PHYSICAL REVIEW B **107**, 235204 (2023)

Exciton ground-state energy with full hole warping structure

Roland Combescot[1] and Shiue-Yuan Shiau[2]
[1]*Laboratoire de Physique, Ecole Normale Supérieure, PSL Université, Sorbonne Université,*
Paris Diderot Université, CNRS, 24 rue Lhomond, F-75005 Paris, France
[2]*Physics Division, National Center for Theoretical Sciences, 10617, Taipei, Taiwan*

(Received 15 February 2023; revised 26 April 2023; accepted 9 June 2023; published 26 June 2023)

Most semiconductors, in particular III-V compounds, have a complex valence band structure near the band edge, due to degeneracy at the zone center. One peculiar feature is the warping of the electronic dispersion relations, which are not isotropic even in the vicinity of the band edge. When the exciton, all important for the semiconductor optical properties, is considered, this problem is usually handled by using some kind of angular averaging procedure, that would restore the isotropy of the hole effective dispersion relations. In the present paper, we consider the problem of the exciton ground-state energy for semiconductors with zinc-blende crystal structure, and we solve it exactly by a numerical treatment, taking fully into account the warping of the valence band. In the resulting four-dimensional problem, we first show exactly that the exciton ground state is fourfold degenerate. We then explore the ground-state energy across the full range of allowed Luttinger parameters. We find that the correction due to warping may in principle be quite large. However, for the considered semiconductors with available data on the band structure, the correction turns out to be in the 10%–15% range.

DOI: 10.1103/PhysRevB.107.235204

I. INTRODUCTION

Optical properties of semiconductors are of utmost interest both for fundamental and applied purposes [1,2]. With respect to these properties, excitonic excitations play a prominent role, especially in the visible light spectrum. Hence, knowing the exciton ground-state properties in these compounds is of fundamental importance. Yet, this is in general not a simple matter theoretically, since generically the valence band of the most relevant semiconductors is degenerate at the band edge.

The exciton is made of one conduction electron and one valence hole bound through Coulomb interaction. To solve for the exciton states, one first has to know the electron and hole dispersions, and then the Coulomb interaction between them. In the first step, in order to obtain the band states, one has to resort to purely numerical first-principles calculations, relying on various hypotheses, such as *GW* method [3]. One can then use, for the specific case investigated, the two-body Bethe-Salpeter equation [3], that incorporates the Coulomb interaction, to resolve the exciton eigenstates. In general, these first-principles treatments are more comprehensive in reflecting the details of specific material band and dielectric properties.

Alternatively, semi-analytical approaches such as the $\mathbf{k} \cdot \mathbf{p}$ perturbation theory [4] and Luttinger-Kohn approach [5,6] provide an easy way to obtain the analytical expressions of the electronic energy dispersions that preserve the crystal symmetry. They make few approximations since they merely assume that only small enough wavevectors are important, which is mostly valid in the case of the exciton we want to investigate here. Needless to say, the momentum matrix elements between bands or the Luttinger parameters that appear in these semi-analytical approaches have to be extracted from first-principles calculations, or from experiments, to reflect realistic properties for a specific material. However these approaches allow a systematic analysis in terms of these parameters. This is the approach we follow in the present paper for semiconductors with cubic zinc-blende crystal structure, a most important class of semiconductors.

In general, the hole dispersion in these zinc-blende structures is not isotropic, and the corresponding departure from isotropy is usually called the warping (see below). Note that this anisotropy is different and more complex than the simpler one where the equal energy surfaces are ellipsoids, instead of simple spheres. Hence the simple ellipsoidal models [7,8] cannot be used in our case. Our goal in this paper is to make a systematic study of the impact of the warping of the hole energy dispersion on the ground-state exciton binding energy. In particular we will understand how, despite the hole warping, the excitonic ground-state structure remains compatible with the cubic symmetry. In this paper, we shall use the Luttinger-Kohn approach, in which the valence band parameters are embodied in the three Luttinger parameters. It is important to note that the paper by Luttinger [5] proves in full generality, by symmetry arguments, that the form of their Hamiltonian is the most general for crystals with zinc-blende structure, provided that one remains within a quadratic expansion for the hole band structure, which is enough for a large exciton with wave function built from small enough wave vectors.

Unfortunately, even if the hole energy dispersion is analytically given by the Luttinger Hamiltonian, it is still not a simple matter to obtain the exciton ground state. Indeed, it has been long recognized that, even in the vicinity of the band edge, the resulting valence band dispersion relations, while respecting the cubic symmetry of the crystal, are not isotropic. The source of this anisotropy (warping) actually comes from the $(k_x^2k_y^2 + k_y^2k_z^2 + k_z^2k_x^2)$ term in the hole energies given in Eq. (5) below. As a result, the equal energy surfaces are warped, and this whole matter makes the hole properties difficult to handle toward the resolution of the exciton ground state. This complexity is usually disposed of by taking an appropriate angular average to go back to an isotropic situation. Once the isotropic solution is obtained, it is then possible to include the warping in a perturbative way [9]. In the present paper, we will not make use of such a simplification, but rather explore the exciton ground-state energy in these zinc-blende semiconductors in full generality.

In this standard case of semiconductors with cubic zinc-blende crystal structure, the states at the valence band edge have a p-wave character, which leads to a sixfold degeneracy when spin is taken into account. When spin-orbit coupling is strong enough to move a twofold degenerate split-off subband far enough in energy from the valence band edge, so that it becomes irrelevant to the excitonic structure, one is left with a fourfold degeneracy. This is the situation we will consider. As for the dielectric constant that appears in the Coulomb potential, it has been pointed out [10] that the static screening is a valid approximation when the exciton binding is a small fraction of the band gap, as in III-V and II-VI semiconductors. To focus on the warping effect, we will not here consider the possible anisotropy of the dielectric constant [11,12].

In the next section, we recall the hole effective kinetic energy Hamiltonian, standard for these compounds. In the following Sec. III, we write the Schrödinger equation for the exciton under the form of an integral equation appropriate for our study. We show in Sec. IV in full generality that, due to the cubic symmetry, the exciton ground state has a fourfold degeneracy (omitting the twofold additional degeneracy coming from the conduction electronic spin). Finally, in Sec. V, we solve numerically our integral equation. We show that in principle the warping leads to an increase of the exciton binding energy, which may be quite important. However, in practice, for the tabulated semiconductors we have found, this increase is limited to a 10%–15% range. Note that these corrections are significantly larger than those obtained from second-order perturbation theory [9]. Nevertheless, markedly higher values are possible, considering the imprecision of the known valence band parameters data, or other semiconductors yet to be investigated.

II. FORMALISM

In the semiconductor compounds we are considering, the bottom of the conduction band located at the zone center is nondegenerate (except for the spin degree of freedom), but the top of the valence band located at the zone center is, in the absence of spin-orbit coupling, threefold degenerate (forgetting the hole spin), with a p-wave-like character for the corresponding wave functions. As the exciton wave function is predominantly made of electron and hole states in the vicinity of the zone center, this degeneracy has to be taken into account in the exciton wave function. Taking spin into account, this threefold degeneracy becomes a sixfold degeneracy. When spin-orbit coupling is taken into account, this degeneracy is lifted into a fourfold degeneracy, plus a twofold one. In standard compounds, these last two degenerate electronic states are shifted to energy low enough, so that they are irrelevant to the building of the exciton. This is the situation we will restrict ourselves to in the following. The corresponding dispersion relations are sketched in Fig. 1.

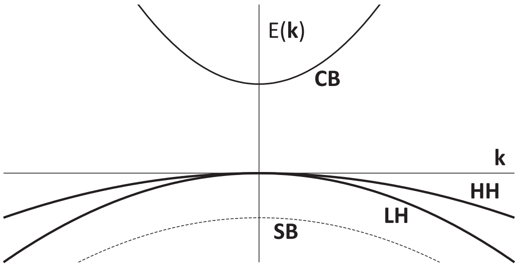

FIG. 1. Sketch of the relevant electronic dispersion relations $E(\mathbf{k})$ involved in our exciton calculation. We have a simple single parabolic dispersion relation CB for the conduction band electron. We have in the valence band two dispersion relations indicated by the two lines HH and LH. However these dispersion relations are coupled and the corresponding hole kinetic energy is given by the 4×4 matrix in Eq. (1) (taking spin into account). Finally the dashed line SB indicates the part of the valence band split-off by spin-orbit coupling, which is low enough in energy to play little role in the excitonic structure. To be complete, one has to take into account the spin degrees of freedom, which are not indicated in our sketch for simplicity.

