Optical properties of Wannier excitons in the linear and weakly nonlinear regime

Jan R. Kuklinski* and Shaul Mukamel
Chemistry Department, University of Rochester, Rochester, New York 14627
(Received 9 April 1990)

We investigate the optical properties of Wannier excitons in bulk semiconductors in the regime of light intensities corresponding to a linear or weakly nonlinear behavior of the active optical medium. Our approach is based on the derivation of equations of motion, which for weak electronic excitation lead to the identification of a quantum-mechanical system of interacting bosons. We calculate the lowest iterative correction to the bosonic limit when the incident light beams are quasiresonant with an excitonic line. Closed expressions for the third-order nonlinear susceptibility $\chi^{(3)}$, the ac Stark shift of the excitonic line, and the effective transition dipole moment of excitons, renormalized by exciton density (phase-space filling), are derived.

I. INTRODUCTION

The interaction of light with semiconductors is of considerable current experimental and theoretical interest.[1-8] The basic principles of this interaction are well understood; the incident photons create electron-hole pairs whose subsequent dynamics and relaxation determine the optical properties. However, the systematic evaluation of many important properties, such as the nonlinearity of the light-matter–interaction process, is a nontrivial task. A particularly interesting situation occurs when the frequency of the incident light is lower than the band-gap energy (i.e., there is no direct creation of free electrons and holes), but matches the energy of excitons, which are bound electron-hole pairs. Recently this type of light-matter–interaction process has been investigated extensively, partly because the nonlinear-optical susceptibilities of semiconductors are related to the development of ultrafast electro-optical switches. In this paper we present a microscopic theory for the optical properties of bulk semiconductors in the linear and weakly nonlinear regime. The implications of the present picture on semiconductor microstructures, such as quantum wells, quantum wires, and quantum dots, will be discussed as well. The nonlinear-optical saturation of semiconductors was extensively studied, particularly in the pump-probe configuration with the pump beam being quasiresonant with an excitonic line.[2-7] The saturation of absorption for increasing density of virtual excitons created by an intense quasiresonant pump was attributed to the phase-space-filling effect,[9-11] i.e., to mutual interaction of excitons occurring when the number of excitons multiplied by the exciton volume is not negligible in comparison to the volume of the sample. Another nonlinear effect, the Stark shift of the excitonic line, was also studied using pump-probe spectroscopy.[12-14] The excitonic ac Stark shift and the third-order nonlinear susceptibility are often evaluated using simple models involving only one or two electron-hole–pair states for the semiconductor.[15-21] We will show that such models are more applicable to geometrically restricted systems such as quantum dots or molecular aggregates rather than to bulk semiconductors.[21] In geometrically restricted systems the energy of the two-exciton state may be largely shifted from twice the energy of a single-exciton state. This implies that creation of a higher number of electron-hole pairs can be blocked by the off-resonance condition, which could justify the use of such few-level models.[21] In contrast, bulk semiconductors are weakly anharmonic, and the excitons are closer to bosons with equally spaced levels.[22,23]

The starting point for our analysis is the observation that the electrons and holes created by the absorption of a photon are uniformly distributed over the medium, and the electron-hole density for a fixed number of excitons is therefore inversely proportional to the size of the medium. In this low-density limit the operators creating such electron-hole pairs obey Bose commutation rules.[24] The low-density approximation rigorously yields the linear-optics limit in both the quantum-mechanical and classical approaches. Hydrogen-atom-like (Wannier) bound-excitonic states appear naturally from the equations of motion, which include two-site dynamical variables. For low electron-hole density, the optically active medium is equivalent to a set of harmonic oscillators, corresponding to excitons with different values of quasimomentum. We next calculate the lowest-order iterative correction to the bosonic limit, which enables us to evaluate nonlinear effects in the regime of weak optical nonlinearity. The relaxation of coherence in the medium is accounted for by incorporating exciton-phonon coupling. We calculate the third-order susceptibility $\chi^{(3)}$ and the Stark shift in a pump-probe configuration. Both are shown to have the characteristic form of a weakly anharmonic oscillator. This result, together with the linear-optics solution, brings us to the conclusion that an excitonic system can be very well modeled as a set of weakly anharmonic oscillators. The light-matter interaction couples each plane-wave component of the electromagnetic field with the excitonic state having the same wave vector. The lowest-order anharmonicity has a form of a quartic potential in the exciton and electric field operators. Calculating linear-optical properties in condensed phases using harmonic-oscillator models for the medium is a well-established procedure that dates back to Lorentz.[25] The addition of anharmonicity in order to account for nonlinear optics is a natural generalization, which was demonstrated by Bloembergen, who used a cubic anharmonic-oscillator model.[26] The present work shows that $\chi^{(3)}$ of bulk semiconductors can be modeled using a quartic anharmonic oscillator [Eq. (4.1)] whose anharmonicity depends on the driving field.

The present theory provides a unified picture which interpolates between the limiting cases whereby the exciton size is small or large compared with the lattice constant (Frenkel and Wannier excitons, respectively). Using the equations of motion developed in this article, the calculation of the linear and the weakly nonlinear-optical response of excitons is straightforward, since it involves only the classical-like, coherent amplitudes of the excitonic states and of the electromagnetic field. We investigate electronic renormalization effects such as the Stark shift and the phase-space filling for a pump-probe experiment with either an off-resonant or resonant pump, and obtain analytical results to first order in exciton density. Previous calculations based on the Keldysh nonequilibrium Green-function technique[2,9–11] were performed only in the limit of the off-resonant pump, and the inclusion of Wannier excitons required a numerical solution of the resulting equations.[10] The theory presented in this paper is for bulk semiconductors; however, some general conclusions can be immediately drawn for quasi-two- and one-dimensional semiconductors (quantum wells and quantum wires), as well as for semiconductor microcrystallites (quantum dots). We show that as the size or temperature is varied, the nonlinear properties of a small semiconductor sample can undergo a transition from the limit of semiclassical anharmonic-oscillator behavior, characteristic for bulk semiconductors, to the limit when only one and two electron-hole–pair states are excited, so that a few-level–system picture is applicable.

II. LINEAR-OPTICAL RESPONSE IN THE LOW-EXCITON-DENSITY REGIME

The analysis presented in this paper is based on the tight-binding model of semiconductors.[3,24] The model Hamiltonian for the electronic system coupled to the electromagnetic field has the form[27]

$$\hat{H}=\int d\mathbf{k}\,(\hat{a}^\dagger_{\mathbf{k}}\hat{a}_{\mathbf{k}}+\tfrac{1}{2})\hbar\omega_{\mathbf{k}}+\int d\mathbf{r}\,\hat{\psi}^\dagger(\mathbf{r})[\hat{p}(\mathbf{r})-e\hat{A}(\mathbf{r})]^2\hat{\psi}(\mathbf{r})$$
$$+\int d\mathbf{r}_1\int d\mathbf{r}_2\,\hat{\psi}^\dagger(\mathbf{r}_1)\hat{\psi}^\dagger(\mathbf{r}_2)V^C(\mathbf{r}_1-\mathbf{r}_2)\hat{\psi}(\mathbf{r}_2)\hat{\psi}(\mathbf{r}_1)+\int d\mathbf{r}\,\hat{\psi}^\dagger(\mathbf{r})V^B(\mathbf{r})\hat{\psi}(\mathbf{r})\;. \tag{2.1a}$$

Here, $\hat{\psi}(\mathbf{r})$ is the field operator representing the electrons, $\hat{a}^\dagger_{\mathbf{k}}$ is the photon-creation operator, $\hbar\omega_{\mathbf{k}}=\hbar kc$ is the energy of the photon, $\hat{p}(\mathbf{r})$ is the electron-momentum operator, e is the electron charge, and $\hat{A}(\mathbf{r},t)$ is the vector potential,

$$\hat{A}(\mathbf{r},t)=\sum_{\mathbf{k}}\left[\frac{2\pi c\hbar}{\mathcal{V}|\mathbf{k}|}\right]^{1/2}[\hat{a}_{\mathbf{k}}(t)+\hat{a}^\dagger_{\mathbf{k}}(t)]e^{i\mathbf{k}\cdot\mathbf{r}}\;, \tag{2.1b}$$

where $\mathcal{V}$ is the quantization volume. Adopting the Coulomb gauge, the electromagnetic potential satisfies $\vec{\nabla}\cdot\vec{\hat{A}}(\mathbf{r})=0$. The electromagnetic interactions are then split into two parts: a minimal coupling part describing the interaction with the transverse modes of the electromagnetic field, and a direct, unretarded Coulomb interaction $V^C(\mathbf{r})$ of pairs of electrons. The interaction of electrons with the nuclei is given by $V^B(\mathbf{r})$.

To proceed further, we need to adopt a more specific model. We assume that the electrons and the holes move among the discrete lattice sites and that each site has two electronic states with a ground-state wave function $\Phi^a(\mathbf{r})$ and an excited-state wave function $\Phi^b(\mathbf{r})$ with corresponding energies ω_a and ω_b, respectively. This leads to the following form of the electronic field operator:

$$\hat{\psi}(\mathbf{r})=\sum_{\mathbf{n}}[\hat{c}_{\mathbf{n}}\Phi^b(\mathbf{r}-\mathbf{r}_{\mathbf{n}})+\hat{d}^\dagger_{\mathbf{n}}\Phi^a(\mathbf{r}-\mathbf{r}_{\mathbf{n}})]\;. \tag{2.2}$$

The system ground state in which all sites are in the lower Φ^a state will be denoted $|\Omega\rangle$. We further introduce operators creating electrons (holes) at site m, $\hat{c}^\dagger_{\mathbf{m}}$ ($\hat{d}^\dagger_{\mathbf{m}}$). The electronic state describing the electron in the conduction band localized on the mth site is given by $\hat{c}^\dagger_{\mathbf{m}}|\Omega\rangle$, and the state describing a hole localized on that site is $\hat{d}^\dagger_{\mathbf{m}}|\Omega\rangle$. These operators satisfy the Pauli commutation rules

$$[\hat{c}_{\mathbf{n}},\hat{c}^\dagger_{\mathbf{m}}]=(\hat{I}-2\hat{c}^\dagger_n\hat{c}_n)\delta_{nm}\,[\hat{d}_{\mathbf{n}},\hat{d}^\dagger_{\mathbf{m}}]=(\hat{I}-2\hat{d}^\dagger_n d_n)\delta_{nm}\;,$$
$$\{\hat{c}_{\mathbf{n}},\hat{c}^\dagger_{\mathbf{n}}\}=\hat{I},\quad\{\hat{d}_{\mathbf{n}},\hat{d}^\dagger_{\mathbf{n}}\}=\hat{I}\;.$$

We next substitute Eq. (2.2) into Eq. (2.1) and recast it in the form of the band-edge, tight-binding Hamiltonian[24] (see Appendix A)

$$\hat{H}=\sum_{s=0}^{3}\hat{H}_s\;, \tag{2.3a}$$

where

$$\hat{H}_0=\int d\mathbf{k}(\hat{a}^\dagger_{\mathbf{k}}\hat{a}_{\mathbf{k}}+\tfrac{1}{2})\hbar\omega_{\mathbf{k}}+\sum_{\mathbf{n}}\hbar(\hat{c}^\dagger_{\mathbf{n}}\hat{c}_{\mathbf{n}}\omega_a+\hat{d}^\dagger_{\mathbf{n}}\hat{d}_{\mathbf{n}}\omega_b)\,, \tag{2.3b}$$

$$\hat{H}_1=\sum_{\mathbf{n},\mathbf{m}'}{}'(T^e\hat{c}^\dagger_{\mathbf{n}}\hat{c}_{\mathbf{n}+\mathbf{m}'}+T^h\hat{d}^\dagger_{\mathbf{n}}\hat{d}_{\mathbf{n}+\mathbf{m}'})\,,$$

$$\hat{H}_2=\int d\mathbf{k}\sum_{\mathbf{n}}g_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}_{\mathbf{n}}}(\hat{a}_{\mathbf{k}}+\hat{a}^\dagger_{-\mathbf{k}})i(\hat{c}_{\mathbf{n}}\hat{d}_{\mathbf{n}}-\hat{c}^\dagger_{\mathbf{n}}\hat{d}^\dagger_{\mathbf{n}})+\int d\mathbf{k}_1\int d\mathbf{k}_2\sum_{\mathbf{n}}\frac{2\pi c\hbar e}{\mathcal{V}\sqrt{|\mathbf{k}_1|\,|\mathbf{k}_2|}}(\hat{a}_{\mathbf{k}_1}+\hat{a}^\dagger_{-\mathbf{k}_1})(\hat{a}_{\mathbf{k}_2}+\hat{a}^\dagger_{-\mathbf{k}_2})e^{i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{r}_{\mathbf{n}}}\,, \tag{2.3c}$$

$$\hat{H}_3=\sum_{\mathbf{n},\mathbf{m}}V^{\mathrm{DD}}(\mathbf{r}_{\mathbf{n}}-\mathbf{r}_{\mathbf{m}})(\hat{c}_{\mathbf{n}}\hat{d}_{\mathbf{n}}+\hat{c}^\dagger_{\mathbf{n}}\hat{d}^\dagger_{\mathbf{n}})(\hat{c}_{\mathbf{m}}\hat{d}_{\mathbf{m}}+\hat{c}^\dagger_{\mathbf{m}}\hat{d}^\dagger_{\mathbf{m}}) \tag{2.3d}$$

$$+\sum_{\mathbf{n},\mathbf{m}}V^{C}(\mathbf{r}_{\mathbf{n}}-\mathbf{r}_{\mathbf{m}})(\hat{c}^\dagger_{\mathbf{n}}\hat{c}_{\mathbf{n}}-\hat{d}^\dagger_{\mathbf{n}}\hat{d}_{\mathbf{n}})(\hat{c}^\dagger_{\mathbf{m}}\hat{c}_{\mathbf{m}}-\hat{d}^\dagger_{\mathbf{m}}\hat{d}_{\mathbf{m}}) \tag{2.3e}$$

$$+\sum_{\mathbf{n},\mathbf{m}}V^{\mathrm{DM}}(\mathbf{r}_{\mathbf{n}}-\mathbf{r}_{\mathbf{m}})(\hat{c}^\dagger_{\mathbf{n}}\hat{c}_{\mathbf{n}}-\hat{d}^\dagger_{\mathbf{n}}\hat{d}_{\mathbf{n}})(\hat{c}^\dagger_{\mathbf{m}}\hat{d}^\dagger_{\mathbf{m}}+\hat{c}_{\mathbf{m}}\hat{d}_{\mathbf{m}})\,, \tag{2.3f}$$

where $\sum'_{\mathbf{n},\mathbf{m}}$ denotes the sum over neighboring sites, and the Coulomb (C), dipole-monopole (DM), and dipole-dipole (DD) potential energies are given by

$$V^C(\mathbf{r})=\frac{e^2}{\epsilon_0 r},\quad V^{\mathrm{DM}}(\mathbf{r})=(\mathbf{e}_d\cdot\nabla_r)\left[\frac{e^2}{\epsilon_0 r}\right]=\frac{\mu_{ab}e^2}{\epsilon_0 r^2}\cos(\theta),\quad V^{\mathrm{DD}}(\mathbf{r})=(\mathbf{e}_d\cdot\nabla_{-r})\cdot(\mathbf{e}_d\cdot\nabla_r)\left[\frac{e^2}{\epsilon_0 r}\right]=\frac{(\mu_{ab}e)^2}{\epsilon_0 r^3}\cos(\theta_1)\cos(\theta_2)\,.$$

For simplicity, we assume that all transition dipoles are parallel and that the incident radiation field is polarized in the same direction given by the unit vector $\mathbf{e}_d$, θ_j is the angle of the jth dipole with the interparticle axis $\mathbf{r}$, and ϵ_0 is the static dielectric constant. H_0 describes the free electromagnetic field and the free electrons and holes, $\hat{H}_1$ describes the mobility of electrons and holes, i.e., the transfer of carriers between sites, and H_2 describes the phonon interactions with electron-hole pairs. Finally, $\hat{H}_3$ describes the various components of the unretarded Coulomb interaction,[28] the interaction between dipoles $[V^{\mathrm{DD}}(\mathbf{r}_{\mathbf{n}}-\mathbf{r}_{\mathbf{m}})]$, between charged sites $[V^C(\mathbf{r}_{\mathbf{n}}-\mathbf{r}_{\mathbf{m}})]$, and the charge-dipole interaction $[V^{\mathrm{DM}}(\mathbf{r}_{\mathbf{n}}-\mathbf{r}_{\mathbf{m}})]$.

The Frenkel-exciton limit is obtained when the electron and hole mobilities vanish (i.e., $T^e=T^h=0$). In this case the second and third terms in $\hat{H}_3$ can be neglected since all sites remain neutral, and the states corresponding to electron-hole pairs located at different sites (i.e., $\hat{c}^\dagger_{\mathbf{n}}\hat{d}^\dagger_{\mathbf{m}}|\Omega\rangle$ with $\mathbf{n}\neq\mathbf{m}$) are never populated.

