Self-trapping in quasi-one-dimensional electron- and exciton-phonon systems

Z. Ivić
Laboratory for Theoretical Physics, Institute of Nuclear Sciences, Vinča, Belgrade, Yugoslavia

D. Kapor and M. Škrinjar
Institute for Physics, Faculty of Sciences, Novi Sad, Yugoslavia

Z. Popović
Laboratory for Theoretical Physics, Institute of Nuclear Sciences, Vinča, Belgrade, Yugoslavia
(Received 23 November 1992; revised manuscript received 31 March 1993)

We study self-trapping of electrons (excitons) in one-dimensional systems with three realistic types of coupling with phonons, applying a variational procedure valid for the whole range of system parameters. Various types of self-trapped states are identified and mapped in the parameter space. Our results are compared to the results of previous studies. The particular case of biological systems is studied and it is shown that the Davydov-soliton concept can be used for the description of electron transport in biological systems, but not for the energy transfer in terms of amide-I vibrations (CO stretching vibration mode).

I. INTRODUCTION

A strong interaction of foreign particles or excitations (electron, hole, exciton, etc.) with lattice vibrations in deformable solids can highly affect their character, resulting in novel features of transport properties and optical spectra of such crystals. This is the result of the trapping of these quasiparticles by a locally distorted surrounding lattice. Such an entity (i.e., a bound state of the quasiparticle and lattice distortion) is now known generally as a *polaron*.[1–3] This phenomenon, commonly called *self-trapping* (ST), is observed in a wide class of substances[4–6] while also predicted in others.[3,7] Examples are numerous, including alkali halide[4] and rare-gas crystals,[5] metal hydrides,[6] biological molecules such as α-helix proteins and DNA,[7,8] and magnetic[9] and molecular crystals,[3] etc.

The character of ST states is highly dependent on the type of coupling between the particle and host phonons, the nature of the host-phonon dispersion, and the dimensionality of the system as well.[10,11] This especially concerns ST on truly one-dimensional (1D) systems where, in the case of a short-ranged electron- or exciton-phonon interaction, the ST state will always be formed and, depending on the values of basic parameters of system, it could expand continuously in spatial extent to form a stable, mobile pulse-shaped excitation: a *soliton*.[7,10–15]

The possibility of realizing solitonlike excitations in quasi-1D systems has caused enormous growth of interest in such materials where the soliton concept has been proposed as a key for understanding the mechanisms of charge and energy transport in Peirels dielectrics [polymer chains of polyacetylene-$(CH_4)_n$ with conjugated bonds[12]] and quasi-1D molecular crystals: α-helix proteins are acetanylide (ACN).[16] However, although theoretical arguments in benefit of soliton existence are rather convincing, the firm experimental evidence is still missing and consequently the validity of the whole soliton concept is doubtful.

This, in particular, concerns the so-called Davydov-soliton (DS) model where a soliton mechanism has been proposed as a basic theoretical framework for the explanation of the way in which biological molecules (α-helix) function in the processes of the charge and energy transfer over long distances. The basic idea of Davydov theory (DT) is that energy released in the hydrolysis of adenosine triphosphate (0.42 eV) can be resonantly transferred to the polypeptide chain of α-helix as a simultaneous absorption of two quanta of CO stretching vibrations (amide-I mode). There it could be stabilized in the form of a solitary wave (DS) due to the interaction with long-wavelength acoustic phonons and then propagate along the chain. Due to its assumed extreme stability against thermal fluctuations and irregularities of crystal lattice, DS could be the ideal candidate for the resolution of the crisis in bioenergetics.[7]

Soon after its appearance, the Davydov concept became the subject of numerous critical reexaminations where various aspects of his model have been discussed. Part of these studies was focused on the formal side of DT only, where the procedure of the derivation of the set of evolution equations for an exciton- (electron-) phonon system (i.e., Davydov *Ansatz*) has been criticized.[17–19] The summary of criticism is that neither of the two versions of the variational states employed in DT reproduces true quantum dynamics of the electron-phonon system since these states do not satisfy the Schrödinger equation for the system.[17–19] Although true, this statement does not offer any deeper insight into the problem of ST in the electron-phonon system, since it just underlines the trivial fact that DT is an approximate one. We believe that any serious effort should be directed towards the analysis of the validity of DT depending on the values of the parameters of the electron-phonon system.

In this paper we shall discuss the phenomenon of ST in 1D media. In particular, we are interested in the circumstances under which the eigenstates of quasiparticles in deformable media can be substantially affected by the interaction with an underlying crystal lattice in the whole range of system parameters and finally we wish to classify various types of ST states.

For that purpose we shall apply the time-independent version of the recently proposed variational method of Brown and Ivić[20–22] (BI) who have studied the relevance of DT for the description of exciton (electron) transport in 1D molecular crystals. This method gives the comprehensive picture of ST in the 1D system covering the whole range of system parameters. In various recent applications of DT, e.g., for the explanation of optical spectra of ACN and other molecules,[16,23,24] the original Davydov model has been modified in the sense that it was assumed that the dispersionless optical phonons could have more significant role than the acoustical ones. For that reason the BI variational method will be applied to three different models which were employed for theoretical description of realistic biologically relevant materials. First we analyze the possibility of ST in the original Davydov model, then we discuss the antisymmetric deformation coupling of an electron (exciton) with neighboring peptide groups, and finally we shall examine the coupling to dispersionless optical phonons.

The structure of this paper is as follows: We introduce the model and variational *Ansatz* in Sec. II. The various types of solutions to variational equations are presented in Sec. III, while the polaron energy is optimized in Sec. IV. The conditions for the formation of the ST states and their classification are formulated in Sec. V. The influence of the nonlinearity parameter is studied in Sec. VI, while the results are compared to the results of previous studies in Sec. VII. The relevance of the approach for some biological problems is studied there, too. Some technical details are exposed in the Appendix.

II. THE MODEL AND VARIATIONAL ANSATZ

Theoretical study of the quasiparticle dynamics in deformable media is based on the well-known Fröhlich Hamiltonian (FH):[2]

$$H=\Delta\sum_n a_n^\dagger a_n - J\sum_n a_n^\dagger(a_{n+1}+a_{n-1})+\sum_{\mathbf{q}}\hbar\omega_{\mathbf{q}} b_{\mathbf{q}}^\dagger b_{\mathbf{q}} + 1/\sqrt{N}\sum_{n\mathbf{q}} F_q e^{iqnR} a_n^\dagger a_n (b_{\mathbf{q}}+b_{-\mathbf{q}}^\dagger) . \quad (2.1)$$

We shall restrict our analysis to the one-dimensional problem where $a_n^\dagger$ (a_n) denotes an operator of creation (annihilation) of an excitation on the nth site, while $b_{\mathbf{q}}^\dagger$ ($b_{\mathbf{q}}$) are phonon creation (annihilation) operators. Δ is the single-site energy, while J is the overlap matrix element of resonant electron (exciton) transfer between adjacent sites. The coupling parameter F_q and phonon frequency ω_q govern the character of ST states. Having in mind the particular interest for the application of the above model in understanding the role of ST states in the functioning of biological substances, we shall consider the following models.

(a) Interaction of a quasiparticle with an acoustic phonon through the short-ranged deformation potential where the excitation sited on the nth peptide group has no preferred orientation interacting symmetrically with the $(n+1)$th and $(n-1)$th molecular groups. Further on, we shall use the following abbreviation: symmetric coupling ADP (acoustic deformation potential) polaron model:

$$F_q=2\chi_1 i(\hbar/2M\omega_q)^{1/2}\sin qR ,\quad \omega_q=\omega_{\mathbf{B}}\sin|qR/2| , \quad (2.2)$$

χ is the electron- (exciton-) phonon coupling strength. $\omega_{\mathbf{B}}=2\sqrt{\kappa/M}$. κ is the stiffness of the chain, while M is the molecular mass (mass of the peptide group in a particular biological context), R is a lattice constant.

(b) Antisymmetric coupling of an excitation with neighboring peptide groups. This model was introduced by Scott (for details see Ref. 25) as a slight modification of the Davydov model with the aim of a more realistic description of the α-helix molecule. The general idea is that stretching of the hydrogen bond immediately adjacent to a C=O oscillatory is of primary significance.

$$F_q=\chi_2(\hbar/2M\omega_q)^{1/2}(e^{iqR}-1) ,\quad \omega_q=\omega_{\mathbf{B}}\sin|qR/2| . \quad (2.3)$$

(c) Excitation point coupling with dispersionless optical phonons:

$$F_q=\chi_3(\hbar/2M\omega_q)^{1/2} ,\quad \omega_q=\omega_0=\text{const} . \quad (2.4)$$

The above model introduced by Holstein[10] is now known as Holstein's molecular crystal model or simply the molecular crystal model (MCM).

Despite its simplicity, the exact eigenstates of FH are yet unknown even in the simplest case defined by (2.1). Exact solutions are known in the transportless limit ($J=0$) only, where appropriate unitary transformation exactly diagonalizes it.[26] For small, but finite J further diagonalization can be performed by means of perturbation theory, usually expanded in terms of a small parameter $2J/\hbar\omega_{\mathbf{B}}$. In this case, lattice distortion is concentrated around a small number of lattice sites (one) and instantaneously follows the slow motion of the particle. Such ST states are usually called small or more precisely nonadiabatic small polarons.

