Burstein shift of the contact exciton

I. Egri
Institut für Theoretische Physik der Rheinisch-Westfälischen Technischen Hochschule Aachen, D-5100 Aachen, West Germany
(Received 10 September 1984)

With the use of the contact exciton model, the influence of a finite concentration of conduction electrons on the exciton energy is studied. Screening, gap renormalization, and the Burstein shift are taken into account. The absolute exciton energy increases with carrier concentration, while the binding energy decreases. However, no Mott transition occurs.

I. INTRODUCTION

The optical properties of semiconductors and insulators are altered drastically when there is a finite concentration of free carriers in the conduction band. This can be achieved by doping as, e.g., in the work of Raz, Gedanken, Even, and Jortner.[1] While the following considerations can be equally well applied to a finite hole concentration in the valence band, the situation is slightly different when holes and electrons are simultaneously present, because then recombination takes place. To avoid this complication, we do not discuss experiments where the carriers are generated optically or by raising the temperature.

In the framework of band theory, the presence of a nonvanishing carrier concentration causes the Burstein shift:[2] As a consequence of the exclusion principle, only unoccupied states are available for interband transitions. The threshold energy for direct absorption is therefore given by

$$\tilde{E}_g = E_g + \frac{\hbar^2 k_F^2}{2\mu} , \tag{1.1}$$

where E_g is the fundamental gap, k_F the Fermi wave vector of the conduction electrons, and μ the reduced mass of valence and conduction band. The situation is sketched in Fig. 1.

As is well known, excitonic effects modify the absorption spectrum. This is due to the Coulomb interaction between electrons and holes and must be taken into account also in the present case of a degenerate semiconductor. It does not only produce bound states (the excitons), but it also gives rise to self-energy corrections of electrons and holes; this leads to a renormalization of the fundamental gap.

In a degenerate semiconductor, the free carriers screen the Coulomb interaction thereby reducing the excitonic binding energy. This effect has been treated by discussing the spectrum of a Yukawa or similar potential.[3,4] In these types of theories, however, the exclusion principle or, more generally, exchange effects are neglected and therefore no Burstein shift is obtained. The most complete treatment of excitons in degenerate semiconductors is due to Mahan.[5] Using a diagrammatic approach, including screening and exchange, he calculated the absorption spectrum for different densities.

In this paper we present a pedestrian version of Mahan's work, using a simplified exciton model, namely, the contact exciton,[6] and the equation of motion method. Our theory can be regarded as an amalgamation of Burstein's theory with conventional exciton theory. This means that we take into account the exclusion principle, screening, and gap renormalization. However, we do not treat effects outside the framework of exciton theory such as, e.g., the relaxation of the conduction electrons or Auger broadening. Therefore, our exciton has an infinite lifetime; we do not obscure this failure by introducing a phenomenological damping constant.

Owing to the simplicity of our model, we can solve the problem analytically. In the weak binding limit, we find an exponentially small binding energy, in agreement with Mahan. An important result of our treatment is the observation that no Mott transition occurs; i.e., we have an exciton with energy below $\tilde{E}_g$ for arbitrarily small potential. However, exciton states with energies in the range between $\tilde{E}_g$ and E_g are degenerate with recombination processes from the Fermi sea in the conduction band (if unoccupied

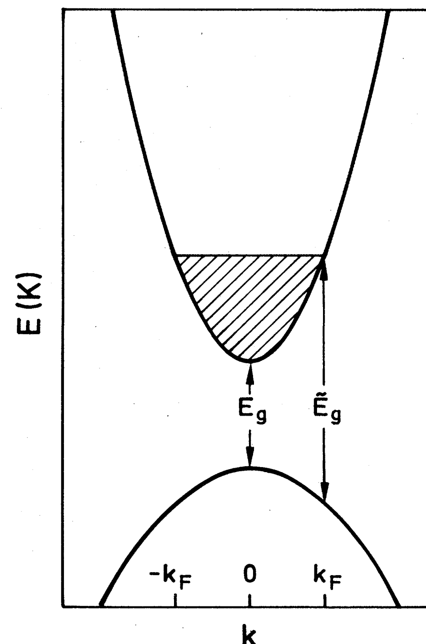

FIG. 1. Energy bands of the model under consideration. The valence band is full (no holes), the conduction band is filled with a degenerate electron gas up to k_F. E_g is the band gap, $\tilde{E}_g$ the threshold energy for optical absorption. Valence electrons with $|\mathbf{k}| < k_F$ cannot make direct transitions.

 ©1985 The American Physical Society

states are available in the valence band). This may lead to a Fano-type damping,[7] but is outside the scope of the present paper.