We will investigate the quantum dynamics of our model system starting with the Heisenberg equations for the operators $\hat{a}_{\mathbf{k}}(t)$ and the two-body electron operators $\hat{Y}_{\mathbf{nm}}(t)\equiv\hat{c}_{\mathbf{n}}(t)\hat{d}_{\mathbf{m}}(t)$. The choice of the two-body operators [rather than the single-electron operators $\hat{c}_{\mathbf{n}}(t)$ and $\hat{d}_{\mathbf{m}}(t)$] as the relevant dynamical variables is natural, since in this system electrons and holes are always created and annihilated in pairs. The commutation relations for the two-body operators $\hat{Y}_{\mathbf{nm}}(t)$ may be derived using Eqs. (B3) and (B4) and have the form

$$[\hat{Y}_{\mathbf{nm}},\hat{Y}^\dagger_{\mathbf{n'm'}}]=\delta_{\mathbf{nn'}}\delta_{\mathbf{mm'}}+\delta_{\mathbf{nn'}}\xi_{\mathbf{m'm}}\hat{C}_{\mathbf{m'm}}+\delta_{\mathbf{mm'}}\xi_{\mathbf{nn'}}\hat{D}_{\mathbf{n'n}}+(\xi_{\mathbf{nn'}}\xi_{\mathbf{mm'}}-1)\hat{C}_{\mathbf{m'm}}\hat{D}_{\mathbf{n'n'}}\,, \tag{2.4}$$

where $\hat{C}_{\mathbf{nm}}\equiv\hat{c}^\dagger_{\mathbf{n}}\hat{c}_{\mathbf{m}}$, $\hat{D}_{\mathbf{nm}}\equiv\hat{d}^\dagger_{\mathbf{n}}\hat{d}_{\mathbf{m}}$, and $\xi_{\mathbf{nm}}\equiv1-2\delta_{\mathbf{nm}}$.

In the limit of low electron and hole density $\langle\hat{C}_{\mathbf{nn}}\rangle\to0$ and $\langle\hat{D}_{\mathbf{nn}}\rangle\to0$, we get[24]

$$\langle[\hat{Y}_{\mathbf{nm}},\hat{Y}^\dagger_{\mathbf{n'm'}}]\rangle\cong\delta_{\mathbf{nn'}}\delta_{\mathbf{mm'}}\,, \tag{2.5}$$

where $\langle\cdots\rangle$ denotes the expectation value, which implies that the operators $\hat{Y}_{\mathbf{nm}}$ satisfy bosonic commutation rules in this limit (note that for N sites, we have N^2 independent boson variables). Using the commutators derived in Appendix B [Eqs. (B5)–(B10)], we obtain the Heisenberg equations of motion for the operators $\hat{Y}_{\mathbf{nm}}(t)$ and $\hat{a}_{\mathbf{k}}(t)$:

$$\dot{\hat{a}}_{\mathbf{k}}(t)=-i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}(t)+g_{\mathbf{k}}\sum_{\mathbf{n}}e^{i\mathbf{k}\cdot\mathbf{r}_{\mathbf{n}}}[\hat{Y}_{\mathbf{nn}}(t)-\hat{Y}^\dagger_{\mathbf{nn}}(t)]-i\sum_{\mathbf{n}}\int d\mathbf{k}'\,v_{\mathbf{kk}'}e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_{\mathbf{n}}}(\hat{a}_{\mathbf{k}'}+\hat{a}^\dagger_{-\mathbf{k}'})\,, \tag{2.6a}$$

$$\dot{\hat{Y}}_{\mathbf{nm}}(t)=-i(\omega_{ba}+V^C(\mathbf{r}_{\mathbf{n}}-\mathbf{r}_{\mathbf{m}}))\hat{Y}_{\mathbf{nm}}(t)+i\sum_{\mathbf{n}'}{}'[T^e+V^{\mathrm{DD}}(\mathbf{r}_{\mathbf{n}'})][\hat{Y}_{\mathbf{n},\mathbf{m}+\mathbf{n}'}(t)+\hat{Y}^\dagger_{\mathbf{n},\mathbf{m}+\mathbf{n}'}(t)]$$
$$+[T^h+V^{\mathrm{DD}}(\mathbf{r}_{\mathbf{n}}')][\hat{Y}_{\mathbf{n}+\mathbf{n}',\mathbf{m}}(t)+\hat{Y}^\dagger_{\mathbf{n}+\mathbf{n}',\mathbf{m}}(t)]-\delta_{\mathbf{nm}}\int d\mathbf{k}\,g_{\mathbf{k}}e^{i\mathbf{k}\cdot\mathbf{r}_{\mathbf{n}}}(\hat{a}_{\mathbf{k}}(t)+\hat{a}^\dagger_{-\mathbf{k}}(t))+\hat{F}_{\mathbf{nm}}(t)\,. \tag{2.6b}$$

Here, $\omega_{ba} \equiv \omega_b - \omega_a$ denotes the band-gap frequency, and the radiation-matter coupling g_k is given by (see Appendix A)

$$g_{\mathbf{k}} = \frac{e}{m_e}\left[\frac{2\pi c\hbar}{\mathcal{V}|\mathbf{k}|}\right]^{1/2}\int d\mathbf{r}\,\Phi^{a*}(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}}(\mathbf{e}_d\cdot\nabla_{\mathbf{r}})\Phi^b(\mathbf{r}) . \tag{2.6c}$$

The coefficient $v_{\mathbf{k}\mathbf{k}'}$ is given by

$$v_{\mathbf{k}\mathbf{k}'} \equiv \frac{2\pi c\hbar e^2}{\mathcal{V}\sqrt{|\mathbf{k}|\,|\mathbf{k}'|}} . \tag{2.6d}$$

It should be noted that the dipole approximation was not invoked in the derivation of Eq. (2.6c). The dipole approximation is made by setting $e^{i\mathbf{k}\cdot\mathbf{r}} \cong 1$ in Eqs. (2.6c). Equation (2.6c) therefore contains a smooth cutoff for the electromagnetic modes with $k \gg R^{-1}$, where R describes the size of electronic wave functions $\Phi^a(r),\Phi^b(r)$. Invoking the dipole approximation, Eq. (2.6c) assumes the form

$$g_{\mathbf{k}} \cong \frac{\omega_{ab}}{\sqrt{|\mathbf{k}|}}\left[\frac{2\pi\hbar c}{\mathcal{V}}\right]^{1/2}\mu_{ab} , \tag{2.6e}$$

where μ_{ab} is the transition dipole matrix element,

$$\mu_{ab} \equiv e\int d\mathbf{r}\,\Phi^{a*}(\mathbf{r})(\mathbf{e}_d\cdot\mathbf{r})\Phi^b(\mathbf{r}) . \tag{2.6f}$$

We have partitioned the right-hand side (rhs) of Eq. (2.6b) into a part linear in $\hat{Y}_{\mathrm{nm}}$ and into other nonlinear terms denoted $\hat{F}_{\mathrm{nm}}(t)$. The nonlinear part contains terms quadratic, cubic, etc. in $\hat{Y}_{\mathrm{nm}}$ and is precisely defined in Appendix D. In the limit of low exciton density, the latter term can be neglected, which results in linearized equations of motion. When $\hat{F}_{\mathrm{nm}}(t)$ is treated perturbatively in the exciton density, we can expand the dynamics in terms of optical susceptibilities. An important property of Eq. (2.6b) is that a part of the Coulomb electron-hole interaction is present in the linearized dynamics. In the following we will employ the rotating-wave approximation (RWA) for the dipole-dipole interaction, and neglect the terms proportional to creation operators $\hat{Y}^{\dagger}_{\mathrm{nn}}$ in Eq. (2.6b). When the electron Bohr radius is large in comparison to the lattice constant, we can treat the site label as a continuous variable, which leads us to the separation of center of mass, $\mathbf{R}_{\mathrm{nm}} \equiv (r_{\mathrm{n}}m_e^* + r_{\mathrm{m}}m_h^*)/(m_e^* + m_h^*)$, and relative variables $\mathbf{r}_{\mathrm{nm}} \equiv \mathbf{r}_{\mathrm{n}} - \mathbf{r}_{\mathrm{m}}$. We further define the electron and hole effective masses $m_e^* = [T^e + \sum'_{\mathrm{n}'} V^{\mathrm{DD}}(\mathbf{r}_{\mathrm{n}})]^{-1}2a^{-2}$ and $m_h^* = [T^h + \sum'_{\mathrm{n}'} V^{\mathrm{DD}}(\mathbf{r}_{\mathrm{n}})]^{-1}2a^{-2}$, where a is the lattice constant. The linear part of Eq. (2.6b), proportional to the operators $\hat{Y}_{\mathrm{nm}}(t)$, can be diagonalized using a coordinate transformation, resulting in

$$\dot{\hat{a}}_{\mathbf{k}}(t) = -i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}(t) + g_{\mathbf{k}}\sum_{\alpha}\int d\mathbf{k}'\,h(\mathbf{k}-\mathbf{k}')\phi_\alpha(0)\hat{Y}_{\alpha\mathbf{k}'}(t) - i\sum_{\mathrm{n}}\int d\mathbf{k}'\,v_{\mathbf{k}\mathbf{k}'}e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_{\mathrm{n}}}[\hat{a}_{\mathbf{k}'}(t) + \hat{a}_{-\mathbf{k}'}] , \tag{2.7a}$$

$$\hat{Y}_{\alpha\mathbf{k}'}(t) = i(E_{\mathbf{k}'} + E_\alpha)\hat{Y}_{\alpha\mathbf{k}'}(t) - \phi_\alpha(0)\int d\mathbf{k}\,g_{\mathbf{k}}h^*(\mathbf{k}-\mathbf{k}')[\hat{a}_{\mathbf{k}}(t) + \hat{a}^{\dagger}_{-\mathbf{k}}(t)] + \hat{F}_{\alpha\mathbf{k}'}(t) , \tag{2.7b}$$

where

$$h(\mathbf{k}-\mathbf{k}') \equiv (2\pi)^{-3}\int_{\mathcal{V}} d\mathbf{r}\,e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} ,$$

and $E_{\mathbf{k}}$ is the center-of-mass kinetic energy of the exciton, i.e., $E_{\mathbf{k}} = |\mathbf{k}|^2/2(m_e^* + m_h^*)$. The diagonalizing transformation used in the derivation of Eqs. (2.7a) and (2.7b) is

$$\hat{Y}_{\alpha\mathbf{k}}(t) = \sum_{\mathrm{n,m}} e^{-i\mathbf{k}\cdot\mathbf{R}_{\mathrm{nm}}}\phi_\alpha(\mathbf{r}_{\mathrm{n}} - \mathbf{r}_{\mathrm{m}})\hat{Y}_{\mathrm{nm}}(t) , \tag{2.8}$$

where $\phi_\alpha(\mathbf{r})$ is the solution of the Schrödinger equation for a hydrogenlike atom,

$$\left[-\frac{1}{2M^*}\Delta_{\mathrm{r}} - \frac{e^2}{\epsilon_0 r}\right]\phi_\alpha(r) = E_\alpha\phi_\alpha(r) . \tag{2.9}$$

Here the effective mass M^* is given by $(M^*)^{-1} = (m_e^*)^{-1} + (m_h^*)^{-1}$, and $\phi_\alpha(\mathbf{r})$ are taken to be real and dimensionless, and normalized as $\sum_{\mathrm{n}}[\phi_\alpha(\mathbf{r}_{\mathrm{n}})]^2 = 1$. The operator $\hat{Y}^{\dagger}_{\alpha\mathbf{k}}(t)$ is the creation operator of a Wannier exciton with hydrogenlike quantum number α and center-of-mass quasimomentum $\hbar\mathbf{k}$. When the RWA with respect to the dipole-dipole interaction is not made, the diagonalizing transformation [Eq. (2.8)] should be modified and $\hat{Y}_{\alpha\mathbf{k}}$ will depend on $\hat{Y}_{\mathrm{nm}}$ as well as on $\hat{Y}^{\dagger}_{\mathrm{nm}}$.

The optical properties of the medium can be expressed in terms of the optical polarization:

$$\hat{P}(\mathbf{k},t) = \int d\mathbf{r}\,\hat{P}(\mathbf{r},t)e^{-i\mathbf{k}\cdot\mathbf{r}} . \tag{2.10a}$$

(Throughout this paper we adopt the same definition for the spatial Fourier transform for all quantities.) For our system we have

$$\hat{P}(\mathbf{k},t) = g_{\mathbf{k}}\sum_{\mathrm{n}}[e^{i\mathbf{k}\cdot\mathbf{r}_{\mathrm{n}}}\hat{Y}_{\mathrm{nn}}(t) + \mathrm{H.c.}] . \tag{2.10b}$$

Using the transformation (2.8), we get

$$\hat{P}(\mathbf{k},t) = g_{\mathbf{k}}\sum_{\alpha}\phi_\alpha(0)[\hat{Y}_{\alpha\mathbf{k}}(t) + \hat{Y}^{\dagger}_{\alpha\mathbf{k}}(t)] . \tag{2.11}$$

At this point we should comment on the limiting case when excitons can be treated as perfect bosons.[24] Consider the case when few excitonic modes denoted $\mathbf{k}=\mathbf{k}'$ are occupied, so that $\langle\hat{Y}^{\dagger}_{\alpha k'}\hat{Y}_{\alpha k'}\rangle \gg 1$, but the exciton density is uniformly distributed in space, i.e., $\langle\hat{Y}^{\dagger}_{\mathrm{nn}}\hat{Y}_{\mathrm{nn}}\rangle \cong N^{-1}\sum\langle\hat{Y}^{\dagger}_{\alpha\mathbf{k}}\hat{Y}_{\alpha\mathbf{k}}\rangle \ll 1$, where N is the total number of sites. In the limit when $N \to \infty$, but the occupation of **k** modes is kept constant (i.e., we fix the total number of excitons), the excitonic system will behave as a set of harmonic oscillators. The bosonic treatment of excitons thus corresponds to zeroth order in a systematic expansion in exciton density.

The first-order polarization is obtained by neglecting $\hat{F}_{\alpha\mathbf{k}}(t)$. The frequency- and wave-vector-dependent dielectric function is given by

$$\hat{P}^{(1)}(\mathbf{k},\omega)=\frac{\epsilon(\mathbf{k},\omega)-1}{4\pi}\hat{E}(\mathbf{k},\omega) ,$$

where $\hat{E}(\mathbf{k},\omega)=i\omega\hat{A}(\mathbf{k},\omega)$. Solving the linear part of Eq. (2.7b), we get

$$\hat{P}^{(1)}(\mathbf{k},\omega)=\sum_{\substack{\alpha\\ \epsilon\to 0^+}}\int d\mathbf{k}'\,h(\mathbf{k}-\mathbf{k}')\frac{|g_{\mathbf{k}'}\phi_\alpha(0)|^2}{(E_{\mathbf{k}'}+E_\alpha)^2-(\omega-i\epsilon)^2}\times\hat{E}(\mathbf{k}',\omega) , \quad (2.12)$$

where $\epsilon\to 0^+$. For a bulk semiconductor we have $h(\mathbf{k}-\mathbf{k}')\cong\delta(\mathbf{k}-\mathbf{k}')$ (the absorption of a photon with momentum $\hbar\mathbf{k}$ leads to the creation of an excitonic state with quasimomentum $\hbar\mathbf{k}'\cong\hbar\mathbf{k}$). We then get

$$\epsilon(\mathbf{k},\omega)=1+4\pi\sum_\alpha\frac{|g_{\mathbf{k}}\phi_\alpha(0)|^2}{(E_\alpha+E_{\mathbf{k}})^2-\omega^2} . \quad (2.13)$$

The Elliott expression of linear absorption for Wannier excitons[24] is obtained from Eq. (2.13) by setting $k=0$, and taking the imaginary part. In agreement with the Elliott theory, we note that the oscillator strength of the exciton transition is inversely proportional to the exciton volume, since $|\phi_\alpha(0)|^2$ scales as the inverse volume. The α summation shows the contribution from bound-excitonic states, corresponding to various discrete states of the hydrogen atom, as well as from continuum states corresponding to free electron-hole pairs.

Equation (2.13) shows that in the low-density limit the optically active medium is equivalent to a set of harmonic oscillators labeled α and **k**. A monochromatic component of the electromagnetic field with wave vector **k** excites only the oscillators with the same wave vector. We further note that in the Frenkel limit ($T_e,T_h\to 0$), when only tightly bound electron and hole states are created, a simple result for the creation operator of an excitonic state can also be derived (since the only value of the relative electron-hole variable is $r=0$, we need consider only the center-of-mass motion). In the Frenkel limit we first consider the discrete problem for the relative variable and change the center-of-mass motion to the continuum limit. We thus have

$$\hat{Y}_{\mathbf{k}}(t)=\sum_{\mathbf{n}}e^{i\mathbf{k}\cdot\mathbf{r}_{\mathbf{n}}}\hat{Y}_{\mathbf{nn}}(t) . \quad (2.14)$$

The dielectric function in this case has the form

$$\epsilon_F(\mathbf{k},\omega)=1+4\pi\frac{|g_{\mathbf{k}}|^2}{(E_f+E_{\mathbf{k}})^2-(\omega-i\epsilon)^2} , \quad (2.15)$$

where E_f is the energy of the excitonic bound state, i.e., $E_f\equiv\omega_{ab}-\int d\mathbf{r}\,V^{DD}(\mathbf{r})|\phi^a(\mathbf{r})|^2|\phi^b(\mathbf{r})|^2$. Note that Eq. (2.15) does not contain the contribution of free electron-hole pairs, which are not included in our model in the Frenkel limit.

