Polarons and solitons

1989 J. Phys.: Condens. Matter 1 5567

(http://iopscience.iop.org/0953-8984/1/33/001)

J. Phys.: Condens. Matter. **1** (1989) 5567–5593. Printed in the UK

REVIEW ARTICLE

Polarons and solitons

A J Fisher, W Hayes and D S Wallace
Clarendon Laboratory, University of Oxford, Parks Road, Oxford OX1 3PU, UK

Received 20 March 1989

Abstract. The idea of a polaron, a charge carrier in a solid, is reasonably familiar. The concept of the bipolaron is less so. A related excitation, the soliton, which may be charged or uncharged and is characterised by large non-linearity, has attracted increasing attention in recent years. Physical properties of these various excitations are discussed, with reference to specific examples.

1. Introduction

When a charge is injected into a solid the atoms in its neighbourhood are both electronically polarised and displaced. When the charge moves under the influence of an electric field both the electronic polarisation and the displacement of the neighbours move with it, and the response to the electric field will be determined by an effective mass rather than by the bare band mass. The charge carrier with its associated polarisation and distortion field is referred to as a polaron and one can regard the polaron as a carrier with an effective mass m_p^*, different from the bare band mass m^*. There are three principal mechanisms by which a free carrier interacts with a lattice:

(i) Fröhlich coupling, in which the carrier couples to the electric field produced by LO phonons; this occurs only in polar crystals such as KCl and GaAs and is the most powerful mechanism.

(iii) Piezoelectric coupling, in which the carrier couples to the electric field produced by acoustic phonons in materials of appropriate symmetry such as ZnO and quartz.

(iii) Deformation coupling, in which the carrier energy is affected by the strain produced by acoustic modes; this occurs in all crystals.

The degree of localisation of carriers in perfect crystals is determined by a balance of opposing tendencies (Hayes and Stoneham 1985). Delocalisation is encouraged by a gain in energy equal to approximately half the band width whereas localisation is encouraged because of the gain in lattice relaxation energy associated with confinement of charge. Delocalised charges are referred to as large polarons and are typical of more covalent materials such as GaAs. Localised charges are referred to as small polarons and occur in more ionic solids such as KCl. In the case of the small polaron, motion is by hopping and the carrier may be immobilised at low temperatures, as in the case of the self-trapped hole in KCl (Hayes and Stoneham 1985). Occasionally there is near-equality of the localising and delocalising tendencies (Toyozawa 1961) and introduction of modest disorder can tip the balance to give charge localisation and a change from

0953-8984/89/335567 + 27 $02.50

high-mobility non-polaronic transport to low-mobility small-polaronic transport (Emin 1983—see below).

In magnetically disordered materials the spin of a carrier can polarise the magnetic moments of neighbouring magnetic ions and the moving charge is accompanied by a cloud of polarised spins. By analogy with the conventional lattice polarons mentioned above, this entity is called a magnetic polaron; if the charge is bound to an impurity or other lattice defect it is called a bound magnetic polaron (BMP). The concept of a BMP was first introduced to explain the metal–insulator transition occurring at the ferromagnetic transition temperature in Eu-rich EuC at $T_c = 50$ K (Torrance *et al* 1972). Here the onset of magnetic order is associated with donor delocalisation and an increase in electrical conductivity by a factor of about 10^{13}. The localisation of the donor-state wavefunction in EuO above T_c is not an example of Anderson localisation but is thermally induced stabilisation of a small-polaron state triggered by electron–magnon interaction.

In recent years there has been considerable interest in the non-linear aspects of lattice excitations (see e.g. Trullinger *et al* 1986), particularly of domain walls. Such non-linear excitations are referred to as kinks or solitons. The word 'soliton' refers to solitary waves corresponding to particular solutions of non-linear equations describing propagation of excitations in media that are dispersive and non-linear (Zabusky and Krustal 1956). Only one-dimensional systems can be described analytically and the solutions preserve shape and velocity after interactions. In a non-linear system the velocity depends not only on the frequency but also on the amplitude, and in some circumstances the effect of amplitude dependence can compensate for that of frequency dependence, resulting in a stable solitary wave; it is well known, for example, that it is possible to transmit soliton-like light pulses in an optical fibre.

The high symmetry of the *trans*-polyacetylene (*t*-PA) chain with its conjugated double bonds makes this polymer an ideal system for the discussion of the soliton concept; perturbation of the conjugation propagates like a solitary wave in this system. A mobile conjugated defect associated with a dangling bond has in the past been referred to as a bond alternation defect, or kink, or domain wall, but is now more generally known in *t*-PA as a soliton (Kivelson 1986, Roth and Bleier 1987, Heeger *et al* 1988). In conducting polymers generally the charge carriers are analogous to the polarons mentioned earlier and the relationship between polarons and solitons is of interest.

In § 2 we shall review ionic polarons and the interactions between them, giving rise to the possibility of bipolaron formation; the significance of bipolarons for superconductivity is briefly discused. In § 3 we discuss magnetic polarons, with some emphasis on mixed magnetic semiconductors. The phenomenon of Jahn–Teller polarons is briefly dealt with in § 4, with some reference to solitons and superconductivity. Finally, in § 5 we discuss solitons and polarons in conducting polymers and again we consider bipolaron formation.

2. Polarons and bipolarons in ionic solids

2.1. Large polarons and small polarons

The simplest model of an electron interacting with a lattice involves a single harmonic oscillator of frequency ω, representing a normal mode of the lattice, to which a force F representing the effect of the electron is applied (see e.g. Hayes and Stoneham 1985, p

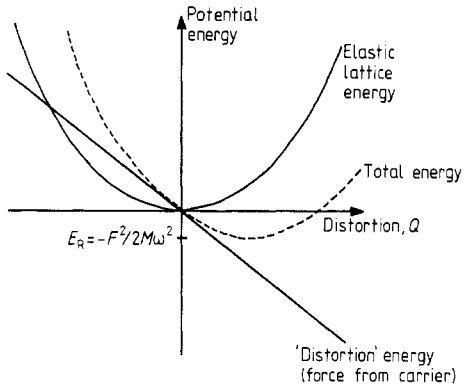

Figure 1. Illustration of the terms involved in producing the relaxation energy E_R when a force F is exerted on a classical harmonic oscillator.

31). Treating the oscillator classically for the moment and representing its configuration by the coordinate Q, the energy is

$$E = \tfrac{1}{2}M[(\mathrm{d}Q/\mathrm{d}t)^2 + \omega^2 Q^2] - FQ. \tag{1}$$

Making the substitution

$$Q' = Q - F/(M\omega^2) \tag{2}$$

we find

$$E = \tfrac{1}{2}M[(\mathrm{d}Q'/\mathrm{d}t)^2 + \omega^2 Q'^2] - F^2/(2M\omega^2). \tag{3}$$

This describes a harmonic oscillator with the original frequency ω but with its equilibrium coordinate shifted by $F/(M\omega^2)$ relative to its value when $F = 0$ and with its ground-state energy lowered by the relaxation energy

$$E_R = -F^2/(2M\omega^2). \tag{4}$$

This behaviour is illustrated schematically in figure 1. For the moment we shall assume that the mass M associated with the oscillator is large and the frequency ω rather small; the lattice cannot therefore follow the instantaneous motion of the electron and responds only to the *probability* that the electron is in a particular place. This limit is known as the *adiabatic* limit. Since the lattice is effectively frozen into position, its kinetic energy vanishes and to find the configuration of minimum energy we only have to minimise the relaxation energy E_R. (The opposite limit, known as the *anti-adiabatic* or *quasi-adiabatic* limit, occurs when the frequency ω is very large and the lattice modes can follow the instantaneous motion of the electron. The lattice is sometimes said, rather confusingly, to follow the electron motion adiabatically in this case.)

In the adiabatic limit, we now ask what kind of electron wavefunction will minimise the total energy of the coupled electron–lattice system and, in particular, what will be the degree of *localisation* of the electron described by this wavefunction (Emin and Holstein 1976)? To answer this question we must stipulate the physics that gives rise to the force F. In particular we shall be interested in *short-range* electron–lattice coupling, where the electron exerts a force on only the adjacent atom, and the *Coulomb* interaction with the deformation-induced polarisation of the lattice. If we compress or expand the whole system by a scale factor L the electron wavefunction follows the scaling law

$$\psi \propto L^{-d/2} \tag{5}$$

in d dimensions because it must remain normalised. Therefore, for *short-range* coupling,

the force F is just proportional to the probability that the electron will be found at that site:

$$F \propto |\psi|^2 \propto L^{-d}. \tag{6}$$

For a *Coulomb* interaction the force on a given atom due to the electron being situated at some other site scales as

$$F \propto L^{-2}|\psi|^2 \propto L^{-d-2}. \tag{7}$$

Integrating equation (4) for the relaxation energy over all lattice modes and (for the Coulomb case) over all positions of the electron and the atom, we find that for a short-range interaction the total relaxation energy scales as

$$E_{\mathrm{R}}^{\text{short range}} \propto L^{-d} \tag{8}$$

and for a Coulomb interaction as

$$E_{\mathrm{R}}^{\text{Coulomb}} \propto L^{-4+d}. \tag{9}$$

If both short-range and Coulomb interactions are present there is an additional cross term

$$E_{\mathrm{R}}^{\text{cross}} \propto L^{-2}. \tag{10}$$

Depending on which of these interactions are present and on their relative magnitudes, a variety of different behaviours of the total energy E as a function of scale length L may be obtained. If there is *no* electron–lattice coupling we have only the kinetic energy of the electron, which is positive (in contrast to the relaxation energy) and scales as L^{-2}. It therefore decreases monotonically as $L \rightarrow \infty$ and the familiar extended Bloch state is formed.

For a three-dimensional system in which only the Coulomb interaction with the lattice is present, or for a one-dimensional system in which there is only a short-range interaction (such as the SSH model of conjugated polymers discussed in § 5) the positive kinetic energy always dominates for small L and prevents the system collapsing to small length scales. The only stable length scale is at some finite L; this corresponds to a *large-polaron* state where the electron wavepacket extends over some finite length L.

If however we have a three-dimensional system with only a short-range interaction, or a one-dimensional system with only a Coulomb interaction, the relaxation energy E_{R} scales as L^{-3} and overwhelms the kinetic energy penalty at small L, causing the electron wavepacket to shrink until it becomes localised around a single lattice site and our continuum scaling arguments break down. This situation, where the electron is trapped at a single lattice site by the lattice distortion it causes, and the potential well is deep enough to produce a bound state, was first envisaged by Landau (1932), who also realised that in the absence of other effects it is separated from the (now metastable) delocalised Bloch state by an energy barrier. The electron is now said to form a *small-polaron* state.

If both the long-range Coulomb *and* the short-range electron–lattice interactions are present, different effects can be produced. A minimum in the energy, corresponding to a metastable large polaron, can appear at finite L, or alternatively the energy barrier between the Bloch state and the small-polaron state can be removed.

2.2. *The polaron effective mass and transport properties*

In reality we cannot neglect the kinetic energy (and hence the inertial effects) associated with the normal modes of a crystal lattice and they must be described as quantum, not

classical, oscillators. We shall review briefly some of the consequences for the theory of large polarons, and then pass on to consider small polarons.