II. EIGENVALUE EQUATION

The contact exciton model is defined by the following Hamiltonian:

$$H=H_0+H_c \ ,$$
$$H_0=\sum_{\mathbf{k}}[E_c(\mathbf{k})c_{\mathbf{k}}^\dagger c_{\mathbf{k}}+E_v(\mathbf{k})v_{\mathbf{k}}^\dagger v_{\mathbf{k}}] \ ,$$
$$H_c=-V_0\sum_j c_j^\dagger v_j v_j^\dagger c_j \ . \tag{2.1}$$

H_0 is the Hamiltonian of the band model. From the Coulomb interaction, only the direct electron-hole part is retained, and is modeled by a contact interaction, effective only for electrons and holes on the same lattice site. It is written down in Wannier representation, where $c_j^\dagger, v_j^\dagger$ create a conduction or valence electron, respectively, in a Wannier state, located at $\mathbf{R}_j$.

Introducing the electron-hole-pair operator

$$S_{\mathbf{k}}(\mathbf{q})=v_{\mathbf{k}}^\dagger c_{\mathbf{k}+\mathbf{q}} \ , \tag{2.2}$$

the interaction term can be written as

$$H_c=-V_0\frac{1}{N}\sum_{\mathbf{k},\mathbf{k}',\mathbf{q}} S_{\mathbf{k}}^\dagger(\mathbf{q})S_{\mathbf{k}'}(\mathbf{q}) \ . \tag{2.3}$$

The exciton operator is defined as a linear combination of electron-hole pairs with the same total wave vector $\mathbf{q}$,

$$S_{\mathbf{q}}=(N)^{-1/2}\sum_{\mathbf{k}}\phi(\mathbf{k},\mathbf{q})S_{\mathbf{k}}(\mathbf{q}) \ , \tag{2.4}$$

and, within exciton theory, it is requested to satisfy the following equation of motion:

$$[S_{\mathbf{q}},H]=E(\mathbf{q})S_{\mathbf{q}} \ . \tag{2.5}$$

Then, $\phi(\mathbf{k},\mathbf{q})$ is the (Fourier transform of the envelope) wave function and $E(\mathbf{q})$ the energy of the exciton.

In order to evaluate the commutator in (2.5), we need $[S_{\mathbf{k}}(\mathbf{q}),H]$. It is given by

$$[S_{\mathbf{k}}(\mathbf{q}),H_0]=[E_c(\mathbf{k}+\mathbf{q})-E_v(\mathbf{k})]S_{\mathbf{k}}(\mathbf{q}) \ , \tag{2.6}$$

$$[S_{\mathbf{k}}(\mathbf{q}),H_c]=-V_0\frac{1}{N}\sum_{\mathbf{k}'',\mathbf{k}',\mathbf{q}'}(v_{\mathbf{k}}^\dagger v_{\mathbf{k}''}v_{\mathbf{k}'}^\dagger c_{\mathbf{k}'+\mathbf{q}'}\delta_{\mathbf{k}+\mathbf{q},\mathbf{k}''+\mathbf{q}'} - c_{\mathbf{k}''+\mathbf{q}'}^\dagger c_{\mathbf{k}+\mathbf{q}}v_{\mathbf{k}'}^\dagger c_{\mathbf{k}'+\mathbf{q}'}\delta_{\mathbf{k},\mathbf{k}''}) \ . \tag{2.7}$$

From (2.7) it is clear that electron-hole pairs are not true eigenstates of the system: they couple to more complicated states, namely, electron-hole pairs dressed by intraband transitions. This is where the dynamics of the conduction electrons comes into play and, in a rigorous theory, one has to take them into account. We stay within the framework of exciton theory and truncate the hierarchy starting with (2.7) by applying the random-phase approximation (RPA). This means that we replace the product of four operators in (2.7) by all possible contributions of two operators times the expectation value of the remaining two. We then get

$$[S_{\mathbf{k}}(\mathbf{q}),H_c]_{\mathrm{RPA}}=-V_0\frac{1}{N}\sum_{\mathbf{k}'}[f_c(\mathbf{k}')+1-f_v(\mathbf{k}')]S_{\mathbf{k}}(\mathbf{q}) - V_0[f_v(\mathbf{k})-f_c(\mathbf{k}+\mathbf{q})]\frac{1}{N}\sum_{\mathbf{k}'}S_{\mathbf{k}'}(\mathbf{q}) \ , \tag{2.8}$$

with

$$\langle v_{\mathbf{k}}^\dagger v_{\mathbf{k}'}\rangle=f_v(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'}=\delta_{\mathbf{k},\mathbf{k}'} \ ,$$
$$\langle c_{\mathbf{k}}^\dagger c_{\mathbf{k}'}\rangle=f_c(\mathbf{k})\delta_{\mathbf{k},\mathbf{k}'}=\theta(k_F-|\mathbf{k}|)\delta_{\mathbf{k},\mathbf{k}'} \ . \tag{2.9}$$

Here, we have made explicit use of the fact that the valence band is full and the electron gas in the conduction band is degenerate.