On the contrary, in the adiabatic limit $2J/\hbar\omega_{\mathbf{B}}\gg 1$, lattice distortion is extended over a large number of sites (large polaron or soliton) and it cannot follow the internal motion of the particle inside the trap. The theory of large polarons is mainly based on a simple Pekar variational approach where adiabaticity allows one to neglect lattice kinetic energy in the lowest order, and to treat it as a small perturbation in further procedure.[1,11]

In most of the practical applications, values of the physical parameters lay in the intermediate region where existing theories are less accurate. Therefore, there exists an obvious need for the development of the comprehensive method applicable in the whole range of values of the physical parameters of the system.

There are three basic parameters featured in the energy

spectrum of an electron- (exciton-) phonon system whose relations determine the character of ST states in the system, if any: $2J$ is the electron bandwidth, $\hbar\omega_{\mathrm{B}}$ (or $\hbar\omega_0$ for MCM) is the phonon bandwidth, and finally the so-called small polaron binding energy $E_{\mathrm{B}}=1/N\sum_q |F_q|^2/\hbar\omega_q$ (practically equal to the negative value of ground-state energy in the transportless limit).

In order to achieve the above proposed goal, we apply the method of Brown and Ivić[20–22] based upon the application of variational ansatz interpolating between two well-known limits: nonadiabatic ($2J/\hbar\omega_{\mathrm{B}}\ll 1$) and adiabatic ($2J/\hbar\omega_{\mathrm{B}}\gg 1$), covering strong-, intermediate-, and weak-coupling limits. We choose normalized *BI* trial state in the form:

$$|\psi\rangle=\sum_n \psi_n a_n^\dagger |0\rangle_{\mathrm{ex}} \exp\left[1/2\sum_{\mathbf{q}}(\alpha_{\mathbf{q}}\alpha_{qn}^*-\alpha_{\mathbf{q}}^*\alpha_{qn})\right] \times\exp\left[\sum_{\mathbf{q}}(\beta_{qn}b_{\mathbf{q}}^\dagger-\beta_{qn}^* b_{\mathbf{q}})\right]|0\rangle_{\mathrm{ph}} , \tag{2.5}$$

where ψ_n represents an electron (exciton) wave function (amplitude) while $\alpha_{\mathbf{q}}$, α_{qn}, and $\beta_{qn}=\alpha_{\mathbf{q}}-\alpha_{qn}$ are variational parameters whose meaning is clear from the expectation value of the lattice distortion in the variational state:

$$\langle\psi|u_n|\psi\rangle=N^{-1/2}\sum_{\mathbf{q}}(\hbar/2M\omega_q)^{1/2}e^{iqnR}(\alpha_{\mathbf{q}}+\alpha_{-\mathbf{q}}^*)-N^{-1/2}\sum_{\mathbf{q}m}(\hbar/2M\omega_q)^{1/2}e^{iqnR}(\alpha_{qm}+\alpha_{-qm}^*)|\psi_m|^2 . \tag{2.6}$$

It is clear that lattice distortion consists of two competing contributions. The first one measured by the magnitude of the coherent amplitude $\alpha_{\mathbf{q}}$ defines the so-called frozen distribution which does not follow internal motion of the particle and there each phonon mode behaves in the maximally classical manner. The second contribution is unfrozen distortion which instantaneously follows the motion of the particle. It is the consequence of the real quantum nature of phonons and it also defines a degree of the dressing of the particles by the virtual phonons. Here we can choose $\alpha_{qn}=N^{-1/2}f_q e^{-iqnR}$, where $f_q=f_{-q}^*$. Such a choice satisfies the unitarity condition which makes our results independent of the order of the successive application of two unitary transformations implicitly involved in the definition of the trial state (2.5). We shall treat $\alpha_{\mathbf{q}}$ and f_q as variational parameters whose optimal values should be determined from the minimalization of the energy under the condition of constant momentum $\mathbf{P}$:

$$\mathcal{H}=\langle\psi|H-\mathbf{v}\cdot\mathbf{P}|\psi\rangle . \tag{2.7}$$

Here we use the fact that although the trial function (2.5) is not the eigenstate of the total momentum operator $\mathbf{P}=\mathbf{P}_{\mathrm{ex}}+\mathbf{P}_{\mathrm{ph}}$ of the system, this functional still gives an upper bound to the lowest energy at the momentum $\mathbf{P}$, i.e., the ground-state energy $E_{\mathrm{g.s.}}$ for the constant momentum. In this way, our approach is analogous to the approach of Norris and Whitfield[27] where the theory of Buimistrov and Pekar,[28] valid in the whole parameter space, was used to study energy-momentum relations for the continual ADP with arbitrary value of coupling constants and adiabaticity.

Here $\mathbf{P}=\mathbf{P}_{\mathrm{ex}}+\mathbf{P}_{\mathrm{ph}}$ defines the operator of the total momentum of the electron- (exciton-) phonon system. $\mathbf{v}$ is a Lagrange multiplier and can be interpreted as the polaron velocity, as will be proven later (v will denote $|\mathbf{v}|$):

$$\mathbf{P}_{\mathrm{ex}}=\hbar\sum_{\mathbf{k}}\mathbf{k}a_{\mathbf{k}}^\dagger a_{\mathbf{k}} , \tag{2.8a}$$

$$\mathbf{P}_{\mathrm{ph}}=\hbar\sum_{\mathbf{q}}\mathbf{q}b_{\mathbf{q}}^\dagger b_{\mathbf{q}} . \tag{2.8b}$$

Here $a_{\mathbf{k}}, a_{\mathbf{k}}^\dagger$ are Fourier transforms of electron (exciton) annihilation and creation operators.

After some algebra, the functional (2.7) becomes

$$\begin{aligned}\mathcal{H}=&\left[\Delta-\frac{1}{N}\sum_q F_q(f_q+f_{-q}^*)+\frac{1}{N}\sum_q\hbar\omega_q|f_q|^2\right]\sum_n|\psi_n|^2-\tilde{J}\sum_n\psi_n^*(\psi_{n+1}+\psi_{n-1})\\&-\hbar\sum_{\mathbf{k}}\mathbf{v}\cdot\mathbf{k}|\psi_{\mathbf{k}}|^2+1/N^{-1/2}\sum_{\mathbf{q}n}F_q e^{iqnR}|\psi_n|^2(\alpha_{\mathbf{q}}+\alpha_{-\mathbf{q}}^*)+\sum_{\mathbf{q}}\hbar(\omega_q-\mathbf{q}\cdot\mathbf{v})|\alpha_{\mathbf{q}}|^2\\&-1/N^{-1/2}\sum_{\mathbf{q}n}\hbar(\omega_q-\mathbf{q}\cdot\mathbf{v})(f_q^*\alpha_{\mathbf{q}}e^{iqnR}+f_q\alpha_{\mathbf{q}}^*e^{-iqnR})|\psi_n|^2 ,\end{aligned} \tag{2.9}$$

where $\tilde{J}=J\exp[-(1/N)\sum_q|f_q|^2(1-\cos qR)]$ represents the effective intersite transfer integral, while $\psi_{\mathbf{k}}$ denotes the Fourier transform of exciton (electron) wave function introduced here just for the sake of convenience. From $\partial\mathcal{H}/\partial\alpha_q=0$, we find

$$\alpha_{\mathbf{q}}=-N^{-1/2}[F_q^*/\hbar(\omega_q-\mathbf{q}\cdot\mathbf{v})-f_q]\sum_n|\psi_n|^2e^{iqnR} . \tag{2.10}$$

Substituting this result into (2.9) we obtain the following functional after applying the continuum approximation (up to order R^2):

$$\mathcal{H}=\left[\Delta-\frac{1}{N}\sum_q F_q(f_q+f^*_{-q})+\frac{1}{N}\sum_q \hbar\omega_q|f_q|^2-2\tilde{J}\right]\int\frac{dx}{R}|\psi(x)|^2+\tilde{J}R^2\int\frac{dx}{R}|\psi_x(x)|^2$$
$$-\frac{v\hbar}{2i}\int\frac{dx}{R}(\psi^*\psi_x-\psi\psi^*_x)-\int\int\frac{dxdy}{R}K(x-y)|\psi(x)|^2|\psi(y)|^2$$
$$+\sum_q F_q(f_q+f^*_{-q})\int\int\frac{dxdy}{R^2}e^{iq(x-y)}|\psi(x)|^2|\psi(y)|^2-\sum_q\hbar\omega_q|f_q|^2\int\int\frac{dxdy}{R^2}e^{iq(x-y)}|\psi(x)|^2|\psi(y)|^2 . \quad (2.11)$$

The kernel $K(x-y)$ is defined as

$$K(x-y)=1/(2\pi\hbar)\int_{-\infty}^{+\infty}dq\, e^{iq(x-y)}\omega_q|F_q|^2/(\omega_q^2-q^2v^2) . \quad (2.12)$$

III. TYPES OF SOLUTIONS

So far we did not specify the form of the electron (exciton) wave function $\psi(x)$ and the discussion is quite general because the above functional allows one to study the possibility of the formation of linear (Bloch band states corresponding to the free excitations) as well as nonlinear excitations—solitons are large polarons and small adiabatic polarons. For that purpose we must find the optimized values of the variational parameter f_q which obviously lies somewhere between two limiting cases: $f_q=0$, which should correspond to large polaron or soliton solutions and $f_q=F_q/\hbar\omega_q$ recovering a small polaron band limit. However, one knows from the literature that $f_q=F_q/\hbar\omega_q$ in the transportless limit only, therefore $0<f_q<F_q/\hbar\omega_q$. A particular value of f_q should, in principle, determine the character of the eigenstates of the system and corresponding energy eigenvalues, but it could also be used as a criterion for the applicability of previous theoretical approaches. Comparing the BI trial state (2.5) with previously utilized variational *Ansätze,* one can see that the *ad hoc* choice $f_q=0$ corresponds to Pekar's variational *Ansatz* or so-called D_2 states according to the currently accepted terminology in the theory of DS.[17–19] On the other hand, the choice $f_q=F_q/\hbar\omega_q$ corresponds to the so-called D_1 *Ansatz* which is exact in the nonadiabatic limit ($J\to 0$) so these states correspond to nonadiabatic small polarons.

