, 1910 (2004).

[17] J. Wunderlich, B. Kaestner, J. Sinova, and T. Jungwirth, Phys. Rev. Lett. **94**, 047204 (2005).

[18] S. D. Ganichev, E. L. Ivchenko, S. N. Danilov, J. Eroms, W. Wegscheider, D. Weiss, and W. Prettl, Phys. Rev. Lett. **86**, 4358 (2001).

[19] S. D. Ganichev, E. L. Ivchenko, V. V. Bel'kov, S. A. Tarasenko, M. Sollinger, D. Weiss, W. Wegscheider, and W. Prettl, Nature (London) **417**, 153 (2002).

[20] S. D. Ganichev, U. Rössler, W. Prettl, E. L. Ivchenko, V. V. Bel'kov, R. Neumann, K. Brunner, and G. Abstreiter, Phys. Rev. B **66**, 075328 (2002).

[21] L. E. Golub, Phys. Rev. B **67**, 235320 (2003).

[22] S. D. Ganichev, P. Schneider, V. V. Bel'kov, E. L. Ivchenko, S. A. Tarasenko, W. Wegscheider, D. Weiss, D. Schuh, B. N. Murdin, P. J. Phillips, C. R. Pidgeon, D. G. Clarke, M. Merrick, P. Murzyn, E. V. Beregulin, and W. Prettl, Phys. Rev. B **68**, 081302(R) (2003).

[23] V. V. Bel'kov, S. D. Ganichev, P. Schneider, C. Back, M. Oestreich, J. Rudolph, D. Häagele, L. E. Golub, W. Wegscheider, and W. Prettl, Solid State Commun. **128**, 283 (2003).

[24] S. D. Ganichev and W. Prettl, J. Phys.: Condens. Matter **15**, R935 (2003).