Having applied the RPA, Eqs. (2.6) and (2.8) contain only the operators $S_{\mathbf{k}'}(\mathbf{q})$, and the eigenvalue equation (2.5) can be solved with the ansatz (2.4). On the other hand, the intraband dynamics has disappeared; the conduction electrons are present only via their occupation number $f_c(\mathbf{k})$.

The first line of (2.8) contains only $S_{\mathbf{k}}(\mathbf{q})$, just as (2.6). Therefore, it can be absorbed into the contribution from H_0 and leads to the gap renormalization. Defining the number density as

$$n=\frac{1}{N}\sum_{\mathbf{k}}f_c(\mathbf{k}) \ , \tag{2.10}$$

we have finally

$$[S_{\mathbf{k}}(\mathbf{q}),H]=[E_c(\mathbf{k}+\mathbf{q})-E_v(\mathbf{k})-nV_0]S_{\mathbf{k}}(\mathbf{q}) + V_0[1-f_c(\mathbf{k}+\mathbf{q})]\frac{1}{N}\sum_{\mathbf{k}'}S_{\mathbf{k}'}(\mathbf{q}) \ . \tag{2.11}$$

Inserting this expression into the eigenvalue equation, the following equation can be derived:[6]

$$1=V_0\frac{1}{N}\sum_{\mathbf{k}}\frac{\theta(|\mathbf{k}+\mathbf{q}|-k_F)}{E_c(\mathbf{k}+\mathbf{q})-E_v(\mathbf{k})-nV_0-E(\mathbf{q})} \ . \tag{2.12}$$

The root of this equation, $E(\mathbf{q})$, is the exciton energy. The Burstein shift is contained in (2.12) through the numerator: states below k_F are excluded. As already mentioned, the term nV_0 in the denominator represents the gap renormalization. In order to include screening, we must specify the dependence of V_0 on the density n. We put

$$V_0=V(1-\lambda k_F) \ , \tag{2.13}$$

where λ is an additional parameter, controlling the importance of screening. The form of (2.13) is motivated by the fact that screening in a degenerate electron gas is described by the squared Thomas-Fermi wave vector q_{TF}^2. This quantity scales as $n^{1/3}$, i.e., as k_F.

We evaluate (2.12) for isotropic parabolic bands

$$E_c(\mathbf{k})=E_g^0+\frac{\hbar^2k^2}{2m_e}, \quad E_v(\mathbf{k})=-\frac{\hbar^2k^2}{2m_h} \ , \tag{2.14}$$

and a spherical Brillouin zone

$$v_B=\frac{4\pi}{3}k_0^3 \ . \tag{2.15}$$

For direct transitions, i.e., $\mathbf{q}=0$, this leads to the following integral:

$$1=\frac{3V_0}{W}\int_\alpha^1 dx\frac{x^2}{x^2+\epsilon} , \quad (2.16)$$

with

$$W=\frac{\hbar^2k_0^2}{2\mu}, \quad \alpha=\frac{k_F}{k_0}=n^{1/3}, \quad \epsilon=\frac{E_g^0-nV_0-E(0)}{W} . \quad (2.17)$$

W is the joint bandwidth, α the filling factor, and ϵ the dimensionless exciton energy measured relative to the renormalized gap

$$E_g(n)=E_g^0-nV_0 . \quad (2.18)$$

III. RESULTS AND DISCUSSION

Solving (2.16) for ϵ, we obtain the energy $E(0)$ of the bound state, if it exists. First of all, we note that the integral diverges logarithmically for $\epsilon=-\alpha^2$, or

$$E(0)=E_g(n)+W\alpha^2=\tilde{E}_g . \quad (3.1)$$

Therefore, $\tilde{E}_g$ is the solution of (2.16) for $V_0=0$. This means that the exciton energy approaches the Burstein-shifted threshold for vanishing potential. In addition, due to this divergence, there is a solution $\epsilon>-\alpha^2$ for every finite $V_0>0$. Since the Mott transition is defined as the point where the exciton energy merges into the continuum, $E(0)=\tilde{E}_g$, we conclude that it takes place at $V_0=0$; i.e., there is no Mott transition for finite V_0. Note also that $\epsilon=0$ plays no particular role.