III. EXCITON-PHONON COUPLING

A complete physical picture of the resonant interaction of light with excitons requires the incorporation of the damping of the excitonic coherence by phonons. The description of highly excited excitonic states coupled to a phonon bath is more complicated than the weak-phonon-coupling model employed for Frenkel excitons,[29] since highly excited Wannier excitons are likely to be dissociated into free electron-hole pairs due to interaction with the phonon bath. However, if we restrict ourselves to strongly bound excitons (i.e., excitons with a low α quantum number), and assume that the electron-phonon coupling is weak and that the density of excitons is sufficiently low, the exciton-phonon coupling may be described by a simple Hamiltonian, quadratic in exciton annihilation-creation operators and linear in the phonon displacement field, as shown in Eq. (C4). When considering only elastic-scattering processes, the exciton-phonon coupling given by (C4) reduces to

$$\hat{H}_4=\int d\mathbf{q}\,\hbar\omega_{\mathbf{q}}\hat{b}^\dagger_{\mathbf{q}}\hat{b}_{\mathbf{q}}+\sum_\alpha\int d\mathbf{q}\int d\mathbf{k}\,F_{\alpha\alpha}(\mathbf{k},\mathbf{q})\hat{Y}^\dagger_{\alpha,\mathbf{k}+\mathbf{q}}\hat{Y}_{\alpha,\mathbf{k}}(\hat{b}_{\mathbf{q}}+\hat{b}^\dagger_{-\mathbf{q}}) , \quad (3.1)$$

where $\hat{b}_{\mathbf{q}}$ is the phonon-creation operator, which satisfies the commutation relation $[\hat{b}_{\mathbf{q}},\hat{b}^\dagger_{\mathbf{q}'}]=\delta_{\mathbf{qq}'}$. Equation (3.1) can be considered the lowest-order expansion in the electron-phonon–coupling strength, which results in elastic processes. We stress that the form of coupling in Eqs. (C4) and (3.1) is found to be a straightforward generalization of the standard model of weak exciton-phonon coupling for Frenkel excitons,[29] where annihilation and creation operators of Frenkel excitons with wave vector **q** are replaced by analogous operators describing Wannier excitons.

The derivation of the equations of motion for the polarization operators of excitons starting with the microscopic electron-phonon coupling is given in Appendix C. The derivation is along the lines of that presented in Ref. 29, employing iterative expansion with respect to the exciton-phonon interaction: This kind of iterative treatment, together with the key sequence of approximations for Frenkel excitons, is known as the Bogoliubov-Tyablikov method.[30] The procedure presented in Appendix C is based on a double expansion of exciton-phonon interaction. The first expansion consists of an iterative treatment of the phonon-bath motion, i.e., to zeroth order the phonons are not perturbed by excitons, and successive orders in the interactions modify the exciton motion. The other iteration consists of expanding this result in powers of the electron-phonon coupling. The derivation of the nonlinear corrections to the bosonic picture of excitons is presented in Appendix D. This calculation consists of inserting the exact commutation relations given by Eqs. (B3)–(B10), instead of the linearized commutator (2.5), and taking the nonlinear terms $\hat{f}^{s}_{1\mathbf{k}}(t)$ into account. The resulting equations of motion including the excitonic nonlinearity are given by Eq. (D17). When the phonon dephasing of the exciton- and phonon-mediated nonlinearity represented in Eq. (C20) is incorporated as well, we obtain the following equations of motion for the coupled dynamics of excitons and photons:

$$\dot{\hat{a}}_{\mathbf{k}}(t)=-i\omega_{\mathbf{k}}\hat{a}_{\mathbf{k}}(t)+g_{\mathbf{k}}\sum_{\alpha}\phi_{\alpha}(0)\hat{Y}_{\alpha\mathbf{k}}(t)-i\sum_{n}\int d\mathbf{k}'\,v_{\mathbf{k}\mathbf{k}'}e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}_n}[\hat{a}_{\mathbf{k}'}(t)+\hat{a}^{\dagger}_{-\mathbf{k}'}(t)]\;, \tag{3.2a}$$

$$\dot{\hat{Y}}_{\alpha\mathbf{k}}(t)=-i(E_1+E_{\mathbf{k}})\hat{Y}_{\alpha\mathbf{k}}(t)-g_{\mathbf{k}}\phi_{\alpha}(0)[\hat{a}_{\mathbf{k}}(t)+\hat{a}^{\dagger}_{-\mathbf{k}}(t)]-\gamma\hat{Y}_{\alpha\mathbf{k}}(t)+\gamma\hat{N}_{\alpha\mathbf{k}}(t)+\hat{\hat{F}}_{\alpha\mathbf{k}}(t)\;, \tag{3.2b}$$

where $\hat{\hat{F}}_{\alpha\mathbf{k}}(t)$ stands for the nonlinear interactions and is a sum of six terms, i.e., $\hat{\hat{F}}_{\alpha\mathbf{k}}(t)=\sum_{s=1}^{6}\hat{f}^{(s)}_{\alpha\mathbf{k}}(t)$. The terms $\hat{f}^{(s)}_{\alpha\mathbf{k}}(t)$ with $s=1,2,\ldots,5$ represent the exciton-exciton and photon-exciton-exciton nonlinear interactions. The phonon effects enter Eq. (3.2b) in three terms.

(i) The $-\gamma\hat{Y}_{\alpha\mathbf{k}}(t)$ term describes the damping of excitonic polarization by elastic scattering.

(ii) The phonon-assisted scattering of light by excitons can be schematically viewed as a combination of two distinct processes: the first is a coherent scattering in which a coherent state of the incoming field is transformed into a coherent excitonic state. The electron and hole may subsequently recombine, giving rise to a coherent scattered wave; the other process is an incoherent contribution in which the coherent excitonic state is converted into an uncorrelated state of excitons which subsequently recombine, leading to the emission of incoherent photons (where the average electric field vanishes). The $\gamma\hat{N}_{\alpha\mathbf{k}}(t)$ term describes the generation of the incoherent component of excitonic polarization. An important difference between the two components of the scattered radiation is that the incoherent emission is isotropic, whereas the coherent component is emitted mostly in a few selected directions that satisfy exciton-phonon momentum conservation. In the following we shall only consider the coherent part of scattered radiation and neglect the $\gamma\hat{N}_{\alpha\mathbf{k}}$ term and the incoherent background.

(iii) Phonon-induced optical nonlinearity $\hat{f}^{(6)}_{\alpha\mathbf{k}}$—the first-order perturbation of the phonon motion by excitons has a feedback effect on the excitonic dynamics; this nonlinear effect is described by $\hat{f}^{(6)}_{\alpha\mathbf{k}}$ [Eq. (C20)].

The nonlinear corrections to the bosonic picture will be discussed in the next section. Here we will only consider the linearized regime, when the nonlinear term $\hat{\hat{F}}_{\alpha\mathbf{k}}(t)$ is neglected. We take the expectation value of the operators in Eqs. (3.2) and obtain a simple equation for the complex amplitude $Y_{\alpha,\mathbf{k}}(\mathbf{k},\omega)=\langle\hat{Y}_{\alpha,\mathbf{k}}(\mathbf{k},\omega)\rangle$. This results in the dielectric function

$$\epsilon(\mathbf{k},\omega)=1+4\pi\sum_{\alpha}\frac{|g_{\mathbf{k}}\phi_{\alpha}(0)|^2}{(E_{\alpha}+E_{\mathbf{k}})^2-(\omega-i\gamma_{\alpha})^2}\;. \tag{3.3}$$

We reiterate that Eq. (3.3) cannot be used for an arbitrary state α, since for highly excited excitonic states the energy spacing between Wannier excitons is sufficiently small to allow inelastic phonon-assisted transitions $\alpha\rightarrow\beta$ and the possible dissociation of excitons, which are not included in Eq. (3.1).

IV. ANHARMONIC-OSCILLATOR PICTURE AND THE NONLINEAR SUSCEPTIBILITY $\chi^{(3)}$

In order to investigated the weak nonlinear corrections to the bosonic limit, we need to perform an averaging of all quantities entering Eqs. (3.2) and derive effective equations involving only expectation values of electromagnetic field and excitonic polarization. The main difficulty lies in the evaluation of the averages of products of operators entering into the nonlinear term $\hat{\hat{F}}_{\alpha\mathbf{k}}(t)$, which result in an infinite hierarchy of coupled dynamical variables. In order to close the hierarchy, we assume that the density matrix of our system is factorized in the form of a product of field and matter coherent states. Within this approximation, products involving excitonic polarization and electromagnetic field operators are factorized as products of averages of $\mathbf{k}$-space exciton polarizations and amplitudes of the modes of the electromagnetic field. We further assume that the frequency of radiation exciting the semiconductor is quasiresonant with the 1s ($\alpha=1$) excitonic line and neglect all other excitonic states. We then obtain nonlinear equations for the amplitude, $Y_{\alpha=1,\mathbf{k}}(t)\equiv\langle\hat{Y}_{\alpha=1,\mathbf{k}}(t)\rangle$ which are truncated next, retaining only terms up to cubic order. These manipulations are performed in Appendix D, resulting in

$$\dot{Y}_{1\mathbf{k}}(t)=-(i\Omega_{\mathbf{k}}+\gamma_1)Y_{1\mathbf{k}}(t)+\mu A(t,\mathbf{k})$$
$$+\int d\mathbf{k}_1\int d\mathbf{k}_2\int d\mathbf{k}_3\,\delta(\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2-\mathbf{k}_3)[\lambda_1 iY_{1\mathbf{k}_1}(t)Y^{*}_{1-\mathbf{k}_2}(t)Y_{1\mathbf{k}_3}(t)+\lambda_2\mu A(t,\mathbf{k}_1)Y^{*}_{1-\mathbf{k}_2}(t)Y_{1\mathbf{k}_3}(t)]\;, \tag{4.1}$$

where

$$\mu\equiv\mu_{ab}\left[\frac{2\pi\hbar c\,\omega_{ab}}{\mathcal{V}}\right]^{1/2}\phi_1(0)\;.$$

The nonlinear coupling coefficients λ_1,λ_2 are given by

$$\lambda_1\equiv\phi_1(0)I_1I_2+\phi_1^2(0)I_3+2\phi_1^2(0)I_1$$
$$+12(T^e+T^h)\phi_1^2(0)+\lambda_{\text{ph}}\;, \tag{4.2a}$$

$$\lambda_2\equiv I_2\;, \tag{4.2b}$$

where I_1, I_2, and I_3 are given by

$$I_1 \equiv \frac{1}{(2\pi)^{9/2}} \int_* d\mathbf{r}\, a^{-3/2} e^{i\bar{\mathbf{k}}\cdot\mathbf{r}} V^{DD}(\mathbf{r}) , \tag{4.2c}$$

$$I_2 \equiv \frac{1}{(2\pi)^{9/2}} \int_* d\mathbf{r}_1 \int_* d\mathbf{r}_2\, a^{-3} \phi_1(\mathbf{r}_1)\phi_1(\mathbf{r}_2)\phi_1(\mathbf{r}_1-\mathbf{r}_2) , \tag{4.2d}$$

$$I_3 \equiv \frac{1}{(2\pi)^{9/2}} \int_* d\mathbf{r}\, V^{DD}(\mathbf{r})\phi_1^2(\mathbf{r}) a^{-3/2} . \tag{4.2e}$$

$\int_* d\mathbf{r}$ denotes the integral over $\mathbf{r}$ with the exclusion of $|\mathbf{r}| \le a$, where a is the lattice constant. $\bar{\mathbf{k}}$ denotes the wave vector corresponding to a light frequency resonant to the $\alpha=1$ exciton (i.e., $|\mathbf{k}c| = \Omega_{k=0}$) and satisfies $\bar{\mathbf{k}}\cdot\mathbf{e}_d = 0$. λ_{ph}, defined in Eqs. (C19) and (C21), is a phonon-induced nonlinearity.

Equations (4.1) and (4.2) constitute the main result of this paper. They describe the dynamics of our system in the limit of small and uniformly distributed exciton density $n_{ex}(\mathbf{r})$ [i.e., the number of excitons is small compared with the number of sites, $n_{ex} \cong N^{-1}\sum_{\alpha,\mathbf{k}} \langle \hat{Y}_{\alpha,\mathbf{k}}(t)\hat{Y}_{\alpha,\mathbf{k}}(t)\rangle \ll 1$]. The physical picture underlying Eq. (4.1) is of that of weakly anharmonic, semiclassical oscillators, which evolve in coherent states with the coherent amplitudes described by averages of annihilation operators, where these amplitudes are governed by the (weakly) nonlinear equation (4.1). This picture is, of course, only valid for sufficiently weak nonlinearity, since in the opposite case the quantum-mechanical density matrix can differ considerably from a product of coherent states. Using the present picture the nonlinear properties of bulk semiconductors are described by two parameters. The first parameter (λ_1) is a sum of contributions originating from the Coulomb interaction between excitons, the mobility of electrons and holes, and a phonon-mediated interaction. The latter contribution has recently attracted considerable attention.[4,6,31] The second parameter (λ_2) represents the nonlinear contribution originating from the interaction of electrons with the transverse part of the electromagnetic field.

The validity of Eq. (4.1) depends crucially on the relative magnitude of the coupling constant $g_{\mathbf{k}}$ and the dephasing rate γ_1. Consider the interaction of a coherent beam with a semi-infinite semiconducting medium in the limit of linear light-matter interaction, with infinite exciton mass. In this case we can apply a one-dimensional analysis, where the excitonic medium is modeled by the means of harmonic oscillators with an oscillator strength $\lambda_0 = |g_{\mathbf{k}}|^2/2E_1$ per unit length. Both the amplitude of the light beam and the excitonic polarization will be exponentially damped, i.e., $|P(r)|^2 \sim \exp[-(2\lambda_0/\gamma_1)r]$, and the condition for uniform distribution of the excitonic population has the form $\lambda_0 L/\gamma_1 \ll 1$, where L is the thickness of the sample. This means that the sequence of approximations leading to Eq. (4.1) is valid when $\lambda_0 L \ll \gamma_1$, corresponding to a temperature high enough for the exciton-photon interaction to be overdamped by phonons. These equations also hold in another limit, when all the incident light beams are far off resonance with respect to the excitonic line, so that $g_{\mathbf{k}}$ is small relative to the detuning of the laser frequency from the excitonic line. As we will show in the following, in this regime the interaction of a coherent beam with wave vector $\mathbf{k}_0$ excites the collective excitonic mode with center-of-mass momentum $\hbar\mathbf{k} \approx \hbar\mathbf{k}_0$, and nonlinear effects can be described using a simple picture of interacting, weakly anharmonic k-space oscillators.

When the damping is weak, $\lambda_0 L \gg \gamma_1$, the $\mathbf{k}$-exciton and $\mathbf{k}$-photon modes are strongly correlated and form new elementary modes, called polaritons.[32–34] In this case, Eq. (4.1) is no longer valid since the excitation of the semiconducting sample with monochromatic coherent light can lead to a nonuniform distribution of excitonic density and, therefore, the analysis of nonlinear effects requires addressing the full nonlinear interaction and propagation problem.

We next consider a four-wave-mixing experiment involving three incident coherent beams with frequencies $\omega_1, \omega_2, \omega_3$ and wave vectors $\mathbf{k}_1, \mathbf{k}_2, \mathbf{k}_3$. We shall calculate the nonlinear polarization with frequency $\omega = \omega_1 - \omega_2 + \omega_3$ and wave vectors $\mathbf{k} = \mathbf{k}_1 - \mathbf{k}_2 + \mathbf{k}_3$. It is given by

$$P^{(3)}(\mathbf{k},\omega) = \chi^{(3)}(-\mathbf{k},-\omega;\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_3,\omega_1,-\omega_2,\omega_3) \times E(\mathbf{k}_1,\omega_1)E^*(\mathbf{k}_2,\omega_2)E(\mathbf{k}_3,\omega_3) . \tag{4.3a}$$

The nonlinear susceptibility $\chi^{(3)}$ is obtained from Eq. (4.1) by solving the linearized equations for the excitonic polarization $Y_{\alpha=1,\mathbf{k}}(t)$, inserting the solution in the nonlinear terms on the rhs of Eq. (4.1), and treating this nonlinear term as an external perturbation. In this way we calculate the nonlinear corrections to the excitonic polarization and the $\chi^{(3)}$ susceptibility, calculated within the rotating-wave approximation,

$$\chi^{(3)}(-\mathbf{k},-\omega;\mathbf{k}_1,-\mathbf{k}_2,\mathbf{k}_3,\omega_1,-\omega_2,\omega_3) = \sum_{1,3} \frac{\mu^4}{\omega_1\omega_2\omega_3} \times \left[\frac{\lambda_1}{(\omega-\Omega_{\mathbf{k}}+i\gamma_1)(\omega_1-\Omega_{\mathbf{k}_1}+i\gamma_1)(\omega_2-\Omega_{\mathbf{k}_2}-i\gamma_1)(\omega_3-\Omega_{\mathbf{k}_3}+i\gamma_1)} + \frac{\lambda_2}{(\omega-\Omega_{\mathbf{k}}+i\gamma_1)(\omega_2-\Omega_{\mathbf{k}_2}-i\gamma_1)(\omega_3-\Omega_{\mathbf{k}_3}+i\gamma_1)} \right] , \tag{4.3b}$$

where $\hbar\Omega_{\mathbf{k}} \equiv E_1 + E_{\mathbf{k}}$ and the summation is over the permutation of $\omega_1, \mathbf{k}_1$ with $\omega_3, \mathbf{k}_3$.