[25] S. D. Ganichev, V. V. Bel'kov, L. E. Golub, E. L. Ivchenko, P. Schneider, S. Giglberger, J. Eroms, J. De Boeck, G. Borghs, W. Wegscheider, D. Weiss, and W. Prettl, Phys. Rev. Lett. **92**, 256601 (2004).
[26] J. Nitta, T. Akazaki, H. Takayanagi, and T. Enoki, Phys. Rev. Lett. **78**, 1335 (1997).
[27] J. Schliemann, J. C. Egues, and D. Loss, Phys. Rev. Lett. **90**, 146801 (2003).
[28] L. Leibler, Phys. Rev. B **16**, 863 (1977).
[29] F. T. Vas'ko, JETP Lett. **30**, 541 (1979).
[30] G. Bastard, Phys. Rev. B **24**, 5693 (1981).
[31] G. Bastard, Phys. Rev. B **25**, 7584 (1982).
[32] R. Lassnig, Phys. Rev. B **31**, 8076 (1985).
[33] G. Lommer, F. Malcher, and U. Rössler, Phys. Rev. B **32**, 6965 (1985).
[34] L. G. Gerchikov and A. V. Subashiev, Sov. Phys. Semicond. **26**, 73 (1992).
[35] B. A. Foreman, Phys. Rev. B **48**, R4964 (1993).
[36] P. Pfeffer and W. Zawadzki, Phys. Rev. B **52**, R14332 (1995).
[37] E. A. de Andrada e Silva, G. C. La Rocca, and F. Bassani, Phys. Rev. B **55**, 16293 (1997).
[38] T. Schäapers, G. Engels, J. Lange, T. Klocke, M. Hollfelder, and H. Lüth, J. Appl. Phys. **83**, 4324 (1998).
[39] R. Winkler, Phys. Rev. B **62**, 4245 (2000).
[40] G. B. Bachelet, D. R. Hamann, and M. Schlüter, Phys. Rev. B **26**, 4199 (1982).
[41] S. Goedecker, M. Teter, and J. Hutter, Phys. Rev. B **54**, 1703 (1996).
[42] C. Hartwigsen, S. Goedecker, and J. Hutter, Phys. Rev. B **58**, 3641 (1998).
[43] C. D. Hu and E. Zaremba, Phys. Rev. B **37**, 9268 (1988).
[44] J. M. Pitarke, R. H. Ritchie, and P. M. Echenique, Phys. Rev. B **52**, 13883 (1995).
[45] A. Bergara, I. Campillo, J. M. Pitarke, and P. M. Echenique, Phys. Rev. B **56**, 15654 (1997).
[46] A. Bergara, J. M. Pitarke, and P. M. Echenique, Phys. Rev. B **59**, 10145 (1999).
[47] V. U. Nazarov, S. Nishigaki, and T. Nagao, Phys. Rev. B **66**, 092301 (2002).
[48] K. Nagao, S. A. Bonev, A. Bergara, and N. W. Ashcroft, Phys. Rev. Lett. **90**, 035501 (2003).
[49] R. Resta, S. Baroni, and A. Baldereschi, Superlattices Microstruct. **6**, 31 (1989).
[50] S. Baroni, R. Resta, and A. Baldereschi, in *Proceedings of the 19th International Conference on the Physics of Semiconductors, Warsaw, 1988*, edited by W. Zawadzki (Institute of Physics, Polish Academy of Sciences, Warsaw, 1988), pp. 525–528.
[51] S. Baroni, R. Resta, A. Baldereschi, and M. Peressi, in *Spectroscopy of Semiconductor Microstructures*, Vol. 206 of NATO Advanced Study Institute Series B: Physics, edited by G. Fasol, A. Fasolino, and P. Lugli (Plenum, New York, 1989), pp. 251–272.
[52] M. Peressi, S. Baroni, A. Baldereschi, and R. Resta, Phys. Rev. B **41**, 12106 (1990).
[53] M. Peressi, S. Baroni, R. Resta, and A. Baldereschi, Phys. Rev. B **43**, 7347 (1991).
[54] L. Colombo, R. Resta, and S. Baroni, Phys. Rev. B **44**, 5572 (1991).
[55] S. Baroni, M. Peressi, R. Resta, and A. Baldereschi, in *Proceedings of the 21st International Conference on the Physics of Semiconductors, Beijing, 1992*, edited by P. Jiang and H.-Z. Zheng (World Scientific, Singapore, 1992), pp. 689–696.
[56] M. Peressi and S. Baroni, Phys. Rev. B **49**, 7490 (1994).
[57] B. Montanari, M. Peressi, S. Baroni, and E. Molinari, Appl. Phys. Lett. **69**, 3218 (1996).
[58] L. J. Sham, Phys. Rev. **150**, 720 (1966).
[59] B. A. Foreman, following paper, Phys. Rev. B **72**, 165345 (2005).
[60] W. A. Harrison, E. A. Kraut, J. R. Waldrop, and R. W. Grant, Phys. Rev. B **18**, 4402 (1978).
[61] D. Vanderbilt and R. D. King-Smith, Phys. Rev. B **48**, 4442 (1993).
[62] F. Bernardini, V. Fiorentini, and D. Vanderbilt, Phys. Rev. B **56**, R10024 (1997).
[63] F. Bernardini, V. Fiorentini, and D. Vanderbilt, Phys. Rev. Lett. **79**, 3958 (1997).
[64] F. Bernardini and V. Fiorentini, Phys. Rev. B **57**, R9427 (1998).
[65] V. Fiorentini, F. Bernardini, F. Della Sala, A. Di Carlo, and P. Lugli, Phys. Rev. B **60**, 8849 (1999).
[66] F. Bechstedt, U. Grossner, and J. Furthmüller, Phys. Rev. B **62**, 8003 (2000).
[67] F. Bernardini, V. Fiorentini, and D. Vanderbilt, Phys. Rev. B **63**, 193201 (2001).
[68] A. Zoroddu, F. Bernardini, P. Ruggerone, and V. Fiorentini, Phys. Rev. B **64**, 045208 (2001).
[69] A. A. Abrikosov, L. P. Gorkov, and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Dover, New York, 1975).
[70] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (Dover, New York, 2003).
[71] J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Westview, Oxford, 1998).
[72] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. 1, pp. 884–886.
[73] B. Friedman, *Principles and Techniques of Applied Mathematics* (Wiley, New York, 1956), pp. 67, 90, and 112.
[74] A. J. Layzer, Phys. Rev. **129**, 897 (1963).
[75] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 2920 (1986).
[76] M. P. Surh, M.-F. Li, and S. G. Louie, Phys. Rev. B **43**, 4286 (1991).
[77] L. A. Hemstreet, C. Y. Fong, and J. S. Nelson, Phys. Rev. B **47**, 4238 (1993).
[78] G. Theurich and N. A. Hill, Phys. Rev. B **64**, 073106 (2001).
[79] P. Nozières, *Theory of Interacting Fermi Systems* (Benjamin, New York, 1964).
[80] L. J. Sham and W. Kohn, Phys. Rev. **145**, 561 (1966).
[81] W. Jones and N. H. March, *Theoretical Solid State Physics* (Dover, New York, 1973), Vol. 1.
[82] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Saunders, Philadelphia, 1976), p. 575.
[83] L. Hedin, Phys. Rev. **139**, A796 (1965).
[84] L. Hedin and S. Lundqvist, in *Solid State Physics*, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic, New York, 1969), Vol. 23, pp. 1–181.
[85] V. Ambegaokar and W. Kohn, Phys. Rev. **117**, 423 (1960).
[86] W. Jones and N. H. March, *Theoretical Solid State Physics* (Ref. 81), Vol. 1, pp. 210–212.
[87] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[88] W. Jones and N. H. March, *Theoretical Solid State Physics* (Ref. 81), Vol. 1, p. 282.

