Understanding semiconductor valence mass

Monique Combescot[1] and Shiue-Yuan Shiau [2]

[1]*Sorbonne Université, CNRS, Institut des NanoSciences de Paris, 75005-Paris, France*
[2]*Physics Division, National Center for Theoretical Sciences, Hsinchu, 30013, Taiwan*

The Bloch theorem mathematically proves that in a periodic crystal, electrons can acquire a negative mass. The present work aims to provide a physical understanding for why this is so. We successively analyze the consequences of the threefold orbital valence-state coupling to (i) a nondegenerate orbital level in the conduction band, (ii) a threefold orbital level in the conduction band, and (iii) spin states through spin-orbit interaction. We show that it is not at all trivial for valence electrons to acquire a negative mass for *whatever* their momentum with respect to the crystal axes: It is necessary to not only have a coupling to a degenerate orbital conduction level but also a symmetry breaking of the threefold valence subspace by the spin quantization axis, as induced by spin-orbit interaction. Due to the relativistic origin of this interaction, the existence of negative valence masses thus constitutes an unexpected signature of quantum relativity.

I. INTRODUCTION

In semiconductors, some electrons have a negative mass. This fact is the prime reason for these materials to have shaped today's technology [1,2]. As a physicist, it is worth understanding why and how this is so, independently from the mathematical derivations based on Bloch theorem [3–5] or group theory [6–8], and also independently from the numerous methods that calculate semiconductor band structures, like the tight-binding approach [9,10], Luttinger's approach [11], and the numerical density functional method [12,13].

A direct consequence of electrons having a negative effective mass is that when such an electron is excited, it leaves a hole, that is, an electron absence in the corresponding state [4,5]. This hole essentially behaves as a quantum particle having a positive mass. As a result, semiconductors host two types of fermionic particles, the conduction electrons and the valence holes. Having opposite charges, they can bind into a bosonic particle, the exciton, which is similar to a hydrogen atom, but with a much larger size due to the electron and hole effective masses, one order of magnitude smaller than the free electron mass, and to the semiconductor dielectric constant of the order of 10. These basics largely explain that in addition to their tremendous technological interest [2,14,15], semiconductors have provided an ideal playground for a large number of exciting many-body effects of fundamental physics [16–19].

The Bloch theorem mathematically proves that the energies of electrons subjected to a periodic ion lattice potential form bands, with minima and maxima. Close to these extrema, the band curvatures are positive or negative, enforcing the resulting electron effective mass to also be positive or negative, as commonly demonstrated in textbooks through the simple Kronig-Penney model [20]. Behind this beautiful but abstruse mathematics, there is a drastic change in physics:

The electron that suffers a periodic potential is a free electron with a positive mass m_0. Due to interaction with the periodic ion lattice, this electron ends by behaving like a free electron but with a much lighter mass, that can even become negative. What are the forces that drive this pathological sign change in the electron mass? Understanding its physics will help tailoring the particle effective mass of other systems in search for new technology [21–23].

The usual approach to the effective masses of semiconductor electrons makes use of group theory. Indeed, the degenerate orbital states of the upper valence band suffer a relativistic spin-orbit interaction [24,25] that mixes orbital and spin degrees of freedom. A standard but simple-minded way to derive the spin-orbit eigenstates for semiconductors follows the procedure developed for atoms [18]. It is based on the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$ of the electron. However, because the concept of orbital momentum $\mathbf{L}$ only has a meaning for electron states that have a spherical symmetry [26], like atomic states, such a derivation cannot be used for crystals, in spite of the validity of the obtained results. Until very recently [25], the procedure based on group theory has been the only correct way to derive the spin-orbit eigenstates in a periodic crystal. Through the double groups that mix orbital and spin symmetries, this approach leads to different dispersion relations for valence electrons in the various spin-orbit states.