Let us consider a simple four-wave-mixing experiment performed using two light beams. We set $\mathbf{k}_3 = \mathbf{k}_1$ and consider the signal at $\mathbf{k} = 2\mathbf{k}_1 - \mathbf{k}_2$. When this experiment is performed on a chromophore perturbed by collisions or by phonons, the signal shows a dephasing-induced extra resonance known as degenerate four-wave mixing.[35,36] As ω_1 is tuned across ω_2 while both ω_1 and ω_2 are kept off resonance, $\chi^{(3)}$ will show a narrow resonance at $\omega_1 = \omega_2$ whose width is determined by level relaxing (T_1 processes).[37] This extra resonance vanishes in the absence of dephasing. A recent calculation of this type for Frenkel excitons has been made.[38] This differs dramatically from the present collective quasiparticle picture that employs weak exciton-phonon coupling. Equation (4.3b) does not show the dephasing induced resonances at $\omega_1 = \omega_2$, since within the present approximations we do not have pure dephasing ($T_1 = 2T_2$).

V. ELECTRONIC RENORMALIZATION EFFECTS INDUCED BY A PUMP BEAM: ac STARK SHIFT AND PHASE-SPACE FILLING

In this section we consider the modification of optical properties of a bulk semiconductor induced by a coherent pump field tuned in the vicinity of an isolated exciton line, and calculate corrections to the 1s exciton energy, dipole moment, and relaxation rate, to first order in exciton density. We will assume that these optical properties are subsequently probed by a second beam that is weak in comparison to the pump beam, i.e., we assume that the number density of excitons created by the probe is negligible compared to that created by the pump. We further assume that the population of excitons created by the pump beam adiabatically follows the pump intensity: this limit holds when the light-matter interaction is overdamped by phonons or when the pump is off resonance with respect to the excitonic line. The total electromagnetic field and the polarization wave associated with the 1s Wannier exciton can be described in terms of slowly varying envelopes. We denote the envelopes of the pump beam $E_s(\mathbf{k},t)$ and those of the probe $E_0(t)$, assuming further that the probe field is nearly monochromatic. Similarly, we define the envelopes describing the excitonic polarization associated with the pump and probe beams as $Y_s(t)$ and $Y_0(t)$, respectively. We thus have

$$E(\mathbf{k},t) = E_0(t)\exp(i\omega_0 t)\delta(\mathbf{k}_0 - \mathbf{k}) + E_s(\mathbf{k},t)\exp(i\omega_{\mathbf{k}} t) , \quad (5.1a)$$

$$Y_{1\mathbf{k}}(t) = Y_0(t)\exp(i\omega_0 t)\delta(\mathbf{k}_0 - \mathbf{k}) + Y_s(\mathbf{k},t)\exp(i\omega_{\mathbf{k}} t) , \quad (5.1b)$$

where $\omega_{\mathbf{k}} \equiv kc$, and $\mathbf{k}_0$ denotes the wave vector of the probe beam.

Following the assumption of adiabatic behavior of excitonic polarization created by the pump beam, we have $Y_s(\mathbf{k},\omega) \cong \chi^{(1)}(\mathbf{k},\omega) E_s(\mathbf{k},\omega)$. When Eqs. (5.1) are inserted into Eq. (4.1) and a linearization is performed with respect to the $Y_0(t)$ variables, we obtain the following equation for excitonic density associated with the probe frequency, $Y_0(t)$:

$$\dot{Y}_0(t) = -[i(\Omega_r - \omega_0) + \gamma_r] Y_0(t) + \frac{\mu_r}{i\omega_0} E_0(t) , \quad (5.2)$$

with the renormalized coefficients

$$\mu_r = \mu(1 - \lambda_2 n_{\mathrm{ex}}(t)) , \quad (5.3a)$$

$$\Omega_r = \Omega_0 + \Delta_s(t) , \quad (5.3b)$$

$$\gamma_r = \gamma(1 + \lambda_2 n_{\mathrm{ex}}(t)) , \quad (5.3c)$$

where

$$n_{\mathrm{ex}}(t) = \int d\mathbf{k} \frac{|E_s(\mathbf{k},t)|^2}{[(\omega_{\mathbf{k}} - \Omega_{\mathbf{k}})^2 + \gamma_1^2]\omega_{\mathbf{k}}^2} , \quad (5.4a)$$

$$\Delta_s(t) = 2\lambda_1 n_{\mathrm{ex}}(t) + \lambda_2 \int d\mathbf{k} \frac{|E_s(\mathbf{k},t)|^2 (\omega_{\mathbf{k}} - \Omega_{\mathbf{k}})}{[(\omega_{\mathbf{k}} - \Omega_{\mathbf{k}})^2 + \gamma_1^2]\omega_{\mathbf{k}}^2} . \quad (5.4b)$$

The properties of our medium as measured by the probe beam can be described by the dielectric function (3.3) with the oscillator strength, the eigenfrequency of the excitonic line, and the damping rate replaced by the renormalized quantities μ_r, Ω_r, and γ_r, respectively. The absorption spectrum, related to the imaginary of $\epsilon(\mathbf{k},\omega)$, is then given by

$$S(\omega) = \frac{\mu_r^2 \gamma_r}{[(\omega - \Omega_r)^2 + \gamma_r^2] 2\Omega_r} . \quad (5.5)$$

Equation (5.2) is a linearized equation describing excitonic polarization, with the parameters being renormalized by terms proportional to the density of excitons created by the pump beam. Note that the renormalization of the dipole moment depends only on λ_2, whereas the Stark shift $\Delta_s(t)$ is a sum of two terms related to λ_1 and λ_2. For a monochromatic pump, we have further that

$$\Delta_s(t) = [2\lambda_1 + \lambda_2(\omega - \Omega_{\mathbf{k}})] n_{\mathrm{ex}}(t) . \quad (5.6)$$

The damping rate [Eq. (5.3c)] is modified by an additional term proportional to the excitonic density. This contribution is a result of a conversion of coherent excitons into incoherent excitons, induced by the nonlinear processes. The coupling strength μ_r is renormalized as well. This renormalization implies that excitation of a finite exciton density by the pump beam effectively reduces the oscillator strength of the medium in comparison with the linear regime. This is in agreement with the phase-space-filling (PSF) picture of the saturation of absorption in semiconductors.[10] Obviously, the results discussed by us are restricted to the limit $n_{\mathrm{ex}} \ll 1$, and for that reason we do not obtain a true saturation. It can be seen that exciton-density corrections (i.e., linear and higher-order terms in n_{ex}) to the excitonic polarization associated with the pump beam, $Y_s(k,t)$, lead to higher-order corrections to μ_r, Ω_r, and γ_r which are of second and higher order in

$n_{ex}(t)$. A simple resummation to higher exciton densities may be obtained using a Padé approximation for Eq. (5.2b) [i.e., we set $\mu_r \approx \mu(1+\lambda_2 n_{ex})^{-1}$]. However, when higher-order effects in n_{ex} are considered, the truncation of the nonlinear equation for the excitonic polarization $Y_{1k}(t)$ to third order [Eq. (4.1)] is no longer valid. This means that the Padé approximant is oversimplified, and for higher exciton densities a more elaborate microscopic picture than that given by Eq. (4.1) should be developed. In addition, for higher pump intensities, the absorption of photons from both pump and probe beams by excitons, leading to dissociation of excitons into free electron-hole pairs, cannot be disregarded. This means that for high light intensity a finite density of free carriers will be created, leading to the screening of electron-hole attraction and finally to the dissociation of bound-excitonic states.[39]

Both types of nonlinearity (λ_1 and λ_2) have a distinctly different character. The contribution to the ac Stark shift governed by λ_1 does not change its sign when a monochromatic pump beam is tuned across the excitonic line, and has its maximum when the pump beam is exactly resonant with the exciton frequency. The λ_2 contribution is dispersive, i.e., it changes its sign when the pump frequency is tuned across the pump-exciton resonance, and vanishes on resonance. The nonlinearity proportional to λ_2 contributes to renormalization of the oscillator strength of the medium, to the ac Stark shift, and to the modification of the damping rate, whereas the nonlinear interaction described by λ_1 contributes only to the ac Stark shift and to the renormalization of the damping rate. It then follows that the nonlinear-optical properties of a bulk semiconductor will depend on whether the λ_1- or λ_2-type nonlinearities are dominant. We expect the λ_1 nonlinearity to be dominant when phonon-induced nonlinearity is significant, as is the case in recent experiments.[6,31] Therefore, in this case a Stark shift of the excitonic line should be observed without renormalization of the oscillator strength of the medium.

When phonon-induced nonlinearity is not dominant, a more careful analysis is needed to estimate the relative strength of both types of nonlinearities. Let us consider the ratio of the two components entering in the $\chi^{(3)}$ nonlinear susceptibility given by Eq. (4.3). In the simple case, when all three laser beams considered in Eqs. (4.2) have the same frequencies, the ratio of the term of $\chi^{(3)}$ proportional to λ_1 to the term proportional to λ_2 is simply equal to $\eta=\lambda_1/(\lambda_2|\gamma_1+i\Delta|)$, where Δ is the detuning of these beams from the exciton frequency $\Omega_{\mathbf{k}}$. This, in turn, can be expressed as

$$\eta \simeq \frac{E_{kin}+E_c}{\hbar|\gamma_1+i\Delta|} , \tag{5.7}$$

where E_{kin} is the characteristic kinetic energy of the exciton translation and E_c is the energy of Coulomb exciton-exciton interaction. Using Eqs. (5.6) and (5.7), we see that the nonlinearity related to λ_2 is dominant when light frequencies are tuned sufficiently far off resonance with respect to the exciton line. In the opposite limit, when the light frequency is closer to resonance with the excitonic line ($|\Delta| \leq \gamma_1$), effects associated with λ_1 become dominant, and if the temperature-dependent relaxation rate γ_1 is small relative to E_{kin} and E_c, the nonlinearity in the vicinity of exciton-photon resonance is dominated by the λ_1 nonlinearity. The latter statement leads us to a conclusion that at low temperatures the optical nonlinearity near resonance will be λ_1. However, this argument has some limitations, since the phonon-dephasing rate does not vanish at zero temperature, and the radiative losses contribute to the damping of excitonic polarization. Therefore, γ_1 cannot attain an arbitrarily low value at low temperatures.

Finally, we investigate the relative magnitude of λ_1 and λ_2 as the size of the exciton is varied. We assume that the static dielectric constant of our system (ϵ_0) is varied, which modifies the exciton radius $R_{ex}=(\hbar^2/M^*e^2)\epsilon_0$. We rescale the integrals [Eqs. (4.2)] by introducing a dimensionless integration variable $\mathbf{x}=\mathbf{r}/R_{ex}$ and modify the potential-energy terms [Eqs. (2.3)], which results in the following scaling, in the limit of large exciton size:

$$\lambda_1 \sim R_{ex}^{-3}, \quad \lambda_2 \sim R_{ex}^{-3/2} .$$

This implies that the λ_2 nonlinearity dominates for large exciton radius R_{ex}.

It is interesting to make a further comparison between these two types of nonlinearities by considering the following oscillator driven by an external electric field:

$$\ddot{q}(t)+2\gamma\dot{q}(t)+\Omega_0^2 q(t)=\mu E(t)+\lambda_1 W_3\{\dot{q}(t),q(t)\}+\lambda_2 \mu E(t) W_2\{\dot{q}(t),q(t)\} , \tag{5.8}$$

where $W_n\{\dot{q},q\}$ denotes an nth-order polynomial in $\dot{q}$ and q. We assume that $\gamma,\Delta \ll \omega_0$, where Δ is the detuning between the driving field $E(t)$ and the oscillator frequency Ω_0. The third-order susceptibility and the ac Stark shift, calculated from Eq. (5.8) employing the RWA approximation following from $\Delta \ll \Omega_0$, have the same form of Eq. (5.4), provided the translational kinetic energy is neglected (i.e., $\Omega_{\mathbf{k}} \cong E_1/\hbar=\Omega_0$). We further note that for off-resonance excitation, when $\gamma \ll \Delta \ll \Omega_0$, the third-order susceptibility and the ac Stark shift originating from the λ_2 term alone have the same form as that of a two-level atom. This result seems to shed some light on the comparison between the $\chi^{(3)}$ susceptibility and the ac Stark shift of a bulk semiconductor given by Eqs. (5.4), and the same quantities calculated using models involving only a few electron-hole–pair states.[16–21] In the off-resonance regime both $\chi^{(3)}$ and Δ_s calculated either using the anharmonic-oscillator model with the λ_2 term alone or the two- or three-level models are identical. However, the λ_1 nonlinearity cannot be derived from the few-level–system model. Also, if we keep only the λ_2 nonlinearity, and consider the adiabatic on-resonance case (i.e., when light-matter interaction is overdamped by phonons), the results of both approaches are different. These arguments do not contradict the phenomenological approach to off-resonance excitation of semiconductors, modeled as a few-level system,[19] since the contribution of λ_2 becomes dominant for off-resonance pumping. In other words, far off resonance, the system of weakly anharmonic excitons can be adequately modeled as a few-level system.[19] However, the excitonic optical nonlinearity in the entire regime of off-resonance and near-resonance pumping cannot be described using such few-level models. The comparison of the nonlinear-optical features of excitonic systems such as ac Stark shift and renormalization of the oscillator strength, to those known from atomic physics, has attracted considerable attention in the literature.[10,12–18,20] Many experimental investigations used off-resonance pumping, and the ac Stark shift was observed simultaneously with reduction of the oscillator strength of the medium (saturation).[9,19] The physical picture of excitonic nonlinearity was therefore found to have similarities to that known from atomic physics.[19] However, new aspects of the excitonic nonlinearity, such as the role of biexcitons,[40] were also recognized.[18,20] Recent experiments on GaAs quantum wells showed a large Stark shift without renormalization of the oscillator strength of the medium.[12] This effect, also confirmed by theoretical calculations,[13] was the source of some controversy in the literature,[20] and demonstrates a situation where the nonlinear-optical properties of semiconductors are very different from those known from atomic physics. Using the present theory, this could be the result of a dominant role of the λ_1 nonlinearity in this system.

VI. DISCUSSION

In this paper we have studied the interaction of light with excitons in bulk semiconductors. In the regime of low exciton density and when phonon-induced dephasing is larger than the radiation-matter coupling, the medium can be described using an interacting-boson model [Eq. (4.1)]. The interaction among these bosons and with the electric field can be represented by a quartic potential [Eq. (4.1)]. The systematic evaluation of these nonlinear corrections allowed us to calculate the third-order optical susceptibility and the ac Stark shift. Our closed-form expressions are governed by two parameters, representing two different types of nonlinearity. The first (λ_1) arises from Coulomb interaction between excitons, nonbosonic corrections to the mobility of electrons and holes, and nonlinear processes mediated by phonons. The second (λ_2) arises from nonbosonic corrections to the dipolar radiation-matter coupling. The following physical picture is obtained from the present theory: the excitonic medium in the regime of weak optical nonlinearity can be modeled as a set of weakly coupled semiclassical anharmonic oscillators. Each oscillator is labeled by the wave vector $\mathbf{k}$ representing the center-of-mass momentum of the electron-hole pair, and an internal quantum number α corresponding to the hydrogenic eigenstate of their relative motion. The oscillators describe both bound and free-electron-hole pairs, i.e., both excitons and free carriers. The anharmonicity couples different oscillators and modifies the dynamics of each oscillator.

The idea that excitons at low densities behave as bosons is well established and was discussed numerous times in the literature.[22,24,41] The present microscopic calculation of optical nonlinearities were performed by considering the lowest-order iterative correction to the bosonic limit in the frame of Heisenberg equations. The present results were obtained using several assumptions and approximations. We have considered a simple tight-binding model admitting two electronic states per site. Effects related to "heavy" and light" holes present in a realistic semiconductor with a more complex band structure are not considered in our model.[19] When calculating the nonlinear properties of a bulk semiconductor, following the assumption that the incoming radiation is resonant or quasiresonant with an excitonic level, the dynamical variables associated with the other excitonic states and free carriers have been neglected, so that we could consider only a single-exciton line. In practice, however, we usually deal with a finite density of free carriers due to incoherent excitation phenomena. The latter effect can be incorporated in our calculations by replacing the Coulomb potential with a screened Yukawa form. Another limitation is that, as a consequence of the iterative treatment of the Coulomb interaction, bound states arising from attractive interactions between excitons, i.e., excitonic molecules, are not included in our calculations. This implies that the present results are valid only when we neglect resonances between the incoming light and exciton-biexciton[40] transitions (the exciton-biexciton transition can result in an additional nonlinearity associated with the exciton-biexciton transition). However, even if we stay far off such resonances, the finite-temperature phonon-mediated processes may lead to the formation of excitonic molecules. The effect of these processes on the excitonic polarization can be modeled phenomenologically by modifying the damping rate γ in Eqs. (3.2). The effects of phonon-induced generation of excitonic molecules is significant only for a limited range of parameters, since for sufficiently high density of free carriers the screening of the Coulomb potential results in the dissociation of bound-excitonic states. Finally, we have invoked the rotating-wave approximation for the exciton-phonon interaction. This is appropriate when the frequencies of all components of the electromagnetic field entering our problem are resonant or quasiresonant with a given excitonic level. Processes such as third-harmonic generation or two-photon absorption require the inclusion of additional terms, and are beyond the scope of the present calculation.