Let us assume the simple form of the dressing parameter $f_q=\delta F_q/\hbar\omega_q$, where δ measures the relative extent of induced lattice distortion and consequently the degree of dressing. This assumption implies equal dressing for all phonon modes and at the first sight it looks like a very strong assumption since the whole set of variational parameters (one for each mode) is substituted by a single one. However, according to previous studies,[13,14,21,22,29] where this assumption was made, one can see that this method gives the same qualitative predictions as the q-dependent one, while the estimates of the ground-state (g.s.) energy are slightly higher. For this reason the "δ approach" was mainly used for the qualitative analyses of ST phenomena, while a detailed discussion of its accuracy was not presented until now. In the present context, we shall estimate the validity of the δ approach by comparing it to the translationally invariant (TI) theory of Emin[14] and Ventzl and Fischer[30] in Sec. VII.

Clearly, optimal values of δ lies somewhere between 0 and 1, where 0 corresponds to soliton or large polaron excitations, while $\delta=1$ recovers the small polaron limit. Substituting this simplified form into the functional (2.11), we obtain

$$\mathcal{H}=[\Delta-\delta(2-\delta)E_B-2\tilde{J}]\int\frac{dx}{R}|\psi(x)|^2+\tilde{J}R^2\int\frac{dx}{R}|\psi_x(x)|^2-\frac{v\hbar}{2i}\int\frac{dx}{R}(\psi^*\psi_x-\psi\psi^*_x)$$
$$-\int\int\frac{dxdy}{R}K(x-y)|\psi(x)|^2|\psi(y)|^2+aE_B\delta(2-\delta)\int\frac{dx}{R}|\psi(x)|^4 . \quad (3.1)$$

Here $E_B=1/N\sum_q|F_q|^2/\hbar\omega_q$ is the small polaron binding energy, while the effective tunneling term now has the form

$$\tilde{J}=Je^{-\delta^2S} , \quad (3.2)$$

where

$$S=1/N\sum_q[|F_q|^2/(\hbar\omega_q)^2](1-\cos qR) \quad (3.3)$$

is the exciton-phonon coupling constant. For models of interest the explicit values of E_B and S are

$$E_B=4\chi_1^2/M\omega_B^2 , \quad S=(8/3\pi)E_B/\hbar\omega_B \quad (3.4)$$

for the ADP symmetric mode,

$$E_B=2\chi_2^2/M\omega_B^2 , \quad S=(8/\pi)E_B/\hbar\omega_B \quad (3.5)$$

for the ADP antisymmetric coupling, and finally

$$E_B=\chi_3^2/2M\omega_B^2\ ,\quad S=E_B/\hbar\omega_B \tag{3.6}$$

for Holstein's MCM model.

Parameter a defines the degree of nonlinearity having values $a=1$ for MCM and the antisymmetric ADP model and $a=2$ for the ADP symmetric model. It relates the coefficient of the nonlinear term to the binding energy.

$$K(x-y)=\begin{cases} aE_B\delta(x-y)/(1-v^2/c^2) & \text{ADP}\ , \\ E_B\int (dq/2\pi)e^{iq(x-y)}/[1-(qv/\omega_0)^2] & \text{MCM}\ . \end{cases} \tag{3.7}$$

The nonlinear functional (2.11) or (3.1) can be used as a basis for the analysis of the existence of the various types of ST states. However, if we wish to examine linear excitations (Bloch band states, small polaron band), we must neglect the nonlinear term since the plane waves are not self-consistent solutions of the NL functional [(2.11), (2.12), and (3.1)], as was explicitly shown by Rashba.[3,31] It is not very hard to prove that neglecting the NL term and choosing the exciton (electron) wave function in the plane-wave form, our functional (2.11) is equivalent to the previously used method of Emin[14] and Venzl and Fischer.[30] We shall use later such a linearized approach in order to compare our δ approach with previous theories.

Let us now discuss the possible polaron (soliton) solutions of the above functional. As the first step we shall find the general expression for the polaron momentum, effective mass and prove that v has the meaning of polaron (soliton) velocity. We choose the electron (exciton) wave function in the form $\psi(x)=e^{ikx}\phi(x)$, where $\phi(x)$ is the real function while $k=\hbar v/(2\tilde{J}R^2)=m^*v/\hbar$ ($m^*=\hbar^2/2\tilde{J}R^2$):

$$\begin{aligned}\mathcal{H}=&[\Delta-\delta(2-\delta)E_B-2\tilde{J}-m^*v^2/2]\int\frac{dx}{R}\phi(x)^2 \\ &+\tilde{J}R^2\int\frac{dx}{R}\phi_x(x)^2-\int\int\frac{dxdy}{R}K(x-y)\phi(x)^2\phi(y)^2 \\ &+aE_B\delta(2-\delta)\int\frac{dx}{R}\phi(x)^4\ .\end{aligned} \tag{3.8}$$

The energy of the system can be easily obtained from the following relation:

$$E=\mathcal{H}+\mathbf{P}\cdot\mathbf{v}\ , \tag{3.9}$$

where $\mathbf{P}$ is the total momentum of the system given by

$$\mathbf{P}_{\rm tot}=m^*\mathbf{v}+\hbar\sum_{\mathbf{q}}\mathbf{q}\left|\alpha_{\mathbf{q}}-\sum_n\alpha_{qn}|\psi_n|^2\right|^2\ . \tag{3.10}$$

Introducing the explicit forms of $\alpha_{\mathbf{q}}$ and α_{qn}, we arrive in the continuum limit at the following expressions for the momentum and energy:

$$\mathbf{P}_{\rm tot}=m^*\mathbf{v}+\mathbf{v}\int\int\int\frac{dq}{2\pi}\frac{dxdy}{R}\frac{2\omega_q^2q^2|F_q|^2}{(\omega_q^2-q^2v^2)}\phi(x)^2\phi(y)^2e^{iq(x-y)}\ , \tag{3.11}$$

$$\begin{aligned}E=&[\Delta-\delta(2-\delta)E_B-2\tilde{J}+m^*v^2/2]\int\frac{dx}{R}\phi(x)^2+\tilde{J}R^2\int\frac{dx}{R}\phi_x(x)^2 \\ &-\int\int\frac{dxdy}{R}K(x-y)\phi(x)^2\phi(y)^2+aE_B\delta(2-\delta)\int\frac{dx}{R}\phi(x)^4+vP_{\rm ph}\ ,\end{aligned} \tag{3.12}$$

where $P_{\rm ph}$ is the second term in (3.11).

Using the above two expressions we can easily prove that

$$v=\partial E/\partial P=(\partial E/\partial v)/(\partial P/\partial v)\ , \tag{3.13}$$

which means that v can be interpreted as soliton (polaron) velocity. The effective mass follows from the relation

$$m_{\rm eff}=(\partial P_{\rm tot}/\partial v)_{v\to 0}\ ,$$

which gives the general expression for the soliton effective mass:

$$\begin{aligned}m_{\rm eff}=m^*+&\int\int\int\frac{dq}{2\pi}\frac{dxdy}{R}\frac{2q^2|F_q|^2}{\omega_q^3} \\ &\times\phi(x)^2\phi(y)^2e^{iq(x-y)}\ .\end{aligned} \tag{3.14}$$

Let us discuss some particular cases.

A. ADP model

Inserting the explicit forms for F_q and ω_q we have the following eigenvalue problem which follows from $\partial\mathcal{H}/\partial\phi=0$:

$$\begin{aligned}\tilde{J}R^2\phi_{xx}+&[\mathcal{E}-\Delta+\delta(2-\delta)E_B+2\tilde{J}-m^*v^2/2]\phi \\ &-2aE_B[\delta(2-\delta)-1/(1-v^2/c^2)]\phi^3=0\ ,\end{aligned} \tag{3.15}$$

which has the well-known soliton solution:

$$\phi(x)=\sqrt{\mu/2}\,\mathrm{sech}(\mu x/R)\ , \tag{3.16}$$

where

$$\mu=\frac{aE_B[(1-\delta)^2+\delta(2-\delta)v^2/c^2]}{2J(1-v^2/c^2)}\exp(\delta^2S)\ , \tag{3.17}$$

while

$$\mathcal{E}=\Delta-\delta(2-\delta)E_B-2\tilde{J}+m^*v^2/2 - \frac{a^2E_B^2[(1-\delta)^2+\delta(2-\delta)v^2/c^2]}{2J(1-v^2/c^2)}e^{\delta^2 S}$$

denotes the internal electron (exciton) energy inside the trap:

$$m_{\text{eff}}=m^*e^{\delta^2 S}[1+\gamma\pi^2S^2(1-\delta)^2] , \tag{3.18}$$

where γ equals $\frac{3}{2}$ for symmetric and $\frac{2}{3}$ for antisymmetric coupling.

It is obvious that the character of ST states is governed by the value of δ which will be determined in the next sections from the minimization of the soliton energy. The *ad hoc* choice $\delta=0$ recovers DS solutions at once, while $\delta=1$ recovers the so-called soliton excitations of a small polaron band obtained first by BI.[20] Properties of both kinds of solutions are well known and there is no need for their extensive analysis here.