We shall consider first the so-called *Fröhlich interaction* (see § 1) between an electron and the polarisation induced by the displacement of a longitudinal optic (LO) phonon mode. The interaction with the polarisation field at a distance r from the electron takes the form (in SI units)

$$V(r) = \frac{e^2}{4\pi\varepsilon_0 r}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon}\right) \tag{11}$$

where ε is the zero-frequency dielectric constant, which includes both electronic and ionic screening, ε_∞ is the dielectric constant at a frequency high compared with phonon frequencies but low compared with inter-band transition energies (which therefore includes only electronic screening) and ε_0 is the permittivity of the vacuum. It is customary to assume that the LO modes are dispersionless, with common frequency ω_{LO}, and to define the dimensionless coupling constant

$$\alpha = \tfrac{1}{2}\left(\frac{1}{\varepsilon_\infty} - \frac{1}{\varepsilon}\right)\frac{e^2}{4\pi\varepsilon_0 r_p \hbar\omega_{LO}}. \tag{12}$$

The quantity r_p is the *polaron radius*

$$r_p = [\hbar/(2m^*\omega_{LO})]^{1/2} \tag{13}$$

where m^* is the bare band mass of the electron (i.e. the effective mass neglecting polaron effects). It is thus the de Broglie wavelength of an electron with the LO phonon energy, and α is half the ratio of the polarisation interaction at this radius to the LO phonon energy. With these definitions and assumptions, the Fröhlich Hamiltonian for the coupled electron–lattice system is

$$H = \sum_k \varepsilon_k c_k^\dagger c_k + \hbar\omega_{LO}\sum_q b_q^\dagger b_q + \sum_{k,q} V_{kq}(b_q^\dagger c_{k-q}^\dagger c_k - b_q c_{k+q}^\dagger c_k) \tag{14}$$

where c_k annihilates an electron of Bloch wavevector k, $\varepsilon_k = \hbar^2k^2/2m^*$, $V_q = i(4\pi\alpha/V)^{1/2}(\hbar/2m\omega_{LO})^{1/4}(\hbar\omega_{LO}/q)$ where V is the sample volume and b_q annihilates a phonon of wavevector q. If the coupling is small, its properties are most conveniently studied by perturbation theory in the parameter α; this shows that the polaron effective mass m_p^* is given by

$$m_p^* = m^*[1 + \alpha/6 + O(\alpha^2)] \tag{15a}$$

and that the expectation value of the number of phonons accompanying the electron is $\alpha/2$ (see Kittel 1963 p 137). For a perfect two-dimensional system, in contrast to the three-dimensional systems discussed above, Das Sarma (1983) found that

$$m_p^* = m^*[1 + \pi\alpha/8 + O(\alpha^2)] \tag{15b}$$

so that reduction in dimensionality from 3D to 2D leads to a nearly three-fold increase in the predicted polaron mass enhancement. We might expect to find such behaviour in systems where the effective size in one direction is less than the polaron radius; however, there is as yet no convincing evidence from experimental work on heterostructures that this enhancement actually occurs (Nicholas *et al* 1988). Note that continued reduction in the dimensionality from two to one leads to the prediction of small-polaron formation from Fröhlich coupling (see above).

Unfortunately for many of the ionic crystals of interest (e.g. the alkali halides), we have $\alpha \geqslant 1$ and perturbation theory is inapplicable. Feynman (1955) showed how this difficulty may be circumvented using a path integral treatment similar to the one he introduced in quantum electrodynamics. He used an approximate form for the electron–phonon coupling and integrated out the phonon modes to leave an effective action function for the electron; he was then able to derive a variational principle and thus to deduce an upper bound to the energy of the coupled electron–lattice system. The dependence of this energy on the electron velocity yields an expression for the polaron effective mass that agrees with (15*a*) for small α. For a review of more recent developments in this field, see Peeters and Devreese (1984).

We now consider the case of the small polaron, for which a review of the theoretical position has recently been given by Stoneham (1989). It is easiest to start from a tight-binding model of the electronic structure and to include only the 'on-site' electron–phonon coupling (Holstein 1959, Lang and Firsov 1963). In other words, we consider only the dependence on the atomic positions of the matrix elements of the electronic Hamiltonian that are *diagonal* in the Wannier representation. This automatically ensures that transitions in which an electron hops from one site to another obey the Condon approximation, i.e. the electronic transition matrix elements do not depend on the phonon coordinates.

The quantum-mechanical analogue of the transformation (2) from Q to Q' that took us from (1) to (3) is a canonical transformation that eliminates the electron–phonon coupling. The Hamiltonian is rewritten in terms of *dressed* electron and phonon operators; each phonon mode has its equilibrium shifted by an amount that depends on the position of the electrons, and the creation of an electron on a particular site is accompanied by its associated lattice distortion. The new transformed Hamiltonian has three features not found in the original. First, there is a uniform shift downwards in energy because of the relaxation of the phonons. Secondly, a phonon-mediated attraction between the dressed electrons is introduced, discussion of which we shall defer until § 2.3. Thirdly, the hopping amplitude for the dressed electrons between lattice sites is no longer a simple number but involves redistribution of the lattice distortion and is therefore a function of the phonon operators. Equivalently we can say that this hopping amplitude is a function of the vibrational quantum numbers of the lattice modes before and after the transition, since it involves the overlap of the vibrational wavefunctions in the initial and final electronic states.

The Huang–Rhys factor $S_{0\alpha}$ (Huang and Rhys 1950, Pekar 1950) for the phonon mode α at zero temperature is defined as the ratio of the change in relaxation energy $\Delta E_{R\alpha}$ for that mode during the hopping process to the phonon energy $\hbar\omega_\alpha$:

$$S_{0\alpha} = \Delta E_{R\alpha}/\hbar\omega_\alpha. \tag{16}$$

The overlap between the ground states of all the phonon oscillators before and after the electronic transition is just $\exp(-\Sigma_\alpha S_{0\alpha})$. At finite temperatures we define

$$S_\alpha = S_{0\alpha} \coth(\hbar\omega_\alpha/2kT). \tag{17}$$

The sum of the overlaps between phonon states with the same quantum numbers n_α, weighted by the probabilities of these quantum numbers occurring at temperature T, is then $\exp(-\Sigma_\alpha S_\alpha)$. Note that it is a strongly *decreasing* function of temperature, because for a given mode the overlaps become smaller with increasing n_α, and with increasing temperature the mean value of n_α rises. Hence the amplitude for the hopping of an

electron between sites without change in the phonon occupation numbers, i.e. without absorption or emission of phonons, also *decreases strongly with increasing temperature*.

Let us consider the consequences of this result. The combined electron–lattice system has translational symmetry, so the Bloch wavevector is a good quantum number. At very low temperatures, however strong the electron–phonon coupling, the eigenstates of the system are extended states with definite wavevectors that are linear combinations of 'small-polaron' states where the electron and its associated lattice distortion are localised at a given site. The result is the formation of a 'polaron band' whose width is proportional to the hopping amplitude discussed above. However, as the temperature rises two things happen: the polaron band width decreases sharply (and therefore the polaron effective mass increases sharply), and the probability of scattering processes involving the emission and absorption of phonons increases. If an electron is prepared in a localised small-polaron state at a certain lattice site, it is very likely to be scattered by one of these inelastic phonon processes before its wavepacket has been significantly broadened. It is therefore most useful to consider transport processes above a certain critical temperature in terms of the thermally assisted hopping of such small-polaron states. While such a cross-over from coherent to diffusive motion has not been observed in electron transport, it has recently been reported in the diffusion of muonium in KCl measured by muon spin relaxation (Kiefl *et al* 1989). In the muon case, the cross-over occurs at about 70 K.

The consequences for the electrical conductivity as a function of temperature and of frequency are discussed by Holstein (1959), Lang and Firsov (1963), Reik (1972) and Mott (1987a). At high temperatures the conductivity has a simple activated form

$$\sigma = \sigma_0 \exp(-E_a/kT) \tag{18}$$

where E_a is the activation energy for the small-polaron hopping process. This type of dependence has been observed in a number of transition-metal oxides and over a wide temperature range in orthorhombic sulphur (Spear 1974) as well as in solid N_2, O_2 and CO (Loveland *et al* 1972). At lower temperatures there is no experimental evidence for *polaron band* transport, but this is not surprising because particles with such large effective masses and such narrow bands would be very readily localised by trapping at impurity sites. There is indirect evidence, however, of a transition from activated transport to some other mechanism in the dramatic reduction of the activation energy for conduction in Li-doped NiO at low temperatures (Bosman and van Daal 1970, Mott 1987a). We note the recent evidence suggesting that the holes in this type of doped oxide reside primarily in oxygen p states and not, as is sometimes supposed, on the metal ions (Kuiper *et al* 1989).

It is possible to discuss the frequency dependence of the conductivity on the basis of a very simple model, the *Polder model*. Figure 2 shows a configuration coordinate diagram for the lattice modes when the small polaron is situated on two neighbouring sites. The most important contribution to the conductivity at frequency ω comes from Franck–Condon hopping processes where the phonon modes do not change during the electronic transition but are left to readjust afterwards. If the effective mass and frequency of the lattice mode are M and ω_0 respectively, and its configuration is again represented by the single coordinate Q, then the potential energy curves for the lattice with the electron on the two different sites are

$$V_1(Q) = (M\omega_0^2/2)(Q - Q_0)^2 \tag{19a}$$

$$V_2(Q) = (M\omega_0^2/2)(Q + Q_0)^2 \tag{19b}$$

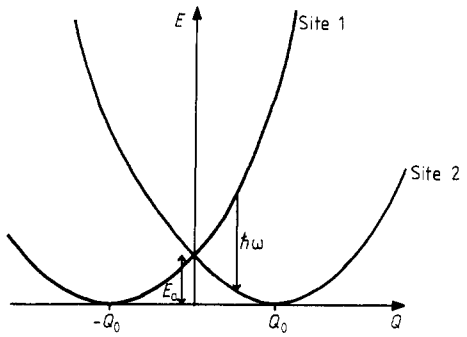

Figure 2. Configuration coordinate diagram for a small polaron hopping between two sites in a lattice. The curves with minima at $-Q_0$ and Q_0 describe the energy of the lattice with the charge localised at positions 1 and 2 respectively. The arrow shows a Franck–Condon transition of energy $\hbar\omega$ as discussed in the text.

where $\pm Q_0$ are the minimum potential energy configurations for the lattice when the electron occupies the different sites. The energy $\hbar\omega$ of the Franck–Condon transition is related to the value of Q for which it occurs by

$$\hbar\omega = 2M\omega_0^2 QQ_0. \tag{20}$$

The most important factor in determining the net rate of this process is the difference in the equilibrium occupancies of the states separated by the arrow in figure 2; thus

$$\sigma(\omega) \propto \exp\left(-\frac{M\omega_0^2(Q^2+Q_0^2)}{2kT}\right)\left[\exp\left(\frac{M\omega_0^2 QQ_0}{kT}\right) - \exp\left(-\frac{M\omega_0^2 QQ_0}{kT}\right)\right]$$
$$\propto \sinh\left(\frac{\hbar\omega}{2kT}\right)\exp\left(-\frac{E_a}{kT} - \frac{(\hbar\omega)^2}{16E_a kT}\right) \tag{21}$$

where $E_a = V_1(0) = V_2(0)$ (see figure 2).