The present analysis provides a simple description of bulk semiconductors. It also allows us to draw several qualitative conclusions concerning semiconductor systems having a lower dimensionality. Let us consider a quantum-well structure in which the confinement in one direction is sufficiently strong to ensure that only one transverse mode is excited. In this case, two-dimensional (2D) electron-hole bound states are formed (2D excitons) and the analysis of this system can be straightforwardly done along the lines presented here. We identify the bosonic exciton states arising from the diagonalization of linearized excitonic dynamics and then consider nonlinear effects by deriving the effective interaction coupling these bosonic modes. The physical picture underlying such an interacting-boson picture is identical to the one considered here; when a photon is absorbed, an excitonic state with the excitonic density being inversely proportional to the volume of the sample is created. Following that, the exciton density is treated as a small expansion parameter which leads to the interacting-boson picture. This results in the same form of nonlinearity governed by two parameters (i.e., the form of the ac Stark shift, renormalized excitonic dipole strength, and third-order susceptibility is qualitatively the same). Similarly, the application of the present theory to 1D semiconductors is straightforward. Finally, we note that a similar method of evaluating the nonlinear-optical response based on the iterative solution of nonlinear equations of motion was developed earlier in the study of polariton effects in molecular crystals.[42]

Some interesting conclusions can be drawn with regard to quantum-dot microstructures. When their size is decreased, the probability that two excitons are found within the range of the nonlinear interaction increases inversely proportional to its volume. This means that the energy of the two-exciton state is shifted in comparison to twice the energy of the single-exciton state. When this shift becomes larger than the broadening associated with phonon relaxation and the radiative width, the description of the nonlinearity of our system by means of quasiclassical anharmonic oscillators is no longer possible. However, when a monochromatic excitation is tuned near a single-exciton resonance, a two- or three-level model involving only the ground state, and one and possibly two exciton states, is applicable. Such a picture is commonly used in theoretical studies of quantum dots.[3] The straightforward conclusion from our analysis is that the transition from the regime when the optical nonlinearity is described by means of a weakly anharmonic system to the one described by a few-level system is not only governed by the size of the semiconductor particle, but also by the temperature. When the temperature is increased, the nonlinear-optical properties of a small semiconducting sample can change from a behavior characteristic of a microstructure to that of a bulk semiconductor. We should stress that our conclusions hold only when the size of the quantum dot is larger than the exciton radius. In the opposite case, we reach the regime of quantum confinement, where excitonic states no longer exist.[21]

ACKNOWLEDGMENTS

We wish to thank Dr. J. Knoester for most useful discussions and critical comments. Support by the U.S. Air Force Office of Scientific Research, the Science and Technology Center for Photoinduced Charge Transfer, the U.S. National Science Foundation, and by the donors of the Petroleum Research Fund, administered by the American Chemical Society, is gratefully acknowledged. One of us (S.M.) is also grateful to the Camille and Henry Dreyfus Foundation (New York, NY) for partial support.

APPENDIX A: DERIVATION OF THE BAND-EDGE HAMILTONIAN

In this appendix we outline the derivation of the band-edge Hamiltonian [Eqs. (2.3)] starting with Eq. (2.1a).

The Hamiltonian (2.1) can be partitioned as follows:

$$\hat{H}=\hat{H}_0+\hat{H}_1+\hat{H}_2+\hat{H}_3 \;, \tag{A1}$$

where

$$\hat{H}_0+\hat{H}_1=\int d\mathbf{r}\,\hat{\psi}^\dagger(\mathbf{r})\hat{p}^{\,2}\hat{\psi}(\mathbf{r})+\int d\mathbf{r}\,\hat{\psi}^\dagger(\mathbf{r})V^B(\mathbf{r})\hat{\psi}(\mathbf{r}) + \int d\mathbf{k}\,\hbar\omega_{\mathbf{k}}(\hat{a}^\dagger_{\mathbf{k}}\hat{a}_{\mathbf{k}}+\tfrac{1}{2}) \;, \tag{A2a}$$

$$\hat{H}_2=e^2\int d\mathbf{r}\,\hat{\psi}^\dagger(\mathbf{r})\hat{A}^{\,2}\psi(\mathbf{r})-2e\int d\mathbf{r}\,\hat{\psi}^\dagger(\mathbf{r})\hat{p}\;\hat{A}\;\hat{\psi}(\mathbf{r}) \;, \tag{A2b}$$

$$\hat{H}_3=\int d\mathbf{r}_1\int d\mathbf{r}_2\,\hat{\psi}^\dagger(\mathbf{r}_1)\hat{\psi}^\dagger(\mathbf{r}_2)V^C(\mathbf{r}_1-\mathbf{r}_2)\hat{\psi}(\mathbf{r}_2)\hat{\psi}(\mathbf{r}_1) \;. \tag{A2c}$$

We now make use of the electronic field operator (2.2). $\Phi^a(r),\Phi^b(r)$ are the wave functions of the two localized electronic states of interest (see Sec. II). We assume

$$\int d\mathbf{r}\,\Phi^{i*}(\mathbf{r}-\mathbf{r}_n)\Phi^j(\mathbf{r}-\mathbf{r}_m)=\delta_{ij}\delta_{nm}+\epsilon_0^{ij}\kappa \;, \tag{A3}$$

$$\int d\mathbf{r}\,\Phi^{i*}(\mathbf{r}-\mathbf{r}_n)(\mathbf{e}_d\cdot\nabla_r)\Phi^j(\mathbf{r}-\mathbf{r}_m)=(1-\delta_{ij})\delta_{n,m}\eta_1^j+\epsilon_1^{ij}\kappa \;, \tag{A4}$$

and

$$\int d\mathbf{r}\,\Phi^{i*}(\mathbf{r}-\mathbf{r}_n)\Delta_r\Phi^j(\mathbf{r}-\mathbf{r}_m)=\delta_{ij}\delta_{n,m}\eta_2^i+\epsilon_2^{ij}\kappa \;, \tag{A5}$$

where $\kappa=1$ if n and m are nearest neighbors and $\kappa=0$ otherwise, and $i,j=a,b$. ϵ_0^{ij} describe the electron and hole mobility effect due to partial overlapping of electronic wave functions describing electronic states localized at neighboring sites (we assume that only the overlap between the nearest-neighbor sites is finite). Following the assumption of the tight-binding limit, the overlap of neighboring states is obviously small, i.e., $\epsilon_0^{ij},\epsilon_1^{ij},\epsilon_2^{ij}\ll 1$ [the quantities η_1,η_2^i represent the integrals in (A4) and (A5) calculated for $\mathbf{n}=\mathbf{m}$]. The simplest realistic picture is obtained when the overlap effect is neglected in Eqs. (A3) and (A4), i.e., we set $\epsilon_0^{ij}=\epsilon_1^{ij}=0$, but we take it into account in (A5) setting $\epsilon_2^{ij}\neq 0$. In this case we get the mobility of electrons and holes (which is absent for $\epsilon_2^{ij}=0$), and the electron-hole pair generated by the absorption of a photon is always located on the same site. The other overlap parameters $\epsilon_0^{ij}\neq 0$ and $\epsilon_1^{ij}\neq 0$ lead to various corrections. For example, there is a small probability that the electron-hole pair created will be located at neighboring sites, but this does not modify the physical picture in a significant way. Therefore we assume $\epsilon_0^{ij}=\epsilon_1^{ij}=0$ and $\epsilon_2^{ij}\neq 0$ in our calculations.

Employing the form of the vector potential $\hat{A}(\mathbf{r},t)$ given by Eq. (2.3b) and of the electronic operator $\hat{\psi}(\mathbf{r},t)$ given by Eq. (2.2), we write $\hat{H}_2$ in terms of creation and annihilation operators of localized states:

$$\hat{H}_2=\sum_{\mathbf{n}}\int d\mathbf{k}\, g_{\mathbf{k}}\frac{e}{m_e}i(\hat{c}_{\mathbf{n}}\hat{d}_{\mathbf{n}}-\hat{c}^{\dagger}_{\mathbf{n}}\hat{d}^{\dagger}_{\mathbf{n}})(\hat{a}_{\mathbf{k}}+\hat{a}^{\dagger}_{-\mathbf{k}})e^{i\mathbf{k}\cdot\mathbf{r}}$$

$$+\int d\mathbf{k}_1\int d\mathbf{k}_2\sum_{\mathbf{n}}\frac{2\pi c\hbar e^2}{\mathcal{V}\sqrt{|\mathbf{k}_1|\,|\mathbf{k}_2|}}(\hat{a}_{\mathbf{k}_1}+\hat{a}^{\dagger}_{-\mathbf{k}_1})(\hat{a}_{\mathbf{k}_2}+\hat{a}^{\dagger}_{-\mathbf{k}_2})e^{i(\mathbf{k}_1+\mathbf{k}_2)\cdot\mathbf{r}_{\mathbf{n}}}\ , \tag{A6}$$

where $g_{\mathbf{k}}$ is given by Eq. (2.6c).

In the following we employ Eqs. (A3)–(A6) and the relations $\hat{c}_{\mathbf{n}}\hat{c}^{\dagger}_{\mathbf{n}}=\hat{I}-\hat{c}^{\dagger}_{\mathbf{n}}\hat{c}_{\mathbf{n}}$ and $\hat{d}_{\mathbf{n}}\hat{d}^{\dagger}_{\mathbf{n}}=\hat{I}-\hat{d}^{\dagger}_{\mathbf{n}}\hat{d}_{\mathbf{n}}$ to rewrite the Hamiltonian as

$$\hat{H}_0+\hat{H}_1=\sum_{\mathbf{n}}\omega_{ab}(\hat{c}^{\dagger}_{\mathbf{n}}\hat{c}_{\mathbf{n}}+\hat{d}^{\dagger}_{\mathbf{n}}\hat{d}_{\mathbf{n}})+\sum_{\mathbf{n}}\sum_{\mathbf{n}'}(T^e\hat{c}^{\dagger}_{\mathbf{n}}\hat{c}_{\mathbf{n}+\mathbf{n}'}+T^h\hat{d}^{\dagger}_{\mathbf{n}}\hat{d}_{\mathbf{n}+\mathbf{n}'})+\int d\mathbf{k}\,\hbar\omega_{\mathbf{k}}(\hat{a}^{\dagger}_{\mathbf{k}}\hat{a}_{\mathbf{k}}+\tfrac{1}{2})\ , \tag{A7}$$

$$\hat{H}_3=\sum_{\mathbf{n},\mathbf{m}}V^{C}(r_{\mathbf{n}}-r_{\mathbf{m}})(\hat{c}^{\dagger}_{\mathbf{n}}\hat{c}_{\mathbf{n}}-\hat{d}^{\dagger}_{\mathbf{n}}\hat{d}_{\mathbf{n}})(\hat{c}^{\dagger}_{\mathbf{m}}\hat{c}_{\mathbf{m}}-\hat{d}^{\dagger}_{\mathbf{m}}\hat{d}_{\mathbf{m}})+\sum_{\mathbf{n},\mathbf{m}}V^{\mathrm{DD}}(r_{\mathbf{n}}-r_{\mathbf{m}})(\hat{c}_{\mathbf{n}}\hat{d}_{\mathbf{n}}+\hat{c}^{\dagger}_{\mathbf{n}}\hat{d}^{\dagger}_{\mathbf{n}})(\hat{c}_{\mathbf{m}}\hat{d}_{\mathbf{m}}+\hat{c}^{\dagger}_{\mathbf{m}}\hat{d}^{\dagger}_{\mathbf{m}})$$

$$+\sum_{\mathbf{n},\mathbf{m}}V^{\mathrm{DM}}(r_{\mathbf{n}}-r_{\mathbf{m}})(\hat{c}_{\mathbf{n}}\hat{d}_{\mathbf{n}}+\hat{c}^{\dagger}_{\mathbf{n}}\hat{d}^{\dagger}_{\mathbf{n}})(\hat{c}^{\dagger}_{\mathbf{m}}\hat{c}_{\mathbf{m}}-\hat{d}^{\dagger}_{\mathbf{m}}\hat{d}_{\mathbf{m}})\ , \tag{A8}$$

where $T^i=(\hbar^2/2m_e)\epsilon_2^{ii}$ (m_e is the electron mass) and the potential energies $V^{C}(\mathbf{r})$, $V^{\mathrm{DD}}(\mathbf{r})$, and $V^{\mathrm{DM}}(\mathbf{r})$ are given by Eqs. (2.3).

APPENDIX B: COMMUTATOR ALGEBRA

In this appendix we present the commutators of operators necessary for deriving the Heisenberg equations of the tight-binding model. These commutators were given in Appendix F of Ref. 24, but the expressions given there for the commutators (B5)–(B7) are incorrect.

In our tight-binding model, each site can be in one of the following four states:

$$|1,\mathbf{n}\rangle\equiv|\Omega\rangle\ , \tag{B1a}$$

$$|2,\mathbf{n}\rangle\equiv\hat{c}^{\dagger}_{n}|\Omega\rangle\ , \tag{B1b}$$

$$|3,\mathbf{n}\rangle\equiv\hat{d}^{\dagger}_{\mathbf{n}}|\Omega\rangle\ , \tag{B1c}$$

$$|4,\mathbf{n}\rangle\equiv\hat{c}^{\dagger}_{\mathbf{n}}\hat{d}^{\dagger}_{\mathbf{n}}|\Omega\rangle\ , \tag{B1d}$$

where the operators $\hat{c}_{\mathbf{n}}$ and $\hat{d}_{\mathbf{n}}$ are defined by

$$\hat{c}_{\mathbf{n}}=|1,\mathbf{n}\rangle\langle\mathbf{n},2|+|3,\mathbf{n}\rangle\langle\mathbf{n},4|\ , \tag{B2a}$$

$$\hat{d}_{\mathbf{n}}=|1,\mathbf{n}\rangle\langle\mathbf{n},3|+|2,\mathbf{n}\rangle\langle\mathbf{n},4|\ . \tag{B2b}$$

The states $|1,\mathbf{n}\rangle,\ldots,|4,\mathbf{n}\rangle$ represent site n in the ground state, with one electron, with one hole, and with an electron-hole pair, respectively.

The basic commutation relations of $\hat{c}^{\dagger}_{\mathbf{n}}$ and $\hat{d}^{\dagger}_{\mathbf{n}}$ are

$$[\hat{c}_{\mathbf{n}},\hat{c}^{\dagger}_{\mathbf{m}}]=(\hat{I}-2\hat{c}^{\dagger}_{\mathbf{n}}\hat{c}_{\mathbf{n}})\delta_{\mathbf{nm}}\ , \tag{B3}$$

$$[\hat{d}_{\mathbf{n}},\hat{d}^{\dagger}_{\mathbf{m}}]=(\hat{I}-2\hat{d}^{\dagger}_{\mathbf{n}}\hat{d}_{\mathbf{n}})\delta_{\mathbf{nm}}\ . \tag{B4}$$

The commutators of matter operators necessary for calculation of Heisenberg equations can be evaluated by a repeated application of Eqs. (B3) and (B4), resulting in

$$[\hat{Y}_{\mathbf{nm}},\hat{Y}^{\dagger}_{\mathbf{n'm'}}]=\delta_{nn'}\delta_{\mathbf{mm'}}+\delta_{\mathbf{nn'}}\xi_{\mathbf{m'm}}\hat{C}_{\mathbf{m'm}}+\delta_{\mathbf{mm'}}\xi_{\mathbf{n'n}}\hat{D}_{\mathbf{n'n}}+(\xi_{\mathbf{nn'}}\xi_{\mathbf{mm'}}-1)\hat{C}_{\mathbf{m'm}}\hat{D}_{\mathbf{n'n}}\ , \tag{B5}$$

$$[\hat{Y}_{\mathbf{mn}},\hat{C}_{\mathbf{n'}l}]=\delta_{\mathbf{nn'}}\delta_{\mathbf{n}l}\hat{Y}_{\mathbf{m}l}-\delta_{\mathbf{nn'}}(1-\delta_{\mathbf{n}l})2\hat{Y}_{\mathbf{m}l}\hat{C}_{\mathbf{nn}}\ , \tag{B6}$$

$$[\hat{Y}_{\mathbf{mn}},\hat{D}_{\mathbf{m'}l}]=\delta_{\mathbf{mm'}}\delta_{\mathbf{m}l}\hat{Y}_{l\mathbf{n}}-\delta_{\mathbf{mm'}}(1-\delta_{\mathbf{m}l})2\hat{Y}_{l\mathbf{n}}\hat{D}_{\mathbf{mm}}\ , \tag{B7}$$

where $\xi_{\mathbf{nm}}\equiv1-2\delta_{\mathbf{nm}}$.