The four-dimensional hole subspace we now focus on has the symmetry character of a $J = 3/2$ angular momentum [5], resulting from the composition of the p-wave-like wave functions and of the $1/2$ spin. For the present purpose, it is more convenient to use the notations $|1\rangle \equiv |3/2\rangle$, $|2\rangle \equiv |1/2\rangle$, $|3\rangle \equiv |-1/2\rangle$ and $|4\rangle \equiv |-3/2\rangle$ to denote the standard fourfold hole basis.

Following the approach of Luttinger and Kohn [5,6] (with slightly different notations [13]) in the absence of magnetic field, the hole kinetic energy part $H_h(\mathbf{k})$ of the exciton Hamiltonian is given by

$$H_h(\mathbf{k}) = \frac{\hbar^2}{2m_0}\begin{pmatrix} G+\Delta & D & F & 0 \\ D^* & G-\Delta & 0 & F \\ F^* & 0 & G-\Delta & -D \\ 0 & F^* & -D^* & G+\Delta \end{pmatrix}$$
$$\equiv \frac{\hbar^2}{2m_0}(G\,\mathbb{1} + h_h), \tag{1}$$

where m_0 is the vacuum electron mass (providing a mass scale), and G, Δ, D, and F are defined from the hole wave vector $\mathbf{k}$ by

$$\begin{aligned} G &= \gamma_1 k^2, \quad \Delta = \gamma_2\left(k^2 - 3k_z^2\right), \\ D &= -2\sqrt{3}\gamma_3(k_x - ik_y)k_z, \\ F &= -\sqrt{3}\gamma_2\left(k_x^2 - k_y^2\right) + 2i\sqrt{3}\gamma_3 k_x k_y, \end{aligned} \tag{2}$$

where x, y, and z are the cubic axes, and γ_1, γ_2, and γ_3 are the Luttinger [5] constants characterizing the semiconductor.

One may notice that, within a multiplicative factor, h_h^2 reduces to the unit matrix

$$\left(\frac{2m_0}{\hbar^2}H_h(\mathbf{k}) - G\,\mathbb{1}\right)^2 = h_h^2 = \lambda^2\,\mathbb{1} \tag{3}$$

with

$$\lambda^2 \equiv \Delta^2 + |D|^2 + |F|^2 = 4\gamma_2^2 k^4 + 12\left(\gamma_3^2 - \gamma_2^2\right)\left[k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2\right] \tag{4}$$

so that the doubly degenerate eigenvalues $E_\pm(\mathbf{k})$ of $H_h(\mathbf{k})$ are given by the well-known result [14]

$$\frac{2m_0}{\hbar^2}E_\pm(\mathbf{k}) = \gamma_1 k^2 \pm 2\left[\gamma_2^2 k^4 + 3\left(\gamma_3^2 - \gamma_2^2\right)\left[k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2\right]\right]^{1/2}. \tag{5}$$

The square root on the right-hand side gives rise to the well-known warping, implying in general a departure from a dispersion relation with simple spherical symmetry. This symmetry is recovered only in the particular case where $\gamma_2 = \pm\gamma_3$. In this specific case, one finds the standard heavy and light holes, with mass m_H and m_L, related to our coefficients by (assuming $\gamma_2 > 0$)

$$\frac{m_0}{m_H} = \gamma_1 - 2\gamma_2, \quad \frac{m_0}{m_L} = \gamma_1 + 2\gamma_2. \tag{6}$$

In order to make sense as a hole dispersion relation, $E_\pm(\mathbf{k})$ which gives the hole kinetic energy should naturally be positive for any $\mathbf{k}$. This implies some limitations on the Luttinger coefficients, from the explicit expression given by Eq. (5). Indeed, considering the negative branch $E_-(\mathbf{k})$, this requires the square root on the right-hand side to be less than $\gamma_1 k^2$. If we consider first the case where $\gamma_3 < \gamma_2$, the square root is maximal when the bracket multiplying $\gamma_3^2 - \gamma_2^2$ is zero, which occurs when $k_x = k_y = 0$, or $k_y = k_z = 0$, or $k_z = k_x = 0$, that is, when $\mathbf{k}$ is on one of the cubic axes. In this case, the positivity condition is merely

$$\gamma_1 > 2\gamma_2. \tag{7}$$

In the opposite case where $\gamma_3 > \gamma_2$, in order to maximize the square root, we instead look for the maximal value of the bracket multiplying $(\gamma_3^2 - \gamma_2^2)$. For fixed k, it is found when $\pm k_x = \pm k_y = \pm k_z = k/\sqrt{3}$, that is, when $\mathbf{k}$ is along one of the principal cubic diagonals. In this case, the square root is merely equal to $\gamma_3 k^2$, and the positivity condition is (assuming $\gamma_3 > 0$)

$$\gamma_1 > 2\gamma_3. \tag{8}$$

In the particular case where $\gamma_3 = \gamma_2$, these conditions merely reduce to the fact that m_H should be positive, which is physically obvious.

III. INTEGRAL EQUATION FOR EXCITONIC EIGENSTATES

The Hamiltonian for the exciton is merely obtained by adding to the hole kinetic energy the electronic kinetic energy $\mathbf{p}_e^2/2m_e$, where $\mathbf{p}_e$ is the conduction band electronic momentum, and m_e the corresponding electronic band mass. We assume that the electronic dispersion relation is isotropic [15]. The electronic spin does not bring any complication, except that naturally all the degeneracies are multiplied by two; so, we do not indicate it explicitly. Finally, we have to include in the Hamiltonian the electron-hole attractive Coulomb interaction $V(\mathbf{r}_e - \mathbf{r}_h)$, where $\mathbf{r}_e$ and $\mathbf{r}_h$ are respectively the electron and hole position, with $V(\mathbf{r}) = -e^2/(4\pi\epsilon_{sc}r)$ where ϵ_{sc} is the semiconductor dielectric constant [16]. Since this interaction depends only on the relative position of the hole and the electron, it is translationally invariant, so the excitonic momentum is conserved. In the present paper, we restrict ourselves to the case where this momentum is zero, so that the electron and hole momentum are opposite $\mathbf{p}_e = -\mathbf{p}_h = -\hbar\mathbf{k}$. Hence, the electron term merely adds a contribution $\hbar^2k^2/2m_e$ to the hole kinetic energy Eq. (1), and we are back to a one-body problem. The potential term is the Coulomb interaction, and the kinetic energy is still given by Eq. (1), provided γ_1 is replaced by

$$\bar{\gamma}_1 = \gamma_1 + \frac{m_0}{m_e}. \tag{9}$$

Considering the complexity of the hole kinetic energy, it is more convenient to write the excitonic Schrödinger equation in momentum space, which leads to an integral equation. This is much easier to handle numerically than the four-dimensional partial differential equation one would find if one worked in real space. We denote the $\mathbf{k}$ representation of the four-components wave function as $a(\mathbf{k}) \equiv (a_1(\mathbf{k}), a_2(\mathbf{k}), a_3(\mathbf{k}), a_4(\mathbf{k}))$. This four-dimensional Schrödinger equation then reads

$$\bar{H}_h(\mathbf{k})\, a(\mathbf{k}) + \int \frac{d\mathbf{k}'}{(2\pi)^3} V(\mathbf{k} - \mathbf{k}')\, a(\mathbf{k}') = E\, a(\mathbf{k}), \tag{10}$$

where $V(\mathbf{q}) = -e^2/(\epsilon_{sc}q^2)$ is the Fourier transform of the Coulomb interaction, and E is the exciton energy to be solved for. $\bar{H}_h(\mathbf{k})$ is just given by Eq. (1), except that γ_1 has to be replaced by $\bar{\gamma}_1$. Note that the Coulomb interaction is diagonal in our representation, while $\bar{H}_h(\mathbf{k})$ is not.