For sure, the group theory formalism is extremely powerful. Yet, through mixing the orbital and spin subspaces within the double groups, as done at the very first line, it lacks a physical transparency. In addition, most semiconductor physicists who do not master group theory find it difficult to follow, and equally hard to accept that learning such a general but heavy formalism is necessary to solve problems dealing with a threefold orbital degeneracy only. This is why, in a first work [25], we reconsidered the energy splitting of the threefold orbital states at the Γ point, i.e., for $\mathbf{k} = \mathbf{0}$ electron

momentum, induced by the spin-orbit interaction. We pinned down the physical origin of the change from the "natural" orbital states $\mu = (x, y, z)$ labeled along the crystal axes, to their linear combinations $\eta = (1, 0, -1)$ similar to the $\ell = 1$ atomic states.

The present work follows the same spirit: by only using conceptually simple arguments, we unravel the physics that drives the positive mass of a free electron toward a negative value when this electron is put in a periodic ion lattice. The sign change in the electron mass, from positive to negative, fundamentally relies on couplings to states that are *opposite* in parity and *higher* in energy than the upper valence states at hand. Yet, this is far from enough. To end with a negative effective mass *whatever* the electron momentum direction **k** with respect to the crystal axes, it is necessary to break the orbital symmetry of the degenerate valence states along these axes. This is done in two ways:

(i) By mixing the orbital states of the degenerate valence level, through its coupling to the orbital states of a *degenerate* level in the conduction band. By contrast, its coupling to a nondegenerate conduction level, like the lowest conduction band, plays no role in getting a negative mass.

(ii) By introducing the electron spin. Through the spin-orbit interaction, the spin quantization axis breaks the orbital degeneracy $\mu = (x, y, z)$ of electrons in a cubic crystal. The resulting orbital eigenstates, labeled as $\eta = \pm 1$, are linear combinations of (x, y) states, while the z state along the spin quantization axis stays unaffected: it just corresponds to the spin-orbit eigenstate labeled as $\eta = 0$.

These two ways of mixing are necessary for valence electrons to end with a negative effective mass whatever the **k** momentum direction. It is then clear that their interplay renders the physics of this negative effective mass quite complicated, and definitely beyond the simple interpretation based on a Kronig-Penney model [20], even if the Bloch theorem mathematically draws the correct conclusion within a few lines. Indeed, the common textbook explanation for negative effective mass involves Bragg reflection: In one-dimensional spinless Kronig-Penney model with a lattice spacing a, negative effective mass occurs when the momentum k approaches the band boundaries $n\pi/a$ from below, with $n = \pm 1, \pm 2, \ldots$. When k is slightly less than $n\pi/a$, the scattering processes from each lattice potential interfere in such a way that the particle wave is largely reflected, that is, propagates backward even as we increase its momentum k by applying a force; the particle transfers more momentum to the lattice than it receives from the applied force [4]. While this physical argument supports the negative effective mass as a general feature in all periodic crystals, it is obvious that the physics of negative effective mass in real materials is more involved.

In order to catch the above physics in the simplest way, we here use a **k** · **p** approach [27–29]. Its big advantage is to possibly analyze, step by step, the various couplings experienced by the valence electrons, to ultimately catch the key ones.

This paper is organized as follows. Section II provides general arguments on the physical origin of a negative valence mass and the possible dependence of the electron energy on the momentum **k**, making use of the fact that in a cubic crystal, the (x, y, z) axes play the same role.

Section III recalls the **k** · **p** procedure to obtain the electron energy up to second order in momentum **k**. We give the possible forms for the degenerate and nondegenerate orbital wave functions of a cubic crystal, which fundamentally are even in the case of the valence band and odd in the case of the conduction band. We also give the form of the eigenstates resulting from spin-orbit interaction. Finally, for later use, we relate operators written in the $\mu = (x, y, z)$ basis and the $\eta = (1, 0, -1)$ basis.

In Sec. IV, we derive the energy of a valence electron in a threefold orbital level, by considering its coupling to an orbital level in the conduction band which is either nondegenerate or threefold, or its coupling to both types of levels.