We further have

$$[\hat{Y}_{\mathbf{mn}},\hat{C}_{\mathbf{n'n'}}\hat{C}_{\mathbf{n''n''}}\delta_{\mathbf{n'n''}}]=\delta_{\mathbf{nn'}}\hat{Y}_{\mathbf{mn}}\hat{C}_{\mathbf{n''n''}}+\delta_{\mathbf{nn''}}\hat{Y}_{\mathbf{mn}}\hat{C}_{\mathbf{n'n'}}\ , \tag{B8}$$

$$[\hat{Y}_{\mathbf{mn}},\hat{D}_{\mathbf{m'm'}}\hat{D}_{\mathbf{m''m''}}\delta_{\mathbf{m'm''}}]=\delta_{\mathbf{mm'}}\hat{Y}_{\mathbf{mn}}\hat{D}_{\mathbf{mm''}}+\delta_{\mathbf{mm''}}\hat{Y}_{\mathbf{mn}}\hat{D}_{\mathbf{m'm}}\ , \tag{B9}$$

$$[\hat{Y}_{\mathbf{mn}},\hat{C}_{\mathbf{n'n}}\hat{D}_{\mathbf{m'm'}}\delta_{\mathbf{n'm''}}]=\delta_{\mathbf{nn'}}\hat{Y}_{\mathbf{mn}}\hat{D}_{\mathbf{m'm'}}+\delta_{\mathbf{mm'}}\hat{Y}_{\mathbf{mn}}\hat{D}_{\mathbf{n'n}}\ . \tag{B10}$$

These commutations were used in the derivation of Eq. (D1).

APPENDIX C: COUPLING OF WANNIER EXCITONS TO PHONONS

The coupling of Wannier excitons to phonons is usually described using two types of interactions: the long-range Fröhlich interaction (mediated by the Coulomb potential), taking place when excitation of a lattice-vibrational mode is accompanied by dielectric polarization, and a short-range interaction associated with the "direct" impact of the deformation amplitude on electronic degrees of freedom.[24] Both effects create effective potentials acting on electrons and holes. Adopting the notation of Ref. 24, we denote the polarization by $\mathbf{Q}\xi(\mathbf{r},t)$, where $\xi(\mathbf{r},t)$ denotes the displacement created by both optical and acoustical phonons and $\mathbf{Q}$ is the polarization tensor. We then have

$$V_e^{\mathrm{eff}}(\mathbf{r},t)=-\frac{1}{\epsilon_0}\int d\mathbf{r}'\frac{\nabla[\mathbf{Q}\xi(\mathbf{r}',t)]}{|\mathbf{r}-\mathbf{r}'|}+d_c\xi(\mathbf{r},t)\ , \tag{C1a}$$

$$V_h^{\rm eff}(\mathbf{r},t)=\frac{1}{\epsilon_0}\int d\mathbf{r}'\frac{\nabla[\mathbf{Q}\xi(\mathbf{r}',t)]}{|\mathbf{r}-\mathbf{r}'|}-d_\vartheta\xi(\mathbf{r},t)\ , \qquad \text{(C1b)}$$

where d_c and d_ϑ are constants.[24] We next consider the total potential acting on the electron pair $V_2(\mathbf{r}_1,\mathbf{r}_2,t)=V_e^{\rm eff}(\mathbf{r}_1,t)+V_h^{\rm eff}(\mathbf{r}_2,t)$, and decompose the displacement $\xi(\mathbf{r},t)$ into plane waves, which leads to

$$V_2(\mathbf{r}_1,\mathbf{r}_2,t)=\int d\mathbf{k}\,S(\mathbf{r}_1,\mathbf{r}_2,\mathbf{k})\xi(\mathbf{k},t)\ . \qquad \text{(C2)}$$

We now calculate the matrix elements between various excitonic states induced by the effective potential $[\mathbf{r}_{12}\equiv\mathbf{r}_1-\mathbf{r}_2,\ \mathbf{R}_{12}\equiv(m_e^*\mathbf{r}_1+m_h^*\mathbf{r}_2)/(m_e^*+m_h^*)]$,

$$\tilde{F}_{\alpha\beta}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{q})=\int d\mathbf{r}_1\int d\mathbf{r}_2\,\phi_\alpha(\mathbf{r}_{12})e^{i\mathbf{k}_1\cdot\mathbf{R}_{12}}\phi_\beta^*(\mathbf{r}_{12})e^{-i\mathbf{k}_2\cdot\mathbf{R}_{12}}S(\mathbf{r}_1,\mathbf{r}_2,\mathbf{q})\ . \qquad \text{(C3)}$$

Finally, we consider a quantized displacement field $\hat{\xi}(\mathbf{q},t)=\hat{b}_\mathbf{q}(t)+\hat{b}^\dagger_{-\mathbf{q}}(t)$, where $\hat{b}_\mathbf{q}$ is the phonon-creation operator, i.e., $[\hat{b}_\mathbf{q},\hat{b}^\dagger_{\mathbf{q}'}]=\delta_{\mathbf{q}\mathbf{q}'}$. This leads to the following Hamiltonian describing the coupling between Wannier excitons and phonons:

$$\hat{H}_{\text{ex-ph}}=\sum_{\alpha,\beta}\int d\mathbf{q}\int d\mathbf{k}\,\tilde{F}_{\alpha\beta}(\mathbf{k}+\mathbf{q},\,\mathbf{k},\mathbf{q})\hat{Y}^\dagger_{\alpha\mathbf{k}+\mathbf{q}}\hat{Y}_{\beta,\mathbf{k}}(\hat{b}_\mathbf{q}+\hat{b}^\dagger_{-\mathbf{q}})\ . \qquad \text{(C4)}$$

$\tilde{F}_{\alpha\alpha}$ describes elastic-scattering processes when only the center-of-mass momentum of excitons is changed, while the terms with $\alpha\neq\beta$ describe interactions where the internal state of the exciton is changed (inelastic scattering). The Hamiltonian given by Eq. (C4) is a straightforward generalization of the Hamiltonian describing the weak-coupling limit of Frenkel excitons coupled to a phonon bath.[29]

We will assume that the energy of phonons is not high enough to allow processes in which α excitons are excited into higher states, dissociated into free carriers, or annihilated due to phonon-assisted recombination, i.e., we will neglect inelastic scattering. The approximation restricting one to elastic-scattering events is more likely to fail for $\alpha>1$ excitons, since for Rydberg-like excitons the phonon-mediated decay to lower excitonic states is expected to be relevant for a correct physical picture.

We next calculate the impact of the phonon bath on the exciton motion assuming that only a single-excitonic line is excited (i.e., that the incident light is resonant with the α exciton). The Heisenberg equations generated by our model Hamiltonian [we denote $F(\mathbf{k},\mathbf{q})=\tilde{F}_{\alpha,\alpha}(\mathbf{k}+\mathbf{q},\,\mathbf{k},\mathbf{q})$],

$$\hat{H}_Q=\int d\mathbf{q}\,\hbar\omega_\mathbf{q}\hat{b}^\dagger_\mathbf{q}\hat{b}_\mathbf{q}+\int d\mathbf{k}\,\hbar\Omega_\mathbf{k}\hat{Y}^\dagger_{\alpha\mathbf{k}}\hat{Y}_{\alpha\mathbf{k}}+\int d\mathbf{q}\int d\mathbf{k}\,F(\mathbf{k},\mathbf{q})\hat{Y}^\dagger_{\alpha,\mathbf{k}+\mathbf{q}}\hat{Y}_{\alpha,\mathbf{k}}(\hat{b}_\mathbf{q}+\hat{b}^\dagger_{-\mathbf{q}})\ , \qquad \text{(C5)}$$

having the following forms (we treat the $\hat{Y}_{\alpha\mathbf{k}}$ as perfect bosons, i.e., $[\hat{Y}_{\alpha\mathbf{k}},\hat{Y}^\dagger_{\alpha'\mathbf{k}'}]=\delta_{\alpha\alpha'}\delta_{\mathbf{k}\mathbf{k}'}$):

$$\dot{\hat{Y}}_{\alpha\mathbf{k}}(t)=-i\Omega_\mathbf{k}\hat{Y}_{\alpha\mathbf{k}}(t)-i\int d\mathbf{q}\,F(\mathbf{k},\mathbf{q})\hat{Y}_{\alpha\mathbf{k}+\mathbf{q}}(t)\hat{\xi}(\mathbf{q},t)\ , \qquad \text{(C6a)}$$

$$\dot{\hat{b}}_\mathbf{q}(t)=-i\omega_\mathbf{q}\hat{b}_\mathbf{q}(t)-i\int d\mathbf{q}\,F^*(\mathbf{k},\mathbf{q})\hat{Y}^\dagger_{\alpha\mathbf{k}+\mathbf{q}}(t)\hat{Y}_\mathbf{k}(t)\ , \qquad \text{(C6b)}$$

where $\hbar\Omega_\mathbf{k}$ now denotes the energy of α excitons with center-of-mass momentum $\hbar\mathbf{k}$. We will consider the dynamics described by Eq. (C6) using several approximations equivalent to that used in the context of the Bogoliubov-Tyablikov method of solving the dynamic systems described by Hamiltonians similar to Eq. (C4) (see Refs. 29 and 30). However, unlike in the standard Bogoliubov-Tyablikov approach, we will not consider equations for consecutive Green's functions obtained by averaging operator products, but we will stay within the operator picture.

The main approximation employed in the following is that the phonon bath remains in a thermal state, e.g., it is not modified by the exciton-phonon interaction:

$$\hat{b}_\mathbf{q}(t)\approx\hat{b}_\mathbf{q}(0)e^{i\omega_q t}\ . \qquad \text{(C7)}$$

This picture is expected to be correct in the low-exciton-density limit,[29] which is exactly the limit of our interest. We should stress also that in a more general case than the one described by Eq. (C5) we get a finite relaxation time of phonon modes into the thermal equilibrium state due to anharmonic phonon-phonon scattering.

We now formally solve Eq. (C6a) $[\hat{\xi}^0(\mathbf{q},t)=\hat{b}_\mathbf{q}(0)e^{i\omega_q t}+\hat{b}^\dagger_\mathbf{q}(0)e^{-i\omega_q t}]$,

$$\hat{Y}_{\alpha\mathbf{k}}(t)=\hat{Y}_{\alpha\mathbf{k}}(0)e^{-i\Omega_\mathbf{k}t}-i\int d\mathbf{q}\,F(\mathbf{k},\mathbf{q})\int_0^t dt'\,e^{-i\Omega_\mathbf{k}(t-t')}\hat{Y}_{\alpha,\mathbf{k}+\mathbf{q}}(t')\hat{\xi}^0(\mathbf{q},t')\ , \qquad \text{(C8)}$$

and insert this solution into the integral on the left-hand-side of Eq. (C6a), which brings us to $[F(\mathbf{k},\mathbf{q})=F^*(\mathbf{k}+\mathbf{q},\,-\mathbf{q})]$

$$\dot{\hat{Y}}_{\alpha\mathbf{k}}(t)=-i\Omega_{\mathbf{k}}\hat{Y}_{\alpha\mathbf{k}}(t)-\int d\mathbf{q}\int d\mathbf{q}'\int_0^t dt'\,F(\mathbf{k},\mathbf{q})F(\mathbf{k}+\mathbf{q},-\mathbf{q}')e^{-i\Omega_{\mathbf{k}+\mathbf{q}}(t-t')}\times\hat{Y}_{\alpha,\mathbf{k}+\mathbf{q}-\mathbf{q}'}(t')\langle\hat{\xi}^0(\mathbf{q},t)\hat{\xi}^0(-\mathbf{q}',t')\rangle+\hat{N}_{\mathbf{k}}(t)\,, \quad (C9a)$$

with

$$\hat{N}_{\alpha\mathbf{k}}(t)=\int d\mathbf{q}\int d\mathbf{q}'\int_0^t dt'\,F(\mathbf{k},\mathbf{q})F(\mathbf{k}+\mathbf{q},-\mathbf{q}')e^{-i\Omega_{\mathbf{k}+\mathbf{q}}(t-t')}\hat{Y}_{\alpha,\mathbf{k}+\mathbf{q}-\mathbf{q}'}(t')\times[\hat{\xi}^0(\mathbf{q},t)\hat{\xi}^0(-\mathbf{q}',t')-\langle\hat{\xi}^0(\mathbf{q},t)\hat{\xi}^0(-\mathbf{q}',t')\rangle]-i\int d\mathbf{q}\,F(\mathbf{k},\mathbf{q})\hat{Y}_{\alpha,\mathbf{k}+\mathbf{q}}(0)e^{-i\Omega_{\mathbf{k}}t}\hat{\varphi}_{\mathbf{q}}(t)\,. \quad (C9b)$$

The thermal averages over the phonons are

$$\langle\hat{\xi}^0(\mathbf{q},t)\rangle=0\,, \quad (C10a)$$

$$\langle\hat{\xi}^0(\mathbf{q},t)\hat{\xi}^0(-\mathbf{q}',t')\rangle=\delta(\mathbf{q}-\mathbf{q}')[n_{\mathbf{q}}e^{i\omega_{\mathbf{q}}(t-t')}+(n_{\mathbf{q}}+1)e^{-i\omega_{\mathbf{q}}(t-t')}]\,, \quad (C10b)$$

where $n_{\mathbf{q}}=(e^{\hbar\omega_{\mathbf{q}}/kT}-1)^{-1}$. Next, we approximate $\langle\hat{N}_{\alpha\mathbf{k}}(t)\rangle\cong0$, which is equivalent to a factorization approximation of the polarization-phonon-operator products (both averages of the phononic operators obviously vanish). We next rewrite Eq. (C9a) as

$$\dot{\hat{Y}}_{\alpha\mathbf{k}}(t)=-i\Omega_{\mathbf{k}}\hat{Y}_{\alpha\mathbf{k}}(t)-\int_0^t dt'\,G_{\mathbf{k}}(t-t')\hat{Y}_{\alpha\mathbf{k}}(t')+\hat{N}_{\mathbf{k}}(t)\,, \quad (C11)$$

where

$$G_{\mathbf{k}}(\tau)=\int d\mathbf{q}\,|F(\mathbf{k},\mathbf{q})|^2e^{-i\Omega_{\mathbf{k}+\mathbf{q}}\tau}[n_{\mathbf{q}}e^{i\omega_{\mathbf{q}}\tau}+(n_{\mathbf{q}}+1)e^{-i\omega_{\mathbf{q}}\tau}]\,. \quad (C12)$$

Finally, we consider the Fermi golden-rule limit, which is expected to hold for weak electron-phonon coupling. This is again equivalent to the lowest-order iteration in the electron-phonon coupling and excitonic density for the exciton mass operator (compare to Ref. 29). To this end, we get $[\gamma_{\mathbf{k}}(T)=\int_0^\infty G_{\mathbf{k}}(\tau)d\tau]$

$$\dot{\hat{Y}}_{\alpha\mathbf{k}}(t)=-i\Omega_{\mathbf{k}}\hat{Y}_{\alpha\mathbf{k}}(t)-\gamma_{\mathbf{k}}(T)\hat{Y}_{\alpha\mathbf{k}}(t)+\hat{N}_{\alpha\mathbf{k}}(t)\,, \quad (C13)$$

where

$$\gamma_{\mathbf{k}}=\pi\int d\mathbf{q}\,|F(\mathbf{k},\mathbf{q})|^2[n_{\mathbf{q}}\delta(\omega_q+\Omega_{k+q})+(n_{\mathbf{q}}+1)\delta(\omega_q-\Omega_{k+q})]\,. \quad (C14)$$

Our final equation, (C13), has the form of a simple dissipative-operator equation (i.e., Langevin operator equation). We can describe it intuitively as a phonon-mediated process which converts the coherent exciton polarization created by a coherent driving electromagnetic wave into an incoherent polarization, which builds up from the noise-source term $\hat{N}_{\alpha\mathbf{k}}(t)$ following the approximation $\langle\hat{N}_{\alpha\mathbf{k}}(t)\rangle\cong0$. The noise term has a relatively complicated from [see Eq. (C9b)]; however, we restrict our investigation to the coherent component of scattered light and therefore we did not need to employ Eq. (C9b). Finally we will assume that the decay rate does not vary much in the regime of the wave vectors of interest, i.e., we denote $\gamma_{\mathbf{k}}\cong\gamma_1$.