It is convenient to rewrite this equation so that only dimensionless quantities appear. The natural energy unit is the Rydberg, so we write the exciton energy as $E = -\bar{E}\, m_0(e^2/4\pi\epsilon_{sc})^2/(2\hbar^2)$ where $\bar{E}$ is the dimensionless exciton binding energy. The corresponding natural length unit is the Bohr radius, so we also go to a dimensionless wave vector $\bar{\mathbf{k}}$ defined by $\mathbf{k} = \bar{\mathbf{k}}\, m_0(e^2/4\pi\epsilon_{sc})/\hbar^2$. Taking into account that $\bar{H}_h(\mathbf{k})$ is a quadratic function of the wave vector, our Schrödinger equation becomes

$$(\bar{H}_r(\bar{\mathbf{k}}) + \bar{E})a(\bar{\mathbf{k}}) = \frac{1}{\pi^2}\int d\bar{\mathbf{k}}' \frac{1}{(\bar{\mathbf{k}} - \bar{\mathbf{k}}')^2}\, a(\bar{\mathbf{k}}'). \tag{11}$$

where we have also defined a reduced dimensionless hole kinetic energy $\bar{H}_r(\bar{\mathbf{k}}) = (2m_0/\hbar^2)\bar{H}_h(\bar{\mathbf{k}})$.

Furthermore, it is clear from Eqs. (1) and (9) that the important parameters are the ratios $\gamma_2/\bar{\gamma}_1$ and $\gamma_3/\bar{\gamma}_1$. Setting

$$\bar{\gamma}_2 = \frac{\gamma_2}{\bar{\gamma}_1} \text{ and } \bar{\gamma}_3 = \frac{\gamma_3}{\bar{\gamma}_1}, \tag{12}$$

it is convenient to have only these parameters appearing. This requires a further rescaling, obtained by setting $\bar{\mathbf{k}} = K/\bar{\gamma}_1$ and $\bar{E} = \bar{\epsilon}/\bar{\gamma}_1$. This leads to

$$((K^2 + \bar{\epsilon})\mathbb{1} + \bar{h}_h(\mathbf{K}))a(\mathbf{K}) = \frac{1}{\pi^2}\int d\mathbf{K}' \frac{1}{(\mathbf{K} - \mathbf{K}')^2}\, a(\mathbf{K}'), \tag{13}$$

where $\bar{h}_h(\mathbf{K})$ has exactly the same expression as $h_h(\mathbf{k})$ in Eq. (1) (with $\mathbf{k}$ replaced by $\mathbf{K}$) except that, in the explicit expressions for its elements given in Eq. (2), one has to replace γ_2 and γ_3 by $\bar{\gamma}_2$ and $\bar{\gamma}_3$, respectively.

We note that corresponding to our above request for the positivity of the hole kinetic energy, we have to require that the kinetic energy for the exciton one-body problem is always positive. Proceeding as in the above derivation of Eqs. (7) and (8), we find that our coefficients must satisfy

$$|\bar{\gamma}_2| < \tfrac{1}{2} \text{ and } |\bar{\gamma}_3| < \tfrac{1}{2}. \tag{14}$$

For numerical work, it is more convenient to apply to Eq. (13) the inverse of the 4×4 matrix $(K^2 + \bar{\epsilon})\mathbb{1} + \bar{h}_h(\mathbf{K})$. Since we have noted in Eqs. (3) and (4) that $\bar{h}_h^2 = \lambda^2 \mathbb{1}$ is proportional to the unit matrix (with $\gamma_{2,3}$ and $\mathbf{k}$ replaced in λ by $\bar{\gamma}_{2,3}$ and $\mathbf{K}$), this inverse is merely $((K^2 + \bar{\epsilon})\mathbb{1} - \bar{h}_h(\mathbf{K}))/((K^2 + \bar{\epsilon})^2 - \lambda^2)$. Hence Eq. (13) becomes

$$a(\mathbf{K}) = \frac{(K^2 + \bar{\epsilon})\mathbb{1} - \bar{h}_h(\mathbf{K})}{(K^2 + \bar{\epsilon})^2 - \lambda^2} \frac{1}{\pi^2} \int d\mathbf{K}' \frac{1}{(\mathbf{K} - \mathbf{K}')^2} a(\mathbf{K}'). \tag{15}$$

Note that the conditions Eq. (14) correspond to the requirement that the denominator $(K^2 + \bar{\epsilon})^2 - \lambda^2$ is always positive, in particular for large K.

To conclude this section, it is useful to consider the particular case where our exciton problem reduces to the case of the hydrogen atom. This corresponds merely to the case where $\bar{h}_h(\mathbf{K}) = 0$ and the integral equation (15) reduces to a one-dimensional equation

$$a(\mathbf{K}) = \frac{1}{\pi^2 (K^2 + \bar{\epsilon})} \int d\mathbf{K}' \frac{1}{(\mathbf{K} - \mathbf{K}')^2} a(\mathbf{K}'). \tag{16}$$

We readily know the solution which is the Fourier transform of the ground-state wave function $\exp(-r/a_0)$ of the hydrogen atom, that is (omitting the unimportant constant prefactor) $a(\mathbf{K}) = 1/(1 + K^2)^2$. Taking into account that, with our reduced units, the ground-state energy is merely $\bar{\epsilon} = 1$, one can easily check analytically that this expression of $a(\mathbf{K})$ is indeed solution of the integral equation (16).

It is important to note that the explicit large K behavior $a(\mathbf{K}) \sim 1/K^4$ of this solution, which insures the large K convergence of the integral on the right-hand side, is actually a generic feature of the general equation (15). Indeed, making use of this convergence consistently allows us to write in this large K regime $1/(\mathbf{K} - \mathbf{K}')^2 \simeq 1/K^2$ in the integral. Then, the resulting $1/K^2$ factor, together with the explicit prefactor of the integral, indeed leads to $a(\mathbf{K}) \sim 1/K^4$ for the large K behavior of $a(\mathbf{K})$. This behavior is naturally of particular interest for the numerical solution of this integral equation.

IV. GROUND-STATE DEGENERACY

One might think at first that Eq. (15) gives for the exciton ground state a single nondegenerate solution, with a specific symmetry. This is not what happens. Instead, we find that quite generally the exciton ground state, at zero total momentum, has an exact fourfold degeneracy (without taking into account the conduction electron spin degeneracy). This comes directly from the cubic symmetry of the crystal. This is fully analogous to what would occur if we had a complete rotational invariance. In this last case, this would naturally imply a full degeneracy, with the eigenstates having in this four-dimensional subspace an angular dependence given by the spherical harmonics $Y_{\ell m}(\theta, \varphi)$ with $\ell = 3/2$ and $m = (\pm 3/2, \pm 1/2)$. In the present case, we have actually the lower cubic symmetry, but it is enough to similarly insure the complete degeneracy of the ground state.

This result may at first look surprising if one naively considers the situation with spherical symmetry $\bar{\gamma}_2 = \bar{\gamma}_3$. Since in this case one has heavy and light holes, one could think of two possible exciton states, made respectively with a heavy or a light hole (each one being doubly degenerate). However, this omits the fact that the Coulomb interaction is not diagonal in this heavy-light hole representation. Indeed, it has been shown [17] that the proper treatment of this specific case leads to a fourfold degeneracy. Here, we extend this result to the general case.

To see this in a more specific way, we have to consider the symmetry properties of the solutions. We rewrite the Schrödinger equation (13) for the exciton, making explicit the wave function and matrix components. It reads, with $(m, n) = 1, 2, 3, 4$

$$\sum_{n=1}^{4} ((K^2 + \bar{\epsilon})\delta_{mn} + \bar{h}_{mn}(\mathbf{K})) a_n(\mathbf{K}) = \frac{1}{\pi^2} \int d\mathbf{K}' \frac{1}{(\mathbf{K} - \mathbf{K}')^2} a_m(\mathbf{K}'), \tag{17}$$

where $\bar{h}_{mn}(\mathbf{K})$ are the matrix elements of $\bar{h}_h(\mathbf{K})$ and δ_{mn} is the Kronecker symbol. The fact that this equation satisfies time-reversal invariance implies that, if $a_n(\mathbf{K})$ is solution of this equation, $(-1)^{n-1} a^*_{5-n}(\mathbf{K})$ is also solution with the same energy. This transformation can be understood from the spin-3/2 nature of our components. This property can be checked directly from the explicit expression of the hole kinetic energy terms $\bar{h}_{mn}(\mathbf{K})$, given by Eq. (2).