This is in contrast with the undoped case ($n=\alpha=0$). There, the continuum starts at E_g^0 and a finite $V_0>W/3$ is necessary to produce a bound state $\epsilon>0$.[8,9] This is due to the singularity in the density of states at $\mathbf{k}=0$. However, $\mathbf{k}=0$ is excluded from the integration as soon as $\alpha>0$. If one used a screened potential like (2.13) but put $\alpha=0$ in (2.16), one would erroneously predict a Mott transition at $V_0=W/3$ and $\epsilon=0$. This mistake is made in the conventional formulation of the problem as a Schrödinger equation with a screened potential.[3,4]

The integral in (2.16) can be carried out analytically and we obtain

$$1=\frac{3V_0}{W}\left[1-\alpha+\sqrt{|\epsilon|}f\left(\frac{(1-\alpha)\sqrt{|\epsilon|}}{\alpha+\epsilon}\right)\right] , \quad (3.2)$$

where

$$f(x)=\begin{cases}\tanh^{-1}(x) & -\alpha^2<\epsilon\leq 0 ,\\ -\tan^{-1}(x) & \epsilon\geq 0 .\end{cases} \quad (3.3)$$

Note again that $f(x)$ is continuous at $\epsilon=0$ and the two branches of the solution $\epsilon(V_0,\alpha)$ join smoothly at

$$\tilde{V}_0=\frac{W}{3(1-\alpha)} . \quad (3.4)$$

The upper branch vanishes as $\alpha\to 0$.

In Fig. 2 we have plotted the solution of (3.2) for some representative parameters.

For low densities, the following approximate, but explicit solution can be derived by expanding $f(x)$. For $V_0>\tilde{V}_0$ we get

$$E(0)=E_g^0-E_B^0+\frac{4W}{\pi}\left(\lambda k_F\sqrt{\epsilon_0}+\frac{n}{3\sqrt{\epsilon_0}}\right) , \quad (3.5)$$

where $E_B^0=W\epsilon_0$ is the exciton binding energy for $n=0$.

For weak binding and $V_0<\tilde{V}_0$ we obtain

$$E(0)=\tilde{E}_g-2\frac{\hbar^2k_F^2}{\mu}\frac{1-\alpha}{1+\alpha}\exp(-1/\delta) , \quad (3.6)$$

where

$$\delta=\frac{\alpha V_0}{2(1-\alpha)(\tilde{V}_0-V_0)} . \quad (3.7)$$

This is essentially Mahan's result.[5]

This completes the discussion of the influence of free carriers on the contact exciton. Details may be due to the model; trends are believed to be general. It must be remembered, however, that the real situation is complicated by the dynamics of the conduction electrons.

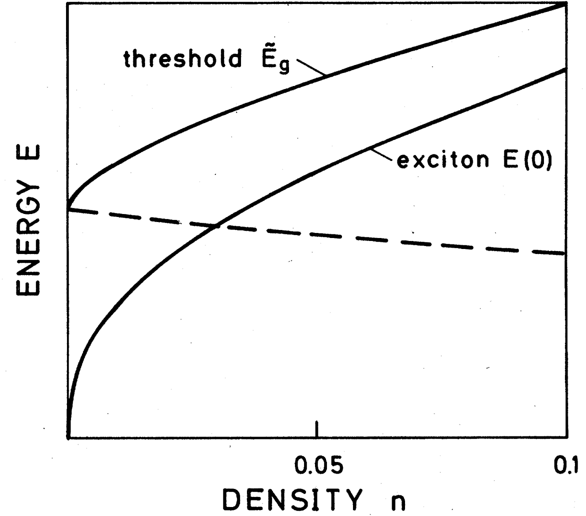

FIG. 2. Exciton energy $E(0)$ and threshold energy $\tilde{E}_g$ as functions of the density according to (3.2). Energies are measured in units of W, the joint bandwidth. A bare potential of $V=0.7W$ and a screening parameter $\lambda k_0=1$ has been chosen. The dotted line represents the renormalized gap $E_g(n)$.

[1] B. Raz, A. Gedanken, U. Even, and J. Jortner, Phys. Rev. Lett. **28**, 1643 (1972).
[2] E. Burstein, Phys. Rev. **93**, 632 (1954).
[3] J. G. Gay, Phys. Rev. B **4**, 2567 (1971).
[4] J. D. Dow, in *Proceedings of the Twelfth International Conference on the Physics of Semiconductors,* edited by M. Pilkuhn (Teubner, Stuttgart, 1974), p. 957.
[5] G. D. Mahan, Phys. Rev. **153**, 882 (1967).
[6] I. Egri, J. Phys. C **12**, 1843 (1979).
[7] U. Fano, Phys. Rev. **124**, 1866 (1961).
[8] R. Ramirez, L. Falicov, and J. Kimball, Phys. Rev. B **2**, 3383 (1970).
[9] I. Egri, Z. Phys. B **42**, 99 (1981).