B. Soliton solutions for Holstein's model

In this case the eigenvalue problem has the form

$$\tilde{J}R^2\phi_{xx}+[\mathcal{E}-\Delta+\delta(2-\delta)E_B-2\tilde{J}-m^*v^2/2]\phi-2E_B\delta(2-\delta)\phi^3+2E_B\int\frac{dq}{2\pi}\int dy\frac{e^{iq(x-y)}}{1-q^2v^2/\omega_0^2}\phi^2(y)\phi(x)=0 . \tag{3.19}$$

The above equation cannot be solved exactly except in the motionless case ($v=0$) and then these solutions are given by (3.16) with μ defined by (3.17) (with $v=0$). Unfortunately, moving Holstein's polaron was not investigated in detail until now. Although (3.19) cannot be solved exactly, some of the properties of the soliton solutions can be analyzed qualitatively even without explicit knowledge of $\phi(x)$. We note first that, as in the case of DS, there exists a velocity which cannot be exceeded by our excitations. This limiting velocity equals the phase speed of sound: $v_f=\omega_q/q=\omega_0/q$. A satisfactory qualitative description of the behavior of slow solitons can be achieved on the basis of direct variational method. For slow solitons, we can adopt the following approximation in (3.1):

$$K(x)=\delta(x)-(v/\omega_0)^2\partial^2/\partial x^2\delta(x) . \tag{3.20}$$

Assuming that the shape of slow solitons remains practically unchanged, we can use (3.11) as a trial state where μ is the parameter to be determined from the minimum condition of the functional

$$\mathcal{H}\cong[\Delta-\delta(2-\delta)E_B-2\tilde{J}-m^*v^2/2]+\tilde{J}\mu^2/3 - E_B\mu(1-\delta)^2/3-4E_Bv^2\mu^3/(15\omega_0^2R^2) , \tag{3.21}$$

which was obtained after substitution of (3.20) into (3.1). So we have

$$\mu=2\mu_0(\omega_0/q_0v)^2[1-\sqrt{1-(1-\delta)^2(q_0v/\omega_0)^2}] . \tag{3.22}$$

$\mu_0=E_B/2\tilde{J}$ is the inverse width of the immobile polaron; ω_0/q_0 represents the maximal polaron velocity, while $q_0=2E_B/\sqrt{5}\tilde{J}R=4\mu_0/\sqrt{5}R$ represents the maximal momentum of virtual phonons engaged in the lattice distortion.

The effective mass is given by

$$m=m^*e^{\delta^2 S}[1+\tfrac{4}{15}S^2\mu^2(1-\delta)^2] . \tag{3.23}$$

Although approximate, the above results can serve as the basis for the analysis of the possibility of realizing soliton excitations of the small polaron band in Holstein's model. In the limit $\delta=1$, $\mu\to0$, which means that such excitations are of the infinite radius so we conclude that in this model "velocity solitons"[20] do not appear.

Let us summarize the results of this section.

(a) Plane-wave solutions (band states) are not self-consistent solutions of 1D exciton-phonon systems which means that they do not arise in the systems under consideration and consequently there is no possibility for the coexistence of free (band) states and ST states (polarons, solitons) characteristic for 3D exciton-phonon systems.

(b) In the case of the ADP model (both versions of coupling) there appear three kinds of soliton states: (1) large polaron, solitonlike states (Davydov solitons $\delta=0$); (2) soliton excitations of a small polaron band $\delta=1$; (3) intermediate soliton states with $0<\delta<1$.

(c) For the MC model, soliton excitations of the small polaron band do not appear.

(d) Finally, let us note that in both models there also appears a so-called adiabatic small polaron a special case of nonlinear excitations in these systems, where $\mu\geq1$ and $J/\hbar\omega_B\gg1$.

Obviously, the character of these states is determined by the value of physical parameters of the systems. In order to distinguish various types of possible ST states of the system, we must find δ whose value controls the value of soliton parameters.

IV. OPTIMIZATION OF POLARON ENERGY AND CONDITIONS FOR THE FORMATION OF SELF-TRAPPED STATES

We are now ready to look for the value of δ which minimizes the energy, i.e., the degree of dressing corresponding to stable ST states. In this section we shall concentrate on some details of energy optimization and the conditions which restrict the range in which the parameters of the system occur. The physical interpretation of the solutions obtained will be presented in the next section.

In order to find optimized values of δ, we shall consider the static ($v=0$) case without loss of generality. Substituting the above-mentioned soliton solutions into the

energy functional and using the explicit expression for μ in terms of δ, we obtain the energy of the system:

$$E=\Delta-\delta(2-\delta)E_{\mathbf{B}}-2J\exp(-\delta^2 S) -a^2\frac{(1-\delta)^4}{12}\frac{E_{\mathbf{B}}^2}{J}e^{\delta^2 S} . \tag{4.1}$$

We are looking for the dressing parameter δ as a function of the parameters of the system. The set of values of the parameters which determines the stationary points of E follows from the stationarity condition $\partial E/\partial\delta=0$. Taking into account

$$E_{\mathbf{B}}^{-1}\partial(E-\Delta)/\partial\delta=-2(1-\delta)+2B\delta\exp(-\delta^2 S) +2a^2 S\frac{(1-\delta)^3}{3B}e^{\delta^2 S}[1-S\delta(1-\delta)/2] , \tag{4.2}$$

where $B=2JS/E_{\mathbf{B}}\cong 2J/\hbar\omega_{\mathbf{B}}$ for all of the mentioned cases, we can find a transcendent equation for δ in terms of system parameters S and B:

$$\delta=\frac{1}{1+[2B/(1\pm K)]e^{-\delta^2 S}} , \tag{4.3}$$

where

$$K=\{1-\tfrac{4}{3}a^2 S\delta(1-\delta)[1-S\delta(1-\delta)/2]\}^{1/2} . \tag{4.4}$$

Relation (4.3) is going to be the initial point of our discussions. Let us first check if it can reproduce well-known results. The limit $\delta\to 1$ is realized either for $B\to 0$ or $S\to\infty$. These are, in fact, two different ways of describing the immobile exciton, i.e., the transportless limit. On the other hand, the case $\delta\to 0$ is achieved for $B\to\infty$ (for small S) and this is the case of strong adiabaticity. We can recall here that we have mentioned that our results will also offer the test for the applicability of various trial states. This discussion already gives hints in that direction and we shall be more specific later. It is also important to notice that both limits are achieved independent on the sign of K.

Depending on the sign in front of K in Eq. (4.3), we can have, in principle, two branches of solutions for $\delta(S,B)$ defining two possible types of ST states. Each of the branches defines in the (δ,S) plane a set of points describing the extreme values of E. The very presence of two branches implies that coexistence of ST stable states, in the same region of parameters, which was not previously noticed. So we shall pay special attention to the behavior of the $(-)$ branch.

Stable states are defined by the minima of energy, that is, by the points where $\partial^2 E/\partial\delta^2>0$. They are separated from the maxima $(\partial^2 E/\partial\delta^2<0)$ (unstable states) by the line called the stability line. This line is the set of points where $\partial^2 E/\partial\delta^2=0$ and $\partial E/\partial\delta=0$, that is, which belong to the lines defined by (4.3). Obviously, if we decide to plot (S,δ) points for a given B, then the stable ST states will occur on one side of this line only. First we derive $\partial^2(E-\Delta)/\partial\delta^2$ and, combining the two above-mentioned conditions, we obtain the equation for the stability line as

$$z^2(1-2\delta^2 S)/\delta^2 S+2z =\tfrac{4}{3}a^2 y^2\{y^2-y[\tfrac{9}{2}-1/(2S\delta)]+3\} , \tag{4.5}$$

where $z=y(1\pm K)$ and $y=S\delta(1-\delta)$.

We are now in the position to determine the conditions for the existence of the possible kinds of stable ST states which we have mentioned in previous sections. Our aim is to determine δ in terms of system parameters. We choose to look for δ as a function of S with B as a parameters. Equations (4.3) and (4.6) can only be solved numerically. Prior to numerical treatment, we should mention several restrictions which follow from certain mathematical and physical demands. The detailed discussion is given in the Appendix and we just list them here, because they largely simplify the numerics, by limiting the range of the solutions.

First of all, there exists the condition of reality of δ, which is important for $a^2>\frac{3}{2}$ and leads to the appearance of two boundaries in the $S-\delta$ plane. For the $(-)$ branch, there is also the condition $\delta>0$, leading to $y<2$. Finally, the continuum approximation is valid for $\mu\ll 1$, so $\mu=1$ is the final boundary for the applicability of the continuum approximation. For the $(+)$ branch, the condition is fulfilled for $a<3$, while for the $(-)$ branch, the condition is $\frac{7}{4}<a<3$ and then the region is bounded by the two following curves:

$$S=(1/2a)/[\delta(1-\delta)]$$

and

$$S=(6/a-2)/[5\delta(1-\delta)] .$$

We shall discuss the effect of these limitations on the value of the parameter of nonlinearity, after we find solutions of $\delta=\delta(S)$. The most convenient way of visualization is to plot $S-\delta$ points corresponding to a fixed value

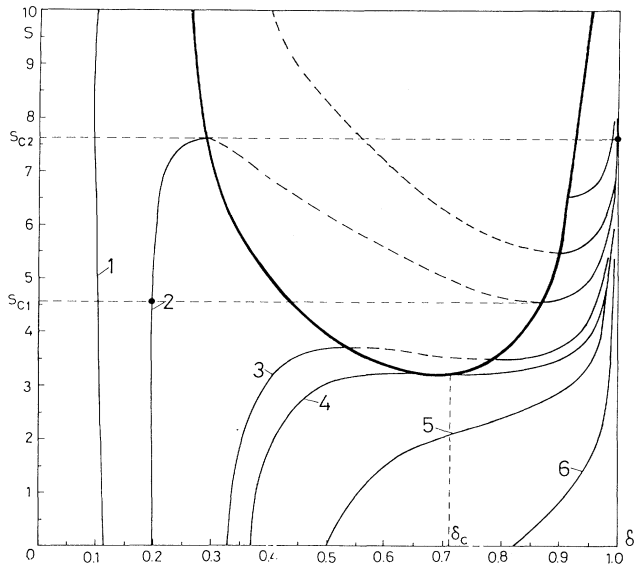

FIG. 1. Adiabates $S=S(\delta)$ for $a=1$ corresponding to the following values of B: 1, 8.0; 2, 4.0; 3, 2.0; 4, $B_c=1.712$; 5, 1.0; 6, 0.20; ——, stability line.

