Unification of polaron and soliton theories of exciton transport

David W. Brown and Zoran Ivić*
Institute for Nonlinear Science, R-002, University of California, San Diego, La Jolla, California 92093
(Received 9 June 1989)

We present a unified theory of polaron and soliton dynamics by combining time-dependent variational methods recently applied to the theory of Davydov solitons with partial-dressing methods well known from polaron theory. We focus on the simplest partial-dressing assumption, applying a common dressing fraction to all phonon modes. Our fundamental result is a system of nonlinear evolution equations in which the tendency of a system to form Davydov solitons is balanced against its tendency to form small polarons. We subsequently apply time-independent variational methods to determine the optimal dressing fraction in a mean-field manner. The characterization of the partially dressed soliton states that results is complete with respect to the system parameter space. Consistent with prior works from polaron theory, we find a self-trapping transition that is only weakly modified by our inclusion of nonlinearities, and we reinterpret this transition in terms of our newly obtained soliton states. Applying our results to a central problem in bioenergetics, we obtain results markedly different from well-known results of Davydov's theory.

I. INTRODUCTION

The presence of foreign particles or excitations (electronic or vibronic) in a deformable solid often causes local distortions of the host material which can profoundly affect the character and dynamics of would-be quasiparticles. Among the physical systems which present such challenging problems are polar crystals, where the problem was first studied;[1,2] metal hydrides, wherein interstitial hydrogen isotopes can cause large volume dilations;[3,4] organic molecular crystals such as anthracene and naphthalene, wherein the mobilities of photoinjected charge carriers exhibit novel temperature dependences;[5,6] organic molecular crystals such as pyrene and α-perylene, wherein exciton spectra are profoundly affected by local distortions;[7,8] and biological molecules such as DNA (Ref. 9) and the α-helix (Refs. 10 and 11) and biological materials such as the molecular crystals acetanilide (Refs. 12 and 13) and l-alanine (Ref. 14), wherein a number of phenomena have been related to conformational excitations.

The body of theory that has evolved to describe energy transport in deformable media necessarily embraces a number of different points of view, some of which appear, at different times and in different ways, to be in conflict. Central to most of the relevant literature is concept of the polaron. When a particle such as an exciton is created in or injected into a solid, the presence of the exciton induces a distortion in the surrounding medium. "Polaron" is generally given to mean the quasiparticle consisting of the original exciton together with the distortion it induces in the host medium. While conceptually simple, this definition is operationally inadequate since it contains no prescription by which to identify the relevant distortion. As immediate refinements there are the more useful concepts of the small polaron and the large polaron. "Small polaron" is generally given to mean a polaron which occupies a minimum number of host lattice sites, usually one. "Large polaron" is generally given to mean a polaron which occupies a number of host lattice sites well in excess of this minimum, usually a number large enough to justify a continuum approximation. Small-polaron theories are usually addressed to the nonadiabatic situation in which the exciton energy bandwidth is small relative to the typical phonon frequency;[1-6,15,16] such treatments are largely based on the use of canonical transformations to identify sets of "small-polaron states" in which perturbation theory is often developed in orders of the ratio of the exciton bandwidth to the typical phonon frequency. On the other hand, large-polaron theories are usually addressed to the adiabatic limit in which the kinetic energy of lattice vibrations can be neglected in lowest order;[5,6,17-22] such treatments are frequently based on the application of variational methods. Small-polaron theory is, historically, a linear theory, while the theory of large polarons contains well-known nonlinearities. The nonlinearity implicit in the description of large polarons has given rise to transport theories based on the somewhat more specialized concept of envelope solitons. Roughly speaking, the greater part of perturbation theory is most appropriate in the small-polaron regime, where effective masses are large, and the greater part of soliton theory is most appropriate in the large-polaron regime, where effective masses are small.

We put forward a theory based on recent advances in the theory of envelope solitons which allows a number of these sometimes conflicting perspectives to be unified in a single formulation. The theory we present covers the entire parameter range from the small-polaron regime to the large polaron regime, and recovers in appropriate limits small-polaron theory and Davydov's theory of envelope solitons. In Sec. II we present the elements of the model on which our theory is based. In Sec. III we apply time-dependent variational methods to obtain equations of motion and special solutions in the small- and large-

polaron regimes. In Sec. IV we apply time-independent variational methods in order to further optimize the solutions obtained in Sec. III. In this context it becomes necessary to address the notion of self-trapping and the concept of a self-trapping transition. The existence or nonexistence of a self-trapping transition is an issue widely discussed in the literature, with conclusions both affirming and denying the existence of a transition. Our findings show that continuous changes in system parameters can produce discontinuous changes in state character consistent with the existence of a self-trapping transition. We argue the consistency of our conclusions with both the affirmative and negative conclusions drawn by others.

II. THE MODEL

Investigations of quasiparticle transport in deformable media traditionally centers on systems modeled by the Fröhlich Hamiltonian[2]

$$H=\sum_m E_m a_m^\dagger a_m-\sum_{m,n} J_{mn} a_m^\dagger a_n+\sum_q \hbar\omega_q b_q^\dagger b_q + \sum_{q,n} \hbar\omega_q(\chi_n^q b_q^\dagger+\chi_n^{q\,*} b_q) a_n^\dagger a_n \ . \tag{2.1}$$

(Notation here and throughout this paper is that of Ref. 22.) In this general form any coupling geometry and any bare band structure can be accommodated. While most of our results can be obtained in this general case, most of our specific results are obtained for the translationally invariant acoustic chain model based on the Hamiltonian

$$H=\sum_m E a_m^\dagger a_m-\sum_n J a_n^\dagger(a_{n+1}+a_{n-1}) + \sum_n \left[\frac{\hat{P}_n^2}{2M}+\frac{w}{2}(\hat{Q}_n-\hat{Q}_{n-1})^2\right] + \sum_n \chi(\hat{Q}_{n+1}-\hat{Q}_{n-1}) a_n^\dagger a_n \ . \tag{2.2}$$

This latter form of H results from choosing the phonon dispersion relation ω_q and the dimensionless coupling function χ_n^q to have the forms

$$\omega_q=\omega_B \sin(\tfrac{1}{2}|ql|) \ , \tag{2.3}$$

$$\chi_n^q=\chi^q e^{-iqR_n}=\frac{2i\chi\sin(ql)}{(2NM\hbar\omega_q^3)^{1/2}} e^{-iqR_n} \ . \tag{2.4}$$

Exact solutions of the Fröhlich Hamiltonian are known only in the limit of infinite effective mass, wherein the transfer integrals J_{mn} between different sites in the medium are all zero. In studying the general transport problem, therefore, one is necessarily dealing with approximate approaches. Davydov has built a theory of envelope solitons for such systems based on the two trial states[20,21]

$$|D_1(t)\rangle\equiv\sum_n \alpha_n(t) a_n^\dagger|0\rangle\otimes|\beta_n(t)\rangle \ , \tag{2.5}$$

$$|\beta_n(t)\rangle\equiv\exp\left[\sum_q [\beta_{qn}(t) b_q^\dagger-\beta_{qn}^*(t) b_q]\right]|0\rangle \ , \tag{2.6}$$

and

$$|D_2(t)\rangle\equiv\sum_n \alpha_n(t) a_n^\dagger|0\rangle\otimes|\beta(t)\rangle \ , \tag{2.7}$$

$$|\beta(t)\rangle\equiv\exp\left[\sum_q [\beta_q(t) b_q^\dagger-\beta_q^*(t) b_q]\right]|0\rangle \ . \tag{2.8}$$

The D_1 ansatz differs from the more widely used D_2 ansatz in that the coherent state amplitudes in the latter case are site independent [i.e., $\beta_{qn}(t)=\beta_q(t)$ for all n]. The D_2 ansatz represents the quantum state of each vibrational normal mode by a single coherent state, and hence has at all times a strongly classical character. The failure of Davydov's theory based on the D_2 ansatz state to recover the exact small-polaron states in the infinite effective-mass limit points to the fact that this strong classical character is too restrictive an assumption in the small-polaron regime, and makes any D_2 formulation inappropriate as a general theory. On the other hand, the D_1 ansatz represents the quantum state of each vibrational normal mode by a dynamically phased superposition of many coherent states (one per site) allowing the quantum nature of the phonon system to play a greater role. Davydov's original treatment of the D_1 ansatz state considered the $\alpha_n(t)$ and $\beta_{qn}(t)$ to be dynamical variables to which Hamilton's equations of motion could be applied.[20] This has since been shown to require revision;[23] however, in the place of these Hamilton equations there now exist optimized equations of motion derived for the D_1 state by time-dependent variational methods.[24,25]