In Sec. V, we perform similar calculations, focusing on the fourfold spin-orbit eigenstates, labeled as $\tilde{j}_z = (\pm 3/2, \pm 1/2)$, which are the physically relevant states in the presence of spin-orbit interaction.

We then conclude.

II. GENERAL ARGUMENTS

We consider a direct-gap semiconductor like GaAs, with band extrema occurring for **k** = **0**, called the Γ point. We want to determine the dispersion relation close to the maximum of the valence band that originates from threefold orbital states. Due to symmetry between the cubic crystal axes (x, y, z), we expect the dispersion relation, that is, the energy difference $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}}$, to depend on the components of the momentum **k** through (k_x^2, k_y^2, k_z^2) taken in a cyclic way. Possible combinations at lowest order in k are $k_x^2 + k_y^2 + k_z^2 = k^2$,

$$k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2 = S_\mathbf{k}, \quad k_x^2 k_y^2 k_z^2 = P_\mathbf{k}. \quad (1)$$

Terms like $S_\mathbf{k}$ or $P_\mathbf{k}$ in the dispersion relation cause the energy to be nonspherical—known as "warping." Handling such a nonspherical energy in problems dealing with Coulomb interaction is fundamentally impossible. This is why the warping is commonly averaged out to produce a spherical mass. We will, however, see that the existence of a warping in the valence band is fundamentally linked to the existence of a negative curvature: it is a crucial feature of the valence effective mass problem.

III. REQUIRED BACKGROUND

The derivation we propose for the valence electron effective mass does not require the knowledge of group theory. It only uses very basic solid-state physics. Nevertheless, to settle the notations properly, we have chosen to briefly recall what we are going to use.

A. The k · p formalism

The **k** · **p** formalism [27] is a simple but powerful tool to understand semiconductor band structures. It starts with the Hamiltonian of a free electron with positive mass m_0, in a potential $V(\mathbf{r})$ having the lattice periodicity. The Bloch states

$|n, \mathbf{k}\rangle$, eigenstates of this Hamiltonian

$$0 = \left[\frac{\hat{\mathbf{p}}^2}{2m_0} + V(\mathbf{r}) - \varepsilon_{n,\mathbf{k}}\right]|n, \mathbf{k}\rangle, \qquad (2)$$

are characterized by a band index n and a momentum $\mathbf{k}$. When $\mathbf{k}$ is small, the $\varepsilon_{n,\mathbf{k}}$ energy can be obtained from the knowledge of all Bloch states for $\mathbf{k} = \mathbf{0}$. Indeed, by writing the Bloch-state wave function in a sample volume L^3 as

$$\langle \mathbf{r}|n, \mathbf{k}\rangle = \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{L^{3/2}}\langle \mathbf{r}|u_{n,\mathbf{k}}\rangle \qquad (3)$$

in Eq. (2), we find that $|u_{n,\mathbf{k}}\rangle$ fulfills

$$0 = (\hat{h}_{\mathbf{k}} - \varepsilon_{n,\mathbf{k}})|u_{n,\mathbf{k}}\rangle, \qquad (4)$$

with $\hat{h}_{\mathbf{k}} = \hat{h}_{\mathbf{0}} + \hbar^2\mathbf{k}^2/2m_0 + \hat{w}_{\mathbf{k}}$. For $\mathbf{k}$ small, the operator $\hat{w}_{\mathbf{k}}$, defined as

$$\hat{w}_{\mathbf{k}} = \frac{\hbar}{m_0}\mathbf{k}\cdot\hat{\mathbf{p}}, \qquad (5)$$

can be treated as a perturbation.