The important frequency dependence of this result is reproduced by a more mathematically rigorous derivation using the Kubo formula (Kubo 1957). It gives a characteristic peak in the frequency-dependent part of the conductivity at $\omega = 4E_a$ (see figure 3). Qualitatively similar frequency dependence is observed in a number of transition-metal oxides and other materials (Bosman and van Daal 1970, Reik 1972) and has been interpreted in terms of polaronic effects, although there are other possibilities. Note the similarity of this type of hopping to the charge-transfer optical transitions observed at defect sites like the V^- centre (hole trapped at a cation vacancy) in MgO and other oxides (Hayes and Stoneham 1985).

We conclude our consideration of small-polaron conduction by discussing the Hall effect in a solid where the conduction is by polaron hopping. The usual arguments for the magnitude and sign of the Hall coefficient do not apply in this case and indeed it is not obvious whether there should be a Hall effect at all. However, it can be shown (Friedman and Holstein 1963)) that in a magnetic field the different possible hopping paths between two sites will interfere to produce a Hall mobility that is *always* n-type, regardless of the sign of the carriers, if the orbitals concerned are s states. For p states the Hall coefficient can be positive for electrons and negative for holes, and it has been suggested that this explains the anomalous sign of the Hall coefficient in amorphous silicon and other materials (Emin 1977a, b).

2.3. Interactions between polarons; bipolaron formation

So far we have not considered how polarons interact with each other. There will be two types of interaction present. First, there will be a Coulomb repulsion between the

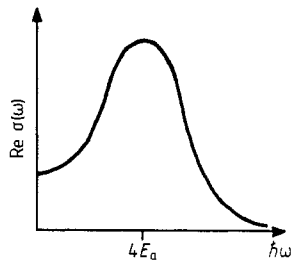

Figure 3. Sketch of the frequency dependence of the conductivity given by equation (21).

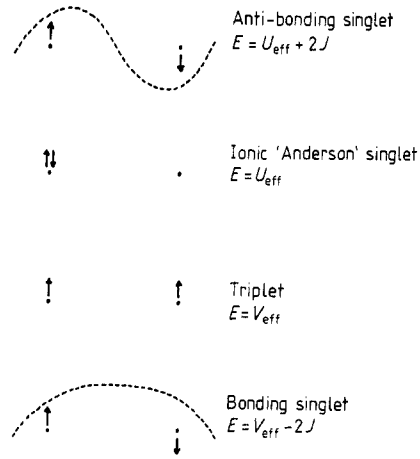

Figure 4. States of a model two-site, two-electron system where the electrons are coupled to the inter-atomic coordinate. Symbols are as defined in the text.

electrons. This may be screened out at large distances but in many oxides it is large when two electrons occupy the same site in the lattice. In a half-filled band, the consequences for the transport properties can be dramatic. If the repulsion rises beyond a certain critical multiple of the band width there is a transition to an insulating state where each site is occupied by only one electron and no current flows; there is a gap in the excitation spectrum, equal to the magnitude of the on-site repulsion, before an excited state can be produced which involves double occupancy of some lattice sites (Mott 1949, Hubbard 1964).

In addition, there may be an attractive interaction between electrons that arises because it is advantageous for them to share the same accompanying cloud of lattice distortion. We can see how this arises from the simple example described in § 2.1. Suppose we have two separate lattice oscillators, both of effective mass M and frequency ω, and each experiences a force from only one of two electrons; we denote the two forces by F_1, F_2. This might correspond to a situation with short-range electron–lattice coupling where the electrons are widely separated. Then the total gain in relaxation energy of the lattice modes is

$$E_R = -(F_1^2 + F_2^2)/2M\omega^2. \tag{22}$$

If, however, we arrange the electrons so that they are coupled to the *same* mode (if the coupling is short-range this will involve bringing them together in real space), then the relaxation energy is

$$E_R = -\frac{(F_1 + F_2)^2}{2M\omega^2} = -\frac{F_1^2 + F_2^2}{2M\omega^2} - \frac{F_1 F_2}{M\omega^2}. \tag{23}$$

In other words, we gain relaxation energy whenever we couple two electrons to the same lattice mode with forces F_1, F_2 of the same sign, and this results in an attractive force between them. The Coulomb repulsion between two carriers on the same site in the crystal is often denoted by U; the interaction with the lattice renormalises this to an effective value U_{eff}. If the attractive interaction mediated by the lattice is large enough

the net interaction may be attractive. U_{eff} will then be negative and this behaviour is therefore referred to as *negative-U behaviour*. In the case of the Jahn–Teller polarons discussed in § 4, the directional dependence of the attractive force can be quite complicated (Stoneham and Bullough 1971).

Negative-U behaviour is postulated to occur in chalcogenide glasses (Anderson 1975, Street and Mott 1975). It is supposed that a single electron in a dangling bond left unsaturated in the random network structure has a net *attraction* for a second electron. A pair of singly occupied dangling bonds is thus assumed to be unstable with respect to disproportionation into an unoccupied and a doubly occupied state. The strong electron–phonon coupling in this case results from the strong occupancy-dependent interaction of the dangling bond with lone pairs of valence electrons on nearby atoms. The dangling-bond state is then described as a *negative-U state*.

This model explains the absence of the Pauli paramagnetic susceptibility characteristic of unpaired electron spins as well as characteristics of the absorption spectrum; the unoccupied dangling bond is predicted to be an acceptor and the doubly occupied state a donor, both with rather small binding energies. The absorption spectrum therefore shows a tail below the band edge and no features in mid-gap (Street *et al* 1975). Further strong experimental support for this idea comes from the results of measurements of the transient photoconductivity of a-As_2Se_3 as a function of temperature (Thio *et al* 1984). In the liquid state above the 'glass temperature', doubly occupied and unoccupied dangling bonds are created thermally in pairs, requiring an energy E_{pair}. An additional energy equal to the modulus of the effective electron–electron attraction U_{eff} must then be supplied to produce two singly occupied states, which can act as recombination centres. The total activation energy for the recombination rate is thus $(E_{\text{pair}} + |U_{\text{eff}}|)/2$. Below the glass temperature, however, a fixed concentration of dangling bonds is frozen into the structure and the activation energy drops to $|U_{\text{eff}}|/2$. The low-temperature activated behaviour of the decay rate of the transient photoconductivity of a-As_2Se_3 therefore provides a direct measure of $|U_{\text{eff}}|$, which is found to be about 0.7 eV.

Recently several theoretical investigations of the negative-U concept have been reported. Baraff *et al* (1979, 1980) and Lannoo *et al* (1981) have calculated the relative energies of the V^0 and V^+ charge states of the silicon vacancy and find that the doubly occupied V^0 is more stable than the singly occupied V^+, while Stoneham and Sangster (1983) have investigated the relative charge-state stabilities for transition-metal ions in ionic materials. Total energy pseudopotential calculations have also been carried out with full atomic relaxation for the Se anti-site defect in c-As_2Se_3 (such calculations involving the complex range of geometries involved in the amorphous material are still not feasible) and showed that a pair of neutral anti-site defects of this type would be unstable with respect to a positively and a negatively charged defect (Tarnow *et al* 1988). The calculated energy release on disproportionation was 0.28 ± 0.08 eV per pair.

The overall net attraction of a pair of electrons can have consequences in the bulk as well as at defect sites. To see the kind of states that may form, it is instructive to consider a model two-site system where the electrons are coupled to the single inter-atomic displacement coordinate (Chakraverty *et al* 1978, Toyozawa 1981). Both the on-site and inter-site Coulomb repulsions of the electrons are reduced from their bare values U and V by the interaction with the lattice to effective values U_{eff} and V_{eff}, and either or both may become negative (attractive). Under the plausible assumption that V_{eff} is more negative than U_{eff}, the electronic ground state in the limit where we can neglect the lattice kinetic energy is a singlet bonding state with even symmetry under inversion and

energy $V_{\text{eff}} - 2J$ relative to two isolated polarons. As the energy rises we have (see figure 4) successively an even triplet state of energy V_{eff}, an odd singlet state of the type originally suggested by Anderson (1975), which is purely ionic and therefore has energy U_{eff}, and an anti-bonding singlet state with energy $U_{\text{eff}} + 2J$. Here J is an effective inter-atomic exchange integral

$$2J = 4t_{\text{eff}}^2/(U_{\text{eff}} - V_{\text{eff}}) \tag{24}$$

where t_{eff} is the effective inter-site tunnelling integral, reduced from its bare value by the reduction factor $\exp(-S)$, S being the Huang–Rhys factor of equation (17). Note how J acts as an antiferromagnetic exchange integral, stabilising the singlet ground state relative to the triplet; it has been conjectured (de Jongh 1988a, b) that this additional *magnetic* stability is important in certain circumstances for polaron pairing (but see § 3).

In a bulk solid, the expected behaviour of the system will depend on the ratio of the binding energy of the lowest two-electron state relative to two isolated polarons, denoted Δ, to the polaron band width $W = 2zt_{\text{eff}}$, where z is the number of nearest neighbours in the lattice. If $\Delta \ll W$ then the polarons interact weakly with each other, whereas if $\Delta \gg W$ we expect the polarons to form pairs that are more or less permanent. These stable bound pairs are known as *bipolarons*. There is some evidence that such a species is responsible for the activated transport observed in the diamagnetic phase of Ti_4O_7 between 130 and 150 K (Lakkis *et al* 1976, Schlenker and Marezio 1980), and for anomalies in the heat capacity of vanadium bronzes (Chakraverty *et al* 1978). These ideas have also been used in the interpretation of the absorption spectrum of $LiNbO_3$ (Koppitz *et al* 1987) and of the conductivity of WO_3 (Gehlig and Salje 1983, Schirmer and Scheffler 1982; see Stoneham 1989 for additional references).

2.4. Superconductivity

Bipolarons are similar to the Cooper pairs that occur in the BCS theory of superconductivity in that they are pairs of electrons bound together by their interaction with the lattice. They differ in that, whereas a Cooper pair consists of two electrons with opposite wavevectors (k-space pairing), a bipolaron consists of two electrons on adjacent lattice sites or even on the same lattice site (real-space pairing). These differences are a consequence of the different assumptions made when deriving their properties, in particular about the form of the force F exerted on the lattice by the electron and about the ratio of the phonon frequencies to the other energy scales in the problem.

In BCS superconductivity, the interaction with the lattice arises from the deformation potential, which expresses the dependence of the electron energy on the sample volume (see § 1). For a wide quasi-free-electron band the effect is to produce a shift in the energy that is almost independent of wavevector but only within energies of the order of the Debye frequency. For electrons whose energies differ by more than this, the lattice cannot respond quickly enough to follow their relative motion and therefore makes no contribution to their interaction. In bipolaron theories, on the other hand, it is usually assumed that the so-called anti-adiabatic limit applies, where the phonon frequencies are large compared with the electron band width and the lattice can therefore follow the electron motion essentially instantaneously. Therefore *all* the electron states in the band can be coupled by the interaction.