The accuracy of any variational method is ultimately limited by one's choice of trial state. In this paper we consider as a trial function an ansatz-state vector intermediate between the D_1 and D_2 states. As we presently show, we do not *generalize* D_1; rather, we choose a subclass of D_1 states with useful properties. To define these states we denote with overtildes quantities defined in a basis of mixed exciton-phonon states, e.g.,[26]

$$\tilde{a}_m\equiv U a_m U^\dagger\equiv a_m\exp\left[\sum_q(\tilde{\beta}_{qm} b_q^\dagger-\tilde{\beta}_{qm}^* b_q)\right] \ , \tag{2.9}$$

$$\tilde{b}_q\equiv U b_q U^\dagger\equiv b_q+\sum_q \tilde{\beta}_{qm} a_m^\dagger a_m \ , \tag{2.10}$$

$$U\equiv\exp\left[-\sum_{q,m}(\tilde{\beta}_{qm} b_q^\dagger-\tilde{\beta}_{qm}^* b_q) a_m^\dagger a_m\right] \ . \tag{2.11}$$

(We note that the parameters of the transformation are time independent.) In terms of these operators we define the trial state vector

$$|\tilde{D}(t)\rangle\equiv\sum_n \tilde{\alpha}_n(t)\tilde{a}_n^\dagger \times\exp\left[\sum_q[\tilde{\beta}_q(t)\tilde{b}_q^\dagger-\tilde{\beta}_q^*(t)\tilde{b}_q]\right]|0\rangle \ , \tag{2.12}$$

wherein the vacuum state is the total vacuum annihilated by any destruction operator. Our motivation for making this definition is that substantial detailed analysis has shown that a degree of phase mixing between the exciton and phonon systems is essential to a faithful description of the quantum nature of the system. Our proposed states accommodate this need by incorporating a degree of phase mixing into dressed basis states. By transforming back into the basis of bare states, our $\tilde{D}$ states can be

put in the form of D_1 states with the identifications

$$\alpha_n(t)=\tilde{\alpha}_n(t)\exp\left[-\tfrac{1}{2}\sum_q[\tilde{\beta}^*_{qn}\tilde{\beta}_q(t)-\tilde{\beta}_{qn}\tilde{\beta}^*_q(t)]\right], \tag{2.13}$$

$$\beta_{qn}(t)=-\tilde{\beta}_{qn}+\tilde{\beta}_q(t)\ . \tag{2.14}$$

Like the general D_1 state, the $\tilde{D}$ states contain as special cases all D_2 states (set $\tilde{\beta}_{qn}=0$) and the exact $J=0$ small-polaron states [set $\tilde{\beta}_q(t)=0$, $\tilde{\beta}_{qn}=\chi^q_n$]. In this the $\tilde{D}$ states have the basic properties desired of a unifying theory of polaron and soliton dynamics. Of course, the Hamiltonian [(2.1) or (2.2)] is not diagonal in these states when $J_{mn}\neq 0$; however, the interaction between dressed excitons $(\tilde{a},\tilde{a}^\dagger)$ and dressed phonons $(\tilde{b},\tilde{b}^\dagger)$ is generally weaker than the interaction between bare excitons $(a,a^\dagger)$ and bare phonons $(b,b^\dagger)$. Because of this weaker interaction, a factored dressed state such as (2.12) affords a more plausible description of the microscopic dynamics than does a factored bare state such as (2.7) in those parameter regimes where phase mixing is important. Since we choose the phase-mixing part of $\beta_{qn}(t)$ to be time independent, we are not considering problems of polaron or soliton formation from bare exciton states; we consider, instead, preformed polarons or solitons. (The appropriate generalization is straightforward.)

III. TIME-DEPENDENT VARIATION

To determine the optimal equations of motion, we apply the time-dependent variational principle[27,28]

$$\delta\int_{t_1}^{t_2}dt\left\langle\psi(t)\left|i\hbar\frac{d}{dt}-H\right|\psi(t)\right\rangle=0\ , \tag{3.1}$$

recently used by Škrinjar *et al.*[24] and Zhang *et al.*[25] to obtain optimized D_1 equations. Clear measures of the accuracy of these optimized equations do not yet exist, though it is known from formal considerations that their solutions deviate from the true quantum solutions in the general case.[29] It is important, therefore, to recognize that as an approximation scheme (3.1) minimizes the deviation of trial solutions from exact, but unknown quantum-mechanical solutions at every instant of time. In applying these methods, one is foregoing the arbitrary short-time precision afforded by perturbation theory in favor of a (possibly) limited precision which is, however, maintained uniformly in time.

Since our proposed state is included in the class of D_1 states, our equations of motion could be obtained from the optimized D_1 equations of Refs. 24 and 25; however, there are sufficient differences between our proposed states and the more general D_1 states to make a direct application of (3.1) the more economical procedure. Thus, applying (3.1) we obtain the formal evolution equations

$$i\hbar\dot{\tilde{\alpha}}_n(t)+i\hbar\sum_q[\tilde{\beta}_{qn}\dot{\tilde{\beta}}^*_q(t)-\tilde{\beta}^*_{qn}\dot{\tilde{\beta}}_q(t)]\tilde{\alpha}_n(t)=\frac{\partial\langle H\rangle}{\partial\tilde{\alpha}^*_n(t)}\ , \tag{3.2}$$

$$i\hbar\dot{\tilde{\beta}}_q(t)-i\hbar\sum_n\tilde{\beta}_{qn}\frac{d}{dt}|\tilde{\alpha}_n(t)|^2=\frac{\partial\langle H\rangle}{\partial\tilde{\beta}^*_q(t)}\ , \tag{3.3}$$

in which the energy functional $\langle H\rangle$ is the expectation value of H in the $\tilde{D}$ state,

$$\langle H\rangle=\sum_m E_m|\tilde{\alpha}_m(t)|^2-\sum_{m,n}\tilde{J}_{mn}\tilde{\alpha}^*_m\tilde{\alpha}_n+\sum_q\hbar\omega_q|\tilde{\beta}_q(t)|^2-\sum_{q,m}\hbar\omega_q(\chi^q_m\tilde{\beta}^*_{qm}+\chi^{q*}_m\tilde{\beta}_{qm}-|\tilde{\beta}_{qm}|^2)|\tilde{\alpha}_m(t)|^2$$
$$+\sum_{q,m}\hbar\omega_q[(\chi^q_m-\tilde{\beta}_{qm})\tilde{\beta}_q(t)^*+(\chi^q_m-\tilde{\beta}_{qm})^*\tilde{\beta}_q(t)]|\tilde{\alpha}_m(t)|^2\ , \tag{3.4}$$

and

$$\tilde{J}_{mn}=J_{mn}\exp\left[-\tfrac{1}{2}\sum_q(|\tilde{\beta}_{qm}|^2+|\tilde{\beta}_{qn}|^2-2\tilde{\beta}^*_{qm}\tilde{\beta}_{qn})\right]. \tag{3.5}$$

Written in this way, the energy functional reflects a weakening of the exciton-phonon interaction and an increase in the exciton effective mass due to the partial dressing of the exciton state.