Let us first consider the symmetry properties under a $\pi/2$ rotation R_z around our quantization axis z. Under this rotation, the component of the wave function $a_n(\mathbf{K})$ becomes $a_n(R_z^{-1}(\mathbf{K}))$, that is, $a_n(K_x, K_y, K_z)$ becomes $a_n(K_y, -K_x, K_z)$. For clarity, we denote as $\bar{a}_n$ the transform of a_n, that is, $\bar{a}_n(K_x, K_y, K_z) = a_n(K_y, -K_x, K_z)$. Similarly, if we perform an additional $\pi/2$ rotation R_z on the wave function, amounting to a total π rotation, we end up with $\bar{\bar{a}}_n(K_x, K_y, K_z) = a_n(-K_x, -K_y, K_z)$. Finally, an additional $\pi/2$ rotation leads to $\bar{\bar{\bar{a}}}_n(K_x, K_y, K_z) = a_n(-K_y, K_x, K_z)$.

However, if we want to rotate the state $|a\rangle = \sum_n a_n |n\rangle$ by a $\pi/2$ rotation R_z, we have to take into account that our basis states $|3/2\rangle \equiv |1\rangle$, $|1/2\rangle \equiv |2\rangle$, $|-1/2\rangle \equiv |3\rangle$ and $|-3/2\rangle \equiv |4\rangle$ change under rotation by acquiring a phase factor through the action of the rotation operator $e^{-i\pi J_z/2\hbar}$. Disregarding the unimportant overall phase factor $e^{-3i\pi/4}$, this leads for the various components a_n of our wave function to an additional phase factor $e^{i(n-1)\pi/2} = i^{n-1}$ coming from our basis states.

If we now make in Eq. (17) the change of variables $K_x \to K_y$ and $K_y \to -K_x$ (and similarly for $\mathbf{K}'$ on the right-hand side), $a_n(\mathbf{K})$ becomes $\bar{a}_n(\mathbf{K})$ while we see from their explicit expressions Eq. (2) that $D(\mathbf{K})$ becomes $i\,D(\mathbf{K})$, $F(\mathbf{K})$ becomes $-F(\mathbf{K})$ and $\Delta(\mathbf{K})$ is unchanged. Comparing with the original equation (17) before the change of variables, one finds the following symmetry property: if a_n is solution of Eq. (17), then $i^{n-1}\,\bar{a}_n$ is also solution of this equation with the same energy. That is, if we have a solution, the wave function obtained by a $\pi/2$ rotation is also a solution. This physically corresponds to the fact that our Hamiltonian is invariant under a $\pi/2$ rotation R_z. Naturally, repeating the $\pi/2$ rotation, we find that $(-1)^{n-1}\,\bar{\bar{a}}_n$ and $(-i)^{n-1}\,\bar{\bar{\bar{a}}}_n$ are also solutions.

This situation is analogous to the simple one-dimensional case where the Hamiltonian is invariant under the transformation $x \to -x$: if $\psi(x)$ is a solution of the Schrödinger equation, $\psi(-x)$ is also a solution with the same energy. So, we can build the two solutions $\psi_S(x) = \psi(x) + \psi(-x)$, and $\psi_A(x) = \psi(x) - \psi(-x)$ which are respectively symmetric and antisymmetric under $x \to -x$.

In the same way, in our case, if $a_n(\mathbf{K})$ is a solution of the Schrödinger equation (17), then by adding the solutions obtained by performing rotations around the z axis by $\pi/2$, π and $3\pi/2$ respectively, we obtain a solution with the same energy which is invariant by $\pi/2$ rotation. This solution is explicitly

$$a_n^{(1)} = a_n + i^{n-1}\,\bar{a}_n + (-1)^{n-1}\,\bar{\bar{a}}_n + (-i)^{n-1}\,\bar{\bar{\bar{a}}}_n. \qquad (18)$$

If we perform on this solution the change of variables $K_x \to K_y$ and $K_y \to -K_x$, we find that $a_n^{(1)}$ is changed into $(-i)^{n-1}\,a_n^{(1)}$, that is $\bar{a}_n^{(1)} = (-i)^{n-1}\,a_n^{(1)}$.

However, as possible solutions of Eq. (17), we can more generally take solutions $a_n^{(p)}$, with $p = 1, 2, 3, 4$, which transform as $\bar{a}_n^{(p)} = (-i)^{n-p}\,a_n^{(p)}$. Generalizing the preceding case $p = 1$, they are obtained explicitly from a solution $a_n(\mathbf{K})$ by

$$a_n^{(p)} = a_n + i^{n-p}\,\bar{a}_n + (-1)^{n-p}\,\bar{\bar{a}}_n + (-i)^{n-p}\,\bar{\bar{\bar{a}}}_n. \qquad (19)$$

Performing the changes $K_x \to K_y$ and $K_y \to -K_x$, we check as for Eq. (18) that they indeed transform as we have indicated. These are the appropriate generalizations to the present case of choosing solutions which are either even or odd under $x \to -x$, when the Hamiltonian is invariant under this transformation $x \to -x$. These four solutions are orthogonal since the scalar product of two different solutions is naturally unchanged when one performs on $\mathbf{K}$ in the integral the change of variables $K_x \to K_y$ and $K_y \to -K_x$ corresponding to the $\pi/2$ rotation, while it is changed due to the different way the two different solutions with $p \neq q$ transform under the rotation

$$\begin{aligned}\sum_{n=1}^{4}\int d\mathbf{K}\,a_n^{(p)*}(\mathbf{K})a_n^{(q)}(\mathbf{K}) &= \sum_{n=1}^{4}\int d\mathbf{K}\,\bar{a}_n^{(p)*}(\mathbf{K})\bar{a}_n^{(q)}(\mathbf{K})\\ &= i^{q-p}\sum_{n=1}^{4}\int d\mathbf{K}\,a_n^{(p)*}(\mathbf{K})a_n^{(q)}(\mathbf{K}) = 0. \qquad (20)\end{aligned}$$

In our one-dimensional case, if we start with an even solution $\psi(x) = \psi_S(x)$, the corresponding odd solution $\psi_A(x)$ is found equal to zero. And conversely, if we start with $\psi_A(x)$, the corresponding even solution is equal to zero. Similarly, in our case with $\pi/2$ rotation, one finds easily that if our starting a_n has a $a_n^{(p)}$ symmetry, the resulting solutions with different symmetry $a_n^{(q)}$, with $q \neq p$, are identically equal to zero.

We note that the time-reversal counterpart of $a_n^{(1)}$, that is $(-1)^{n-1}a_{5-n}^{(1)*}$, transforms like $a_n^{(4)}$, and similarly $a_n^{(2)}$ and $a_n^{(3)}$ are linked in the same way. A solution and its time-reversed function can not be identical because they have different symmetries. Hence, the time reversability automatically implies that there is a twofold degeneracy for the exciton eigenstates (this is Kramer's degeneracy theorem with respect to hole states). In order to reach the conclusion that the degeneracy is actually fourfold, we have to search for a further symmetry of the Hamiltonian.

For this purpose, let us now consider the symmetry corresponding to a $\pi/2$ rotation R_x around the x axis. With respect to the orbital variables, the situation is analogous to the one we had above with R_z: $a_n(\mathbf{K})$ becomes $a_n(R_x^{-1}(\mathbf{K}))$, that is, $a_n(K_x, K_y, K_z)$ becomes $\tilde{a}_n(K_x, K_y, K_z) = a_n(K_x, K_z, -K_y)$. However, in order to obtain the transform of our four-dimensional wave function, we again have naturally to take into account the fact that our basis states are strongly modified under this rotation, that is, to calculate $e^{-i\pi J_x/2\hbar}|n\rangle$. This is a standard calculation, that we will not detail here. It can be done for example by finding the evolution of the matrix elements $f_n(\varphi) = \langle n|e^{-i\varphi J_x/\hbar}|m\rangle$ as a function of φ. For example, for $m = 1$, one writes the first-order differential equations obtained by evaluating $\hbar\partial f_n(\varphi)/\partial\varphi = -i\langle n|J_x\,e^{-i\varphi J_x/\hbar}|1\rangle$ (for example $\partial f_1(\varphi)/\partial\varphi = (-i\sqrt{3}/2)f_2(\varphi)$ from the explicit expression of J_x) and one integrates these equations.