(i) For $\varepsilon_{n_0,\mathbf{0}}$ nondegenerate, the $\varepsilon_{n_0,\mathbf{k}}$ energy follows from

$$\tilde{\varepsilon}_{n_0,\mathbf{k}} = \varepsilon_{n_0,\mathbf{k}} - \varepsilon_{n_0,\mathbf{0}} - \frac{\hbar^2\mathbf{k}^2}{2m_0}$$

$$\simeq \langle u_{n_0,\mathbf{0}}|\hat{w}_{\mathbf{k}}P_\perp\frac{1}{\varepsilon_{n_0,\mathbf{0}} - \hat{h}_{\mathbf{0}}}P_\perp\hat{w}_{\mathbf{k}}|u_{n_0,\mathbf{0}}\rangle, \qquad (6)$$

where $P_\perp$ is the projector over the subspace orthogonal to $|u_{n_0,\mathbf{0}}\rangle$, namely $P_\perp = \sum_{n\neq n_0}|u_{n,\mathbf{0}}\rangle\langle u_{n,\mathbf{0}}|$. Note that the first-order term $\langle u_{n_0,\mathbf{0}}|\hat{w}_{\mathbf{k}}|u_{n_0,\mathbf{0}}\rangle$ disappears due to parity.

(ii) When $\varepsilon_{n_0,\mathbf{0}}$ is degenerate, that is, when $\varepsilon_{n_0,\mathbf{0}}$ is the energy of $|u_{n_0,\mathbf{0}}^{(r)}\rangle$ states with $r = (1, 2, \ldots, N_0)$, the $\hat{w}_{\mathbf{k}}$ operator shifts and splits these N_0 states. The $\varepsilon_{n_0,\mathbf{k}}$ eigenvalues then follow from the cancellation of the determinant of a $N_0 \times N_0$ matrix $\hat{\mathcal{D}}$ defined as

$$\mathcal{D}_{r',r} = -\tilde{\varepsilon}_{n_0,\mathbf{k}}\delta_{r',r} + \mathcal{W}_{r',r}, \qquad (7)$$

$$\mathcal{W}_{r',r} = \langle u_{n_0,\mathbf{0}}^{(r')}|\hat{w}_{\mathbf{k}}P_\perp\frac{1}{\varepsilon_{n_0,\mathbf{0}} - \hat{h}_{\mathbf{0}}}P_\perp\hat{w}_{\mathbf{k}}|u_{n_0,\mathbf{0}}^{(r)}\rangle, \qquad (8)$$

where $P_\perp$ now is the projector over the subspace orthogonal to the $|u_{n_0,\mathbf{0}}^{(r)}\rangle$ subspace. In practice, calculations are made with a restriction of this orthogonal subspace to states whose energy is close to $\varepsilon_{n_0,\mathbf{0}}$, due to the $(\varepsilon_{n_0,\mathbf{0}} - \hat{h}_{\mathbf{0}})$ denominator in the above equation.

Without going into detailed calculation, some useful observations can be drawn from the above equations.

(i) The states coupled to $|u_{n_0,\mathbf{0}}^{(r)}\rangle$ through $\hat{w}_{\mathbf{k}}$, that enter the $P_\perp$ operator, lead to negative $\mathcal{W}_{r',r}$'s when they are higher in energy than $\varepsilon_{n_0,\mathbf{0}}$. They produce a negative $\tilde{\varepsilon}_{n_0,\mathbf{k}}$ value, that makes the resulting $\varepsilon_{n_0,\mathbf{k}}$ curvature, i.e., the inverse effective mass, smaller than $1/2m_0$, and possibly negative when the coupling is large.