To obtain compact results, we immediately specialize to the acoustic chain model posed by the Hamiltonian (2.2). We further assume that the fraction by which each phonon mode dresses the exciton is the same for each mode, such that[30–32]

$$\tilde{\beta}_{qm}=\delta\chi^q_m\ . \tag{3.6}$$

This simplification significantly streamlines the calculations and allows the energy functional $\langle H\rangle$ to be expressed in the more convenient form

$$\langle H\rangle=\sum_m[E-\delta(2-\delta)E_b]|\tilde{\alpha}_m(t)|^2-\sum_m\tilde{J}\tilde{\alpha}^*_m(t)[\tilde{\alpha}_{m+1}(t)+\tilde{\alpha}_{m-1}(t)]+\sum_q\hbar\omega_q|\tilde{\beta}_q(t)|^2$$
$$+(1-\delta)\sum_{q,m}\hbar\omega_q[\chi^q_m\tilde{\beta}_q(t)^*+\chi^{q*}_m\tilde{\beta}_q(t)]|\tilde{\alpha}_m(t)|^2\ , \tag{3.7}$$

in which E_b is the small-polaron binding energy

$$E_b=\sum_q |\chi^q|^2\hbar\omega_q \ , \tag{3.8}$$

and $\tilde{J}$ is the renormalized matrix element *partially* reduced from the bare value J by interactions with phonons

$$\tilde{J}=Je^{-\delta^2 s} \tag{3.9}$$

$$S=4\sum_q |\chi^q|^2\sin^2(\tfrac{1}{2}ql) \ . \tag{3.10}$$

[It is important to note that the small-polaron binding energy (3.8) and the soliton binding energy from Davydov's theory differ significantly in origin and value; for the acoustic chain model we consider the latter is given in terms of the former as $E_b^2/3J$.] Integrating (3.3), we can obtain an integro-differential equation for the dressed-exciton probability amplitudes in the form

$$i\hbar\dot{\tilde{\alpha}}_n(t)=[E-\delta(2-\delta)E_b]\tilde{\alpha}_n(t)-\tilde{J}[\tilde{\alpha}_{n+1}(t)+\tilde{\alpha}_{n-1}(t)]+\sum_q\hbar\omega_q[\chi_n^q\tilde{\beta}_q^*(t)+\chi_n^{q*}\tilde{\beta}_q(t)]\tilde{\alpha}_n(t)$$
$$+\delta(1-\delta)\sum_{qm}\hbar\omega_q(\chi_n^q\chi_m^{q*}+\chi_n^{q*}\chi_m^q)|\tilde{\alpha}_m(t)|^2\tilde{\alpha}_n(t) \ , \tag{3.11}$$

$$\tilde{\beta}_q(t)=\tilde{\beta}_q(0)e^{-i\omega_q t}-i\omega_q(1-\delta)\int_0^t d\tau\, e^{-i\omega_q(t-\tau)}\sum_m\chi_m^q|\tilde{\alpha}_m(\tau)|^2+\delta\int_0^t d\tau\, e^{-i\omega_q(t-\tau)}\sum_m\chi_m^q\frac{d}{d\tau}|\tilde{\alpha}_m(\tau)|^2 \ . \tag{3.12}$$

Using (3.12) to eliminate the explicit phonon variables $\tilde{\beta}_q(t)$, we find

$$i\hbar\dot{\tilde{\alpha}}_n(t)=[E-\delta(2-\delta)E_b]\tilde{\alpha}_n(t)-\tilde{J}[\tilde{\alpha}_{n+1}(t)+\tilde{\alpha}_{n-1}(t)]$$
$$-(1-\delta)^2\sum_m K_{mn}(0)|\tilde{\alpha}_m(t)|^2\tilde{\alpha}_n(t)+(1-\delta)\sum_m K_{mn}(t)|\tilde{\alpha}_m(0)|^2\tilde{\alpha}_n(t)$$
$$+\int_0^t d\tau\sum_m K_{mn}(t-\tau)\frac{d}{d\tau}|\tilde{\alpha}_m(\tau)|^2\tilde{\alpha}_n(t)+\tilde{f}_n(t)\tilde{\alpha}_n(t) \ , \tag{3.13}$$

where

$$K_{mn}(t)=2\sum_q\chi_m^q\chi_n^{q*}\hbar\omega_q\cos\omega_q t \ , \tag{3.14}$$

$$\tilde{f}_n(t)=\sum_q\hbar\omega_q[\chi_n^q\tilde{\beta}_q^*(0)e^{i\omega_q t}+\chi_n^{q*}\tilde{\beta}_q(0)e^{-i\omega_q t}] \ . \tag{3.15}$$

This is the fundamental result of our time-dependent analysis, which, subject only to the simplifying restriction (3.6), is the general and exact consequence of applying the variational principle (3.1) to the trial state (2.12). This equation reflects a dynamic balancing of the system's soliton and small-polaron characteristics. This balance is perhaps most clearly reflected in the sum rule

$$\delta(2-\delta)+(1-\delta)^2=1 \ , \tag{3.16}$$

which holds between the coefficients of the small-polaron binding energy E_b and the standard cubic nonlinearity.

If we specialize to the *large-polaron limit* by setting $\delta=0$, the small-polaron binding energy E_b drops out of the equation and $\tilde{J}$ reverts to the bare value J, exactly recovering Davydov's D_2 theory in the form given by Wang *et al.*[22] This was to be expected, of course, since our $\tilde{D}$ ansatz state reduces to Davydov's D_2 ansatz state in this limit; however, if we specialize to the *small-polaron limit* by setting $\delta=1$, we obtain an altogether new equation of motion,[33]

$$i\hbar\dot{\tilde{\alpha}}_n(t)=(E-E_b)\tilde{\alpha}_n(t)-\tilde{J}[\tilde{\alpha}_{n+1}(t)+\tilde{\alpha}_{n-1}(t)]$$
$$+\int_0^t d\tau\sum_m K_{mn}(t-\tau)\frac{d}{d\tau}|\tilde{\alpha}_m(\tau)|^2\tilde{\alpha}_n(t)$$
$$+\tilde{f}_n(t)\tilde{\alpha}_n(t) \ , \tag{3.17}$$

in which $\tilde{J}$ assumes the fully reduced small-polaron value. This nonlinear evolution equation has a number of interesting properties. If we consider the special solution

$$\tilde{\alpha}_n(t)=\tilde{u}(k)e^{i\{kR_n-[\tilde{E}(k)/\hbar]t\}} \ , \tag{3.18}$$

$$\tilde{E}(k)=(E-E_b)-2Je^{-S}\cos(kl) \ , \tag{3.19}$$

with $\tilde{\beta}_q(0)=0$ (zero temperature), the equations of motion (3.17) automatically linearize, and we find that the Bloch states of the small-polaron energy band are exact stationary solutions, confirming the agreement of our results with small-polaron theory. On the other hand, it is clear from the form of (3.17) that *nonstationary* solutions of (3.17) must have nonlinear character, implying that polaron wave packets cannot propagate in the simple dispersive manner characteristic of energy band theory *even in the small-polaron limit.*

Passing to the continuum limit, we find the general continuum equation of motion resulting from (3.13) to be

$$i\hbar\dot{\tilde{\alpha}}(x,t)=-\frac{\hbar^2}{2\tilde{m}}\frac{\partial^2}{\partial x^2}\tilde{\alpha}(x,t)+\tilde{E}(0)\tilde{\alpha}(x,t)+\tilde{f}(x,t)\tilde{\alpha}(x,t)+\tfrac{1}{2}\tilde{G}(0)\int_0^t d\tau\frac{d}{d\tau}|\tilde{\alpha}(y,\tau)|^2\Big|_{y=x+v_a(t-\tau)}\tilde{\alpha}(x,t)$$
$$+\tfrac{1}{2}\tilde{G}(0)\int_0^t d\tau\frac{d}{d\tau}|\tilde{\alpha}(y,\tau)|^2\Big|_{y=x-v_a(t-\tau)}\tilde{\alpha}(x,t)\ , \tag{3.20}$$

in which $\tilde{m}_\delta$ is the linear effective mass,

$$\tilde{m}=\frac{\hbar^2}{2Jl^2}e^{\delta^2 S}\ , \tag{3.21}$$

and $\tilde{E}(0)$ the bottom of the linear energy band,

$$\tilde{E}(0)=E-\delta(2-\delta)E_b-2Je^{-\delta^2 S}\ . \tag{3.22}$$

$\tilde{G}(v)$ plays the same role as the velocity-dependent nonlinearity parameter in Davydov's theory, but generally has a reduced value

$$\tilde{G}(v)=G(v)-\delta(2-\delta)G(0)$$
$$=\frac{4\chi^2 l}{w}\left[(1-\delta)^2+\frac{v^2}{v_a^2-v^2}\right]\ . \tag{3.23}$$

Seeking special solutions having the D'Alembert property $|\tilde{\alpha}(y,\tau)|=|\tilde{\alpha}(y-v\tau)|$, and making a special choice of $\tilde{\beta}_q(0)$ (in accordance with the corresponding discussion in Ref. 22, we find a nonlinear Schrödinger equation,

$$i\hbar\dot{\tilde{\alpha}}(x,t)=-\frac{\hbar^2}{2\tilde{m}}\frac{\partial^2}{\partial x^2}\tilde{\alpha}(x,t)+\tilde{E}(0)\tilde{\alpha}(x,t)$$
$$-\tilde{G}(v)|\tilde{\alpha}(x,t)|^2\tilde{\alpha}(x,t)\ . \tag{3.24}$$