One obtains in this way the transform $|n\rangle_x \equiv e^{-i\pi J_x/2\hbar}|n\rangle$ of our basis states by the R_x rotation as

$$\begin{aligned}|1\rangle_x &= \frac{1}{2\sqrt{2}}[\,|1\rangle - i\sqrt{3}\,|2\rangle - \sqrt{3}\,|3\rangle + i\,|4\rangle],\\ |2\rangle_x &= \frac{1}{2\sqrt{2}}[-i\sqrt{3}\,|1\rangle - |2\rangle - i\,|3\rangle - \sqrt{3}\,|4\rangle],\\ |3\rangle_x &= \frac{1}{2\sqrt{2}}[-\sqrt{3}\,|1\rangle - i\,|2\rangle - |3\rangle - i\sqrt{3}\,|4\rangle],\\ |4\rangle_x &= \frac{1}{2\sqrt{2}}[i\,|1\rangle - \sqrt{3}\,|2\rangle - i\sqrt{3}\,|3\rangle + |4\rangle]. \qquad (21)\end{aligned}$$

Hence, the solutions $\sum_n a_n^{(p)}|n\rangle$ are transformed into $\sum_{m,n}\tilde{a}_n^{(p)}|m\rangle\langle m|n\rangle_x$, where the matrix elements $\langle m|n\rangle_x$ are given by Eq. (21). These transformed solutions are naturally also solutions of the initial Schrödinger equation (17). If the solutions $a_n^{(p)}$ are fourfold degenerate, this implies that the transformed solutions are linear combinations of the $a_n^{(p)}$, that is, there exists a set of coefficients $\lambda_p^{(q)}$ such that

$$\sum_n \tilde{a}_n^{(q)}(\mathbf{K})\langle m|n\rangle_x = \sum_p \lambda_p^{(q)}a_m^{(p)}(\mathbf{K}). \qquad (22)$$

These simple linear relations between the solutions of Eq. (17) are fairly remarkable, since the solutions of Eq. (17) are not expected to be simple. Since we have no analytical solutions of Eq. (17), we have no way to check them in the general case.

However, there is a limiting situation where we can solve analytically Eq. (17) and check these relations. This is the limit of large wave vectors $\mathbf{K}$, corresponding physically to short distances. In this case, since on the right-hand side of Eq. (17) the $a_m(\mathbf{K}')$ have a limited range, we can write in the integral $1/(\mathbf{K}-\mathbf{K}')^2 \simeq 1/\mathbf{K}^2$, which provides an explicit solution for $a_n(\mathbf{K})$ in terms of the integrals $I_n = \int d\mathbf{K}\, a_n(\mathbf{K})$, by merely inverting the matrix $K^2\delta_{mn} + \bar{h}_{mn}(\mathbf{K})$ as we have done in Eq. (15) ($\bar{\epsilon}$ is negligible in this large $\mathbf{K}$ limit). This gives explicitly

$$a_m(\mathbf{K}) = \sum_{n=1}^{4} \frac{K^2\delta_{mn} - \bar{h}_{mn}(\mathbf{K})}{K^4 - \lambda^2}\, \frac{I_n}{\pi^2 K^2}. \tag{23}$$

If we take the particular case of the $a_n^{(p)}$ solutions, we notice that, when we make in $I_n^{(p)} = \int d\mathbf{K}\, a_n^{(p)}(\mathbf{K})$ the change of variables on $\mathbf{K}$ corresponding to the R_z rotation, the integral is naturally unchanged. On the other hand, taking into account that $a_n^{(p)}(\mathbf{K})$ is transformed into $\bar{a}_n^{(p)}(\mathbf{K}) = (-i)^{n-p}\, a_n^{(p)}(\mathbf{K})$ in this change of variables, we find that $I_n^{(p)}$ is multiplied by $(-i)^{n-p}$. Hence $I_n^{(p)}$ is zero unless $n = p$, so that $I_n^{(p)} \equiv I^{(p)}\,\delta_{n,p}$. Actually we can even choose to have $I^{(p)} = 1$ by taking an appropriate norm for the $a_n^{(p)}$ solutions. Hence, in this case, Eq. (23) simplifies into

$$a_m^{(p)}(\mathbf{K}) = \frac{K^2\delta_{mp} - \bar{h}_{mp}(\mathbf{K})}{\pi^2 K^2 (K^4 - \lambda^2)}. \tag{24}$$

In particular, one can check on these explicit expressions that the symmetry properties $\bar{a}_n^{(p)} = (-i)^{n-p}\, a_n^{(p)}$ are indeed satisfied.

Integrating Eq. (22) over $\mathbf{K}$, and noting that $\int d\mathbf{K}\, \tilde{a}_n^{(q)}(\mathbf{K}) = \int d\mathbf{K}\, a_n^{(q)}(\mathbf{K}) = \delta_{n,q}$ by change of variables, we obtain the coefficients $\lambda_p^{(q)}$ in Eq. (22) as $\lambda_m^{(n)} = \langle m|n\rangle_x$, and in particular $\lambda_n^{(n)} = \langle n|n\rangle_x$. One can then substitute in Eq. (22) the explicit solutions $a_n^{(q)}(\mathbf{K})$ given by Eq. (24), and check after a simple but tedious calculation that it is satisfied. For example, taking the case $q = 1$ and the component $m = 1$, we have to check from Eq. (21) that

$$\begin{aligned}\tilde{a}_1^{(1)}(\mathbf{K}) &- i\sqrt{3}\tilde{a}_2^{(1)}(\mathbf{K}) - \sqrt{3}\tilde{a}_3^{(1)}(\mathbf{K}) + i\tilde{a}_4^{(1)}(\mathbf{K})\\ &= a_1^{(1)}(\mathbf{K}) - i\sqrt{3}a_1^{(2)}(\mathbf{K}) - \sqrt{3}a_1^{(3)}(\mathbf{K}) + ia_1^{(4)}(\mathbf{K})\end{aligned} \tag{25}$$

where we have omitted the common denominator $2\sqrt{2}$ occuring in all the terms from Eq. (21). Using the definition $\tilde{a}_n(K_x, K_y, K_z) = a_n(K_x, K_z, -K_y)$, and noting that $\tilde{a}_4^{(1)}(\mathbf{K}) = a_1^{(4)}(\mathbf{K}) = 0$, we are left with checking that

$$\begin{aligned}&\left[K^2 - \bar{\gamma}_2\left(K^2 - 3K_y^2\right)\right] + 6i\bar{\gamma}_3(K_x + iK_z)K_y\\ &\quad -\left[3\bar{\gamma}_2\left(K_x^2 - K_z^2\right) + 6i\bar{\gamma}_3 K_x K_z\right]\\ &= \left[K^2 - \bar{\gamma}_2\left(K^2 - 3K_z^2\right)\right]\\ &\quad -6i\bar{\gamma}_3(K_x - iK_y)K_z - \left[3\bar{\gamma}_2\left(K_x^2 - K_y^2\right) - 6i\bar{\gamma}_3 K_x K_y\right],\end{aligned} \tag{26}$$

where we have omitted the denominator $\pi^2 K^2(K^4 - \lambda^2)$ common to all the terms. We see that this equality (26) is indeed satisfied, but in a nontrivial way.

It is now clear that the degeneracy of all the four solutions $a_n^{(p)}$ is necessary in order to satisfy such an identity. Indeed, if on the contrary $a_n^{(1)}$ (together with its time-reversed conjugate $a_n^{(4)}$) and $a_n^{(2)}$ (together with $a_n^{(3)}$) had different energies, the transformed of $a_n^{(1)}$ by the rotation R_x would have to be a linear combination of $a_n^{(1)}$ and $a_n^{(4)}$ only, since the two other solutions had a different energy. This would lead to relations analogous to Eq. (25), but with only two solutions ($a^{(1)}$ and $a^{(4)}$) on the right-hand side instead of four as in Eq. (25). This is a much more stringent algebraic requirement, that clearly can not be satisfied. Indeed, one can check in the large $\mathbf{K}$ limit that the corresponding relation is algebraically incompatible with the explicit solution. This provides an explicit proof that the four solutions $a_n^{(p)}$ are indeed degenerate [18].