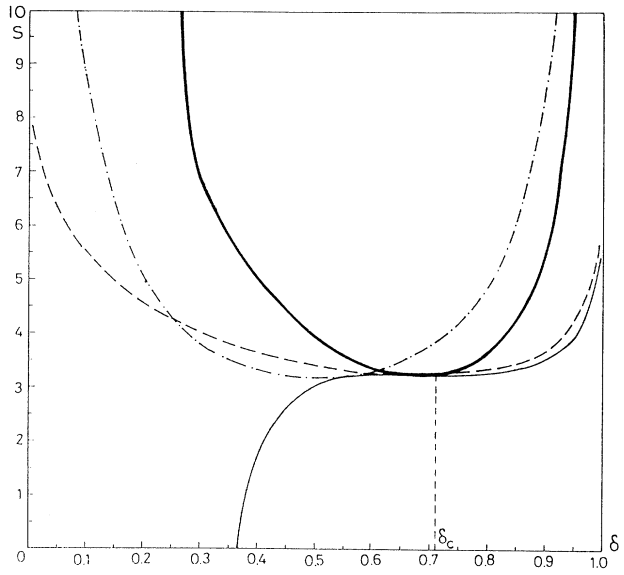

FIG. 2. Important boundaries in the parameter space: ——, critical adiabate; ——, stability line; —·—·—·, metastability boundary; — — —, continuum approximation validity limit.

of B in the $S-\delta$ diagram. We have plotted $S(\delta)$ (Figs. 1–3) for a few chosen values of B spanning the whole range of adiabaticity. Please notice that we use $S=S(\delta)$ instead of $\delta=\delta(S)$ which is a more natural choice.[21] The most important practical reason for this is that the points where both derivatives of the energy vanish correspond to the points on the $S=S(\delta)$ curve where the first derivative $\partial S/\partial\delta$ vanishes. [This means that stability lines pass through stationary points of $S=S(\delta)$.]

V. CLASSIFICATION OF STABLE SELF-TRAPPED STATES

Let us now review the results of our analysis. We have considered two values of the NL parameter a relevant for the models studied: $a=1$ and 2. Before going to these

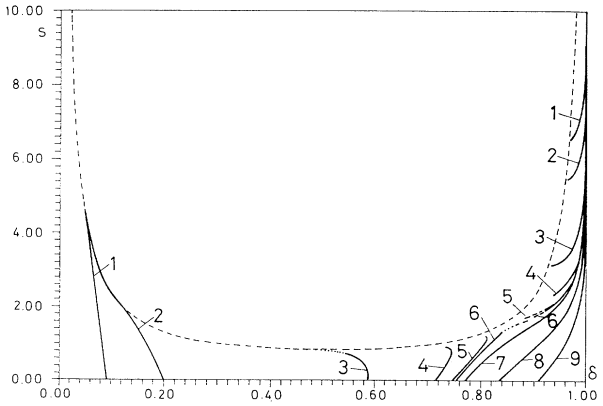

FIG. 3. Adiabates $S=S(\delta)$ for $a=2$ corresponding to the following values of B: 1, 10.0; 2, 4.0; 3, $\frac{2}{3}$; 4, 0.55; 5, 0.4; 6, 0.34; 7, 0.30; 8, 0.20; 9, 0.10; — — —, continuum approximation validity limit.

particular cases, let us look for some general properties. One of the general properties is the behavior for $\delta\to 0$. We see that for $S\to 0$, $\delta\to 1/(1+B)$, which gives us a good estimate of the parameter space region in which the curve lies.

There is another region with a-independent behavior: that is the limit $\delta\to 1$ occurring for each adibate either for $S\to\infty$ or for $B\ll 1$, independent of the value of S. This limit corresponds to the highly dressed, practically immobile (i.e., self-trapped) polaron whose mass tends to infinity, while its radius is smaller than lattice spacing. The divergence of the polaron effective mass is the consequence of the vanishing of the effective hopping integral: $\tilde{J}=J(1-\delta)(1+K)/2B\delta$ when $\delta\to 1$. For that reason we refer to such an entity as a nonadiabatic small polaron rather than a "highly dressed" soliton of infinite radius, as one could expect, at first glance, according to the expression for inverse soliton width where $\mu\to 0$ for $\delta\to 1$. Substituting the value $\delta=\delta_0\cong 1$ in the stability condition $\partial^2 E/\partial\delta^2>0$, we find the condition for the existence and stability of these states:

$$\delta_0\cong 1\to 1-Be^{-S}(1-2S)>0\to 3\delta_0^2 S\exp(-\delta_0^2 S)\ll 1\ ,$$

which is quite analogous to the one previously obtained by Emin.[14]

On the other hand, in the weak-coupling regime ($S\ll 1$), there appear δ values which, depending on the value of B, can be significantly lower than unity, even in the nonadiabatic case, yet they are also far from zero. Such states should be called partially dressed soliton states. We shall now study them for the two particular cases of interest.

Let us study first the case $a=1$. Since here K is always real, no restrictions come from that demand. On the other hand, the $(-)$ branch does not satisfy the condition for the applicability of the continuum limit so there appears no $(-)$ branch, which agrees with the fact that the $(-)$ branch was never detected in the previous studies. As for the $(+)$ branch, it is detected only below the curve $S=0.8/[\delta(1-\delta)]$, where the continuum limit is valid.

It is important to notice that for small B, the curve includes only minima. Yet, above a certain value $B_c=1.712$, the curve has a region of maxima corresponding to unstable states. The critical adibate corresponding to B_c is determined from the condition of the tangent to the stability line. Looking at Fig. 1 as the graph of $\delta=\delta(S)$, one can see that it is not a single-valued function, since for given S there appear new values, two of them corresponding to minima, so there occurs an abrupt transition between stable states.

We can classify stable states according to their spatial extent and the degree of adiabaticity (Fig. 2). First of all, close to $\delta=1$, we definitely deal with a nonadiabatic small polaron. In the region bounded from above by the curve $B=B_c$ and $\delta\simeq 0.9$, we still notice nonadiabatic behavior, but these states lie well below the $\mu=1$ curve, so they can be classified as large but nonadiabatic polarons, whose existence was not predicted previously.

For $B>B_c$ and left from the $\mu=1$ curve, we deal with an adiabatic large polaron. Since there exist two stable states corresponding to given S and given B, we can con-

struct the metastability boundary, by drawing a horizontal line ($S=$const) through the point where the given adiabate cuts the stability line. The point where this horizontal line crosses the adiabate on the other side of the stability line belongs to the metastability boundary (mb) line. In Fig. 1, we have denoted these two points by S_{c1} and S_{c2} on curve 2 corresponding to $B=4.0$.

We notice three separate regions in the $S-\delta$ plane. Below the mb line only stable states exist, which are uniquely defined for a given S and B. Above the stability line, there are no stable states. Between mb and the stability line, the states could be defined as metastable, since there exist two stable states on each adiabate for a given S, one corresponding to $\delta<\delta_c$ and the other one for $\delta>\delta_c$ ($\delta_c\cong 0.71$). The curve $\mu=1$ goes through the middle of the region on the left-hand side, so we can separate large and small adiabatic polarons. On the right-hand side ($\delta>\delta_c$), we encounter only small nonadiabatic polarons.

Let us now consider the case $a=2$ (Fig. 3). Here we encounter both branches in principle. Each branch must satisfy the condition of the reality of δ which, in this particular case, has the form

$$S<1-\sqrt{5/8}=0.21, \quad S>1+\sqrt{5/8}=1.79 .$$

These two curves bound a forbidden region in $S-\delta$ plane, where none of the branches should appear.

The condition $K<1$ implies $y<2$, so there is another upper boundary for the $(-)$ branch. The most important restriction comes from the demand for the validity of the continuum approximation. We remember that there exists a necessary condition $y<1/(2a)=\frac{1}{4}$, which is taken care of by the reality condition $y<0.21$. On the other hand, (A9), $y>(\frac{6}{2}-2)/5=\frac{1}{5}$. We see that the $(-)$ branch exists, yet it is defined in a small section of the $S-\delta$ plane between the curves:

$$S>0.20/[\delta(1-\delta)] , \quad S<0.21/[\delta(1-\delta)] .$$

So for all practical purposes, its existence could be neglected.