This equation has soliton solutions of the form

$$\tilde{\alpha}(x,t)=\left[\frac{\kappa}{2}\right]^{1/2}\frac{e^{-i[\tilde{E}(0)/\hbar]t}e^{i(kx-\omega t)}}{\cosh[\kappa(x-vt)]}\ , \tag{3.25}$$

where

$$\hbar k\equiv\tilde{m}v\ , \tag{3.26}$$

$$\hbar\kappa\equiv\frac{\tilde{m}\tilde{G}(v)}{2\hbar}\ , \tag{3.27}$$

$$\hbar\omega\equiv-\frac{\tilde{m}\tilde{G}(v)^2}{8\hbar^2}+\tfrac{1}{2}\tilde{m}_\delta v^2 \tag{3.28}$$

In keeping with the distinctions which can be drawn between the small- and large-polaron regimes, this nonlinear Schrödinger equation and its soliton solution in the small-polaron limit ($\delta=1$) differ from corresponding results of Davydov's theory ($\delta=0$) in a number of ways. In the small-polaron limit the plane-wave solutions of (3.24) are not bare exciton Bloch states as in Davydov's theory, but instead are small-polaron Bloch states which already reflect an increased effective mass ($\tilde{m}>m$) and energy lowering [$\tilde{E}(0)<E(0)$] which results from the dressing of the bare exciton. In compensation for this renormalization of the exciton band structure, the nonlinear term in (3.24) is substantially weakened relative to the nonlinearity in Davydov's theory; indeed the linearity of the zero-velocity equation implies the *absence* of a static-soliton solution. Slow-soliton solutions of (3.24) exist for all values of δ; however, slow solitons in Davydov's theory have a limiting, finite size (given by the static-soliton solution), while slow-soliton solutions of (3.24) in the small-polaron limit broaden into plane waves (small-polaron Bloch states) as the group velocity goes to zero. Slow solitons in Davydov's theory are *lower* in energy than the plane-wave solutions of the same nonlinear Schrödinger equation (excitons), and are separated from these plane waves by a finite energy gap. On the other hand, slow-soliton solutions of (3.24) in the small-polaron limit are *higher* in energy than the plane-wave solutions of the same nonlinear Schrödinger equation (small polarons) and deform continuously into these plane waves with decreasing velocity without an energy gap. Finally, the soliton effective mass in Davydov's theory differs from (is greater than) the effective mass of the plane-wave solutions (excitons), while the soliton solutions of (3.24) in the small-polaron limit have the same effective mass as the plane-wave solutions (small polarons).

When δ takes on values intermediate between 0 and 1, state characteristics intermediate between those of the small polaron and Davydov solitons are found.

IV. OPTIMIZING THE DRESSING FRACTION

The time-dependent variation of the preceding section was carried out for a fixed, but arbitrary value of the dressing fraction δ which can be used to "tune" the equations of motion. For a given set of system parameters, there should be an optimal choice for δ corresponding to a minimum-energy state. For such a choice of δ the equations of motion (3.13) describe the evolution of the system near this minimal state. The optimal dressing fraction should vary with the temperature, but as a time- and space-independent quantity δ can depend only on average properties of the system. The dressing fraction is thus a mean-field quantity, quite different from other parameters appearing in (3.13); our equations of motion are not mean-field equations since no thermal averaging is invoked in their derivation and thermal fluctuations appear directly in the equations of motion. Estimations of the optimal dressing fraction based on mean-field energy-minimization arguments have been previously considered in the context of linear theory by Toyozawa,[30] Yarkony and Silbey,[31] and Venzl and Fischer,[34] and in the context of Davydov's theory by Venzl and Fischer,[34] Alexander and Krumhansl,[13] and Satarić *et al.*[32] Our present calculations generalize these analyses by considering the full nonlinear dependence of the system energy on exciton probability amplitudes, and by identifying the optimal states with solutions of the equations of motion (3.13).

Temperature enters our dynamical equations only through the fluctuations $\tilde{f}_n(t)$, in which the ambient tem-

perature is implicit in the phonon initial data $\{\tilde{\beta}_q(0)\}$. The phonon initial data generally decompose into incoherent $[\tilde{\beta}_q^{\rm th}(0)]$ and coherent $[\tilde{\beta}_q^{\rm sol}(0)]$ contributions reflecting the coexistence of thermal fluctuations and coherent structures in the initial state. Explicit temperature dependences are obtained by averaging results over the distribution

$$P_T[\tilde{\beta}_q^{\rm th}(0)]=\prod_q \frac{e^{-|\tilde{\beta}_q^{\rm th}(0)|^2/\langle n_q\rangle}}{\langle n_q\rangle} , \tag{4.1}$$

where $\langle n_q\rangle$ is the Bose distribution $[\exp(\hbar\omega_q/k_BT)-1]^{-1}$. In this way we can compute, for example, the average energy of the system at the temperature T, and determine the dressing fraction by minimizing this average energy against variations in δ.

Self-consistency requires that our procedure preserve as substantially as possible the properties of those solutions already derived by the time-dependent variational method. However, since we are choosing a single dressing fraction to characterize many solutions of our dynamical equations, it is possible that our choice of δ will not apply equally well to all solutions of those equations. (For example, the optimal dressing fraction may depend on the group velocity.) For this reason we minimize the average energy using properties of those solutions of (3.13) having the greatest physical importance, namely those solutions characterized by a small group velocity (small k). This limitation enters our calculation when we approximate $\tilde{\beta}_q(t)$; using (3.9), we integrate once by parts and make a quasistatic approximation $(d/dt)|\alpha_l(t)|^2\approx 0$ twice, resulting in

$$\tilde{\beta}_q(t)\approx\tilde{\beta}_q^{\rm th}(0)e^{-i\omega_q t}-(1-\delta)\sum_m \chi_m^q|\tilde{\alpha}_m(t)|^2 . \tag{4.2}$$

With this single approximation, the *unaveraged* energy is given by

$$\langle H\rangle\approx\sum_n[E-\delta(2-\delta)E_b]|\tilde{\alpha}_n(t)|^2-\sum_n\tilde{J}\tilde{\alpha}_n^*(t)[\tilde{\alpha}_{n+1}(t)+\tilde{\alpha}_{n-1}(t)]$$
$$+\sum_q\hbar\omega_q|\tilde{\beta}_q^{\rm th}(0)|^2-\tfrac{1}{2}(1-\delta)^2\sum_{m,n}K_{mn}(0)|\tilde{\alpha}_m(t)|^2|\tilde{\alpha}_n(t)|^2 . \tag{4.3}$$

In averaging this energy over thermal fluctuations, we assume that the system is in global equilibrium, so that the thermal average of each term in (4.3) is independent of time.

In averaging the kinetic energy we must account for the fact that thermal fluctuations may have a significant impact on phase relations. Extending our quasistatic approximation to the present case, we assume that the principal effect of thermal fluctuations on the kinetic energy can be accommodated by a phase transformation. Recalling the relations (2.13) and (2.14) which map our factored "$\tilde{D}$-type" state onto a phase-correlated "D_1-type" state, we observe that it is not the full strength of the thermal fluctuations which is effective in modulating these phase correlations, but only a fraction equal to the dressing fraction δ. Thus we approximate

$$\tilde{\alpha}_n^*(t)\tilde{\alpha}_{n\pm1}(t)\approx|\tilde{\alpha}_n(t)|\,|\tilde{\alpha}_{n\pm1}(t)|\times\exp\left[\frac{\delta}{i\hbar}\int_{-\infty}^{t}d\tau[\tilde{f}_n(\tau)-\tilde{f}_{n\pm1}(\tau)]\right] . \tag{4.4}$$

Carrying out the thermal average as in Ref. 35, we write

$$\left\langle\sum_n\tilde{J}\tilde{\alpha}_n^*(t)\tilde{\alpha}_{n\pm1}(t)\right\rangle_T\approx\left\langle\sum_n\tilde{J}|\tilde{\alpha}_n(t)|\,|\tilde{\alpha}_{n\pm1}(t)|\right\rangle_T e^{-\delta^2[S(T)-S(0)]} , \tag{4.5}$$

where

$$S(T)=4\sum_q|\chi^q|^2\sin^2(\tfrac{1}{2}ql)\coth(\hbar\omega_q/2k_BT) , \tag{4.6}$$

$$S(0)=\frac{8}{3\pi}\frac{E_b}{\hbar\omega_B},\quad S(\infty)\approx 8\frac{k_BTE_b}{\hbar^2\omega_B^2} . \tag{4.7}$$