(ii) When the $\hat{\mathcal{W}}$ determinant is equal to zero, the equation $\det\hat{\mathcal{D}} = 0$ must contain a solution $\tilde{\varepsilon}_{n_0,\mathbf{k}} = 0$, that is, $\varepsilon_{n_0,\mathbf{k}} = \varepsilon_{n_0,\mathbf{0}} + \hbar^2\mathbf{k}^2/2m_0$, which corresponds to an effective mass equal to the positive free electron mass m_0. As a result, when $\det\hat{\mathcal{W}} = 0$, we immediately deduce that the amount of states included in the truncated $P_\perp$ operator is not enough to produce a negative effective mass, as required for valence electrons.
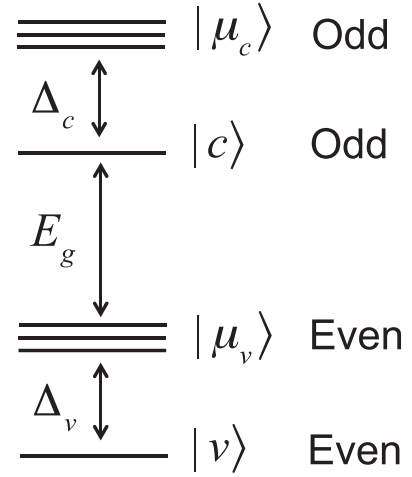

FIG. 1. The physically relevant orbital states are the ones close to the band gap E_g. In zinc-blend semiconductors like GaAs, these are the nondegenerate and threefold degenerate conduction states $|c\rangle$ and $|\mu_c\rangle$ separated by an energy Δ_c, and the nondegenerate and threefold degenerate valence states $|v\rangle$ and $|\mu_v\rangle$ separated by an energy Δ_v. Conduction states are fundamentally odd in parity, while valence states are even.

B. Valence and conduction states with crystal periodicity

The physically relevant states of the valence and conduction bands are the ones close to the band gap. They are shown in Fig. 1. Their orbital parts are either nondegenerate as in the case of the $|v\rangle$ and $|c\rangle$ states, or threefold degenerate as in the case of $|\mu_v\rangle$ and $|\mu_c\rangle$ states, with $\mu = (x, y, z)$ along the crystal axis [28]. Using the lattice periodicity, we can expand their wave functions in terms of the reciprocal lattice vectors $\mathbf{K}$.

Conduction orbital states have a parity which is fundamentally odd. Since the cubic axes play the same role, we can write them [25] as

$$-\langle -\mathbf{r}|c\rangle = \langle \mathbf{r}|c\rangle = \sum_{\mathbf{K}}\frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{L^{3/2}}K_xK_yK_z F_{K,c}, \qquad (9)$$

$$-\langle -\mathbf{r}|\mu_c\rangle = \langle \mathbf{r}|\mu_c\rangle = \sum_{\mathbf{K}}\frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{L^{3/2}}K_\mu G_{K,c}. \qquad (10)$$

Similarly, the valence orbital states, which are fundamentally even, can be written as

$$\langle -\mathbf{r}|v\rangle = \langle \mathbf{r}|v\rangle = \sum_{\mathbf{K}}\frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{L^{3/2}} F_{K,v}, \qquad (11)$$

$$\langle -\mathbf{r}|\mu_v\rangle = \langle \mathbf{r}|\mu_v\rangle = \sum_{\mathbf{K}}\frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{L^{3/2}}\frac{K_xK_yK_z}{K_\mu} G_{K,v}. \qquad (12)$$

These F and G functions only depend on $K = |\mathbf{K}|$.

The cyclic symmetry of these orbital states is broken by the spin quantization axis when spin-orbit interaction is introduced. The six states $|\mu_v\rangle \otimes |\pm 1/2\rangle$ then split into four and two states (see Fig. 2) in a way similar to their atomic counterpart with quantum number j. This is why we label them as $\tilde{j}$, not j, in order to keep the reader reminded that orbital angular momentum has no meaning for electrons in a periodic crystal.
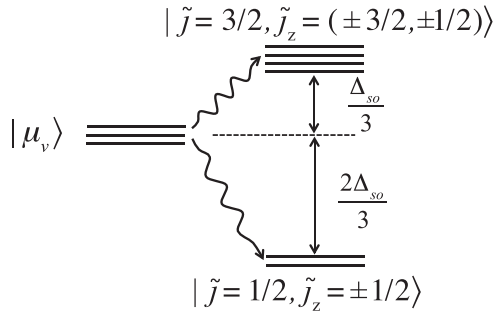
FIG. 2. The spin-orbit interaction splits the six valence states $|\mu_v\rangle \otimes |\pm 1/2\rangle$ into four and two states, separated by Δ_{so}. These states are given in Eqs. (13), (14), and (15).