As for the $(+)$ branch, the condition for the validity of continuum approximation is $y<\frac{1}{5}=0.20$. This means that the only region where the curve can appear is the part of the $S-\delta$ plane below this curve, since it is lower than the boundary following from the reality condition. The problem is two asymptotic regions: $\delta\to 0$ and $\delta\to 1$. If one substitutes the value $y=y_l$ in Eq. (4.4) one obtains

$$1+2B\exp[-\delta y_l/(1-\delta)]=1/\delta .$$

This equation has no solution $\delta=0$ but can have the solution $\delta=1$. This implies that no curve defined in the region below the continuum curve can cross this curve and appear again in the region for $\delta=0$. On the other hand, there can appear small polaron solutions.

These states are all stable since one can simply evaluate the value of $\partial^2 E/\partial\delta^2$ and show that it is positive in the whole area below the continuum limit boundary. So there is no need to discuss the stability curve, etc., because they would obviously fall in the forbidden region. It seems that the systems of lower value of a possess a larger variety of possible states.

VI. THE INFLUENCE OF THE NONLINEARITY PARAMETER

We have seen that the dependence of optimal δ on system parameters is strongly influenced by the value of nonlinearity parameter a. Let us first remind the reader that a represents the ratio of the coefficient of the nonlinear term to the linear small polaron binding energy E_B. We are going to analyze the effect of a, following the qualitative analysis proposed in Ref. 21.

We return first to the NL functional (3.1) and substitute the explicit value of the ADP kernel $K(x-y)$ (3.7) which gives

$$\mathcal{H}=[\Delta-\delta(2-\delta)E_B-2\tilde{J}]\int\frac{dx}{R}|\psi(x)|^2 + \tilde{J}\tilde{R}^2\int\frac{dx}{R}|\psi_x(x)|^2-aE_B(1-\delta)^2\int\frac{dx}{R}|\psi(x)|^4 . \tag{6.1}$$

Let us now compare the leading linear contribution and nonlinear term. Both of these terms lower the energy, so they both stabilize the system. We define a function $b(\delta)$ as the sum of the two coefficients which stand in front of E_B:

$$b(\delta)=\delta(2-\delta)+a(1-\delta)^2 . \tag{6.2}$$

This expression has the boundary values $b(0)=a$ and $b(1)=1$ for any value of a. It is important that b is always lower than 1 for $a<1$, while it is higher than 1 for $a>1$. Only for $a=1$, $b=1$, so only then are the two terms in perfect balance. In fact, for $a<1$, the linear term is a more stabilizing factor, while for $a>1$ it is the nonlinear term which contributes mostly to the stabilization.

VII. DISCUSSION

After presenting the above analysis, we are now finally in the position to turn our attention to the problem of the occurrence of various types of ST states in more realistic conditions, especially in some relevant biological substances where the assumption of a possible model of solitonic (polaronic) energy and charge transfer is a relevant one. Since the δ approach predicts the existence of several types of NL ST states, except small polaron band states, we must compare our results with the results of previous theories: solitonic ones, as well as TI theories, so that we could determine the limit to which our predictions can be accepted as reliable. We are going to estimate the ground-state energy $E_{g.s.}$ for various approaches.

We have already mentioned that our *Ansatz* can reproduce the results of the original Davydov (or better Pekar) D_z *Ansatz* with the choice $\delta=0$. Substituting $\delta=0$ into the expression for the energy (4.1), we obtain

$$E_{g.s.}(\delta=0)=-2J-\frac{a^2}{12}\frac{E_B^2}{J} , \tag{7.1}$$

so the difference in energies is

$$\Delta E_{\text{g.s.}} = E_{\text{g.s.}}(\delta) - E_{\text{g.s.}}(\delta=0)$$
$$= -\delta(2-\delta)E_{\text{B}} - 2J[\exp(-\delta^2 S)-1]$$
$$-\frac{a^2}{12}\frac{E_{\text{B}}^2}{J}[(1-\delta)^4 e^{\delta^2 S}-1] . \quad (7.2)$$

We expect soliton solutions in the range of large adiabaticity and for $S < B^2$. This allows us to expand the exponent up to S. Dividing expression (7.2) by E_{B},

$$\Delta E_{\text{g.s.}}/E_{\text{B}} = -\delta(2-\delta) - \frac{2JS\delta^2}{E_{\text{B}}}$$
$$+\frac{a^2}{12}\frac{E_{\text{B}}}{J}\delta(2-\delta)(2-2\delta+\delta^2) . \quad (7.3)$$

We have neglected $\delta^2 S$ in the last term, since E_{B}/J is already proportional to S. Rearrangement of the terms leads to

$$\Delta E_{\text{g.s.}}/E_{\text{B}} = -2+(1+B)\delta + \frac{a^2 E_{\text{B}}}{3J} - \frac{a^2 E_{\text{B}}}{6J}\delta(3-2\delta^2)$$
$$\cong -2 + a^2 E_{\text{B}}/3J . \quad (7.4)$$

This is the lowest-order estimate, which shows that our approach always gives lower energy, since a pure soliton demands $E_{\text{B}}/J \ll 1$, sufficient to compensate even for $a^2=4$.

Now, let us compare $E_{\text{g.s.}}(\delta)$ to $E_{\text{g.s.}}$ obtained by linear (TI) theories. Naturally, in the region where $E_{\text{g.s.}}(\delta)$ is lower, NL excitations described in the previous section give a better description of the system, while in the opposite case, the band picture is more favorable. For that purpose, let us look only at the linear part of the energy functional [(2.11), (2.12), and (3.1)], before introducing the assumption of equal dressing of each phonon mode. Taking ψ_n in the form of the plane wave and remembering that the ground state corresponds to a vanishing wave vector, we have

$$E_{\text{g.s.}}^{\text{lin}} = -\frac{1}{N}\sum_q F_q(f_q + f_{-q}^*) + \frac{1}{N}\sum_q \hbar\omega_q |f_q|^2 - 2JA , \quad (7.5)$$

$$A = \exp\left[-\frac{1}{N}\sum_q |f_q|^2(1-\cos qR)\right] . \quad (7.6)$$

From here on, we shall neglect the constant term Δ because it has no influence on our further analysis.

If we vary $E_{\text{g.s.}}^{\text{lin}}$ over f_q, we obtain

$$f_q^* = \frac{F_q}{\hbar\omega_q + 2JA(1-\cos qR)} . \quad (7.7)$$

As we have previously mentioned, this is the equation which was previously derived for various particular cases by Emin[14] (in the context of the MC model), Venzl and Fischer[30] (symmetric ADP model), and finally Köngeter and Wagner[32] in the special case when only one variational parameter is retained. The result is

$$E_{\text{g.s.}}^{\text{lin}} = -2I_2 + I_3 - 2JA , \quad (7.8)$$

$$A = \exp(-I_1) . \quad (7.9)$$

Here, we have

$$I_1 = \frac{1}{N}\sum_q \frac{|F_q|^2}{(\hbar\omega_q)^2}\frac{1-\cos qR}{[1+2JA/(\hbar\omega_q)(1-\cos qR)]^2} , \quad (7.10)$$

$$I_2 = \frac{1}{N}\sum_q \frac{|F_q|^2}{\hbar\omega_q}\frac{1}{1+2JA/(\hbar\omega_q)(1-\cos qR)} , \quad (7.11)$$

$$I_3 = \frac{1}{N}\sum_q \frac{|F_q|^2}{\hbar\omega_q}\frac{1}{[1+2JA/(\hbar\omega_q)(1-\cos qR)]^2} . \quad (7.12)$$

There exist formal relations between three integrals and we are going to use them in particular cases.

Let us first study the MC model as an example. Using the notation of Secs. III and IV, we obtain

$$I_1 = S(1+2BA)^{-3/2} , \quad (7.13)$$

$$I_2 = E_{\text{B}}(1+2BA)^{-1/2} , \quad (7.14)$$

$$I_3 = (1+BA)(1+2BA)^{-3/2} . \quad (7.15)$$

The self-consistent equation for the determination of A (or S) is

$$A = \exp[-S(1+2BA)^{-3/2}] . \quad (7.16)$$

Using (7.8), (7.13), and (7.14), we obtain

$$E_{\text{g.s.}}^{\text{lin}}/\hbar\omega_0 = -2S(1+2BA)^{-3/2}$$
$$+S(1+BA)(1+2BA)^{-1/2} - BA$$
$$= (1+3BA)\ln A - BA . \quad (7.17)$$

Expression (7.17) defines energy in terms of B and A. One can treat B as a parameter and study the dependence on A. On the other hand, A is just an auxiliary variable, which is related to the proper parameter of the system S by (7.16).

Before studying the energy, we must look at $A = A(S)$, which is represented for several values of B on Fig. 4. A varies in the range between 0 ($S \to \infty$) and 1 ($S=0$). For small B, their relation is unique, but for higher B, there appear three values of A for a given value of S. There exists a critical value $B_c = e^{5/3}/3 = 1.764\,83$, where the behavior of adiabates changes. If one looks at $S = S(A)$, one finds two minima, one for $A \cong 0$ and the other one for $A \cong 1$ and a maximum between them. We have also plotted the stability curve which is the locus of the points where the first derivative of $S(A)$ vanishes because that corresponds to the set of points where the second derivative of $E_{\text{g.s.}}$ vanishes, so that the points below the curve correspond to stable states (minima) while points above the curve correspond to the maxima of energy (unstable states).