This has the effect of transforming the partially reduced $\tilde{J}$ of earlier sections into the corresponding mean-field quantity

$$\tilde{J}=Je^{-\delta^2S}\rightarrow\tilde{J}(T)=Je^{-\delta^2S(T)} . \tag{4.8}$$

To complete our approximation of the kinetic-energy term, we must obtain an approximate expression for the amplitude contribution which remains on the right-hand side of (4.5). While we cannot determine in advance the detailed dependence of this contribution on either δ or T, we must make some assumption about its form if we are to account for the implicit nonlinear dependence of the kinetic energy on the exciton probability distribution. Observing that the kinetic energy of an excitation localized on a single site is zero, and taking into consideration the properties of the soliton solutions derived in Sec. III, we arrive at the interpolation formula

$$\left\langle\sum_n\tilde{J}\tilde{\alpha}_n^*(t)[\tilde{\alpha}_{n+1}(t)+\tilde{\alpha}_{n-1}(t)]\right\rangle_T\approx 2\tilde{J}(T)\left[1-\frac{(\kappa l)^2}{6+(\kappa l)^2}\right] , \tag{4.9}$$

where κ is given by (3.27) in which $\tilde{J}(T)$, rather than $\tilde{J}$, is used to determine the linear effective mass.

Considering now the explicitly nonlinear potential energy, we define a dimensionless quantity $\bar{\rho}$ by the relation

$$\bar{\rho}E_b\equiv\left\langle\tfrac{1}{2}\sum_{m,n}K_{mn}(0)|\tilde{\alpha}_m(t)|^2|\tilde{\alpha}_n(t)|^2\right\rangle_T . \tag{4.10}$$

$\bar{\rho}$ is a measure of the mean amplitude of density fluctuations; $\bar{\rho}=1$ for a state localized on a single site, $\bar{\rho}=\frac{1}{6}\kappa l$ for solitons in the continuum approximation, and $\bar{\rho}=0$ for plane waves. The coherence of a state affects $\bar{\rho}$ in that a fragmented probability distribution generally results in

a smaller value of $\bar{\rho}$ than does an unfragmented pulse. Taking into account the properties of the soliton solutions we obtained in Sec. III, we assume

$$\bar{\rho} \approx \frac{\kappa l}{6+\kappa l} , \tag{4.11}$$

where, as above, κ is given by (3.27) using $\tilde{J}(T)$ to determine the effective mass.

The interpolation formulas (4.9) and (4.11) are chosen to agree with the disparate values the kinetic energy and nonlinear potential energy should assume for both broad and narrow exciton probability distributions. These interpolations are purely phenomenological, and certainly not unique; however, comparable choices are essential if self-consistency is to be maintained throughout the full parameter space of the problem.

Combining these results, and using (3.27) to eliminate κl, we obtain, for the average energy in thermal equilibrium,

$$\langle H \rangle_T \approx E - \delta(2-\delta)E_b - 2Je^{-\delta^2 S(T)} + \sum_q \hbar\omega_q \langle n_q \rangle + 2Je^{-\delta^2 S(T)} \left[\frac{\frac{1}{4}(1-\delta)^4 E_b^2}{\frac{1}{4}(1-\delta)^4 E_b^2 + 24Je^{-2\delta^2 S(T)}} \right]$$
$$-(1-\delta)^2 E_b \left[\frac{\frac{1}{2}(1-\delta)^2 E_b}{\frac{1}{2}(1-\delta)^2 E_b + 12Je^{-\delta^2 S(T)}} \right] , \tag{4.12}$$

in which all dependence of the average energy on the dressing fraction is now explicit.

We determine the optimal dressing fraction by minimizing $\langle H \rangle_T$ against variations in δ. In order to clearly understand the manner in which nonlinearity affects the optimal dressing fraction, we proceed with this determination in two stages: First, we neglect the nonlinear corrections to the energy; this "linear" estimation closely follows previous estimations of δ by Toyozawa[30] and Yarkony and Silbey.[31] Second, we reintroduce the nonlinear corrections and show how these prior results must be generalized.

Imposing the minimum conditions

$$\frac{\partial \langle H \rangle_T}{\partial \delta} = 0, \quad \frac{\partial^2 \langle H \rangle_T}{\partial \delta^2} > 0 , \tag{4.13}$$

we find the optimal dressing fraction to be

$$\delta = \frac{E_b}{E_b + 2\tilde{J}(T)S(T)} , \tag{4.14}$$

subject to the restriction

$$\delta^2(1-\delta) \leq \frac{1}{2S(T)} . \tag{4.15}$$

Equation (4.14) is a transcendental equation for δ, owing to the exponential dependence of $\tilde{J}(T)$ on the dressing fraction. To discuss the general dependence of the dressing fraction on the important system parameters, we define the dimensionless quantity

$$B(T) = \frac{2J}{E_b} S(T) = \frac{8}{3\pi} \frac{2J}{\hbar\omega_B} \frac{S(T)}{S(0)} , \tag{4.16}$$

in terms of which

$$\delta = \frac{1}{1 + B(T)e^{-\delta^2 S(T)}} . \tag{4.17}$$

This scaling relation admits a phase diagram exhibiting characteristics of a first-order phase transition.[30] The dotted curves in Fig. 1 show the dependence of δ on $S(T)$ for various values of $B(T)$ as given by the "linear" formula (4.17). As a scaling plot, every Fröhlich-type system is represented, subject only to the constancy of the temperature. Since $B(T)$ is independent of the exciton-phonon coupling constant χ [cf. (4.16)] and $S(T)$ is independent of the resonance integral J [cf. (4.6)], it is convenient to consider variations in $B(T)$ to be variations in J relative to a fixed phonon bandwidth $\hbar\omega_B$, and variations in $S(T)$ to be variations in the polaron binding energy E_b relative to $\hbar\omega_B$. The critical point of the phase transition is given by the critical parameter values

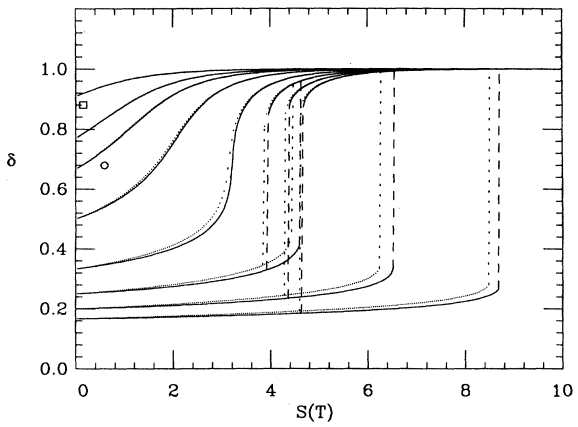

FIG. 1. Dressing fraction plotted against parameter space. Dotted lines, $\delta[B(T),S(T)]$ resulting from the formula (4.17) which neglects nonlinear corrections to the energy. Solid lines, $\delta[B(T),S(T)]$ which results from including nonlinear corrections to the average energy. Each dotted or solid pair of curves corresponds to a fixed value of $B(T)$; from top to bottom, $B(T)=0.1$, 0.3, 0.5, 1.0, 2.0, 3.0, 4.0, and 5.0. Short-dashed and long-dashed lines connect the upper and lower branches of the linear and nonlinear $\delta[B(T),S(T)]$ curves, respectively, forming hysteresis loops. □ and ○ indicate the dressing fraction predicted for the α-helix at $T=0$ and 300 K, respectively, using the standard parameters in Table I.

$B_c=\frac{1}{2}e^{3/2}$, $S_c=\frac{27}{8}$, and $\delta_c=\frac{2}{3}$. All variations of $B(T)$ and $S(T)$ with $S(T)$ below the critical point result in smooth changes in the dressing fraction; for $S(T)$ sufficiently small, the full range of dressing fractions from 0 to 1 is possible as a function of $B(T)$. Similarly, all variations of $B(T)$ and $S(T)$ with $B(T)$ below the critical point result in smooth changes in δ; in this regime, $\delta>\delta_c$ for all values of $S(T)$. On the other hand, for fixed values of $B(T)$ greater than B_c, increases in $S(T)$ along B isopleths eventually encounter instabilities due to the existence of two relative minima in the dependence of the total energy on the dressing fraction.