The upper fourfold states read as

$$|\tilde{\jmath} = 3/2, \tilde{\jmath}_z = \pm 3/2\rangle = |\pm 1_v\rangle \otimes |\pm 1/2\rangle, \qquad (13)$$

$$|\tilde{\jmath} = 3/2, \tilde{\jmath}_z = \pm 1/2\rangle$$
$$= \frac{1}{\sqrt{3}}(|\pm 1_v\rangle \otimes |\mp 1/2\rangle + \sqrt{2}|0_v\rangle \otimes |\pm 1/2\rangle), \quad (14)$$

while the lower twofold states read as

$$|\tilde{\jmath} = 1/2, \tilde{\jmath}_z = \pm 1/2\rangle$$
$$= \frac{1}{\sqrt{3}}(\sqrt{2}|\pm 1_v\rangle \otimes |\mp 1/2\rangle - |0_v\rangle \otimes |\pm 1/2\rangle). \quad (15)$$

For a spin quantization axis taken along **z**, the orbital states $|\eta_v\rangle$ with $\eta = (1, 0, -1)$ read in terms of the $|\mu_v\rangle$ orbital states as

$$|\pm 1_v\rangle = \frac{\mp i|x_v\rangle + |y_v\rangle}{\sqrt{2}}, \quad |0_v\rangle = i|z_v\rangle, \qquad (16)$$

following the Landau-Lifshitz choice [30] for the arbitrary phase factor.

C. From (x, y, z) to $(1, 0, -1)$ orbital states

The above results show that the relevant orbital indices to handle spin-orbit interaction are not the ones linked to the crystal axes because their cyclic symmetry is broken by the spin quantization axis. Yet, this cyclic symmetry renders the (x, y, z) indices quite convenient for calculations before introducing the spin. This is why we will start with operator written in the (x, y, z) basis and then turn to its representation in the $(1, 0, -1)$ basis.

The 3×3 matrix that represents the operator $\hat{A}$ in the $\mu = (x, y, z)$ state basis reads as

$$\hat{A}_\mu = \begin{pmatrix} A_{x,x} & A_{x,y} & A_{x,z} \\ A_{y,x} & A_{y,y} & A_{y,z} \\ A_{z,x} & A_{z,y} & A_{z,z} \end{pmatrix}, \qquad (17)$$

with $A_{\mu',\mu} = \langle \mu'|\hat{A}|\mu\rangle$, equal to $A^*_{\mu,\mu'}$ when $\hat{A} = \hat{A}^\dagger$. In the problem we here consider, these matrix elements are real; so $A_{\mu',\mu} = A_{\mu,\mu'}$, as taken in the following.

In the $\eta = (1, 0, -1)$ basis, the same operator reads as

$$\hat{A}_\eta = \begin{pmatrix} A_{1,1} & A_{1,0} & A_{1,-1} \\ A_{0,1} & A_{0,0} & A_{0,-1} \\ A_{-1,1} & A_{-1,0} & A_{-1,-1} \end{pmatrix}, \qquad (18)$$

with $A_{\eta',\eta} = \langle \eta'|\hat{A}|\eta\rangle$ still equal to $A^*_{\eta,\eta'}$ because $\hat{A} = \hat{A}^\dagger$. But $A_{\eta',\eta}$ is not necessarily equal to $A_{\eta,\eta'}$ because the matrix elements of $\hat{A}$, real in the μ basis, are not necessarily real in this η basis.