V. NUMERICAL RESULTS

We numerically find the ground-state energy (and the corresponding wave function) by applying an iterative procedure to this Schrödinger equation under the form written in Eq. (15). Let us call A the linear operator that acts on the four-dimensional wave function $a(\mathbf{K})$ on the right-hand side of Eq. (15). Solving Eq. (15) is equivalent to finding an eigenvector of A with the eigenvalue $\Lambda = 1$. In the case of a very large binding energy $\bar{\epsilon}$, the operator A goes to zero and all its eigenvalues are quite small. Hence, none of them can be equal to 1, and there is no state with very large binding energy, as expected. If we decrease $\bar{\epsilon}$, the largest positive eigenvalue $\Lambda_{\max}$ of A will grow. When this eigenvalue of A reaches 1, we will have obtained the largest possible value for $\bar{\epsilon}$ corresponding to an eigenstate. In other words, we will have found the ground-state energy.

It is easy to obtain the largest eigenvalue of A by applying iteratively A to some convenient starting wave function $a(\mathbf{K})$. A practical choice for the starting wave function is to take the first component $a_1(\mathbf{K})$ equal to the hydrogen atom ground-state wave function we have discussed above, and the other components essentially equal to zero. Indeed, iterating n times is equivalent to applying the operator A^n to $a(\mathbf{K})$. But for large n values, A^n is dominated by its largest eigenvalue $(\Lambda_{\max})^n$ and is essentially equivalent to a projection on the corresponding eigenvector and multiplication by $(\Lambda_{\max})^n$. This allows us to identify conveniently $\Lambda_{\max}$ and the corresponding eigenvector. Actually this procedure works only if the spectrum of A does not have nasty features, such as closely spaced largest and second largest eingenvalues, or large negative eigenvalues. Fortunately, in our case, this procedure happens to work quite nicely. We have found that, in practice, typically 20 iterations, or less, already gave a satisfactory convergence for the precision we considered. It is also convenient, in order to find the ground-state energy, to start from the hydrogen atom situation $\bar{\gamma}_2 = \bar{\gamma}_3 = 0$ where the solution is known, and to crank up $\bar{\gamma}_2$ and $\bar{\gamma}_3$ progressively. In this way, the range where the ground-state energy lies is fairly well known at each stage of the calculation.

In the practical task of performing the integral on the right-hand side of Eq. (15), it is much better to get rid of the Coulomb interaction term and its singular behavior. This is done conveniently by performing the change of variables $\mathbf{K}' = \mathbf{K} + \mathbf{Q}$. In this way the integral becomes

$$\int d\mathbf{K}' \frac{1}{(\mathbf{K}-\mathbf{K}')^2}\, a(\mathbf{K}') = \int_0^{\infty} dQ \int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\varphi\; a(K_x + Q\sin\theta\cos\varphi, K_y + Q\sin\theta\sin\varphi, K_z + Q\cos\theta), \tag{27}$$

where θ and φ are the polar and azimuthal angles for $\mathbf{Q}$. However, this change of variables makes it necessary, in order to conveniently perform numerically the integrals, to evaluate $a(K_x, K_y, K_z)$ for any values of K_x, K_y and K_z. This can be done by sampling a number of K_x, K_y, K_z values and infer any $a(K_x, K_y, K_z)$ value by interpolation. In practice, it is rather $(1+K^2)^2\, a(K_x, K_y, K_z)$ that we have interpolated, since in the case of the hydrogen atom, this function is just a constant equal to 1. So, in our exciton case, we do not expect it to have strong variations. This has allowed us to use a simple three-dimensional linear interpolation. Naturally, we have also restricted the range of variation of our variables by the change of variables $K_{x,y,z} = \tan t_{x,y,z}$, with the regular discretization being on the $t_{x,y,z}$'s.

It is convenient to reduce the range of the $\mathbf{K}$ variables by making use of symmetries, for example planar reflexions. If we consider first the change $K_z \to -K_z$, it produces a change of sign for D in Eq. (2), all the other terms being unchanged. As a result, one checks easily that, if $a_n(K_x, K_y, K_z)$ is solution of Eq. (15), then $(-1)^{n-1} a_n(K_x, K_y, -K_z)$ is also solution. As above, we can build a solution that is even under this transform (and is expected to correspond to the ground state). It satisfies $a_n(K_x, K_y, -K_z) = (-1)^{n-1} a_n(K_x, K_y, K_z)$. Note that the time-reversed solution $a_n^T(\mathbf{K}) = (-1)^{n-1} a_{5-n}^*(\mathbf{K})$ transforms as the odd combination in this transform $a_n^T(K_x, K_y, -K_z) = (-1)^n a_n^T(K_x, K_y, K_z)$.

Similarly, in the change $K_x \to -K_x$, F is changed into F^* and D into $-D^*$. One checks that, if $a_n(K_x, K_y, K_z)$ is solution, then $a_{5-n}(-K_x, K_y, K_z)$ is solution. Hence, we can take a solution with the symmetry $a_n(-K_x, K_y, K_z) = a_{5-n}(K_x, K_y, K_z)$. Finally $K_y \to -K_y$ changes F into F^* and D into D^*. So, if $a_n(K_x, K_y, K_z)$ is solution, then $(-1)^{n-1} a_{5-n}(K_x, -K_y, K_z)$ is solution, and we can take a solution with the symmetry $a_n(K_x, -K_y, K_z) = (-1)^{n-1} a_{5-n}(K_x, K_y, K_z)$. Note that, with this choice of solution satisfying simultaneously these three symmetries, if we perform a symmetry with respect to the origin $K_x \to -K_x$, $K_y \to -K_y$ and $K_z \to -K_z$, we find that our solution is invariant, as we expect from the ground-state wave function. These symmetries allow us to restrict the range of our variables to $K_x \geqslant 0$, $K_y \geqslant 0$, $K_z \geqslant 0$.

We want to explore the whole range of $\bar{\gamma}_2$ and $\bar{\gamma}_3$ values allowed by the condition Eq. (14). However, we note that if we change the sign of both $\bar{\gamma}_2$ and $\bar{\gamma}_3$, from Eq. (2) Δ, D and F change sign, that is, h_h defined by Eq. (1) is changed into $-h_h$. But h_h and $-h_h$ must have the same eigenvectors, and moreover have also the same doubly degenerate eigenvalues $\pm\lambda$ as seen in Eqs. (3) and (4). In other words, going from h_h to $-h_h$ amounts to merely relabeling the eigenvectors corresponding to the two positive eigenvalues as corresponding to the two negative eigenvalues, and conversely. It is clear that such a relabeling, which is merely a basis change, does not change the eigenvalue spectrum of the whole Hamiltonian we are interested in. Accordingly, in particular, the exciton ground-state energy is unchanged when changing the sign of both $\bar{\gamma}_2$ and $\bar{\gamma}_3$. If we have a solution for a given sign of the $\bar{\gamma}$'s, we are able to obtain a solution having the same groud-state energy for the $\bar{\gamma}$'s with opposite sign.

Let us now consider what happens if we change only the $\bar{\gamma}_2$ sign, $\bar{\gamma}_3$ being unchanged. From Eq. (2), one finds that Δ is changed into $-\Delta$, and F into $-F^*$, while D is unchanged. Moreover, as we have seen above, in the change of variables $K_x \to K_y$ and $K_y \to -K_x$, $D(\mathbf{K})$ becomes $i\,D(\mathbf{K})$, while $F(\mathbf{K})$ becomes $-F(\mathbf{K})$. But we notice that a further simple change of functions $a_2 \to -ia_2$, $a_4 \to -ia_4$ is equivalent to changing D into iD. Combining these three changes, we have changed Δ into $-\Delta$, F into F^* and D into $-D$. But the resulting equations for Eq. (13), or Eq. (15), are identical to the original ones provided we make the additional change of functions $a_1 \leftrightarrow a_3$, $a_2 \leftrightarrow a_4$. Hence, we conclude that $\bar{\gamma}_2 \to -\bar{\gamma}_2$ leaves the energy spectrum invariant. Combining now the two changes of signs for the γ's, we obtain that the ground-state energy depends only on $|\bar{\gamma}_2|$ and $|\bar{\gamma}_3|$, so we can restrict our study to $0 \leqslant \bar{\gamma}_2 < 1/2$, $0 \leqslant \bar{\gamma}_3 < 1/2$.