As is usual in first-order phase transitions, there exists a region of the parameter space in which the corresponding states of the system are absolutely unstable; i.e., no value of δ in the unstable region corresponds to a relative minimum in the energy as a function of δ. Surrounding this region at both high and low δ are regions metastability wherein there exist nonabsolute relative minima in the energy. In the metastable region, the corresponding states of the system may be long lived, allowing for the possibility of significant hysteresis phenomena. The lower and upper bounds of a hysteresis loop are marked by parameters S_l^* and S_u^*, which are functions of $B(T)$. The thermodynamic transition between high- and low-δ states occurs at a parameter $S^*\in(S_l^*,S_u^*)$, where the two relative minima in the average energy are degenerate. Regardless of where in the hysteresis region a particular excitation undergoes a transition, the quantum state of the excitation undergoes an abrupt change in character marked by a discontinuous increase in dressing fraction and a concomitant jump in the effective mass associated with the state. The discontinuity we find is well known, and its appearance is known to be somewhat dependent on one's analytical method and the nature of the approximations used; however, the appearance of a *rapid* increase (possibly by orders of magnitude) in the effective mass at the indicated point is a robust conclusion, and is the hallmark of a self-trapping transition.[36]

The formula analogous to (4.17) which results from minimizing the total average energy (4.12) including nonlinear corrections is quite cumbersome and so we do not display it here; however, representative results have been determined numerically and are indicated by the solid curves in Fig. 1. Clearly, the nonlinear corrections to the "linear" result based on (4.17) are quite small. Our calculations thus support the conventional picture of self-trapping as a robust one, with good quantitative agreement over a large region of the phase diagram.

When it is useful to do so, the self-trapping transition can be viewed from a quantum and/or linear perspective as transition between two kinds of translationally invariant superposition states: below the transition, superpositions of excitons uncorrelated with the lattice (light, free); above the transition, superpositions of excitons highly correlated with the lattice (heavy, self-trapped). Below the transition, classical resonance between free-exciton waves lead to the well-known large-polaron or soliton states which break translational symmetry and are essentially nonlinear. To the extent that coherent superpositions of self-strapped states can be maintained, similar resonances should exist above the self-trapping transition, allowing for the possible existence of a new class of solitons. A complete view of the self-trapping transition thus involves not only the transition between essentially linear states (free exciton, self-trapped exciton), but a transition between soliton states as well.

Since ours is a fully nonlinear theory, we can address this broader question of self-trapping as a transition between soliton states by directly computing the change in width of the minimum-energy soliton state across the transition. To do this we note that the width λ (in units of the lattice constant) of a static soliton according to (3.27) is given by

$$\lambda=\frac{J}{E_b}\frac{e^{-\delta^2 S(T)}}{(1-\delta)^2}\ . \tag{4.18}$$

Noting that the dependence of δ on $S(T)$ below the transition is very weak, we find that the soliton width decreases approximately exponentially with increasing $S(T)$ as the transition is approached from below. Moreover, considering also the formula (4.15) for the spinodal curve, one can show that on the lower branch the soliton narrows from its infinite width at $S(T)=0$ to a width of the order of the lattice constant at $S^*\in(S_l^*,S_u^*)$. The progressive narrowing of the soliton width and the steady increase in its effective mass as the transition is approached from below are consistent with the notion of self-trapping, but occur continuously as functions of $S(T)$ from the weak coupling limit right up to the transition. Using (3.21) and (4.18), this lower-branch behavior can be summarized by the approximate relation

$$\lambda\propto\tilde{m}^{-1},\quad S<S^*\ . \tag{4.19}$$

On the other hand, considering the asymptotic $[S(T)\gg S^*]$ formula for δ on the upper branch,

$$(1-\delta)\approx B(T)e^{-S(T)}\ , \tag{4.20}$$

we see that the soliton width increases exponentially as δ approaches unity. Again, using (3.21) and (4.18), this behavior can be summarized by the approximate relation

$$\lambda\propto\tilde{m}^{+1},\quad S>S^*\ . \tag{4.21}$$

Denoting with subscripts l and u quantities computed on the lower and upper branches of $\delta(S)$, respectively, we can relate the change in soliton width across the self-trapping transition to the corresponding change in effective mass through the relation

$$\frac{\lambda_u}{\lambda_l}=\frac{m_l}{m_u}\frac{(1-\delta_l)^2}{(1-\delta_u)^2}\ . \tag{4.22}$$

Were one to focus only on the effective mass by neglecting the explicit dependence of the soliton width on the dressing fraction, one might conclude that the jump in effective mass across the transition causes an abrupt narrowing of the soliton; however, since the relation of the soliton width to the effective mass *changes* at the transition, this conclusion would be in error. In fact, computing λ above and below the transition shows that in crossing the transition the soliton experiences a sharp *increase*

in width rather than the sharp *decrease* in width one might expect from the name commonly given to this transition.

The change in the relationship of the soliton width to the effective mass at the transition indicates that the self-trapping transition occurs at that point where the system switches (abruptly) between the two mechanisms through which the system minimizes its total energy. The first mechanism is that of soliton formation through which energy reduction is achieved by classical resonance phenomena that have the consequence of localizing energy in real space; the resulting entity is essentially nonlinear in its characteristics, and a natural representative of this state is a static pulse soliton. The second mechanism is that of small-polaron formation through which energy reduction is achieved by the binding of a localized exciton and localized lattice distortion; the resulting entity is essentially linear in its characteristics, and a natural representative of this state is a Bloch wave.

Since the system is essentially linear both when $S(T)$ is near zero and when δ is near unity, the self-trapping transition marks an abrupt end to the steady increase in nonlinear character which begins in the weak-coupling limit and continues with increasing exciton-phonon coupling strength up to S^*. Beyond S^*, the nonlinear character of the self-trapped state diminishes with increasing exciton-phonon–coupling strength. From this perspective, the transition occurs when the progressive narrowing of the soliton can no longer be supported by the lattice, at which point the soliton collapses into a small-polaron state. Since a small polaron is a bound state of a localized exciton and a localized lattice distortion, this can be viewed as self-trapping; however, the small-polaron *state* into which the soliton collapses is a (generally nonlinear) superposition of such localized states, and, in general, is *not* localized. This small-polaron state becomes progressively more delocalized as exciton-phonon–coupling strength continues to be increased past the transition.

In concluding this discussion, it must be noted that the mean-field characterization of soliton states given above must be severely limited when spatial and/or temporal fluctuations are sufficiently strong. The tendency of soliton states above the self-trapping transition to become more delocalized with increasing coupling strength [decreasing $\tilde{J}(T)$] is in direct competition with the tendency of quantum states to localize in a random field. Even in perfectly crystalline solids, thermal fluctuations present random fields which have similar consequences. The soliton widths we quote must therefore be viewed as upper bounds, maximum widths which are subject to erosion by coherence-degrading mechanisms. This observation is not unique to the self-trapped regime, of course, but it is most important here since extended self-trapped states are easy prey for localization via disorder mechanisms. Regardless of whether solitons exist above the self-trapping transition in a particular system at a particular temperature, our conclusions show that weak nonlinearities persist above the self-trapping transition. Since our time-dependent approach is not of mean-field type with respect to either static or thermal fluctuations, our theory provides a consistent framework in which to study a broad range of phenomena involving such weakly nonlinear self-trapped states.

V. SOLITONS IN BIOENERGETICS

Davydov's theory of envelope solitons is widely discussed as a model of energy transport in biological systems. A specific form of this model has come to be viewed as a test bed for the viability of the soliton concept in biological applications. The standard system consists of a single spine of the three-spine backbone of the α-helix. Excitons are taken to represent energy quanta of the amide-I vibration (C=O stretch) and the transport of these excitons is made possible by the dipole-dipole coupling of amide-I oscillations in adjacent turns of the helix. Elastic properties of the chain are attributed to the hydrogen bonds bridging successive turns of the helix, and the exciton-phonon interaction is derived from the dipole-dipole coupling of the amide-I and hydrogen bonds. The standard model is formulated in terms of the translationally invariant acoustic chain Hamiltonian (2.2) with the system parameters estimated from known properties of the α-helix. While some variability exists in estimates of these parameters, the most widely accepted values[37] are those shown in Table I.