Despite these differences, it is possible to construct theories of superconductivity involving polarons and bipolarons. Following Alexandrov *et al* (1986b) we distinguish two cases. If the electron–phonon coupling is not too large, the binding energy Δ of

bipolarons, as discussed above, is small compared with the band width (Alexandrov 1983). The interaction between polarons can then be treated just like the electron–electron interaction in conventional BCS theory. There are two important differences: first, we assume we are in the anti-adiabatic limit so all the states interact and contribute to the superconductivity, not just those close to the Fermi energy; secondly, it is likely that the important attractive interactions are the *inter*-site ones (like V_{eff} in the example in § 2.3), so the fourier transform of the interaction into k-space has a k-dependence that we cannot ignore. These differences lead to small differences in the formula for the critical temperature from that in BCS theory; basically, the Debye energy is replaced by the polaron band width wherever it appears, and some extra terms arise. The polarons are now bound in pairs with opposite k-vectors, just as in BCS theory; the transition temperature is a strongly increasing function of the attractive interaction between them and hence of the strength of the electron–phonon coupling. Alexandrov (1983) refers to this type of superconductivity as *polaronic*.

If, however, the bipolaron binding energy Δ is large compared with the band width (Alexandrov and Ranninger 1981), the bipolarons behave as stable entities, similar to bosons except for a restriction that they may not occupy the same lattice site. They may therefore be regarded as a lattice gas of bosons interacting with a hard-core repulsion (Alexandrov *et al* 1986a), and show a Bose condensation into a superfluid state at low temperature. The real-space pairing of the electrons means that the coherence length is much smaller than for BCS superconductivity. The effective mass of the bipolarons increases with electron–phonon coupling and therefore the transition temperature *decreases* with increasing coupling strength. Furthermore the spectrum of the superfluid state appears at first sight to be gapless. However, a closer study of the system (Nasu 1987) reveals that the gap decreases exponentially with the electron–phonon coupling in the anti-adiabatic limit. Studied in the random-phase approximation, the collective excitation mode of the bipolaron condensate goes over smoothly to the pair-breaking excitation of the BCS-like state as the bipolaron binding is reduced. In addition, an insulating charge-density-wave state is possible in which the lattice distorts to open up an energy gap at the Fermi energy (Chakraverty 1981); this state competes with the superconducting one and is favoured for strong electron–phonon coupling away from the anti-adiabatic limit.

Following the discovery of the high-temperature superconducting oxides (Bednorz and Müller 1986, Chu *et al* 1987), a number of authors have suggested that a bipolaronic mechanism could be responsible for their high transition temperatures and novel properties (de Jongh 1988a, b, Prelovšek *et al* 1987, Mott 1987b, Ray 1987, Alexandrov *et al* 1987, Alexandrov 1988, Stoneham 1987, Catlow *et al* 1988). We shall not attempt to discuss here the relative merits of these pictures, but point out a few of their strengths and weaknesses. All have some difficulty in relating the superconducting and magnetic parts of the phase diagram of the materials, but all predict (correctly) that superconductivity at moderately high temperatures should be possible in materials where antiferromagnetic order is absent. Some theories predict that the maximum critical temperature will occur at the transition between the polaronic and bipolaronic regimes, where the bipolaron binding energy and the polaron band width are of the same magnitude and neither can be treated as small. This leads to grave difficulties in extracting concrete quantitative predictions from the theory. There are also problems with the apparent observation of a superconducting energy gap in the high-T_c materials. One notable success of these theories is the derivation of an upward curve in the critical magnetic field as a function of temperature in the superconducting state (Alexandrov *et al* 1987).

3. Magnetic polarons

In semiconductors containing magnetic ions there is a large exchange interaction between carrier spins and the spins of the magnetic ions, which gives rise to ferromagnetic alignment of the moments of the magnetic ions surrounding the carrier spin. This combination of carrier and magnetic polarisation cloud is referred to as a magnetic polaron. Free magnetic polarons have not yet been firmly identified. However, carriers localised by defects (donors, acceptors) can induce sizable ferromagnetic moments centred on the defect site; these bound magnetic polarons (BMP) can have ferromagnetic moments as large as about 25 μ_B.

It is well known that EuO undergoes a metal–insulator transition at $T_c = 50$ K associated with a transition from ferromagnetic order to magnetic disorder (Torrance *et al* 1972). Below 50 K conduction in EuO is metallic owing to the presence of ionised donors (oxygen vacancies). Above 50 K collapse of the donor wavefunction occurs, giving an insulator. Below 50 K the uniform magnetisation does not contribute to localisation but the localisation energy associated with the onset of magnetic disorder triggers collapse of the donor wavefunction at T_c with loss of conductivity (Emin *et al* 1987). Application of a magnetic field to the disordered magnetic state reduces the tendency of a carrier to form localised polarons and increases conductivity.

In dilute magnetic semiconductors (DMS) containing magnetic ions, such as $Cd_{1-x}Mn_xTe$, the magnetic interaction between nearest-neighbour Mn^{2+} ions is antiferromagnetic in sign. For $x < 0.7$ this material has the zincblende structure with Mn^{2+} replacing Cd^{2+} randomly. The magnetic spins are distributed in clusters of various sizes. For $x = 0.05$, for example, the probabilities of singles, pairs (spinless), open triples and closed triples are 0.54, 0.24, 0.09 and 0.02 respectively, accounting for 89% of Mn^{2+}; the rest is in larger clusters. The ground state of Mn^{2+} is $3d^5\ {}^6S_{5/2}$ and exchange interaction between the $3d^5$ electrons and spins in the s-like conduction band and p-like valence band of $Cd_{1-x}Mn_xTe$ gives rise to anomalously large Zeeman splittings of band-edge extrema; this results in unusual magneto-optical effects such as giant band-edge Faraday rotation (Gaj *et al* 1978).

Optical experiments have provided the most convincing evidence for BMP in DMS. The most direct evidence has been provided by spin-flip Raman scattering (Ramdas and Rodriguez 1988), but this has only been observed for donor BMP where magnetic polaron effects are relatively weak; the acceptor BMP where the more localised charge is expected to give greater magnetic binding energy has so far proved elusive. The spin-flip line is readily identified by its magnetic field dependence (figure 5). For $Cd_{1-x}Mn_xSe$ ($x = 0.1$) the large initial value of the slope and saturation value of the spin splitting imply an effective g-value of 100 or more. Figure 5 also shows that there is a finite spin-flip energy for $B = 0$ that is not strongly temperature-dependent, indicating a finite resultant magnetic field at the donor site. Even though, on average, when $B = 0$ the Mn^{2+} magnetisation within the donor orbit vanishes, the donor spin can relax sufficiently rapidly to align with fluctuating Mn^{2+} magnetisation to produce the net spin–spin correlation responsible for the zero-field splitting (see Wolff 1988 for a useful review). In general, temperature-independent energy shifts are characteristic of the fluctuation regime.

The model Hamiltonian for a donor BMP in a DMS is

$$H = \frac{p^2}{2m^*} - \frac{e^2}{4\pi\varepsilon_0\varepsilon r} - \alpha \sum_j (\boldsymbol{s}\cdot\boldsymbol{S}_j)\delta(\boldsymbol{r}-\boldsymbol{R}_j) \tag{25}$$

where m^* is the effective mass, ε is the dielectric constant and α is the exchange constant

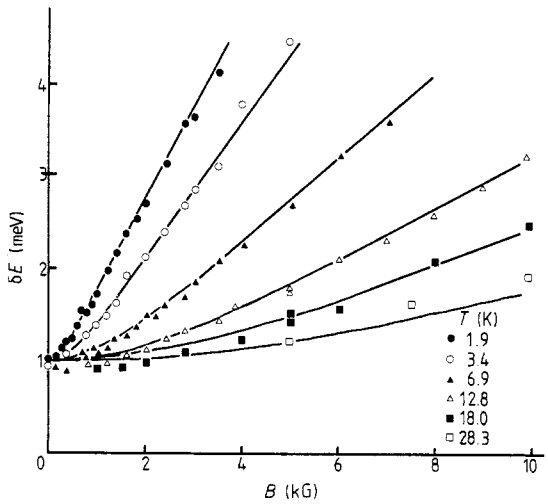

Figure 5. Spin-flip energy in $Cd_{0.9}Mn_{0.1}Se$ as a function of temperature and magnetic field; the full curves are theoretical predictions (after Heiman *et al* 1983).

for the conduction band. Note the change in notation from § 2, where α was used to denote the Fröhlich coupling constant between the carrier and the LO phonons. Usually the small anisotropy (about 5%) of m^* and ε is ignored. However, the neglect of Mn^{2+}–Mn^{2+} interactions in equation (25) can cause problems. With $x = 0.05$ one finds Curie-like behaviour with an effective x, called $\bar{x}$, of about 0.03 due to the presence of antiferromagnetically coupled Mn^{2+} pairs, which are non-magnetic. For $x \leqslant 0.1$ one can use equation (25) with reasonsable confidence by replacing x by $\bar{x}$. However, for $x \geqslant 0.1$ the $\bar{x}$ approximation becomes increasingly unsatisfactory.

One can write the spin term in equation (25) in the form $-\Sigma_j K_j(s \cdot S_j)$ and the parameter $W_0^2 \propto K^2$ is determined from fits to data of the type shown in figure 5 (see Wolff 1988), giving $W_0 = 0.56$ meV for $Cd_{0.9}Mn_{0.1}Se$. In general, the internal energy U of the BMP has two distinct temperature regimes, for $\beta W_0 \ll 1$ and $\beta W_0 \gg 1$. As temperature is reduced the carrier spin, which relaxes rapidly compared to the Mn^{2+} spins, gradually aligns with instantaneous Mn^{2+} fluctuations, causing a decrease of U; this is the fluctuation regime. When alignment of the carrier is nearly complete ($\beta W_0 \simeq 2$) the system can further reduce U by forcing Mn^{2+} spins to adopt energetically favourable configurations, leading at low T to spin saturation ($\langle s \cdot S_j \rangle \rightarrow 5/4$). When $\beta W_0 \ll 1$ (fluctuation regime) each BMP can have moment fluctuations in time over a wide range of values and this is reflected in the linewidth observed in spin-flip Raman scattering. When $\beta W_0 \gg 1$ (collective regime) the BMP has a well defined moment described by a classical Langevin formula. At $T < 2$ K, $\beta W_0 \simeq 4$ so that for $T > 2$ K fluctuations control behaviour whereas for $T < 2$ K the system is beginning to enter the collective regime.

In a typical wide-gap DMS $W_0 \simeq 6$ meV for an acceptor, an order of magnitude greater than for a donor, and the transition from the fluctuation to the collective regime is expected to occur at around 30 K. Although acceptor BMP have proved elusive, donor–acceptor pair recombination provides a means of studying acceptors; the donors and acceptors are well separated and the observed magnetic polaron effects can be assigned largely to the acceptor (Nhung and Planel 1983). A similar situation occurs for exciton luminescence in DMS (see e.g. Wong *et al* 1989). Very large magnetic-field-induced shifts of luminescence are observed and there is a magnetic contribution to the binding energy when $B = 0$. This is largely due to the more localised acceptor but there are difficulties

of interpretation because of the simultaneous presence of localisation induced by ionic disorder.