Equation (C13) represents the limit of linearized exciton dynamics. The next step of our analysis is the evaluation of low-order nonlinear corrections. One type of nonlinear contribution is obtained when the linearized commutator $[\hat{Y}_{\alpha\mathbf{k}},\hat{Y}^\dagger_{\alpha\mathbf{k}}]=\delta(\mathbf{k}-\mathbf{k}')$ is replaced by the full expression given by Eq. (B5), when deriving the time derivative of $\hat{Y}_{\alpha\mathbf{k}}(t)$. However, this contribution only modifies the noise term $\hat{N}_{\alpha\mathbf{k}}(t)$; such a conclusion is reached since in the electron-phonon factorization approximation this term has a zero average. Another nonlinear correction is obtained when the lowest-order perturbation of the phononic motion is considered, e.g.,

$$\hat{b}_{\mathbf{q}}(t)=\hat{b}_{\mathbf{q}}(0)e^{-i\omega_q t}-i\int_0^t dt'\int d\mathbf{k}\,F(\mathbf{k},\mathbf{q})e^{-i\omega_q(t-t')}\hat{Y}^\dagger_{\alpha,\mathbf{k}-\mathbf{q}}(t')\hat{Y}_{\alpha\mathbf{k}}(t')\,. \quad (C15)$$

We next insert Eq. (C15) into Eq. (C6a), which eventually leads us to a modified form of Eq. (C13):

$$\dot{\hat{Y}}_{\alpha\mathbf{k}}(t)=-i\Omega_{\mathbf{k}}\hat{Y}_{\alpha\mathbf{k}}(t)-\gamma_1(T)\hat{Y}_{\alpha\mathbf{k}}(t)+\hat{N}'_{\mathbf{k}}(t)+\hat{f}^{(6)}_{\alpha\mathbf{k}}(t)\,, \quad (C16)$$

where

$$\hat{f}^{(6)}_{\alpha\mathbf{k}}(t)=-\int_0^t dt'\int d\mathbf{q}\int d\mathbf{k}'\,F(\mathbf{k},\mathbf{q})[F(\mathbf{k}',-\mathbf{q})\hat{Y}_{\alpha,\mathbf{k}+\mathbf{q}}(t)\hat{Y}^\dagger_{\alpha,\mathbf{k}'+\mathbf{q}}(t')\hat{Y}_{\alpha\mathbf{k}'}(t')e^{-i\omega_q(t-t')}-F^*(\mathbf{k}',-\mathbf{q})\hat{Y}_{\alpha,\mathbf{k}+\mathbf{q}}(t)\hat{Y}^\dagger_{\alpha\mathbf{k}'}(t')\hat{Y}_{\alpha,\mathbf{k}'-\mathbf{q}}(t')e^{i\omega_q(t-t')}]\,. \quad (C17)$$

In the following we will employ the Markovian approximation for the memory kernel in Eq. (C17). We insert $\hat{Y}_{\alpha k}(t') \cong \hat{Y}_{\alpha k}(t) e^{-i\omega_k(t-t')} e^{\gamma_1(t-t')}$ into (C17):

$$\hat{f}^{(6)}_{1k}(t) = \int d\mathbf{q} \int d\mathbf{k}' D(\mathbf{k},\mathbf{k}',\mathbf{q}) \times \hat{Y}_{\alpha,\mathbf{k}+\mathbf{q}}(t) \hat{Y}^{\dagger}_{\alpha,\mathbf{k}'+\mathbf{q}}(t) \hat{Y}_{\alpha\mathbf{k}'}(t) , \quad \text{(C18)}$$

where [from the hermicity of $\hat{H}_4$, we have $F(\mathbf{k},\mathbf{q}) = F^*(\mathbf{k}+\mathbf{q}, -\mathbf{q})$]

$$D(\mathbf{k},\mathbf{k}',\mathbf{q}) = -F(\mathbf{k},\mathbf{q}) F(\mathbf{k}',\mathbf{q}) \times [(i\omega_q + i\Omega_{\mathbf{k}'} - i\Omega_{\mathbf{k}'+\mathbf{q}} + 2\gamma_1)^{-1} + (i\omega_{-q} + i\Omega_{\mathbf{k}'} - i\Omega_{\mathbf{k}'-\mathbf{q}} + 2\gamma_1)^{-1}] . \quad \text{(C19)}$$

The function $D(\mathbf{k},\mathbf{k}',\mathbf{q})$ can be described as the (nonlinear) coupling between excitonic modes with different center-of-mass momentum, mediated by phonons. In the following we will assume that $D(\mathbf{k},\mathbf{k}',\mathbf{q})$ is approximately constant in the regime of the wave vectors of interest. In this case the phonon-assisted nonlinear interaction between excitons can be described by an additional term in Eq. (D7),

$$\hat{f}^{(6)}_{\mathbf{k}}(t) = i\lambda_{\mathrm{ph}} \int d\mathbf{k}_1 \int d\mathbf{k}_2 \hat{Y}_{\alpha\mathbf{k}_1}(t) \hat{Y}^{\dagger}_{\alpha\mathbf{k}_2}(t) \times \hat{Y}_{\alpha,\mathbf{k}-\mathbf{k}_1+\mathbf{k}_2}(t) , \quad \text{(C20)}$$

with

$$\lambda_{\mathrm{ph}} = -iD(\mathbf{k}_0,\mathbf{k}_0,\mathbf{q}=0) . \quad \text{(C21)}$$

The nonlinear term described by Eq. (C20) has the form of a weakly anharmonic oscillator with fourth-power nonlinearity.

APPENDIX D: LINEAR AND WEAKLY NONLINEAR DYNAMICS OF EXCITONIC POLARIZATION

In this appendix we present the explicit form of the nonlinear terms in the Heisenberg equation for the excitonic creation operator. We then derive the equation of motion describing averaged excitonic polarization in the weakly nonlinear regime.

The equations of motion for the operators $\hat{a}_{\mathbf{k}}(t)$ and $\hat{Y}_{\mathrm{mn}}(t)$ are given by Eqs. (2.6) ($\dot{\hat{x}}(t) = i[\hat{H}, \hat{x}(t)]$, see Eqs. (B5)–(B10), where $\bar{\delta}_{\mathrm{mn}} = 1 - \delta_{\mathrm{mn}}$),

$$\hat{f}^{(1)}_{\mathrm{nm}}(t) = i \sum_j [V^{\mathrm{DD}}(\mathbf{r}_\mathrm{n} - \mathbf{r}_\mathrm{j}) \hat{Y}_{\mathrm{jj}}(t) \hat{C}_{\mathrm{nm}}(t) + V^{\mathrm{DD}}(\mathbf{r}_\mathrm{m} - \mathbf{r}_\mathrm{j}) \hat{Y}_{\mathrm{jj}}(t) \hat{D}_{\mathrm{mn}}] \xi_{\mathrm{nm}} - i2\bar{\delta}_{\mathrm{nm}} \sum_j [V^{\mathrm{DD}}(\mathbf{r}_\mathrm{n} - \mathbf{r}_j) \hat{C}_{\mathrm{nm}} \hat{D}_{\mathrm{nn}} + V^{\mathrm{DD}}(\mathbf{r}_\mathrm{m} - \mathbf{r}_\mathrm{j}) \hat{D}_{\mathrm{mn}} \hat{C}_{\mathrm{mm}}] , \quad \text{(D1a)}$$

$$\hat{f}^{(2)}_{\mathrm{nm}}(t) = -i \sum_j [V^C(\mathbf{r}_\mathrm{n} - \mathbf{r}_j) \hat{C}_{\mathrm{nm}}(t) - V^C(\mathbf{r}_\mathrm{m} - \mathbf{r}_\mathrm{j}) \hat{D}_{\mathrm{mn}}(t)] \hat{Y}_{\mathrm{jj}}(t) \xi_{\mathrm{mn}} - i2\bar{\delta}_{\mathrm{nm}} \sum_j [V^C(\mathbf{r}_\mathrm{n} - \mathbf{r}_j) \hat{C}_{\mathrm{nm}}(t) \hat{D}_{\mathrm{nn}}(t) - V^C(\mathbf{r}_\mathrm{m} - \mathbf{r}_\mathrm{j}) \hat{D}_{\mathrm{mn}}(t) \hat{C}_{\mathrm{mm}}(t)] , \quad \text{(D1b)}$$

$$\hat{f}^{(3)}_{\mathrm{nm}}(t) = -i \left[T^e \hat{C}_{\mathrm{mm}}(t) \sum_{n'}{}' \hat{Y}_{\mathrm{n},\mathrm{m}+\mathrm{n}'}(t) + T^h \hat{D}_{\mathrm{nn}}(t) \sum_{n'}{}' Y_{\mathrm{n}+\mathrm{n}',\mathrm{m}}(t) \right] , \quad \text{(D1c)}$$

$$\hat{f}^{(4)}_{\mathrm{nm}}(t) = -i\delta_{\mathrm{nm}} \sum_j V^{\mathrm{DM}}(\mathbf{r}_\mathrm{n} - \mathbf{r}_\mathrm{j}) [\hat{C}_{\mathrm{jj}}(t) - \hat{D}_{\mathrm{jj}}(t)] + 2i\bar{\delta}_{nm} \sum_j [V^{\mathrm{DM}}(\mathbf{r}_\mathrm{n} - \mathbf{r}_\mathrm{j}) \hat{D}_{\mathrm{mn}}(t) + V^{\mathrm{DM}}(\mathbf{r}_\mathrm{m} - \mathbf{r}_\mathrm{j}) \hat{C}_{\mathrm{nm}}(t)] [\hat{C}_{\mathrm{jj}}(t) - \hat{D}_{\mathrm{jj}}(t)] , \quad \text{(D1d)}$$

$$\hat{f}^{(5)}_{\mathrm{nm}}(t) = -\mu(\hat{A}(t,\mathbf{r}_\mathrm{n}) \hat{D}_{\mathrm{mn}}(t) + \hat{A}(t,\mathbf{r}_\mathrm{m}) \hat{C}_{\mathrm{nm}}(t)) \xi_{\mathrm{nm}} - 2\bar{\delta}_{\mathrm{nm}} \mu(\hat{A}(t,\mathbf{r}_\mathrm{n}) \hat{C}_{\mathrm{nn}}(t) \hat{D}_{\mathrm{mn}}(t) + \hat{A}(t,\mathbf{r}_\mathrm{m}) \hat{D}_{\mathrm{mm}}(t) C_{\mathrm{nm}}(t)) . \quad \text{(D1e)}$$

As shown in Sec. II, Eqs. (2.6)–(2.9), the equation of motion for operators $\hat{Y}_{\mathrm{nm}}(t)$ can be recast in the following form:

$$\dot{\hat{Y}}_{\alpha\mathbf{k}}(t) = -i(E_\alpha + E_\mathbf{k}) \hat{Y}_{\alpha\mathbf{k}}(t) - g_\mathbf{k} \phi_\alpha(0) \hat{A}(\mathbf{k},t) + \sum_{s=1}^{5} \hat{f}^{(s)}_{\alpha\mathbf{k}}(t) , \quad \text{(D2)}$$

where

$$\hat{q}_{\alpha\mathbf{k}}(t) = \sum_{\mathrm{n,m}} \exp[i\mathbf{k}(\mathbf{R}_{\mathrm{nm}})] \phi_\alpha(r_\mathrm{n} - r_\mathrm{m}) \hat{q}_{\mathrm{nm}}(t) .$$

In order to evaluate nonlinear effects in the regime of weak nonlinearity, we will calculate the nonlinear forces $\hat{f}^{(s)}_{\alpha\mathbf{k}}(t)$ assuming that the operators entering into these terms obey linearized dynamics. This corresponds to an iterative treatment in the finite-exciton-density nonlinearity. We will further neglect terms corresponding to higher than third order in the electromagnetic field and excitonic polarization. We also restrict ourselves to $\alpha = 1$ excitons so that $\phi_1(0) g_\mathbf{k} \simeq \mu$.

The equation describing the quantum average of the α-exciton polarization has the form (in the following calculations of excitonic nonlinearity, only $\alpha = 1$ will be considered)

$$\dot{Y}_{\mathbf{k}}(t)=-(E+E_{\mathbf{k}})Y_{\mathbf{k}}(t)+\phi(0)g_{\mathbf{k}}A(t,\mathbf{k})+\sum_{s=1}^{5}\langle\hat{f}_{\mathbf{k}}^{(s)}(t)\rangle . \tag{D3}$$

The average of $\hat{f}_{\alpha\mathbf{k}}^{(s)}(t)$ will be approximated as follows:

$$\langle\hat{f}_{1\mathbf{k}}^{(s)}(t)\rangle\cong\mathrm{Tr}[\hat{f}_{1\mathbf{k}}^{(s)}(0)\rho_L(t)] , \tag{D4}$$

where $\rho_L(t)$ is the density matrix calculated assuming a partially linearized interaction (i.e., its time evolution is governed by a quadratic bosonic Hamiltonian). When dissipation is neglected, the density matrix of our system corresponds to a superposition of coherent states, i.e.,

$$\rho_L(t)\cong|\psi_{\mathrm{coh}}(t)\rangle\langle\psi_{\mathrm{coh}}(t)| , \tag{D5}$$

where the coherent state of the matter is defined as[24]

$$|\psi_{\mathrm{coh}}(t)\rangle=\exp\left[\sum_{\alpha}\int d\mathbf{k}\,\varphi_{\alpha\mathbf{k}}(t)\hat{Y}_{\alpha\mathbf{k}}^{0}\right]|\Omega\rangle , \tag{D6a}$$

$$\varphi_{\alpha\mathbf{k}}(t)=\langle\hat{Y}_{\alpha\mathbf{k}}(t)\rangle . \tag{D6b}$$

We rewrite

$$\varphi_{\mathrm{nm}}(t)=\frac{1}{(2\pi)^3}\sum_{\alpha}\int d\mathbf{k}\,e^{i\mathbf{k}\cdot\mathbf{R}_{\mathrm{nm}}}\phi_\alpha(\mathbf{r}_{\mathrm{n}}-\mathbf{r}_{\mathrm{m}})\varphi_{\alpha\mathbf{k}}(t) , \tag{D7}$$

and get

$$|\psi_{\mathrm{coh}}(t)\rangle=\exp\left[\sum_{\mathrm{n,m}}\varphi_{\mathrm{nm}}(t)\hat{Y}_{\mathrm{nm}}^{0}\right]|\Omega\rangle . \tag{D8}$$

Following the linear-optics limit result $[\hat{Y}_{\mathrm{nm}},\hat{Y}_{\mathrm{n'm'}}^{\dagger}]=\delta_{\mathrm{nn'}}\delta_{\mathrm{mm'}}$, we get

$$|\psi_{\mathrm{coh}}(t)\rangle=\mathcal{N}^{-1}\bigotimes_{\mathrm{n,m}}[\hat{I}+\varphi_{\mathrm{nm}}(t)\hat{Y}_{\mathrm{nm}}^{0}]|\Omega\rangle . \tag{D9}$$

Here we employed $\exp(\sum_\alpha\hat{A}_\alpha)=\prod_\alpha\exp(\hat{A}_\alpha)$ for commuting $\hat{A}_\alpha,\hat{A}_\beta$ $(\alpha\neq\beta)$ and $\exp(\epsilon\hat{A})=a_1\hat{I}+a_2\hat{A}$ if $\hat{A}^2=0$. Since the bosonic picture relies on the assumption than all sites are weakly excited [i.e., $\varphi_{nm}(t)\ll1$], we can approximate $a_1=\mathcal{N}^{-1}a_2=\epsilon\mathcal{N}^{-1}$, where $\mathcal{N}$ is a normalization factor.

Equation (D9) describes the evolution of the excitonic system in the limit of linearized dynamics. Such linearization is obviously valid only when $\varphi_{\mathrm{nm}}(t)\ll1$, which means that the **k**-space oscillator can be highly excited [i.e., $\langle\hat{Y}_{\alpha\mathbf{k}}(t)\rangle>1$] only when the excitation distribution is sharply peaked around some values of **k**.

In order to evaluate the quantum averages of operators represented by Eqs. (D1a)–(D1e), we need to calculate averages of various operator products ($\hat{C}_{\mathrm{nm}}\hat{Y}_{\mathrm{n'm'}}$, $\hat{D}_{\mathrm{nm}}\hat{Y}_{\mathrm{n'm'}}$, etc.) with the time-dependent density matrix $\rho_L(t)$ describing the evolution of the system governed by linearized equations. If the phonon-induced dissipation is disregarded, and the RWA approximation toward exciton-photon interaction is applied, the density matrix has a simple form [Eq. (D5)]. We will approximate the average describing the linearized phonon-perturbed system by the coherent-state result plus a phenomenological correction describing the "incoherent" excitonic population (i.e., excitons scattered by phonons) and the density of free carriers:

$$\langle\hat{C}_{\mathrm{nm}}\rangle\cong\langle\psi_{\mathrm{coh}}(t)|\hat{C}_{\mathrm{nm}}|\psi_{\mathrm{coh}}(t)\rangle+\delta_{\mathrm{nm}}n_T . \tag{D10}$$

The averages of interest have the form [we neglect in our calculations higher than third-order products of the amplitudes $\langle\hat{Y}_{\alpha\mathbf{k}}(t)\rangle$]

$$\begin{aligned}\langle\psi_{\mathrm{coh}}(t)|(\hat{C}_{\mathrm{nm}}+\hat{D}_{\mathrm{mn}})|\psi_{\mathrm{coh}}(t)\rangle &\cong\sum_l[\varphi_{\mathrm{n}l}^*(t)\varphi_{\mathrm{m}l}(t)+\varphi_{l\mathrm{m}}^*(t)\varphi_{l\mathrm{n}}(t)]\\ &=\sum_{\alpha_1,\alpha_2}\int d\mathbf{k}_1\int d\mathbf{k}_2\left[\sum_l e^{-i\mathbf{k}\cdot\mathbf{R}_{\mathrm{n}l}+i\mathbf{k}_2\cdot\mathbf{R}_{\mathrm{m}l}}\phi_{\alpha_1(r_{\mathrm{n}}-r_l)}^*\phi_{\alpha_2(r_{\mathrm{m}}-r_l)}\right.\\ &\qquad\left.+e^{-i\mathbf{k}\cdot\mathbf{R}_{\mathrm{n}l}+i\mathbf{k}_2\cdot\mathbf{R}_{\mathrm{m}l}}\phi_{\alpha_1}^*(r_l-r_{\mathrm{m}})\phi_{\alpha_2}(r_l-r_{\mathrm{n}})\right]\langle\hat{Y}_{\alpha_1\mathbf{k}_1}^{\dagger}(t)\rangle\langle\hat{Y}_{\alpha_2\mathbf{k}_2}(t)\rangle .\end{aligned} \tag{D11}$$