The central question in this area of study is whether nonlinearities in the dynamics of energy transport are sufficiently strong to organize energy into coherent and stable nondispersing excitations which may serve as energy carriers in bioenergetic processes. A quantity which plays a central role in these considerations is the soliton width in the continuum approximation. In Davydov's theory, the number of sites spanned by a static soliton is given by

$$\lambda \propto \frac{J}{E_b} . \tag{5.1}$$

Using the α-helix parameters from Table I, this number is less than unity. The obvious conclusion is that Davydov solitons in the α-helix would have to be confined to one, or at most a few, lattice sites, a con-

TABLE I. First group, system parameters for the α-helix as given in Ref. 37; center group, relevant energy scales in common units; lower group, ratios of intrinsic energy scales.

Quantity	Value
J (cm^{-3})	7.8
l (Å)	5.4
M	$114m_p$
w ($N\,m^{-1}$)	13
χ (N)	6.2×10^{-11}
J (J)	1.549×10^{-22}
E_b (J)	2.957×10^{-22}
$\hbar\omega_B$ (J)	1.742×10^{-21}
300 K (J)	4.141×10^{-21}
$E_b/\hbar\omega_B$	0.1698
$2J/\hbar\omega_B$	0.1779
$2J/E_b$	1.0477

clusion which is supported by analytical studies including discreteness corrections[22] and by numerical studies of Davydov's theory in its discrete-lattice formulation.[38–40] Beyond this semiquantitative conclusion, a qualitative conclusion is frequently drawn that the strong localization found in Davydov's theory points to an essential role for nonlinear dynamics in biological energy-transport processes.

Our theory points to a rather different conclusion. (For the reader's convenience, we have assembled in Table II a number of quantities relevant to the present discussion.) In our theory, the number of sites spanned by a static soliton is given by (4.18). For the same α-helix parameters given in Table I, this results in a soliton width of some 33 lattice constants at $T=0$ K and of six lattice constants at $T=300$ K.

The disparity between the results of Davydov's theory and the theory we put forward is due entirely to the fact that the parameters characterizing the α-helix place it in a region of the system phase diagram in which nonlinearities are weak. This statement is quantified by a straightforward calculation of the dressing fraction δ [the "linear" formula (4.17) is adequate], which shows the optimal state to be approximately 88% dressed at $T=0$ K and 68% dressed at $T=300$ K. This substantial degree of dressing has the immediate consequence that the strength of static nonlinearities (in the nonlinear Schrödinger equation, for example) are reduced relative to Davydov's theory by a factor of 70 at $T=0$ K and by a factor of 10 at $T=300$ K. Since the exciton-phonon coupling is weak [$S(0)\approx 0.04 S_c$], this weakening of nonlinearity is not significantly compensated for by reductions in the renormalized transfer matrix element $\tilde{J}$, which amount to only 11% at $T=0$ K and 24% at $T=300$ K.

We have concluded here on the basis of a mean-field characterization of the soliton state that α-helix solitons may be relatively broad—broad enough at low temperatures to justify a continuum approximation. At present, we do not have a quantitative estimate of the impact fluctuations may have on these conclusions; however, we can compare our mean-field characterizations to the recent quantum Monte Carlo (QMC) results of Wang *et al.*[41] obtained for exactly the same model using the same system parameters from Table I. Wang *et al.* found evidence of structures two to three sites wide at $T=0$ K (compare 33 sites) and found essentially complete localization on a single site above $T=11$ K (compare six sites at $T=300$ K). In order to clearly understand this comparison, it is crucial that we examine the *same* quantity computed in the QMC simulation. Wang *et al.* computed the thermal-equilibrium expectation value of the operator

$$\hat{C}_l=\sum_n \chi(\hat{Q}_{n+l+1}-\hat{Q}_{n+l-1})a_n^\dagger a_n \ . \tag{5.2}$$

(Units here are chosen for notational convenience; we note that this diagnostic operator is expressed in terms of *bare* exciton and phonon operators.) Using (2.12)–(2.15), (4.2), and (4.4) to compute the expectation value of $\hat{C}_l$ in our quasistatic approximation, we find the result

$$\langle \hat{C}_l \rangle_T \approx -2\sum_q |\chi^q|^2 e^{-iqR_l}\hbar\omega_q[\delta+(1-\delta)\langle |P^q|^2\rangle_T] \ , \tag{5.3}$$

in which P^q is the Fourier transform of the exciton probability distribution

$$P^q=\sum_n e^{-iqR_n}|\tilde{\alpha}_n|^2 \tag{5.4}$$

(note $|\tilde{\alpha}_n|^2=|\alpha_n|^2$). The observed deformation function decomposes into a small-polaron part (set $\delta=1$) having weight δ and a soliton part (set $\delta=0$) having weight $1-\delta$. The small-polaron part is wholly independent of exciton probability amplitudes, while the soliton part provides an image of the average shape of exciton probability distribution in space.

At room temperature the weights we predict for the small-polaron and soliton parts of the deformation function (0.68 and 0.32, respectively) differ by only a factor of 2, and our mean-field result suggests a soliton shape which should be resolvable in the room-temperature QMC data. This structure is *not,* in fact, observed; the room-temperature QMC data reflect nothing more structured than a small-polaron superposition state. The reason for this discrepancy lies not in a shortcoming of our theory, but in extrapolating mean-field results (such as our soliton width estimate) into a regime where they are no longer valid. The equations of motion (3.13) are *not* so limited since thermal fluctuations retain their full impact in real time; what we require for an interpretation of the room-temperature QMC data is a reasonable estimate of the effect these fluctuations have on the evaluation of the deformation function. Consider the following argument: Given the large ratios of thermal energies at room temperature to both the small-polaron binding energy E_b and the transfer matrix element J (cf. Table II), it is unlikely that any significant spatial or temporal correlations can be maintained. Perforce, at room temperature, the Boltzmann factor decays across the bare exciton energy band by less than 14%, across the δ-dressed energy band by less than 11%, and across the fully dressed energy band by merely 8%. It is reasonable to anticipate on the basis of these ratios alone that the exciton amplitudes evolve in a nearly stochastic manner. In such a situation the Fourier transform (5.4) must be of order unity, yielding the conclusion

TABLE II. Upper group, ratios of thermal to intrinsic energy scales; lower group, quantities related to the determination of the dressing fraction and soliton width.

Quantity	$T=0$ K	$T=300$ K
$k_BT/\hbar\omega_B$	0	2.378
k_BT/E_b	0	14.00
k_BT/J	0	26.73
$S(T)$	0.1441	0.5848
$B(T)$	0.1510	0.6126
δ	0.8810	0.6791
$e^{-\delta^2 S(T)}$	0.8942	0.7636
λ	33.08	6.365

$$\langle |P^q|^2 \rangle_{T=300\ K} \approx 1 \ . \tag{5.5}$$

Inserting this result in the deformation function, we find that the deformation function converges to the small-polaron result, in complete agreement with the QMC simulation.

Noting the generality of the foregoing argument, we conclude that regardless of the value taken by δ the deformation function of any system approaches the small-polaron form at high temperatures. This includes the $\delta=0$ limit corresponding to Davydov's theory, and helps to explain the consonance of the room-temperature QMC results of Wang *et al.* and of the room-temperature dynamical simulations of Lomdahl and Kerr[41] based on Davydov's theory.

At zero temperature the weights we predict for the small-polaron and soliton parts of the deformation function (0.88 and 0.12, respectively) differ by a factor of 7.3, and our mean-field result suggests a soliton shape which should be lost in the noise of the zero-temperature QMC data. It is clear, however, that the zero-temperature QMC data contain structure not ascribable to the small-polaron part of the deformation function. To help resolve this difficulty, we observe that provided the soliton part of deformation function is small relative to the small-polaron part, an estimate of the dressing fraction can be obtained directly from the QMC data in the form

$$\frac{\langle \hat{C}_0 \rangle_{T=0.27\ K}}{\langle \hat{C}_0 \rangle_{T=300\ K}} \approx \delta_{T=0\ K} \approx 0.7 \ .$$

This "empirical" estimate of the low-temperature dressing fraction is approximately 20% smaller than the predicted value of 0.88. To this smaller dressing fraction there corresponds a smaller soliton width of approximately five to six sites, in rough agreement with the apparent width of the background in the QMC data which our method would attribute to the state's soliton character. Noise in the background of the QMC result precludes a serious estimate of the degree of agreement between these results; we do not anticipate quantitative agreement since quantum fluctuations, which were found to be quite strong in the QMC simulation, are incompletely accounted for in our present analysis. It is likely that the neglected quantum fluctuations would have the effect of reducing the dressing fraction, bringing our analytical and empirical estimates into closer agreement, but at present we have no means of identifying the source of the 20% discrepancy.