Let us finally examine the limiting cases $\bar{\gamma}_2 \to 1/2$, $\bar{\gamma}_3 \to 1/2$. As we have seen below Eq. (6), when either $\bar{\gamma}_2 = 1/2$ or $\bar{\gamma}_3 = 1/2$, there are some $\mathbf{K}$ directions for which $K^2 - \lambda^2 = 0$. As a result, if in Eq. (15) $\bar{\epsilon}$ is fixed, for large values of K, the denominator in the first factor of the right-hand side will go to zero. This implies that $a(\mathbf{K})$ will not behave for all $\mathbf{K}$ directions as $1/K^2$, as we have found in Eq. (16), which leads to a divergent behavior in the integral of Eq. (15). The natural way to escape this problem is to let $\bar{\epsilon}$ be very large, so this divergent behavior of $a(\mathbf{K})$ is pushed to higher values of K, and one avoids an actual singularity in the limits $\bar{\gamma}_2 \to 1/2$, $\bar{\gamma}_3 \to 1/2$.

Unfortunately, an explicit handling of this singular behavior is not so easy in general because it arises from the behavior of $a(\mathbf{K})$ along some specific directions of $\mathbf{K}$. Nevertheless, when we let both $\bar{\gamma}_2 \to 1/2$, $\bar{\gamma}_3 \to 1/2$ at the same time, it is possible to extract the divergent behavior of $\bar{\epsilon}$ by a simple rescaling. Let us set $2\,\bar{\gamma}_2 = 1 - \eta_2$, $2\,\bar{\gamma}_3 = 1 - \eta_3$, and consider $\eta_2 \to 0$, $\eta_3 \to 0$ with $r = \eta_3/\eta_2$ fixed to a finite value. We perform the rescaling $\mathbf{K} = \bar{\mathbf{K}}/\eta_2$, while at the same time setting $\bar{\epsilon} = \alpha/\eta_2$. This leads, for the denominator on the right-hand side of Eq. (15), to $(K^2 + \bar{\epsilon})^2 - \lambda^2 = (2/\eta_2^3)[\bar{K}^4 + \alpha\bar{K}^2 + 3(r-1)(\bar{K}_x^2\bar{K}_y^2 + \bar{K}_y^2\bar{K}_z^2 + \bar{K}_z^2\bar{K}_x^2)]$. In this rescaling, the numerator provides a factor $1/\eta_2^2$, while another factor $1/\eta_2$ comes from the rescaling of the variable $\mathbf{K}'$ into $\bar{\mathbf{K}}'$ in the integral. Hence, the scaling factor η_2 disappears completely

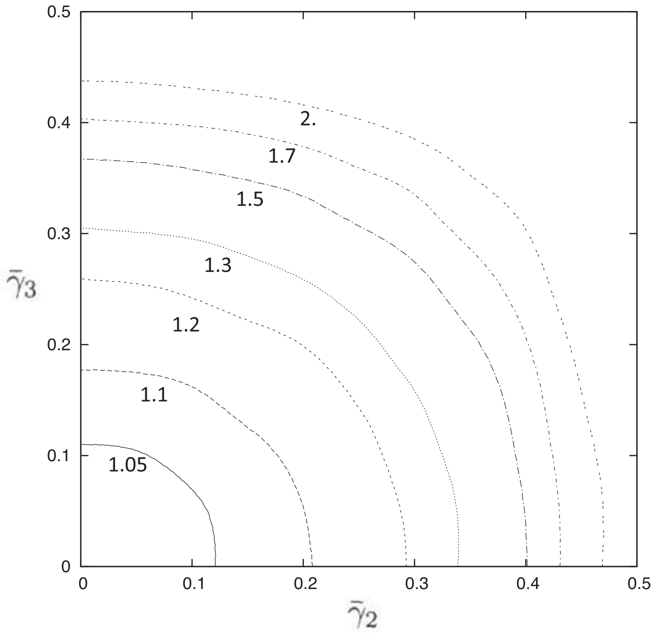

FIG. 2. Reduced ground-state exciton binding energy $\bar{\epsilon}$ in the $\bar{\gamma}_2 - \bar{\gamma}_3$ plane. The level lines for $\bar{\epsilon} = 1.05,\ 1.1,\ 1.2,\ 1.3,\ 1.5,\ 1.7$, and 2. are represented, as indicated in the figure.

and we are left with the following equation, free of singularity:

$$b(\bar{\mathbf{K}}) = \frac{\bar{K}^2 \mathbb{1} - \bar{h}_h(\bar{\mathbf{K}})}{2\left[\alpha \bar{K}^2 + \bar{K}^4 + 3(r-1)\left(\bar{K}_x^2 \bar{K}_y^2 + \bar{K}_y^2 \bar{K}_z^2 + \bar{K}_z^2 \bar{K}_x^2\right)\right]} \times \frac{1}{\pi^2} \int d\bar{\mathbf{K}}' \frac{1}{(\bar{\mathbf{K}} - \bar{\mathbf{K}}')^2}\, b(\bar{\mathbf{K}}') \tag{28}$$

where $b(\bar{\mathbf{K}}) = a(\bar{\mathbf{K}}/\eta_2)$. In the case $r = 1$ we have solved numerically this equation for the scaling factor α in the energy, and we have found $\alpha = 0.43$. This is in fair agreement with the corresponding reduced energy $\bar{\epsilon} = 4.77$ we have found by a direct numerical solution of Eq. (15) for $\eta_2 = \eta_3 = 0.1$.

Following the procedure laid out above, we have calculated the reduced exciton ground-state energy $\bar{\epsilon}$ from Eq. (15) as a function of $\bar{\gamma}_2$ and $\bar{\gamma}_3$, on a grid with spacing 0.05 for these two parameters. Rather than presenting these extensive results as a table, we have used them to evaluate $\bar{\epsilon}$ as a function of $\bar{\gamma}_2$ and $\bar{\gamma}_3$, for any value of these parameters, by a two-dimensional spline interpolation, and then obtain the level lines for various values of $\bar{\epsilon}$ in the $\bar{\gamma}_2 - \bar{\gamma}_3$ plane. They are displayed in Fig. 2. We have chosen not to smooth out the small irregularities which are visible along these level lines, since they result from the imprecision of our calculations and accordingly give a fairly direct information on it. In particular, the results for $\bar{\epsilon}$ close to 1 are very sensitive to this imprecision, so the resulting line for $\bar{\epsilon} = 1.05$ would be pretty bad. Hence, to draw this line, we have rather relied on a quadratic interpolation from the $\bar{\gamma}_2 = 0.1$ and $\bar{\gamma}_3 = 0.1$ results, rather than using our $\bar{\gamma}_2 = 0.05$ and $\bar{\gamma}_3 = 0.05$ results.

The numerical results are fairly regular, with the represented level lines being not so far from quarter circles, provided we do not go too close to the boundaries $\bar{\gamma}_2 = 0.5$ and $\bar{\gamma}_3 = 0.5$. This means that, roughly speaking, $\bar{\epsilon}$ depends only on $\bar{\gamma}_2^2 + \bar{\gamma}_3^2 \equiv 2\bar{\gamma}^2$. Correspondingly, we find that $\bar{\epsilon} - 1 \simeq 2.5(1/\sqrt{1 - 4\bar{\gamma}^2} - 1)$ provides an approximate representation of the numerical results. In particular, it is quadratic for small $\bar{\gamma}$ and it diverges for $\bar{\gamma} \to 1/2$. It gives $\bar{\epsilon} \simeq 4.23$ for $\bar{\gamma} = 0.45$, not so different from the result from our numerical $\bar{\epsilon} = 4.77$ mentioned above.

TABLE I. Relative conduction electron mass m_e/m_0 and Luttinger parameters γ_1, γ_2 and γ_3 of various zinc-blende semiconductor compounds, obtained from Ref. [2], Tables (6.1), (6.2), (6.3), (6.11), and (7.27). The following columns give $\bar{\gamma}_1 = \gamma_1 + m_0/m_e$ from Eq. (9), $\bar{\gamma}_2 = \gamma_2/\bar{\gamma}_1$ and $\bar{\gamma}_3 = \gamma_3/\bar{\gamma}_1$ from Eq. (12), and the corresponding value of $\bar{\epsilon}$, obtained from Fig. 2 by solving Eq. (15).