Emin and Hillery (1988) have considered the formation of magnetic polarons and magnetic bipolarons in an antiferromagnet and have concluded that the conditions necessary for magnetic bipolaron formation are hard to fulfill (see also § 2.4).

4. Polarons and the Jahn–Teller effect

If electron–lattice coupling is introduced into a system where otherwise the electronic ground state would be orbitally degenerate then the Jahn–Teller theorem (Jahn and Teller 1937) states that there always exists a distortion of the lattice that lowers the total energy while lowering the symmetry and removing the orbital degeneracy. The only exception to this rule is a linear system, which is stable against bending; note also that the Kramers degeneracy of a system containing an odd number of electrons, which arises because of spin effects, is not removed by the distortion.

This has the consequence that the Born–Oppenheimer approximation breaks down. The electron and lattice motions cannot be separated, since which distortion of the lattice is the energetically favourable one depends on which of the manifold of orbitally degenerate states the electron occupies. The energy gained by the system as a result of the distortion is known as the Jahn–Teller energy E_{JT}; it is analogous to the relaxation energy E_R in the electron–phonon coupling of non-orbitally degenerate states that was discussed in § 2.

The Jahn–Teller effect is best understood in molecular systems and in defects in solids (Englman 1972, Stoneham 1985). In the case of a defect, the discussion is usually simplified by assuming that the degenerate electronic system, often a rare-earth or transition-metal ion, couples to a single local mode of the lattice only. This enables qualitative solutions for the coupled electronic and vibrational (so-called *vibronic*) states to be found relatively easily (Sturge 1967). We note in passing that, while many of the best-known examples involve cation impurities, Jahn–Teller effects are also observed at anion impurities or intrinsic defects (Hayes and Stoneham 1985).

More complicated, however, is the question of the *cooperative Jahn–Teller effect* (Elliott *et al* 1972, Gehring and Gehring 1975). Here one considers a whole array of structural units, each of which would be in isolation an orbitally degenerate system such as a transition-metal ion of the type discussed above plus its surroundings, but arranged so that a distortion in one system exerts a force on its neighbours. Such an array typically shows a phase transition from a high-temperature disordered phase where the macroscopically averaged expectation value of the distortion at some wavevector is zero to a low-temperature ordered phase where it is not. It is usual to consider such a transition by analogy with the spin systems whose statistical mechanics has been extensively studied; if neighbouring Jahn–Teller systems are coupled so that it is energetically favourable for them to distort in the *same* orientation then the analogy is with a ferromagnetic spin system, whereas if neighbouring units tend to distort in opposite directions the assembly behaves like an antiferromagnetic spin system.

Now we consider the effects of charge transport in such a system. In general, if one of the structural units is orbitally degenerate when occupied by n electrons, it will not also be orbitally degenerate when occupied by $n + 1$ or $n - 1$ electrons. For example, Cu^{2+} is a Jahn–Teller ion but Cu^{+} and Cu^{3+} are not. Thus any full theory of charge transport in a Jahn–Teller system must include the interactions of Jahn–Teller ions with

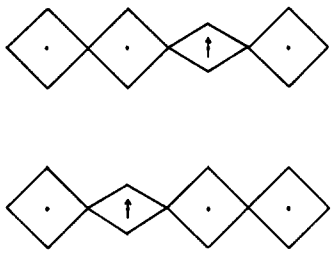

Figure 6. Schematic illustration of the mobility of a Jahn–Teller polaron or soliton due to the translational symmetry of the lattice.

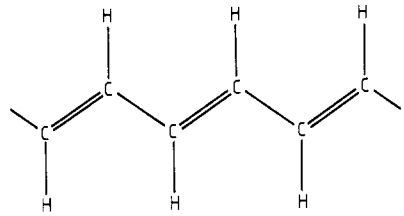

Figure 7. Structure of *trans*-polyacetylene (*t*-PA).

non-Jahn–Teller ions, and there does not seem to be any comprehensive theoretical treatment of this problem. When considering the limits of a lightly doped insulator or semiconductor, whether n-type or p-type, there are two distinct cases to consider. First, the material may consist in the undoped state of Jahn–Teller ions, whose orbital degeneracy will be removed by doping. Secondly, the undoped material may contain no Jahn–Teller units and these may be created by the doping. This second case has been given some theoretical attention. Höck *et al* (1983) have performed a variational treatment at zero temperature of a single carrier in a narrow band with an Einstein oscillator model for the phonon spectrum. They refer to the resulting quasi-particle containing a carrier plus its associated asymmetric lattice distortion as a *Jahn–Teller polaron*. The results are similar to those from treatments of the conventional polaron; as the Jahn–Teller relaxation energy E_{JT} is increased, the width of the resulting polaron band is reduced.

The Jahn–Teller polaron may be considered as an excitation of the (non-linear) field theory representing the coupled electron–phonon system, quite distinct from the linear excitations of the non-interacting system (the Bloch electrons and phonons). This type of excitation is often referred to as a *soliton* (see § 1). This usage is similar to that adopted in our discussion of conducting polymers in § 5; note that, just as in *trans*-polyacetylene, the Jahn–Teller polaron may be regarded as an excitation of a state with an underlying degeneracy. We illustrate this, for the case where the presence of a carrier produces a single Jahn–Teller ion, in figure 6. The mobility of the soliton at low temperatures arises from the property already discussed in § 2.2; since the coupled electron–lattice system still possesses translational invariance, the ground state of the system is a coherent superposition of the soliton on all possible sites.

The Jahn–Teller relaxation energies associated with the Cu^{2+} ion, configuration $3d^9\ ^2D$, in octahedral environments are exceptionally large, of the order of 0.1–0.3 eV. Some of the novel polaron theories of superconductivity invoke this to explain the high transition temperatures of the new copper oxide superconductors (de Jongh 1988a, b, Stoneham 1987). The large Jahn–Teller energies are supposed to result in large forces between polarons, which then bind strongly to form the bipolarons that condense to form a superfluid state.

5. Defects in conjugated polymers

Most conjugated polymers behave as strongly anisotropic (i.e. approximately one-dimensional) semiconductors, with band gaps typically in the range 1.4–3.0 eV. Whereas

in traditional (three-dimensional) semiconductors, the charge carriers are free electrons and holes, with energies close to the band edges, conduction in conjugated polymers is believed to be mediated by charged defects, which are the species formed on doping (the electron–lattice coupling is sufficient to create localised, polaronic states). At low dopant levels, conductivity appears to be largely due to the mobility of these defects themselves along a chain; at higher concentrations, various more complicated mechanisms, involving hopping between chains or between defects as the conductivity-limiting process, have been proposed (and reviewed by Kivelson 1986). The possibility of applications has fuelled the study of the these defect states; for example, the doping process in a number of polymers can be caried out electrochemically and reversibly (Chung *et al* 1984, and numerous others), leading to applications both as electrodes in lightweight rechargeable batteries (Shacklette *et al* 1989) and in semiconductor devices (Burroughes *et al* 1989). The conductivities attainable by doping can be as high as that of copper (Basescu *et al* 1987), and the consequent conductivity : mass ratio (which is an important quantity when considering applications in the aerospace industries) is substantially higher than that of any metal.

In the past, much of the investigation into the properties of conjugated polymers has been carried out within the framework of the model Hamiltonian of Su, Schrieffer and Heeger (SSH) (Su *et al* 1980) and its continuum approximation due to Takayama, Lin-Liu and Maki (TLM) (Takayama *et al* 1980). An extensive review of this work has been given by Heeger *et al* (1988). The SSH Hamiltonian is based on extended Hückel theory, and will be outlined below; the TLM version is often used, as it lends itself to analytic solutions in some simple cases. Other authors, such as Stafstrom and Bredas (1988), have used methods based on semi-empirical Hartree–Fock theory, which have been standard tools of quantum chemistry for many years (see, for example, Pople and Beveridge 1970). In our self-consistent studies of the electronic behaviour and of the distortion of the molecular geometry of these systems (Wallace 1989a, b) we have carried out self-consistent numerical calculations using both the SSH Hamiltonian and the CNDO (complete neglect of differential overlap) and INDO (intermediate neglect of differential overlap) parametrisations of Hartree–Fock theory; the work of Stafstrom and Bredas was performed within the MNDO (modified neglect of differential overlap) parametrisation, which retains some of the two-centre integrals neglected by CNDO and INDO, but is consequently rather more difficult to parametrise. Sadlej (1985) has outlined, and to some extent compared, all three methods.

5.1. trans-Polyacetylene and the SSH Hamiltonian

The archetypal example of a conjugated polymer is *trans*-polyacetylene (*t*-PA), whose structure is shown in figure 7. The carbon atoms are sp^2 hybridised, with one p_z (π) electron per carbon atom not being included in the σ-bond backbone (taking the plane of the polymer to be the xy plane). In simple Hückel theory, these π electrons would be expected to form a half-filled band of width $4t_0$, where t_0 is the hopping integral $\langle n|H|n+1\rangle$ between adjacent carbon atoms, giving a one-dimensional metal. However, a Peierls distortion of the *t*-PA chain occurs, doubling the size of the unit cell and introducing a band gap in the middle of the band (for a more extensive discussion of the bonding in *t*-PA and of the nature of the Peierls distortion in one-dimensional metals, see e.g. Hayes (1985)). This was first proposed for *t*-PA by Longuet-Higgins and Salem (1959); in chemical terms, the backbone consists of alternating single and double bonds (although the bonds are actually rather more similar in terms of length and of bond order

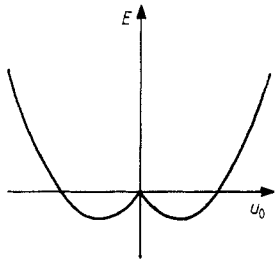

Figure 8. Total energy E of t-PA against dimerisation u_0.

than is implied by this representation, it does form a useful basis from which to discuss the behaviour of the system).

The SSH Hamiltonian sought to quantify this dimerisation, and to provide a framework for the study of states other than the uncharged ground state. It assumed that the only electrons of interest were the π-electrons, i.e. that the σ (bonding) and σ^* (antibonding) bands would lie respectively well below and above the π bands, and that the only atomic displacement of interest was along the chain direction (this displacement, for the nth atom, was labelled u_n). Only nearest-neighbour interactions were considered, but the hopping integral was expanded linearly, as a function of inter-atomic separation, about t_0. The distortion of the σ bonds was given a simple quadratic form and the Hamiltonian thus became

$$H = -\sum_n \sum_s [t_0 - \alpha(u_n - u_{n+1})](c_{ns}^+ c_{n+1,s} + c_{n+1,s}^+ c_{ns}) + \tfrac{1}{2}K \sum_n (u_{n+1} - u_n)^2 \tag{26}$$

where c_{ns}^+ (c_{ns}) is the creation (annihilation) operator for a p_z orbital at site n and with spin s. The parameters t_0, α and K were derived from experimental results.