VI. CONCLUSION

We have constructed a unified theory of polaron and soliton dynamics by combining time-dependent variational methods recently applied to the theory of Davydov solitons with partial-dressing methods well known from polaron theory. In the interest of compactness and clarity, we have focused in this paper on the simplest partial-dressing assumption, applying a common dressing fraction δ to all phonon modes. Our fundamental result is a system of nonlinear evolution equations in which the tendency of a system to form Davydov solitons is balanced against its tendency to form small polarons.

To determine the optimal dressing fraction, we subsequently applied time-independent variational methods in a manner similar to prior determinations by Toyozawa,[30] Yarkony and Silbey,[31] and Venzl and Fischer.[34] The characterization of the partially dressed soliton states which results is complete in that every set of system parameters is represented by a point on the system phase diagram and thus determines a unique dressing fraction. Consistent with prior works from polaron theory, we have found a self-trapping transition which is only weakly modified by our consideration of nonlinearities, and we have reinterpreted this transition in terms of our newly obtained soliton states.

In the specific context of bioenergetics, we have considered the standard model of the α-helix in order to compare our results with results known from other approaches. Our mean-field characterization of the static soliton in the α-helix differs significantly from the well-known result from Davydov's theory. The dressing fraction was found to be approximately 88% at zero temperature, characterizing the α-helix as a *weakly nonlinear* system close to the small-polaron limit. We have found complete agreement with the room-temperature results of Wang *et al.*[41] by viewing room-temperature exciton dynamics as a nearly stochastic process. Our zero-temperature soliton solutions are in substantial agreement with the low-temperature quantum Monte Carlo results of Wang *et al.*[41] when empirical as well as analytical estimates of the dressing fraction are considered.

Our results offer no direct indications that the basic method of applying the time-dependent variational principle (3.1) to the D_1 ansatz state imposes any limitation on our results. The soliton solutions we have obtained can be shown to follow also from a time-independent procedure, showing them to be a robust result not peculiar to the time-dependent method employed. Our principal approximation, the assumption of a single dressing fraction δ for all states, can be lifted at the cost of increased complexity. The dressing fraction can be allowed to be velocity dependent, for example, or to vary from one phonon mode to another;[34] preliminary calculations of this type support the conclusions of this paper. The difficulty we encounter with respect to quantum fluctuations is more subtle. Ours is a real-time analysis; quantum fluctuations, while manifesting themselves in real time through interference phenomena, etc., are not fluctuations in real time. Nonetheless, it can be hoped that some approximate accounting for the additional randomness due to intrinsic quantum noise can be made by introducing judiciously chosen noise terms into our equations of motion. The results of these refinements will be presented elsewhere.

ACKNOWLEDGMENTS

It is a pleasure for the authors to acknowledge numerous useful discussions with Katja Lindenberg and Xidi Wang. This work was supported in part by the U.S. Defense Advanced Research Projects Agency (DARPA) under ARPA order 7048 and by the Serbian Science Foundation.

*Permanent address: Department of Theoretical Physics, Boris Kidric Institute of Nuclear Sciences, P.O. Box 522, Belgrade 11001, Yugoslavia.

[1]L. D. Landau, Phys. Z. Sowjetunion **3**, 664 (1933); S. I. Pekar, Zh. Eksp. Teor. Fiz. **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, Proc. R. Soc. London, Ser. A **215**, 291 (1952); Adv. Phys. **3**, 325 (1954).

[3]K. W. Kehr, in *Hydrogen in Metals I,* Vol. 28 of *Topics in Applied Physics,* edited by G. Alefeld and J. Völkl (Springer-Verlag, Berlin, 1978).

[4]M. I. Klinger, Phys. Rep. **94**, 183 (1983).

[5]Charles B. Duke and L. B. Schein, Phys. Today **33**, 42 (1980).

[6]D. Emin, Phys. Today **35** (2), 34 (1982).

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

[8]M. Ueta, H. Kanzaki, K. Kobayashi, Y. Toyozawa, and E. Hanamura, in *Excitonic Processes in Solids,* Vol. 60 of *Springer Series in Solid-State Sciences,* edited by M. Cardona, P. Fulde, K. von Klitzing, and H.-J. Quisser (Springer-Verlag, Berlin, 1986).

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

[10]A. S. Davydov, *Solitons in Molecular Systems* (Reidel, Boston, 1985).

[11]A. C. Scott, Phys. Rev. A **26**, 578 (1982); **27**, 2767 (1982).

[12]G. Careri, in *Cooperative Phenomena,* edited by H. Haken and M. Wagner (Springer-Verlag, Berlin, 1973); G. Careri, U. Buontempo, F. Carta, E. Gratton, and A. C. Scott, Phys. Rev. Lett. **51**, 304 (1983); G. Careri, U. Buontempo, I. Galuzzi, A. C. Scott, E. Gratton, and E. Shyamsunder, Phys. Rev. B **30**, 4689 (1984); J. C. Eilbeck, P. S. Lomdahl, and A. C. Scott, *ibid.* **30**, 4703 (1984); A. C. Scott, E. Gratton, E. Shyamsunder, and G. Careri, *ibid* **32**, 5551 (1985).

[13]D. M. Alexander, Phys. Rev. Lett. **54**, 138 (1985); D. M. Alexander and J. A. Krumhansl, Phys. Rev. B **33**, 7172 (1986).

[14]A. Migliori, P. M. Maxton, A. M. Clogston, E. Zirngiebl, and M. Lowe, Phys. Rev. B **38**, 13464 (1988).

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

[16]D. Emin, Adv. Phys. **22**, 57 (1973).

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

[18]A. S. Davydov, Phys. Status. Solidi **36**, 211 (1969).

[19]A. D. Davydov and N. I. Kislukha, Phys. Status Solidi **59**, 465 (1973); Zh. Eksp. Teor. Fiz. **71**, 1090 (1976) [Sov. Phys.—JETP **44**, 571 (1976)].

[20]A. S. Davydov, Zh. Eskp. Teor. Fiz. **51**, 789 (1980) [Sov. Phys.—JETP **78**, 397 (1980)].

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

[22]Xidi Wang, David W. Brown, Katja Lindenberg, and Bruce J. West, Phys. Rev. A **37**, 3557 (1988).

[23]David W. Brown, Katja Lindenberg, and Bruce J. West, Phys. Rev. A **33**, 4104 (1986); David W. Brown, Bruce J. West, and Katja Lindenberg, *ibid.* **33**, 4110 (1986).

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

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

[26]T. D. Lee, F. E. Low, and D. Pines, Phys. Rev. **90**, 297 (1953).

[27]P. Langhoff, S. T. Epstein, and M. Karplus, Rev. Mod. Phys. A **44**, 602 (1972).

[28]J. Frenkel, *Wave Mechanics, Advanced General Theory* (Clarendon, Oxford, 1974).

[29]David W. Brown, Phys. Rev. A **37**, 5010 (1988).

[30]Yutaka Toyozawa, Prog. Theor. Phys. **26**, 29 (1961); Yutaka Toyozawa, in *Organic Molecular Aggregates,* edited by P. Reineker, H. Haken, and H. C. Wolf (Springer-Verlag, Berlin, 1983).

[31]David Yarkony and Robert Silbey, J. Chem. Phys. **65**, 1042 (1976); **67**, 5818 (1977).

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

[33]Zoran Ivic and David W. Brown, Phys. Rev. Lett. **63**, 426 (1989).

[34]Gerd Venzl and Sighart F. Fischer, J. Chem. Phys. **81**, 6090 (1984); Phys. Rev. B **32**, 6437 (1985).

[35]David W. Brown, Katja Lindenberg, and Bruce J. West, Phys. Rev. B **37**, 2946 (1988).

[36]For extensive discussions of self-trapping phenomena, see Refs. 7 and 8.

[37]A. C. Scott, Philos. Trans. R. Soc. London, Ser. A **315**, 423 (1985).

[38]P. S. Lomdahl and W. C. Kerr, Phys. Rev. Lett. **55**, 1235 (1985).

[39]Albert F. Lawrence, James C. McDaniel, David B. Chang, Brian M. Pierce, and Robert R. Birge, Phys. Rev. A **33**, 1188 (1986).

[40]L. Cruzeiro, J. Halding, P. L. Christiansen, O. Skøvgaard, and A. C. Scott, Phys. Rev. A **37**, 880 (1988).

[41]Xidi Wang, David W. Brown, and Katja Lindenberg, Phys. Rev. Lett. **62**, 1796 (1989).