Compound	m_e/m_0	γ_1	γ_2	γ_3	$\bar{\gamma}_1$	$\bar{\gamma}_2$	$\bar{\gamma}_3$	$\bar{\epsilon}$
GaAs	0.0675	6.98	2.06	2.93	21.8	0.094	0.134	1.08
AlAs	0.15	3.76	0.82	1.42	10.4	0.079	0.137	1.07
InAs	0.026	19.7	8.4	9.3	58.2	0.144	0.156	1.12
GaP	0.13	4.05	0.49	1.25	11.74	0.042	0.106	1.05
AlP	0.22	3.35	0.71	1.23	7.89	0.090	0.156	1.09
InP	0.0795	5.08	1.6	2.1	17.66	0.091	0.119	1.07
GaSb	0.039	13.4	4.7	6.0	39.04	0.120	0.154	1.11
AlSb	0.14	5.18	1.19	1.97	12.3	0.097	0.160	1.10
InSb	0.0135	34.8	15.5	16.5	108.9	0.142	0.151	1.12
GaN	0.193	2.506	0.636	0.977	7.69	0.083	0.127	1.07
AlN	0.316	1.45	0.349	0.597	4.61	0.076	0.129	1.06
InN	0.054	6.817	2.81	3.121	25.3	0.111	0.123	1.08
BN	0.195	5.13	1.23	2.04	10.26	0.120	0.199	1.15

Regarding the size of the results, we note that taking our reduced ground-state exciton binding energy as unity $\bar{\epsilon} = 1$ corresponds to making use, in the standard expression $E = -\mu_X (e^2/4\pi\epsilon_{sc})^2/(2\hbar^2)$, of a reduced mass μ_X given by $1/\mu_X = 1/m_e + 1/\mu_h$, where from Eqs. (6) and (9) the reduced hole mass μ_h itself is taken equal to the half-sum of the reduced heavy and light hole masses $1/\mu_h = (1/m_H + 1/m_L)/2$. This is a frequently done approximation, for lack of better knowledge. Figure 2 shows that in principle we may have a fairly important departure from this approximation.

In practice, we have investigated a few semiconductors where reliable data for the valence band and the conduction band parameters [1,2,19,20] are available. However, one can still see sizable theoretical and experimental uncertainties in the knowledge of these band parameters [2,20]. In Table I, we display, for various semiconductors, the parameters obtained from Ref. [2], together with our corresponding parameters and results. The results obtained from other references are somewhat similar, so we do not display them. Remarkably, as can be seen from Table I, the strong variations among compounds in the basic Luttinger parameters γ_1, γ_2, γ_3 and conduction electron mass m_e provide a much reduced variation in our reduced Luttinger parameters $\bar{\gamma}_2$ and $\bar{\gamma}_3$. One reason is that the small electron mass increases $\bar{\gamma}_1$ from Eq. (9), which leads to a reduction of the size of $\bar{\gamma}_2$ and $\bar{\gamma}_3$. As seen from Table I, we find that $\bar{\gamma}_2$ is mostly in the range 0.1–0.15, while $\bar{\gamma}_3$ takes slightly higher values in the range 0.1–0.2. For these compounds, as seen from Fig. 2, the warping leads to an increase of the exciton binding energy by about 10%. This should be taken into account as soon as one looks for some precision in the exciton binding energy. The largest correction we have found for a compound with specifically known parameters is in the case of BN [2], where $\bar{\gamma}_2 = 0.12$ and $\bar{\gamma}_3 = 0.2$ leads to a 15% increase from the warping in the exciton binding energy. However, it should be kept in mind that this binding energy is quite sensitive to the parameters values in this range. For example, having $\bar{\gamma}_2 = \bar{\gamma}_3 = 0.2$ would lead to a 20% increase in the binding energy, and $\bar{\gamma}_2 = \bar{\gamma}_3 = 0.25$ would give a 35% increase. Hence, keeping in mind the uncertainties on the valence band parameters, the effect of warping on the exciton binding energy may happen to be quite important.

VI. CONCLUSION

In this paper, we have addressed the effect on the exciton ground-state energy of the warping of the valence band near its edge. We have specifically considered the case of bulk exciton for semiconductors with zinc-blende crystal structure. Assuming as usual that the split-off subband due to spin-orbit coupling is energetically far enough to play a negligible role, we have considered the four-dimensional problem for the hole kinetic energy. We have shown that, due to the cubic symmetry, the exciton ground state has a fourfold degeneracy, despite the valence band warping.

We have introduced, by a simple scaling, reduced Luttinger parameters $\bar{\gamma}_2$ and $\bar{\gamma}_3$, whose absolute values are less than $1/2$. We have studied systematically the exciton ground-state energy as a function of these reduced Luttinger parameters, by solving numerically the integral equation corresponding to the Schrödinger equation in Fourier transform. We find that, compared to its standard expression in the absence of warping (corresponding to the case $\bar{\gamma}_2 = \bar{\gamma}_3 = 0$), the exciton binding energy can in principle increase without limitations. For moderate increases, we have provided an approximate analytical expression for the increase in terms of Luttinger parameters. Going through the values of band structure parameters, and in particular the Luttinger parameters, for tabulated semiconductors, we find a typical 10%–15% increase in the exciton binding energy. However, this energy is fairly sensitive to the values of the Luttinger parameters, and a slight increase in their values due to experimental or/and theoretical uncertainties could lead to a markedly larger increase in the exciton binding energy.

ACKNOWLEDGMENTS

The authors are grateful to Monique Combescot for interesting them in this warping problem and for discussions.

---

[1] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors* (Springer, Berlin, 1996).

[2] I. Vurgaftman, M. P. Lumb, and J. R. Meyer, *Bands and Photons in III-V Semiconductor Quantum Structures* (Oxford University Press, Oxford, 2021).

[3] M. Rohlfing and S. G. Louie, Phys. Rev. Lett. **80**, 3320 (1998); **81**, 2312 (1998).

[4] M. Combescot and S.-Y. Shiau, Phys. Rev. B **101**, 195203 (2020).

[5] J. M. Luttinger, Phys. Rev. **102**, 1030 (1956).

[6] J. M. Luttinger and W. Kohn, Phys. Rev. **97**, 869 (1955).

[7] A. Schindlmayr, Eur. J. Phys. **18**, 374 (1997).

[8] M. Suzuki, T. Uenoyama, and A. Yanase, Phys. Rev. B **52**, 8132 (1995).

[9] A. Baldereschi and N. O. Lipari, Phys. Rev. Lett. **25**, 373 (1970); Phys. Rev. B **3**, 439 (1971).

[10] G. Strinati, Phys. Rev. B **29**, 5718 (1984).

[11] W. Hanke and L. J. Sham, Phys. Rev. B **12**, 4501 (1975).

[12] S. G. Louie, J. R. Chelikowsky, and M. L. Cohen, Phys. Rev. Lett. **34**, 155 (1975).

[13] In terms of those of Ref. [6], our coefficients are given by $\hbar^2\gamma_1/m_0 = 2(A+2B)/3$, $\hbar^2\gamma_2/m_0 = (A-B)/3$ and $\hbar^2\gamma_3/m_0 = C/3$, and we have an additional $\pi/2$ rotation around the z axis. These γ_i are just the coefficients introduced by Luttinger [5].

[14] G. Dresselhaus, A. F. Kip, and C. Kittel, Phys. Rev. **98**, 368 (1955).

[15] The warping of the conduction band is expected to be very weak in the relevant **k** range for the exciton; see *e*.g. U. Rössler, Solid State Commun. **49**, 943 (1984).

[16] In the most general case, we should take into account that the dielectric constant, which describes the reaction of the semiconductor to a bare electric charge, is both frequency and wave vector dependent $\epsilon_{sc}(\omega, \mathbf{q})$. For example, if the exciton size is small enough, an appreciable part of the wave function would lie within distances comparable to atomic lengths, and correspondingly we should take into account the dependence of the dielectric constant on the wavevector **q**. However, for the compounds, we are considering in this paper, the exciton size is large enough so that contributions from these small distances can not be significant, and we may neglect them and accordingly consider the dielectric constant as independent of the wave vector. Similarly the frequency dependence should be taken into account if the exciton ground-state energy is comparable to characteristic energies of the semiconductor, for example the semiconductor gap. However our exciton ground-state energy is much smaller than the gap and we may safely take our dielectric constant as frequency independent.

[17] S.-Y. Shiau, B. Eble, and M. Combescot, Phys. Rev. B **107**, L081203 (2023).

[18] In the simple case, $\bar{\gamma}_3 = 0$ where this 4×4 problem reduces to two 2×2 problems, we have checked numerically this degeneracy (within our numerical precision).

[19] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, J. Appl. Phys. **89**, 5815 (2001).

[20] S. Zh. Karazhanov and L. C. Lew Yan Voon, Semiconductors **39**, 161 (2005).