[89] S. G. Louie, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. Lett. **34**, 155 (1975).
[90] G. Ortiz, R. Resta, and A. Baldereschi, J. Phys.: Condens. Matter **2**, 10217 (1990).
[91] M. T. Yin and M. L. Cohen, Phys. Rev. B **26**, 3259 (1982).
[92] One could also write $\varphi(\mathbf{q})=[v_{\mathrm{an}}(\mathbf{q})+v_c(\mathbf{q})\bar{n}(\mathbf{q})]/\epsilon(\mathbf{q})$ if the second term in Eq. (6.20) is omitted, but it is more convenient to let the term that is not proportional to $v_c(\mathbf{q})$ be an analytic function of $\mathbf{q}$.
[93] If one wishes to go beyond the stated level of approximation, note that Eq. (7.10) is incomplete for class I perturbations, since the nonanalytic part of $\bar{n}_A^{(2)}(\mathbf{q})$ is $O(q^2)$.
[94] M. Tinkham, *Group Theory and Quantum Mechanics* (McGraw-Hill, New York, 1964).
[95] R. G. Dandrea, C. B. Duke, and A. Zunger, J. Vac. Sci. Technol. B **10**, 1744 (1992).
[96] J. Sak, Phys. Status Solidi **27**, 521 (1968).
[97] G. P. Kerker, J. Phys. C **13**, L189 (1980).
[98] N. Troullier and J. L. Martins, Phys. Rev. B **43**, 1993 (1991).
[99] L. Kleinman, Phys. Rev. B **21**, 2630 (1980).
[100] G. B. Bachelet and M. Schlüter, Phys. Rev. B **25**, 2103 (1982).
[101] More precisely, on p. 725 of Ref. 58, Sham says that the contribution from the quadratic density response to the self-energy is "short-ranged," meaning that it contains no terms with a $1/r$ or $1/r^2$ dependence at large r.
[102] W. Seidel, O. Krebs, P. Voisin, J. C. Harmand, F. Aristone, and J. F. Palmier, Phys. Rev. B **55**, 2274 (1997).
[103] H. M. Evjen, Phys. Rev. **39**, 675 (1932).
[104] V. Heine, Proc. Phys. Soc. London **81**, 300 (1963).
[105] L. Kleinman, Phys. Rev. B **24**, 7412 (1981).
[106] E. Merzbacher, *Quantum Mechanics*, 3rd ed. (Wiley, New York, 1998), p. 612.
[107] J. J. Sakurai, *Modern Quantum Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, 1994), p. 273.
[108] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (Ref. 72), pp. 835 and 858.
[109] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products*, 5th ed. (Academic, San Diego, 1994).