For small B, we can expand the exponent

$$A \cong e^{-S}(1+3Be^{-S}) \quad (7.18)$$

giving

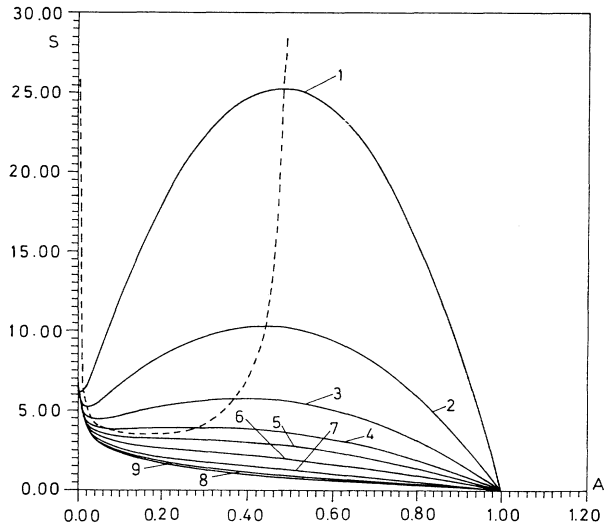

FIG. 4. Adiabates $S=S(A)$ corresponding to the following values of B: 1, 10; 2, 5; 3, 3; 4, 2; 5, 1.5; 6, 1.0; 7, 0.5; 8, 0.2; 9, 0.01; — — —, stability curve.

$$E_{\text{g.s.}}^{\text{lin}}/\hbar\omega_0=-S-Be^{-S}-3SB^2e^{-2}/2 \ . \tag{7.19}$$

For the sake of comparison, let us look at our expression for energy (4.1) with the nonlinear term neglected. Expanding the self-consistent equation for δ (4.3) in terms of small B, one obtains

$$E_{\text{g.s.}}(\delta)/\hbar\omega_0=-S-Be^{-S}-SB^2e^{-S} \ . \tag{7.20}$$

The relative difference is

$$[E_{\text{g.s.}}^{\text{lin}}-E_{\text{g.s.}}(\delta)]/E_{\text{g.s.}}\cong-\tfrac{1}{2}B^2e^{-2S} \ . \tag{7.21}$$

The highest value is obtained for $e^{-2S}\cong 1$ and $B_{\max}\cong\frac{1}{2}$ $(4J/\hbar\omega_0=1)$ giving 12.5% better value for the linear theory.

For intermediate and large B, we have to apply a different treatment. We expand (7.13)–(7.15) in terms of $S(1+2B)^{-3/2}$ as a small parameter, while in (4.1) we cannot neglect the nonlinear term and we use $\delta\cong(1+B)^{-1}$. In that case, one can look for a crossing point where two approaches lead to the same result. In this way, we obtain the following value for the crossing value S_0, knowing that for $S>S_0$ the results of the δ approach give a lower energy:

$$B=10 \ , \quad S_0=10.25 \ , \quad B=8 \ , \quad S_0=9.1 \ ,$$

$$B=5 \ , \quad S_0=7.1 \ , \quad B=4 \ , \quad S_0=6.6 \ .$$

For the symmetric ADP model, one can repeat the whole procedure or use directly the results of Venzl and Fischer[30] (there occur some minor numerical discrepancies). The qualitative behavior is the same. So, for example, for B small we find for both types of ADP couplings a maximal relative difference of ground-state energies of order up to 7%.

In the region of intermediate B, we have found the following crossing points:

(a) ADP symmetric:

$$B=10 \ , \quad S_0=1.477 \ , \quad B=9 \ , \quad S_0=1.4278 \ ,$$

$$B=5 \ , \quad S_0=1.1742 \ , \quad B=4 \ , \quad S_0=0.936 \ ,$$

(b) ADP antisymmetric:

$$B=10 \ , \quad S_0=1.75 \ , \quad B=8 \ , \quad S_0=1.275 \ ,$$

$$B=5 \ , \quad S_0=0.1445 \ ,$$

For a highly adiabatic region $(B\gg 1)$, we obtain the following analytic values for crossing points:

(a) ADP symmetric:

$$S_c=16/\pi^2\ln(3\pi B/2)-24/\pi^2-\tfrac{3}{2} \ ,$$

(b) ADP antisymmetric:

$$S_c=96\pi^2\ln(\pi B/2)-48/\pi^2-6 \ .$$

With proper redefinition of the parameters, these equations agree with the ones obtained by Venzl and Fischer,[30] while the last term(s) represent the correction due to the δ approach.

The reason that we have presented here a rather detailed analysis is because we find it surprising that no one has previously performed it. One can conclude that there exists major similarity in the behavior of $E_{\text{g.s.}}$ in both approaches, but with an essential difference. Let us be more precise. TI theories predict that for $B<B_c$ with increasing S there occurs the continual transition of partially dressed polaron band states towards ST (immobile) small polarons of infinite mass. For $B>B_c$ for each adiabate there exists a metastable region bounded by two critical coupling constants where the coexistence is predicted and there is an abrupt transition from the free band (partially dressed) state into ST state. This change is expressed more in the case of large B values (highly adiabatic region) where a slightly dressed, practically free band state transits into a ST state. Obviously, there exists a large analogy with the predictions of the δ approach, but there, instead of partially dressed band states, we obtain corresponding partially dressed solitonic solutions.

Comparing $E_{\text{g.s.}}$ for both approaches, it follows that in the region of small B $(B\ll 1)$, there is no major difference since self-trapping occurs rather soon. For intermediate and high B values, for each adiabate there exists a crossing value S_0, such that for $S>S_0$, the δ approach gives better results, while TI theories lead to lower energy for $S<S_0$. Yet, since even in the strongly linear regime $E_{\text{g.s.}}(\delta)$ is only a few percent higher, we are not sure about the justification of the application of a linear picture. We feel that if we improve the δ approach by not specifying equal dressing, one might expect the improvement of the NL result with respect to the TI approach in the spirit of the Buimistrov-Pekar theory[28] (where even translational invariance can be incorporated), such that there will be no need for linear theory any more. Such a strong statement is based on the results of Venzl and Fischer[30] who repeated Buimistrov and Pekar's[28] calculations for the MCM and demonstrated

that both approaches converge to the same result for large B and strong coupling.

From the point of view of qualitative behavior, our results allow us to look for the solution of a particular problem: the applicability of the soliton concept, i.e., Davidov-soliton theory for the explanation of the mechanism of the charge and energy transport in biological systems (α-helix, for example).

Using often cited values of the parameters for α-helix or ACN, we can estimate the relevance of the soliton model. We shall use here the values summarized in Ref. 25, remembering that they refer to the problem of the transport of the energy of amide-I oscillations through the resonant coupling of adjacent peptide groups. We must calculate the values of relevant parameters and then test the stability of ST states by checking if the condition $3\delta^2 S e^{-\delta^2 S} \ll 1$ is satisfied.

Under these circumstances, band theory breaks down and polaron transport should have activation character by random hopping between neighboring sites rather than being achieved *via* band mechanism. This practically means that the initiation of the polaron motion (hopping) in realistic conditions needs certain activation energy which is temperature dependent. According to the early papers of Holstein[10] and others,[26,33,34] hopping arises when temperature increases over some critical value T_c which depends on system parameters. Unfortunately, we cannot estimate T_c within the framework of the present temperature-independent variational approach. Details concerning the validity of band theory for small polaron transport with respect to the activation (hopping) mechanism can be found in the above-mentioned papers of Holstein[10] and in numerous review articles dealing with polaron dynamics.[33,34]

In the context of biological applications, the main problem is the smallness of the adiabaticity parameter. For α-helix $J=1.55\times10^{-22}$ J and $\hbar\omega_B=(18-21)\times10^{-22}$ J (independent of the factor 3 especially discussed in Ref. 25) we obtain $B=0.14-0.16$. This adiabate lies very close to the $\delta=1$ axis (0.84–0.88). For a given value of coupling $\chi=35-62$ pN, we obtain $E_B\cong10^{-22}-10^{-23}$ J, so $S\cong0.01-0.1$. (The difference of ground-state energies calculated by linear theories and the δ approach is less than 1%.) On the basis of presented estimates, it seems that we are dealing with Bloch band states which would rapidly turn into ST states with increasing S. This effect is enhanced by the temperature increase.[10,33,34]

On the other hand, for ACN $J=4$ cm^{-1} and $\hbar\omega_0$ to 75 cm^{-1} giving $B\cong0.1$ leading to $\delta>0.92$, which should correspond to strong localization, so we conclude that the energy transport due to amide I in ACN should have predominantly activation character.

On the contrary, it seems that α-helix supports the existence of DS consisting of an electron and the lattice, since model-independent estimates[35–37] for electron hopping terms give $J\cong1$ eV $\cong10^{-19}$ J. In this case obviously $B\gg1$, so the theoretical description based on DS theory (factorization of wave function) seems to be justified.

Let us finally summarize the features of our approach. Its main advantages are high simplicity (single variational parameter) and the ability to give good qualitative estimates and reproduce most of the previous theories. The main disadvantage is its "insensitivity" to the type of coupling because it treats various models in mostly the same way. Yet we think that all the above-listed features qualify this approach to be a good candidate for the extension to finite temperatures and anisotropic problems (where the electron and phonon subsystems have different dimensionalities).

We must mention that this study is inspired by, and in many ways complementary to, the one presented in Refs. 21 and 29. The general approach is the same, but, of course, there are some essential differences. Previous authors take as the most important result for the comparison the results obtained by the linearization. We avoid linearization and study the constraints which appear completely analytically and not as the result of numerical analysis. (We have reason to believe that some of our boundaries would show up in the analysis of previous papers if the space of parameters was larger.) The appearance of the negative branch is essential and it is important to prove that these solutions are spurious in the present context. On the other hand, relaxing the condition of the strict validity of the continuum approximation would lead to the appearance of excitations belonging to the negative branch. Also, the spectrum of detected excitations is much broader. Our approach pointed to the importance of nonlinearity parameter a, a fact not expected from previous calculations because they were performed on a qualitative level giving good behavior for $a=1$.