We next make use of the assumption that the incoming light beams are resonant with the $\alpha=1$ excitonic line; this allows us to consider the term in Eq. (D14) with $\alpha_1=\alpha_2=1$ only (i.e., $\langle Y_{\alpha\mathbf{k}}\rangle=0$ for $\alpha\neq1$). The next important approximation is that wavelengths of interest are much longer than the radius of the $\alpha=1$ (1s) exciton, this allows us to approximate $\phi_1(r_{\mathrm{a}}-r_{\mathrm{b}})e^{i\mathbf{k}\cdot\mathbf{r}_{\mathrm{a}}}\cong\phi_1(r_{\mathrm{a}}-r_{\mathrm{b}})e^{i\mathbf{k}\cdot\mathbf{r}_{\mathrm{b}}}$. Altogether, we obtain

$$\langle\psi_{\mathrm{coh}}(t)|(\hat{C}_{\mathrm{nm}}+\hat{D}_{\mathrm{mn}})|\psi_{\mathrm{coh}}(t)\rangle\cong\int d\mathbf{k}_1\int d\mathbf{k}_2\,G_{\mathrm{nm}}^{a}(\mathbf{k}_1,\mathbf{k}_2)\langle\hat{Y}_{1,-\mathbf{k}_1}^{\dagger}(t)\rangle\langle\hat{Y}_{1,\mathbf{k}_2}(t)\rangle , \tag{D12a}$$

$$G_{\mathrm{nm}}^{a}(\mathbf{k}_1,\mathbf{k}_2)=\sum_l e^{-i(\mathbf{k}_1-\mathbf{k}_2)\cdot r_l}[\phi_1(r_{\mathrm{n}}-r_l)\phi_1(r_{\mathrm{m}}-r_l)+\phi_1(r_l-r_{\mathrm{m}})\phi_1(r_l-r_{\mathrm{n}})] . \tag{D12b}$$

Similarly, we get

$$\langle\psi_{\mathrm{coh}}(t)|\hat{C}_{\mathrm{nn'}}\hat{D}_{\mathrm{mm'}}|\psi_{\mathrm{coh}}(t)\rangle\cong\varphi_{\mathrm{mn}}^*(t)\varphi_{\mathrm{m'n'}}(t)=\int d\mathbf{k}_1\int d\mathbf{k}_2\,G_{nn'mm'}^{b}(\mathbf{k}_1,\mathbf{k}_2)\langle\hat{Y}_{1,-\mathbf{k}_1}^{\dagger}(t)\rangle\langle\hat{Y}_{1,\mathbf{k}_2}(t)\rangle , \tag{D13a}$$

$$\begin{aligned}G_{\mathrm{nn'mm'}}^{b}(\mathbf{k}_1,\mathbf{k}_2)&=e^{-i(\mathbf{k}_1\mathbf{r}_{\mathrm{n}}+\mathbf{k}_2\mathbf{r}_{\mathrm{n'}})}\phi_1(r_{\mathrm{m}}-r_{\mathrm{n}})\phi_1(r_{\mathrm{m'}}-r_{\mathrm{n'}})\langle\psi_c(t)|\hat{Y}_{\mathrm{nm}}(\hat{C}_{ll}+\hat{D}_{ll})|\psi_c(t)\rangle\delta_{\mathrm{n}l}\delta_{\mathrm{m}l}\\ &=\varphi_{\mathrm{nm}}(t)\sum_{l'}[\varphi_{ll'}^*(t)\varphi_{ll'}(t)+\varphi_{l'l}^*(t)\varphi_{l'l}(t)]-\varphi_{\mathrm{nm}}(t)[\varphi_{l\mathrm{m}}^*(t)\varphi_{l\mathrm{m}}(t)+\varphi_{\mathrm{n}l}^*(t)\varphi_{\mathrm{n}l}(t)] ,\end{aligned} \tag{D13b}$$

$$\langle \psi_{\rm coh}(t)|\hat{Y}_{\rm nm}(\hat{C}_{ll}+\hat{D}_{ll})|\psi_{\rm coh}(t)\rangle \delta_{nl}\delta_{ml} \cong \int d\mathbf{k}\, e^{i\mathbf{k}\cdot\mathbf{r}_n}\phi_1(\mathbf{r}_n-\mathbf{r}_m)n_{\rm ex}(\mathbf{r}_l) - \int d\mathbf{k}_1\int d\mathbf{k}_2\int d\mathbf{k}_3\, G^c_{\rm nm\mathit{l}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\langle \hat{Y}_{1,\mathbf{k}_1}(t)\rangle\langle \hat{Y}^\dagger_{1,-\mathbf{k}_2}(t)\rangle\langle \hat{Y}_{1,\mathbf{k}_3}(t)\rangle \;, \tag{D14a}$$

$$G^c_{\rm nm\mathit{l}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)=e^{i(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\cdot\mathbf{r}_n}\phi_1(\mathbf{r}_n-\mathbf{r}_m)[\phi_1^2(\mathbf{r}_l-\mathbf{r}_m)+\phi_1^2(\mathbf{r}_n-\mathbf{r}_l)] \;, \tag{D14b}$$

$$\langle \psi_{\rm coh}(t)|\hat{Y}_{\rm nm}(\hat{C}_{ll}-\hat{D}_{ll})|\psi_{\rm coh}(t)\rangle \equiv \int d\mathbf{k}_1\int d\mathbf{k}_2\int d\mathbf{k}_3\, G^d_{\rm nm\mathit{l}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)\langle \hat{Y}_{1,\mathbf{k}_1}(t)\rangle\langle \hat{Y}^\dagger_{1,-\mathbf{k}_2}(t)\rangle\langle \hat{Y}_{1,\mathbf{k}_3}(t)\rangle \;, \tag{D15a}$$

$$G^d_{\rm nm\mathit{l}}(\mathbf{k}_1,\mathbf{k}_2,\mathbf{k}_3)=e^{i(\mathbf{k}_1+\mathbf{k}_2+\mathbf{k}_3)\cdot\mathbf{r}_n}\phi_1(\mathbf{r}_n-\mathbf{r}_m)[\phi_1^2(\mathbf{r}_l-\mathbf{r}_m)+\phi_1^2(\mathbf{r}_n-\mathbf{r}_l)] \;, \tag{D15b}$$

where $n_{\rm ex}(\mathbf{r},t)$ is the exciton density,

$$n_{\rm ex}(r_n,t)=\sum_l \varphi^*_{ln}(t)\varphi_{ln}(t)=\int d\mathbf{k}_1\int d\mathbf{k}_2\, e^{-i(\mathbf{k}_2-\mathbf{k}_1)\cdot\mathbf{r}_n}\langle \hat{Y}^\dagger_{1,-\mathbf{k}_1}(t)\rangle\langle \hat{Y}_{1,\mathbf{k}_2}(t)\rangle \;. \tag{D16}$$

We next employ Eqs. (D12)–(D15) to express the nonlinear terms (D1a)–(D1e) in the form of products of the amplitudes $\langle \hat{Y}_{1\mathbf{k}}(t)\rangle$ and $\langle \hat{A}(\mathbf{k},t)\rangle$,

$$\dot{Y}_{1\mathbf{k}}(t)=i(\Omega_k+\Delta_T)Y_{1\mathbf{k}}(t)+\mu(1-\delta_T)A(t,\mathbf{k}) + \int d\mathbf{k}_1\int d\mathbf{k}_2\int d\mathbf{k}_3\,\delta(\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2-\mathbf{k}_3)[\lambda_1' iY_{1,\mathbf{k}_1}, Y^*_{1,-\mathbf{k}_2}Y_{1,\mathbf{k}_3}+\lambda_2\mu A(t,\mathbf{k}_1)Y^*_{1,-\mathbf{k}_2}Y_{1,\mathbf{k}_3}] \;, \tag{D17}$$

where λ_1' and λ_2 have the form

$$\lambda_1'=\phi_1(0)I_1I_2+\phi_1^2(0)I_3+2\phi_1^2(0)I_1+12(T^e+T^h)\phi_1^2(0) \;, \tag{D18a}$$

$$\lambda_2=I_2 \;, \tag{D18b}$$

where the integrals I_1, I_2, and I_3 are (α is the lattice constant)

$$I_1=\frac{1}{(2\pi)^{9/2}}\int_* d\mathbf{r}\, a^{-3/2}e^{i\bar{\mathbf{k}}\cdot\mathbf{r}}V^{\rm DD}(\mathbf{r}) \;, \tag{D19a}$$

$$I_2=\frac{1}{(2\pi)^{9/2}}a^{-3}\int_* d\mathbf{r}_1\int_* d\mathbf{r}_2\,\phi_1(\mathbf{r}_1)\phi_1(\mathbf{r}_2)\phi_1(\mathbf{r}_1-\mathbf{r}_2) \;, \tag{D19b}$$

$$I_3=\frac{1}{(2\pi)^{9/2}}a^{-3/2}\int_* d\mathbf{r}\, V^{\rm DD}(\mathbf{r})\phi_1^2(\mathbf{r}) \;. \tag{D19c}$$

These integrals are calculated from discrete sums by switching to the continuum limit. The asterisk implies that the integral is performed with an exclusion of small vector lengths corresponding to $|\mathbf{r}|\le a$. The latter restriction is necessary to get correctly the correspondence between the discrete sum and its integral representation. The vector $\bar{\mathbf{k}}$ is chosen to be perpendicular to the dipole direction $\mathbf{e}_d$ and it corresponds to a frequency of light resonant to the 1s excitonic line, i.e., $|\bar{\mathbf{k}}c|\cong\Omega_{\mathbf{k}=0}$.

We next consider the interaction with the phonon bath. Equation (D17) should then be modified by three additional terms representing damping of excitonic polarization, phonon-induced noise, and phonon-mediated nonlinearity [see Appendix C, Eq. (C17)]. Since the average of the phonon noise vanishes, we obtain after a factorization of the product of excitonic polarization operators corresponding to Eq. (C20) a simple equation of motion describing the dynamics of the averaged excitonic polarization,

$$\dot{Y}_{1\mathbf{k}}(t)=[i(\Omega_k+\Delta_T)+\gamma_1]Y_{1\mathbf{k}}(t)+\mu(1-\delta_T)A(t,\mathbf{k}) + \int d\mathbf{k}_1\int d\mathbf{k}_2\int d\mathbf{k}_3\,\delta(\mathbf{k}-\mathbf{k}_1-\mathbf{k}_2-\mathbf{k}_3)[\lambda_1 iY_{1,\mathbf{k}_1}(t)Y^*_{1,-\mathbf{k}_2}(t)Y_{1,\mathbf{k}_3}(t)+\lambda_2\mu(1-\delta_T)A(t,\mathbf{k}_1)Y^*_{1,-\mathbf{k}_2}(t)Y_{1,\mathbf{k}_3}(t)] \;. \tag{D20}$$

This equation describes the excitonic polarization in the linear and weakly nonlinear regime since it was derived under the weak-nonlinearity approximation when higher than third-order nonlinear terms can be disregarded. The terms Δ_T and δ_T describe the effect of "incoherent" excitons as well as free carriers. We will assume that this density is much lower than that of excitons associated with the coherent polarization of the medium and will therefore disregard these terms. Equation (D20) corresponds to Eq. (4.1) in the main text. The parameter λ_1 in Eq. (D20) is given by

$$\lambda_1=\lambda_1'+\lambda_{\rm ph} \;, \tag{D21}$$

where $\lambda_{\rm ph}$ is given by Eq. (C21).

*On leave from Institute for Theoretical Physics, Polish Academy of Sciences, Aleja Lotników 32/46, PL-02-668 Warsaw, Poland.

[1] H. Haug, J. Lumin. **30**, 171 (1985).

[2] *Optical Nonlinearities and Instabilities in Semiconductors*, edited by H. Haug (Academic, New York, 1988).

[3] M. Bawendi, M. L. Steigerwald, and L. Brus, Annu. Rev. Phys. Chem. (to be published).

[4] A. L. Ivanov and L. V. Keldysh, Zh. Eksp. Teor. Fiz. **84**, 404 (1982) [Sov. Phys.—JETP **57**, 234 (1982)].

[5] E. Yablonovitch, J. P. Heritage, D. E. Aspnes, and Y. Yaffet, Phys. Rev. Lett. **63**, 976 (1989).

[6] B. J. Greene, J. F. Mueller, J. Orenstein, D. H. Rapkine, S. Schmitt-Rink, and M. Thakur, Phys. Rev. Lett. **61**, 325 (1988).

[7] W. Wu, Phys. Rev. Lett. **61**, 1119 (1988).

[8] R. S. Knox, *Theory of Excitons* (Academic, New York, 1963).

[9] S. Schmitt-Rink, D. S. Chemla, and D. A. B. Miller, Phys. Rev. B **32**, 6601 (1985).

[10] S. Schmitt-Rink and D. S. Chemla, Phys. Rev. Lett. **57**, 2752 (1986); S. Schmitt-Rink, D. S. Chemla, and H. Haug, Phys. Rev. B **37**, 941 (1988).

[11] S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla, Phys. Rev. B **35**, 8113 (1987).

[12] W. H. Knox, D. S. Chemla, D. A. B. Miller, J. B. Stark, and S. Schmitt-Rink, Phys. Rev. Lett. **62**, 1189 (1989).

[13] C. Ell, J. F. Müller, K. ElSayed, and H. Haug, Phys. Rev. Lett. **62**, 304 (1989).

[14] D. Frölich, A. Notle, and K. Reimann, Phys. Rev. Lett. **55**, 1335 (1985).

[15] A. Mysorowicz, D. Hulin, A. Antonetti, A. Migus, W. T. Mosselink, and H. Thorlog, Phys. Rev. Lett. **56**, 2748 (1988).

[16] M. Lindberg and S. W. Koch, Phys. Rev. B **38**, 7607 (1988); J. P. Sokoloff, M. Joffre, B. Fluegel, D. Hulin, M. Lindberg, S. W. Koch, A. Migus, A. Antonetti, and N. Penghambrian, *ibid.* **38**, 7615 (1988).

[17] M. Joffre, D. Hulin, A. Migus, and A. Antonetti, J. Mod. Opt. **35**, 1951 (1988).

[18] M. Combescot and R. Combescot, Phys. Rev. Lett. **61**, 117 (1988); Phys. Rev. B **40**, 3788 (1989).

[19] M. Joffre, Ph.D. thesis, University of Paris, 1989.

[20] Ch. Tanguy and M. Combescot, Phys. Rev. B **40**, 12 562 (1989).

[21] L. Banyai, Y. Z. Hu, M. Lindberg, and S. W. Koch, Phys. Rev. B **38**, 8142 (1988).

[22] L. V. Keldysh and A. N. Kozlov, Zh. Eksp. Teor. Fiz. **54**, 978 (1968) [Sov. Phys.—JETP **27**, 521 (1968)].

[23] D. Hulin, A. Mysorowicz, and C. Benoît à la Guillaume, Phys. Rev. Lett. **45**, 1970 (1980).

[24] A. Stahl and D. Baslev, *Electrodynamics of the Semiconductor Band Edge* (Springer-Verlag, Berlin, 1987).

[25] H. A. Lorentz, *The Theory of Electrons* (Teubner, Leipzig, 1916).

[26] N. Bloembergen, *Nonlinear Optics* (Benjamin, New York, 1982).

[27] H. Haken, *Quantum Field Theory of Solids* (North-Holland, Amsterdam, 1983).

[28] The dielectric constant ϵ_0 in the Coulomb potential takes into account the interaction with degrees of freedom not considered in our tight-binding model. If we restrict ourselves to only a few bound excitonic levels, and the interaction with the finite-temperature phonon bath is incorporated, the Coulomb potential should be replaced by a Yukawa potential due to a finite density of free carriers.

[29] A. S. Davydov, *Theory of Molecular Excitons* (Plenum, New York, 1987).

[30] N. N. Bogoliubov and S. V. Tyablikov, DAN SSSR **126**, 53 (1959).

[31] G. J. Blanchard, J. P. Heritage, A. C. VonLehmen, M. K. Kelly, G. L. Baker, and S. Etemad, Phys. Rev. Lett. **63**, 887 (1989).

[32] J. J. Hopfield, Phys. Rev. **182**, 945 (1969).

[33] A. Bosacchi, B. Bosacchi, and S. Franchi, Phys. Rev. Lett. **36**, 1086 (1976).

[34] R. G. Ulbrich and G. W. Fehrenbach, Phys. Rev. Lett. **43**, 963 (1979).

[35] L. J. Rothberg and N. Bloembergen, Phys. Rev. A **30**, 820 (1984).

[36] J. R. Andrew, R. M. Hochstrasser, and H. P. Trommsdorff, Chem. Phys. **62**, 87 (1981).

[37] S. Mukamel and R. F. Loring, J. Opt. Soc. Am. B **3**, 595 (1986).

[38] F. C. Spano and S. Mukamel, Phys. Rev. A **40**, 5783 (1989).

[39] L. Banyai and S. W. Koch, Z. Phys. B **63**, 283 (1986).

[40] There is some ambiguity in the usage of the term "biexciton" in the literature. In this article we use it to denote a bound state of two excitons (also denoted the excitonic molecule). Alternatively, it may denote two interacting (but not bound) excitons. The latter definition is used in the context of small quantum systems such as quantum dots, where the energy of the two-exciton state could be significantly shifted due to exciton-exciton interaction.

[41] V. F. Elesin and Yu. V. Kopaev, Zh. Eksp. Teor. Fiz. **63**, 1447 (1972) [Sov. Phys.—JETP **36**, 767 (1973)].

[42] J. Knoester and A. Mukamel, Phys. Rev. A **41**, 3812 (1990).