A uniformly dimerised chain corresponds to a displacement pattern $u_n = (-1)^n u_0$, where u_0 is constant, and the one-electron energies (the eigenvalues of the Hamiltonian matrix) are now given by

$$E_k^2 = [2t_0 \cos(ka)]^2 + [4\alpha u_0 \sin(ka)]^2. \tag{27}$$

The total energy of the system (replacing the sum of electronic energies by a suitable integral) is given by

$$E_{\text{tot}} = -(4Nt_0/\pi)\mathrm{E}(1 - z^2) + NKt_0^2 z^2/2\alpha^2 \tag{28}$$

where z is the dimensionless quantity $2\alpha u_0/t_0$ and $\mathrm{E}(1 - z^2)$ is an elliptic integral. For small z^2 (z is a directly observed quantity, the ratio of band gap to band width, and is approximately 0.14), the elliptic integral can be approximated by

$$\mathrm{E}(1 - z^2) \simeq 1 + \tfrac{1}{2}(\ln 4/|z| - \tfrac{1}{2})z^2 \tag{29}$$

which has a component linear in z, so the minimum energy state will have non-zero u_0 (since z is known, α can be fitted so that the minimum energy does indeed occur at this value). A graph of total energy against u_0 is shown in figure 8. Note that there are two equal minima, corresponding to positive and negative u_0; the two degenerate states are

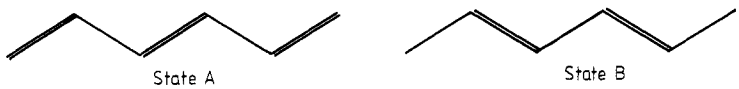

Figure 9. Degenerate states of *t*-PA.

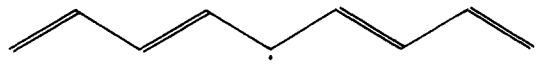

Figure 10. Topological representation of a soliton in *t*-PA. The defect consists of a change in the backbone of the chain from state A to state B, with one localised electronic state associated with an unpaired electron at the point where this change occurs. In reality, both the change of state and the localised electronic state are spread out over a longer distance than implied by this representaion.

labelled A and B in figure 9 and are clearly related by symmetry. The values chosen by SSH for the empirically fixed parameters t_0, α and K gave $u_0 \simeq 0.04$ Å, although we use values that give $u_0 \simeq 0.03$ Å, in better agreement with experimental evidence (Yannoni and Clarke 1983).

5.2. Solitons and the dimerisation parameter

In order properly to study the distortion around a defect, it is helpful to define some parameter that varies smoothly along a polymer chain. SSH used the *staggered displacement* $u_n \equiv (-1)^n u_n$, but this is not always a smoothly varying quantity, since the equilibrium bond length around a charged defect is altered from its uncharged value, and this will lead to rapid apparent fluctuations in u_n. Instead, we define a *dimerisation parameter* d_n by

$$d_n = (-1)^n (b_{n,n+1} - b_{n-1,n}) \tag{30}$$

where $b_{n,n+1}$ is the length of the bond joining atoms n and $n+1$. In the ground state the bond lengths will, like u_n, alternate along the chain, and d_n will be equal to $4u_0$ (which is positive in state A and negative in state B). In a real chain, without periodic boundary conditions, the degree of dimerisation actually increases slightly towards the ends, but this effect is neglected for the purposes of this discussion.

Now we can consider the basic defect in *t*-PA. Imagine that we constrain one end of a polymer chain to be in state A and the other to remain in state B (how this may be done is outlined below). The simplest model of the intervening (defect) region is that of figure 10, where the transition from state A to state B is abrupt, and there is an unpaired electron (free radical) completely localised on the central carbon atom. In fact, the change in dimerisation parameter is more extended; in the TLM approximation, the dimerisation parameter is proportional to $\tanh(x/l)$, where the half-width l is approximately seven C–C spacings, and the wavefunction associated with the defect extends over a similar range. Conditions of bonding (from the chemical viewpoint) or orthogonality between the wavefunctions (in the quantum-mechanical framework) dictate that there remain one localised state, and within the Hückel approximation this state has energy at mid-gap (the mid-gap state may be regarded as a local suppression of the Peierls distortion). This type of defect, whose existence was originally proposed by Longuet-Higgins and Salem (1959) and which was first considered in a quantitative manner (albeit in a very simple form) by Pople and Walmsley (1962), is known as a *soliton* (originally

kink soliton or *bond-alternation defect*), in view of its obvious non-linearity and of the similarity of the TLM continuum model to more familiar equations of field theory, which display more traditional solitonic solutions. In addition, its large size means that it can move freely along a chain without changing its shape, another important soliton property (the effective mass of a soliton was estimated by SSH to be approximately six times the free-electron mass). Its spin–charge relations are slightly unusual. In the neutral chain it is associated with an unpaired electron and thus has spin $S = \frac{1}{2}$; this electron can be removed, or another electron added, to give defects with spin $S = 0$ and charge $\pm e$.

The reason why these spin–charge relations do not violate Kramers' theorem, and the means by which we were able to enforce the change from state A to state B above, are closely allied. If we start with a defect-free chain, with an even number of carbon atoms, each end of the chain is terminated by a double bond. Now if we attempt to create just one soliton, one end of the chain is forced to end with a single bond, thereby leading to a second unpaired electron on a final carbon atom (just as in the discussion of the soliton, this electron will not be associated entirely with this one carbon atom, but its wavefunction will extend into the chain over a few carbon atoms, and the distortion will rather resemble that of half a soliton). We cannot create just one localised state on an even chain: we must create them in pairs. In practice, the localised state we created at the end of the chain will move into the chain's interior, in order to maximise the reduction in σ-bond energy associated with the change in dimerisation amplitude around the defect, and will become another soliton (this was first predicted by Su and Schrieffer (1980), and has been verified dynamically within the CNDO framework (Wallace 1989a,b)). The dynamics of this process are in accordance with the property of a soliton (in mathematical terms) that any distortion injected at one end of an infinite non-linear system will resolve itself into a number of solitons by the time it reaches the other end. If instead of a chain with an even number of carbon atoms we consider a chain with an odd number of carbon atoms, and no defects, the bonding pattern dictates that it is forced to end with a single bond; this state is unstable, and the localised electronic state associated with this single bond will move up the chain as a soliton. Thus one end of the chain will be state A and the other state B; the total number of p_z electrons in the chain is odd (and all the other electrons are paired), so a single localised solitonic state is a necessary property of the ground state.

5.3. General (non-degenerate) polymers

trans-Polyacetylene is, however, not a typical conjugated polymer: the degeneracy between states A and B makes it a very special case (although a few other polymers that display this same degeneracy can be found, they have much larger repeat units and the concept of an extended defect becomes somewhat tenuous). Other polymers, such as *cis*-polyacetylene (*c*-PA) and polypyrrole (PPY) are more typical; in figure 11 they are drawn in their stable forms, the *cis–transoid* and *trans–cisoid* isomers respectively (if we were to define a parameter u_0 to represent the amplitude of the Peierls distortion as for *t*-PA above, then one of these polymers would have a ground state corresponding to positive u_0 and the other to negative u_0; graphs of total energy against u_0, analogous to figure 8, are shown in figure 12). The reasons why these particular isomers are stable is not difficult to see: in *c*-PA, steric repulsion between neighbouring hydrogens is reduced in the *cis–transoid* isomer, while in PPY it is the fact that the C–N bond length necessary to complete the heterocycle is smaller in the *trans–cisoid* isomer than in the *cis–transoid* isomer that leads to the stability of the former. It is immediately clear, therefore, that

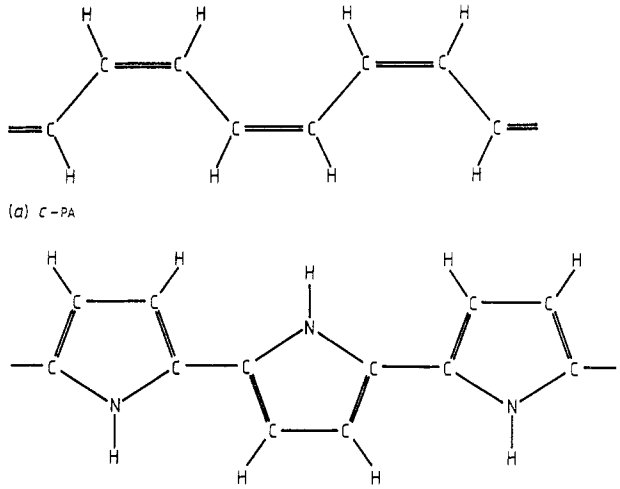

Figure 11. Ground states of (*a*) *cis*-polyacetylene (*c*-PA) and (*b*) polypyrrole (PPY).

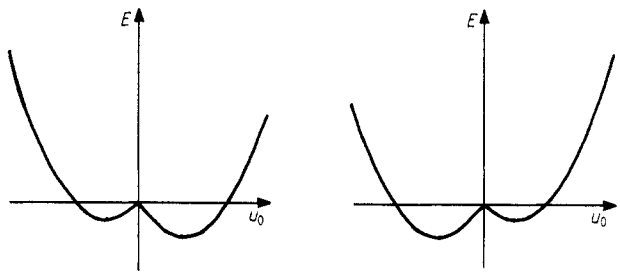

Figure 12. Total energy against dimerisation for non-degenerate polymers (see text).

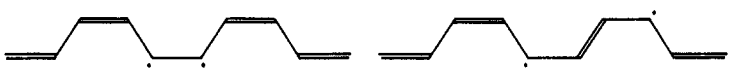

Figure 13. Topological representations of a polaron in *c*-PA. In contrast to the case of a soliton, the amplitude of the defect is free to vary instead of being fixed by the boundary conditions, and this is illustrated here. The right-hand diagram shows how a polaron may, in one limit, be considered as a bound soliton pair, as the chain changes from *cis–transoid* to *trans–cisoid* and back again. There are two localised electronic states; the defect will be stable if at least one is unoccupied or doubly occupied.

solitons cannot exist in *c*-PA or PPY, since their existence implies extended regions of both A and B states. The only simple defect type that can occur in these polymers is represented by a reduction in the dimerisation parameter over a limited region of the chain. It is not so easy to represent in chemical bonding terms as the soliton, especially as the actual amount by which the chain is distorted is variable and not dictated by topology, but two versions, on a schematic carbon backbone, are shown in figure 13. The first represents a minimum dimerisation of zero whereas the second has a *negative* dimerisation in the centre of the defect. This form of defect is associated with a pair of

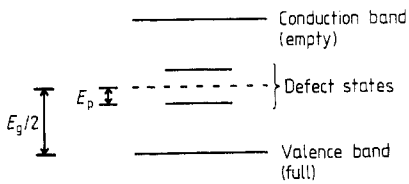

Figure 14. Energy level diagram for a polaron.