ACKNOWLEDGMENT

This work was supported in part by the Serbian Scientific Foundation under the project Physics of Condensed Matter and New Materials, Grant No. 3 and No. 27.

APPENDIX

Here we present the detailed discussion of the limitations which follow from the mathematical and physical demands imposed on the solution of our basic equation (4.3).

(a) Reality of δ demands the reality of the square root K (4.4), i.e., positive definiteness of the algebraic expression $f(y)=2/3\,a^2y^2-4/3\,a^2y+1$. $f(y)$ treated as a quadratic form is positive always for $a^2<\frac{3}{2}$. There are no restrictions on the values for δ and S then, arising from this demand. However, when the nonlinearity parameter a^2 exceeds the value $\frac{3}{2}$, $f(y)$ is positive only in the regions $0<y<y_1$ and $y>y_2$ where y_1 (>0) and y_2 ($>y_1$) are zeros of $f(y)$, $y_{1/2}=1\mp[1-3/(2a^2)]^{1/2}$. This condition introduces restrictions on the allowed area in the $S-\delta$ plane. We must exclude the part lying between the curves:

$$S=y_1/[\delta(1-\delta)]\,,\quad S=y_2/[\delta(1-\delta)]\,. \tag{A1}$$

(b) Condition $\delta>0$ which immediately implies $1\pm K>0$. This demand, in fact, concerns only the negative sign (since $K>0$). So, for the ($-$) branch, we have

$K<1$ which results in the condition

$$y<2 \ . \tag{A2}$$

(c) Continuum approximation is valid if the soliton occupies a large number of lattice sites. This demands $\mu \ll 1$. Therefore, $\mu=1$ is the final boundary for the applicability of the continuum approximation.

Combining (4.3) and (3.22) (with $v=0$), we have

$$\mu=\frac{2aS\delta(1-\delta)}{1\pm K}=\frac{2ay}{1\pm K} \ . \tag{A3}$$

The validity of continuum approximation demands

$$2ay-1<\pm K \ , \tag{A4}$$

which must be analyzed for each sign separately. For the $(+)$ branch, we have the following situation: if $2ay-1>0$, then (A4) can be squared and we arrive at the following condition:

$$y<(6/a-2)/5 \ , \tag{A5}$$

which is meaningful if $3/a-1>0$, i.e., for

$$a<3 \ . \tag{A6}$$

As for the $(-)$ branch, $2ay-1<-K$ cannot be satisfied unless

$$y<1/(2a) \ . \tag{A7}$$

In case it is satisfied, we have $-(1-2ay)<-K$, which gives

$$1-2ay>K \ , \tag{A8}$$

leading to $y>(6/a-2)/5$. As before, positivity of y gives $a<3$. Therefore, besides this one, necessary, but not sufficient, conditions for the existence of polaron solutions defined by the $(-)$ branch are $y<2$ [see (4.8)] and the stronger condition $y<1/(2a)$.

Therefore, physically meaningful solutions should be looked for in the region between the curves:

$$S=(1/2a)/[\delta(1-\delta)]$$

and

$$(6/a-2)/[5\delta(1-\delta)] \ . \tag{A9}$$

These two curves bound a region in the $S-\delta$ plane where the $(-)$ branch satisfies the conditions for the validity of the continuum limit. However, it is possible that the inequalities (A7)–(A9) are not satisfied anywhere. It is a simple algebraic problem to show that inequalities can have simultaneous solutions only for $a>\frac{7}{4}$, so the proper set of values is $\frac{7}{4}<a<3$.

[1]L. D. Landau, Phys. Z. Sowjet Union **3**, 64 (1933); S. I. Pekar, Zh. Eksp. Theor. Phys. **16**, 335 (1946); L. D. Landau and S. I. Pekar, *ibid.* **18**, 419 (1948).

[2]H. Fröhlich, H. Pelzer, and S. Zienau, Philos. Mag. **41**, 221 (1950); H. Fröhlich, Adv. Phys. **3**, 325 (1954).

[3]For a review, see E. I. Rashba, in *Excitons,* edited by E. I. Rashba and M. D. Sturge (North-Holland, Amsterdam, 1982); M. Ueta, H. Kanzaki, K. Kobayashi, Y. Toyozawa, and E. Hanamura, in *Excitonic Processes in Solids,* edited by M. Cardona, P. Fulde, K. von Klitzing, and H.-J. Quisser, Springer Series in Solid-State Sciences Vol. 60 (Springer-Verlag, Berlin, 1986).

[4]B. Luschik, in *Excitons,* edited by E. I. Rashba and M. D. Sturge (North-Holland, Amsterdam, 1982).

[5]R. E. Packard, F. Reif, and C. M. Surko, Phys. Rev. Lett. **25**, 1435 (1970); Ya. I. Fugol, Adv. Phys. **27**, 1 (1978).

[6]K. W. Kehr, in *Hydrogen in Metals I,* edited by. G. Alefeld and J. Völkl, Topics in Applied Physics Vol. 28 (Springer-Verlag, Berlin, 1978); M. I. Klinger, Phys. Rep. **94**, 183 (1983).

[7]A. S. Davydov, Phys. Scr. **20**, 387 (1979).

[8]S. Englander, N. Kallenbach, A. Heeger, J. Krumhansl, and S. Litwin, Proc. Natl. Acad. Sci. (USA) **77**, 7222 (1980).

[9]P. G. de Gennes, Phys. Rev. **118**, 141 (1960); L. Nagaev, Usp. Fiz. Nauk **117**, 437 (1975) [Sov. Phys.Usp. **18**, 863 (1975)].

[10]T. Holstein, Ann. Phys. (N.Y.) **8**, 326 (1959); **8**, 343 (1959).

[11]S. V. Tyablikov, Zh. Eksp. Teor. Fiz. **23**, 381 (1952).

[12]S. A. Brazovskii, Pis'ma Zh. Eksp. Teor. Fiz. **28**, 656 (1978) [JETP Lett. **28**, 606 (1978)]; W. P Su, J. R. Schrieffer, and A. S. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).

[13]D. Emin and T. Holstein, Phys. Rev. Lett. **36**, 323 (1976).

[14]D. Emin, Adv. Phys. **22**, 57 (1973); Phys. Rev. Lett. **28**, 604 (1972).

[15]H. Schüttler and T. Holstein, Ann. Phys. (N.Y.) **166**, 93 (1986).

[16]G. Carreri, U. Buontempo, F. Carta, E. Gratton, and E. Shyansunder, Phys. Rev. Lett. **51**, 304 (1983).

[17]D. W. Brown, K. Lindenberg, and B. J. West, Phys. Rev. A **33**, 4104 (1988).

[18]M. Škrinjar, D. Kapor, and S. Stojanović, Phys. Rev. A **38**, 6402 (1988).

[19]Q. Zhang, V. Romero-Rochin, and R. Silbey, Phys. Rev. A **38**, 6409 (1988).

[20]Z. Ivić and D. W. Brown, Phys. Rev. Lett. **63**, 426 (1989).

[21]D. W. Brown and Z. Ivić, Phys. Rev. B **40**, 9876 (1989).

[22]M. Satarić, Z. Ivić, Z. Shemsedini, and R. Žakula, J. Mol. Electron. **4**, 223 (1988).

[23]A. C. Scott, I. J. Bigio, and C. T. Johnston, Phys. Rev. **39**, 12 833 (1988).

[24]D. M. Alexander and J. A. Krumhansl, Phys. Rev. B **33**, 7172 (1986).

[25]A. Scott, Phys. Rep. **217**, 1 (1992).

[26]I. G. Lang and Yu. A. Firsov, Zh. Eksp. Teor. Fiz. **43**, 1832 (1962) [Sov. Phys. JETP **16**, 1293 (1962)].

[27]G. B. Norris and G. Whitfield, Phys. Rev. B **21**, 3658 (1980).

[28]V. M. Buimistrov and S. I. Pekar, Zh. Eksp. Teor. Fiz. **32**, 1193 (1957) [Sov. Phys. JETP **5**, 970 (1957)]; **33**, 1271 (1957) [**6**, 977 (1958)].

[29]D. W. Brown, K. Lindenberg, and X. Wang, in *Davydov's Soliton Revisited,* edited by P. Christiansen and A. C. Scott (Plenum, New York, 1990).

[30]G. Venzl and S. Fischer, J. Phys. Chem. **81**, 6090 (1984); Phys. Rev. B **32**, 6437 (1985).

[31]E. I. Rashba, Opt. Spektr. **2**, 88 (1957).

[32]M. Wagner and A. Köngeter, J. Chem. Phys. **91**, 3036 (1989).

[33]J. Appel, in *Solid State Physics,* edited by H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1968), Vol. 21, p. 193.

[34]Yu. A. Firsov, in *Polarons,* edited by Yu. A. Firsov (Nauka, Moscow, 1975).

[35]E. Turner and V. E. Alexander, Phys. Rev. **174**, 81 (1968).

[36]E. G. Petrov, *Physics of Charge Transfer in Biosystems* (Naukova Dumka, Kiev, 1984), Chap. 3, p. 168.

[37]D. K. Campbel, A. R. Bishop, and K. Fesser, Phys. Rev. B **26**, 6862 (1982).