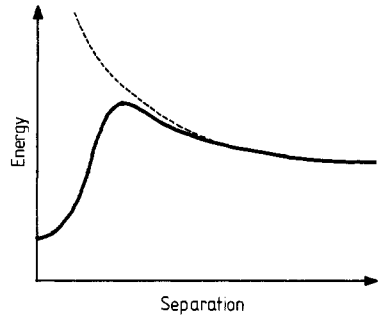

Figure 15. Schematic representation of the energy of a pair of like-charged polarons as a function of their separation (the broken curve is the value obtained by ignoring the attractive component due to interaction between the two lattice distortion patterns).

localised states, which are (within the SSH model) symmetrically located about mid-gap, at energies $\pm E_p$ relative to mid-gap (figure 14) (the value of E_p depends on the geometry of the defect: a very shallow distortion of the lattice will only move the states out of their respective bands by a small amount, while at the opposite extreme the defect will split into a pair of solitons and the two states will meet at mid-gap). Clearly, this type of defect will not be stable in an uncharged, unexcited singlet state, as it will decay trivially into the ground state, but addition of a single charge to the defect will leave no such decay route available and will stabilise the defect. Again, the defect as it actually occurs will be substantially more extended than shown in figure 13. Its form is entirely analogous to that of a polaron in more familiar materials such as the III–V semiconductors or the alkali halides (§ 2), in that it consists of a localised charge coupled to the longitudinal optic modes (i.e. the dimerisation) of the host. Sometimes, this class of defect in conjugated polymers is known as a *polaron soliton*, but there seems no reason for qualifying the description and it is now more generally known as a polaron. The charge : - spin relationship of such polarons is traditional; they carry both charge and spin.

Since the defect of figure 13 has two unpaired electrons, we are free to remove them both, leading to a doubly charged state of spin zero. If this is stable with respect to decay into two polarons, we have a *negative-U* effect (see also § 2.3), and the resulting defect is known as a bipolaron (in the case of *t*-PA, stability with respect to two-soliton decay also needs to be considered, but it should be understood that the negative-*U* effect as such is used specifically to account for stability with respect to two-polaron decay, and that two-soliton decay is an entirely separate process). The localised electronic states in a bipolaron are, like those of a polaron, symmetrically spaced about mid-gap, and the energy level diagram of figure 14 applies equally to polarons and bipolarons.

5.4. Quantitative study of defect formation

We can now start to investigate quantitatively the energies of each of the defects we have been considering. The SSH model provides an easily comprehensible framework and will be used in the first instance, but our results from CNDO will be introduced on occasions as they include effects that are neglected by SSH. The SSH model, as originally proposed, does not consider electron–electron interaction in any form, for example,

and the more commonly used empirical corrections to the model fail to do so in any reliable manner, so results where electron–electron interaction is likely to be important, and in particular the question of the negative-U effect, are best handled by CNDO.

Consider t-PA first, and assume that we are dealing with a chain of even length (to make the boundary conditions as simple as possible). The band gap induced by the Peierls distortion is approximately 1.4 eV, and so (taking the centre of the gap, i.e. the localised p_z-orbital energy, as the zero of energy) the purely electronic energy required to create a soliton is 0.7 eV, since one electron must be raised from the valence band edge to mid-gap. The dimerisation amplitude is reduced around a soliton, though, and there is a gain of approximately 0.25 eV from this source, so the total cost of creating a soliton (more precisely, half the cost of creating a soliton pair) is about 0.45 eV (the exact analytic result from calculations within the TLM model is E_g/π, where E_g is the band gap). If we include electron–electron interaction in the simplest manner possible, i.e. via an on-site Hubbard U, with an extra term in the Hamiltonian given by

$$H_{e-e} = U\Sigma(c^+_{n+}v_{n+} - \tfrac{1}{2})(c^+_{n-}c_{n-} - \tfrac{1}{2}) \tag{31}$$

where the second suffix refers to the electron spin as before, we increase the energy of a charged soliton by about 0.11 eV, a value derived indirectly from the results of photo-induced absorption on t-PA. It can be shown (Kivelson and Heim 1982, Wallace 1989a) that this increase in energy is $\frac{1}{4}U_{eff}$, where

$$U_{eff} = U\Sigma\,|a^{soli}_n|^4 \tag{32}$$

and a^{soli}_n is the coefficient of the nth p_z orbital in the soliton wavefunction, and that the photo-induced absorption peak corresponding to absorption from charged solitons is shifted from mid-gap by $\frac{1}{2}U_{eff}$. This peak is observed at 0.48 eV (Blanchet *et al* 1983), implying $U_{eff} = 0.44$ eV. The energy of an uncharged soliton is decreased by $\frac{1}{4}U_{eff}$.

In the case of polaron formation (see § 5.3), the gain in energy due to the localised reduction in dimerisation amplitude is strongly dependent on the depth of the distortion, as is the loss of energy due to the creation of the localised electronic states; this is why self-consistent calculations are necessary to study defects in conjugated polymers with any accuracy. Analytic solution of the TLM model yields a value of $E_g/2\sqrt{2}$ for E_p (see § 5.3) (Fesser *et al* 1983), while numerical self-consistent solution of the SSH model gives states slightly closer to the centre of the band gap, with $E_p = \pm 0.48$ eV. The consequent lattice distortion is fairly slight, but extends over a slightly larger region than that of a soliton, and leads to a gain in energy of $E_{relax} = 0.29$ eV. The energy required to form a hole polaron is

$$E_{pol} = E_g/2 + (E_g/2 - E_p) - E_{relax} \tag{33}$$

where $E_g/2 = 0.7$ eV is the energy required to remove an electron from the top of the valence band and $(E_g/2 - E_p) = 0.22$ eV is the energy required to raise the now unpaired electron from the top of the valence band to the localised state at $-E_p$. Hence $E_{pol} = 0.63$ eV, which represents a saving of 0.07 eV over the simple removal of an electron from the valence band without lattice distortion. The equivalent calculation for an electron polaron is slightly different, but yields the same energy E_{pol}. CNDO predicts a slightly greater lattice distortion than SSH, but the results are qualitatively similar. Inclusion of a Hubbard U into the SSH model has no effect on the polaron energy.

The calculated energies required to form charged solitons and charged polarons in t-PA suggest that solitons will be formed preferentially. Polarons will still be formed, of course, where boundary conditions prohibit soliton formation (e.g. on adding a single

charge to a chain of even length), since the calculated energy required to create a pair of solitons, one charged and one uncharged (which is the only way to store a single charge between solitons within these boundary conditions) is 0.9 eV.

We now consider the polaron creation energy associated with a single charge on the chain of a polymer other than *c*-PA. Modifying the SSH model to represent *t*-PA gives

$$E_g = 1.94\,\text{eV} \qquad E_p = 0.77\,\text{ev} \qquad E_{relax} = 0.28\,\text{eV}$$

and hence (see equation (33)) $E_{pol} = 0.89$ eV, a saving of 0.8 eV on the direct removal of an electron from the valence band. Again, CNDO predicts a slightly greater lattice distortion, presumably because it takes into account the repulsion of the now positively charged carbon atoms and aims to spread the defect out further as a result.

The situation becomes more interesting when a second charge is added to the polaron, as bipolar formation becomes possible. Again, we deal with *t*-PA first. Self-consistent calculations within both the SSH and CNDO models agree that the charged bipolaron in *t*-PA is unstable with respect to decay into two solitons; we cannot calculate the energy of a bipolaron as such, since the geometry relaxation routines will simply lead to this decay (the energy required to add the second charge is, in the SSH model, rather less than that required to add the first, since the first charge will form a polaron, of energy 0.63 eV, and the second will convert this into a soliton pair, of energy 1.12 eV).

Now consider the possibility of bipolaron formation in *c*-PA, where this route for instability is forbidden. Within SSH, the energy saving due to the σ-bond distortion is 0.43 eV, and the new states are at ± 0.45 eV. As expected, they are closer to the band centre than the states due to the single polaron, as the lattice distortion is greater, but in fact the energy required to form a bipolaron is independent of the energies of these states, since they will either be empty, in the case of a hole bipolaron, or completely full, for an electron bipolaron. Thus the total energy required to form a bipolaron is $(E_g - E_{relax}) = 1.51$ eV, a saving of 0.27 eV on the formation of a pair of polarons.

It is not surprising that bipolarons are predicted to be stable in the SSH model, since electron–electron interaction is not considered and the main term acting to destabilise the bipolaron is therefore ignored. Adding a Hubbard U term again gives a Coulomb energy of about 0.25 eV for the bipolaron, a figure, like its equivalent in *t*-PA, derived indirectly from the results of photo-induced absorption, in this case on polythiophene by Vardeny *et al* (1986). However, the uncertainty in this figure is substantial and the evidence for the stability, or otherwise, of bipolarons in the SSH model should not be regarded as conclusive.

The results of CNDO calculations on *c*-PA should be more reliable, however, since electron–electron interaction is considered in a self-consistent manner. Relative to the energies required to remove electrons from a chain at fixed molecular geometry, we find that the energy required to create a polaron in a chain of length 32 (i.e. a molecule of $C_{32}H_{34}$) is -0.60 eV and that the corresponding bipolaron energy is -3.11 eV, substantially more stable than a polaron pair. This result is in good agreement with the experimental evidence that bipolarons are the species formed on doping at moderate to high concentrations (this evidence will be discussed in § 5.5). At lower concentrations, polarons should still be observed under normal conditions, owing to the significant barrier to polaron recombination (the Coulomb repulsion of a polaron pair is a much longer-range effect than their mutual attraction due to the lattice distortion); a graph representing how the energy of a polaron pair depends on the separation of the polarons is shown in figure 15. Crude numerical estimates of the barrier height (modelling the polaron charge density by a Gaussian distribution, and the Coulomb integral by a term

proportional to $1/(R^2 + a^2)^{1/2}$, with a about two lattice spacings) lead to values of the order of 1 eV; the situation is complicated by the presence of the (charged) dopant ions, however.

5.5. Experimental evidence concerning defect types

Evidence for the nature of the defects formed on doping conjugated polymers has been obtained from electron spin resonance (ESR) measurements, relying on the different charge : spin relationships of the various defects discussed above. Ikehata *et al* (1980) measured the magnetic susceptibility of *t*-PA as a function of dopant level, up to 13.8% doping (expressed as the ratio of dopant molecules to carbon atoms). They found that *t*-PA exhibited a small Curie susceptibility in the undoped state (about 4×10^{-4} spins per carbon atom), which they attributed to the presence of a small number of neutral solitons (presumably imposed by the boundary conditions, since the energy required to create a soliton pair is sufficiently large that thermally created solitons should not exist in appreciable quantities). On doping, the spin concentration *decreased*, corresponding to the ionisation of these solitons (recall that charged solitons are spinless, while uncharged solitons do carry spin), and no Curie susceptibility is seen at dopant concentrations higher than about 2–3% (as charged solitons are formed directly from the ground-state chain). Above about 7%, a significant Pauli contribution to the susceptibility is seen, representing an effective semiconductor–metal transition.

Similar experiments have been performed on PPY (Scott *et al* 1983) and the susceptibility is observed to increase initially, until at about 1% doping it reaches a maximum and starts to tail off. This is consistent with the formation of polarons initially, to be replaced by bipolarons at higher dopant levels as the polarons combine. Chen and Heeger (1986) repeated these experiments on polythiophene, but at each doping level measured how the susceptibility varied with time. They found that, even when significant susceptibility (and hence appreciable polaron concentration) was initially observed, it would decline to a very small value after a few days, as the polarons overcame the kinetic barrier to recombination. How this decay in the susceptibility varied with temperature was not investigated. The timescale over which the decay process occurs will presumably depend on the precise polymer under consideration, as well as on the sample temperature, and results akin to those of Scott *et al* would be expected whenever such an experiment is carried out over a time-scale smaller than the characteristic polaron–bipolaron conversion time (which will chiefly depend on the barrier height of figure 15). Simple numerical models of this competitive polaron–bipolaron formation are easy to write, and yield results in good qualitative agreement with the experimental data.

Acknowledgments

The authors are grateful to Dr G A Gehring, Dr A M Stoneham and Mr S V Traven for their comments on the manuscript.

References

Alexandrov A S 1983 *Zh. Fiz. Khim.* **57** 273 (Engl. Transl. 1983 *Russ. J. Phys. Chem.* **57** 167)
Alexandrov A S and Ranninger J 1981 *Phys. Rev.* B **24** 1164

Alexandrov A S 1988 *Phys. Rev.* B **38** 925
Alexandrov A S, Ranninger J and Robaszkiewicz S 1986a *Phys. Rev. Lett.* **56** 949
—— 1986b *Phys. Rev.* B **33** 4526
Alexandrov AS, Samarchenko D A and Traven S V 1987 *Zh. Eksp. Teor. Fiz.* **93** 1007 (Engl. Transl. 1987 *Sov. Phys.-JETP* **66** 567)
Anderson P W 1975 *Phys. Rev. Lett.* **34** 953
Baraff G A, Kane E O and Schlüter M 1979 *Phys. Rev. Lett.* **43** 956
—— 1980 *Phys. Rev.* B **21** 5662
Basescu N, Liu Z-X, Moses D, Heeger A J, Haarman H and Theophilou N 1987 *Nature* **327** 403
Bednorz G and Müller K A 1986 *Z. Phys.* B **64** 189
Blanchet G B, Fincher C R, Chung T C and Heeger A J 1983 *Phys. Rev. Lett.* **50** 1938
Bosman A J and van Daal H J 1970 *Adv. Phys.* **19** 1
Burroughes J H, Jones C A and Friend R H 1989 *Synth. Met.* **28** C735
Catlow C R A, Tomlinson S M, Islam M S and Leslie M 1988 *J. Phys. C: Solid State Phys.* **21** L1085
Chakraverty B K 1981 *J. Physique* **42** 1351
Chakraverty B K, Sienko M J and Bonnerot J 1978 *Phys. Rev.* B **17** 3781
Chen J and Heeger A J 1986 *Solid State Commun.* **58** 251
Chu C W, Hor P H, Meng R L, Gao L, Huang Z J and Wang Y Q 1987 *Phys. Rev. Lett.* **58** 405
Chung T-C, Kaufman J H, Heeger A J and Wudl F 1984 *Phys. Rev.* B **30** 702
Das Sarma S 1983 *Phys. Rev.* B **27** 2590
de Jongh L J 1988a *Solid Commun.* **65** 963
—— 1988b *Physica* C **152** 171
Elliott R J, Harley R T, Hayes W and Smith S R P 1972 *Proc. R. Soc.* A **328** 217
Emin D 1977a *Phil. Mag.* **35** 1189
—— 1977b *Solid State Commun.* **22** 409
—— 1983 *Comments Solid State Phys.* **11** 35
Emin D and Hillery M S 1988 *Phys. Rev.* B **37** 4060
Emin D, Hillery M S and Liu N H 1987 *Phys. Rev.* B **35** 641
Emin D and Holstein T 1976 *Phys. Rev. Lett.* **36** 323
Englman R 1972 *The Jahn–Teller Effect in Molecules and Crystals* (New York: Wiley)
Fesser K, Bishop A R and Campbell D K 1983 *Phys. Rev.* B **27** 4804
Feynman R P 1955 *Phys. Rev.* B **97** 660
Friedman L and Holstein T 1963 *Ann. Phys.*, **21** 494
Gaj J A, Galazka R R and Uawrocki M 1978 *Solid State Commun.* **25** 193
Gehlig R and Salje E 1983 *Phil. Mag.* B **47** 229
Gehring G A and Gehring K A 1975 *Rep. Prog. Phys.* **38** 1
Hayes W 1985 *Contemp. Phys.* **26** 421
Hayes W and Stoneham A M 1985 *Defects and Defect Processes in Nonmetallic Solids* (New York: Wiley)
Heeger A J, Kivelson S, Schrieffer J R and Su W-P 1988 *Rev. Mod. Phys.* **60** 781
Heiman D, Wolff P and Warnock J 1983 *Phys. Rev.* B **27** 4848
Höck K H, Nickisch H and Thomas H 1983 *Helv. Phys. Acta* **56** 237
Holstein T 1959 *Ann. Phys., NW* **8** 343
Huang K and Rhys A 1950 *Proc. R. Soc.* A **204** 406
Hubbard J 1964 *Proc. R. Soc.* A **281** 401
Ikehata S, Kaufer, J, Woerner T, Pron A, Druy M A, Sivak A, Heeger A J and MacDiarmad A G 1980 *Phys. Rev. Lett.* **45** 1123
Jahn H A and Teller E 1937 *Proc. R. Soc.* A **161** 220
Kiefl R F, Kadono R, Brewer J H, Luke G M, Yen H K, Celio M and Ansaldo E J 1989 *Phys. Rev. Lett.* **62** 792
Kittel C 1963 *The Quantum Theory of Solids* (New York: Wiley)
Kivelson S 1986 *Solitons* ed. S E Trallinger, V E Zakharov and V L Pokrovsky (Amsterdam: North-Holland) p 301
Kivelson S and Heim D E 1982 *Phys. Rev.* B **26** 4278
Koppitz J, Schirmer O F and Kuznetsov A I 1987 *Europhys. Lett.* **4** 1055
Kubo R 1957 *J. Phys. Soc. Japan* **12** 570
Kuiper P, Kruizanger G, Ghijsen J, Sawatzky G A and Verweij H 1989 *Phys. Rev. Lett.* **62** 221
Lakkis S, Schlenker C, Chakravarty B K, Buder R and Marezio M 1976 *Phys. Rev.* B **14** 1429
Landau L D 1932 *Phys. Z. Sowjet.* **3** 664

Lang I G and Firsov Y A 1963 *Zh. Teor. Eksp. Fiz.* **43** 1843 (Engl. Transl. 1963 *Sov. Phys.–JETP* **16** 1301)
Lannoo M, Baraff G A and Schlüter M 1981 *Phys. Rev.* B **24** 955
Longuet-Higgins H C and Salem L 1959 *Proc. R. Soc.* A **251** 172
Loveland R J, Le Comber P and Spear W E 1972 *Phys. Rev.* B **6** 3121
Mott N F 1949 *Proc. R. Soc.* **62** 416
—— 1987a *Conduction in Non-Crystalline Materials* (Oxford: OUP)
—— 1987b *Nature* **327** 185
Nasu K 1987 *Phys. Rev.* B **35** 1748
Nhung T and Planel R 1983 *Proc. 16th Int. Conf. Semiconductors; Physica* B **117–118** 488
Nicholas R J, Hopkins M A, Brummell M A and Leadley D R 1988 *Interfaces, Quantum Wells and Superlattices* Nato ASI Series B, vol. 179, ed. C R Leavens and R Taylor (New York: Plenum) p 243
Peeters F M and Devreese J T 1984 *Solid State Phys.* **38** 81
Pekar S I 1950 *Zh. Teor. Eksp. Fiz.* **20** 510 (in Russian)
Pople J A and Beveridge D L 1970 *Approximate Molecular Orbital Theory* (New York: McGraw-Hill)
Pople J A and Walmsley S H 1962 *Mol. Phys.* **5** 15
Prelovšek P, Rice T M and Zhang F C 1987 *J. Phys. C: Solid State Phys.* **20** L229
Ramdas A K and Rodriguez S 1988 *Semicond. Semimet.* **25** 345
Ray D K 1987 *Phil. Mag. Lett.* **55** 251
Reik H G 1972 *Polarons in Ionic Crystals and Polar Semiconductors* ed. J T Devreese (Amsterdam: North-Holland)
Roth S and Bleier B 1987 *Adv. Phys.* **36** 385
Sadlej J 1985 *Semi-Empirical Methods in Quantum Chemistry* (Chichester: Ellis Horwood)
Sasai M and Fukutome H 1986 *Solid State Commun.* **58** 735
Schirmer O F and Scheffler M 1982 *J. Phys. C: Solid State Phys.* **15** L645
Schlenker C and Marezio M 1980 *Phil. Mag.* B **42** 453
Scott J C, Pfluger P, Krounbi M T and Street G B *Phys. Rev.* B **28** 2140
Shacklette L W, Jow T R, Maxfield M and Hatami R 1989 *Synth. Met.* **28** C655
Spear W E 1974 *Adv. Phys.* **23** 523
Stafstrom S and Bredas J L 1988 *Phys. Rev.* B **38** 4180
Stoneham A M 1985 *Theory of Defects in Solids* (Oxford: Clarendon)
—— 1987 *AERE Report* M3639 (unpublished)
—— 1989 *J. Chem. Soc. Faraday Trans.* II **85** 505
Stoneham A M and Bullough R 1971 *J. Phys. C: Solid State Phys.* **3** L195
Stoneham A M and Sangster M J L 1983 *Radiat. Eff.* **73** 267
Street R A and Mott N F 1975 *Phys. Rev. Lett.* **35** 1293
Street R A, Searle T M and Austin I G 1975 *Phil. Mag.* **32** 431
Sturge M D 1967 *Solid State Phys.* **20** 92
Su W P and Schrieffer J R 1980 *Proc. Natl. Acad. Sci.* **77** 5626
Su W P, Schrieffer J R and Heeger A J 1980 *Phys. Rev.* B **22** 2099 (erratum 1983 **28** 1138)
Takayama H, Lin-Liu Y R and Maki K 1980 *Phys. Rev.* B **21** 2388
Tarnow E, Payne M C and Joannopoulos J D 1988 *Phys. Rev. Lett.* **61** 1772
Thio T, Monroe D and Kastner M A 1984 *Phys. Rev. Lett.* **52** 667
Torrance J B, Shafer M W and McGuire T R 1972 *Phys. Rev. Lett.* **27** 1168
Toyozawa Y 1961 *Prog. Theor. Phys.* **26** 29
—— 1981 *J. Phys. Soc. Japan* **50** 1861
Trullinger S E, Zakharov V E and Pokrovsky V L (ed.) 1986 *Solitons (Modern Problems in Condensed Matter Physics* vol. 17) (Amsterdam: North-Holland)
Vardeny Z, Ehrenfreund E, Brafman O, Nowak M, Schaffer H, Heeger A J and Wudl F 1986 *Phys. Rev. Lett.* **56** 671
Wallace D S 1989a *DPhil Thesis* Oxford University (unpublished)
—— 1989b *Synth. Met.* **28** D457
Wolff P A 1988 *Semicond. Semimet.* **25** 413
Wong K S, Hayes W and Ryan J F 1989 *J. Phys. Condens. Matter* **1** 3115
Yannoni C S and Clarke T C 1983 *Phys. Rev. Lett.* **51** 1191
Zabusky N J and Krustal M D 1956 *Phys. Rev. Lett.* **15** 240