We can relate these two sets of matrix elements by inserting the closure relation $\sum_\mu |\mu\rangle\langle\mu|$ on both sides of the operator $\hat{A}$ as

$$A_{\eta',\eta} = \langle \eta'|\hat{A}|\eta\rangle = \langle \eta'| \sum_{\mu'} |\mu'\rangle\langle\mu'|\hat{A} \sum_\mu |\mu\rangle\langle\mu|\eta\rangle. \quad (19)$$

The link between the $|\eta\rangle$ and $|\mu\rangle$ states given in Eq. (16) leads to

$$A_{1,1} = \frac{i\langle x_v| + \langle y_v|}{\sqrt{2}} \hat{A} \frac{-i|x_v\rangle + |y_v\rangle}{\sqrt{2}} = \frac{A_{x,x} + A_{y,y}}{2}, \quad (20)$$

$$A_{\pm 1,0} = \frac{\pm i\langle x_v| + \langle y_v|}{\sqrt{2}} \hat{A} i|z_v\rangle = \frac{\mp A_{x,z} + iA_{y,z}}{\sqrt{2}}, \quad (21)$$

and so on. So, the $\hat{A}$ operator in the $(1, 0, -1)$ basis reads as

$$\hat{A}_\eta = \begin{pmatrix} \frac{A_{x,x}+A_{y,y}}{2} & \frac{-A_{x,z}+iA_{y,z}}{\sqrt{2}} & \frac{-A_{x,x}+A_{y,y}+2iA_{x,y}}{2} \\ \cdot & A_{z,z} & \frac{A_{x,z}-iA_{y,z}}{\sqrt{2}} \\ \cdot & \cdot & \frac{A_{x,x}+A_{y,y}}{2} \end{pmatrix}, \quad (22)$$

the other elements being obtained from $A^*_{\eta',\eta} = A_{\eta,\eta'}$.

The determinants of the two matrices $\hat{A}_\mu$ and $\hat{A}_\eta$ are equal because these two matrices represent the same operator $\hat{A}$ in different bases of the same 3×3 subspace. Yet, the cyclic symmetry of the (x, y, z) indices renders the calculation of this determinant far simpler for $\hat{A}_\mu$.

IV. VALENCE ELECTRON EFFECTIVE MASS IN THE ABSENCE OF SPIN

We first forget spin and look for the effect of the $\hat{w}_\mathbf{k}$ operator on the threefold orbital states $|\mu_v\rangle$ given in Sec. III. Since the $\hat{\mathbf{p}} = -i\hbar\nabla$ operator in $\hat{w}_\mathbf{k}$ is odd, it couples the even valence states $|\mu_v\rangle$ to odd states only; so, it does not couple $|\mu_v\rangle$ to other valence states, neither $|v\rangle$ nor $|\mu'_v\rangle$ with $\mu' \neq \mu$, but to the conduction state $|c\rangle$ and possibly to the $|\mu_c\rangle$ states slightly above in energy (see Fig. 1).

A. Coupling to $|c\rangle$ only

Using the periodic wave functions given in Sec. III B, we find

$$\langle c|\hat{p}_x|\mu_v\rangle = \int d^3r \sum_{\mathbf{K}'} \frac{e^{-i\mathbf{K}'\cdot\mathbf{r}}}{L^{3/2}} K'_x K'_y K'_z F^*_{\mathbf{K}',c}$$

$$\times \frac{\hbar}{i} \frac{\partial}{\partial x}\left(\sum_\mathbf{K} \frac{e^{i\mathbf{K}\cdot\mathbf{r}}}{L^{3/2}} \frac{K_x K_y K_z}{K_\mu} G_{\mathbf{K},v}\right). \quad (23)$$

As $\frac{\partial}{\partial x}e^{i\mathbf{K}\cdot\mathbf{r}} = iK_x e^{i\mathbf{K}\cdot\mathbf{r}}$, the integral over $\mathbf{r}$ gives $L^3\delta_{\mathbf{K}',\mathbf{K}}$; so, the RHS of the above equation reduces to

$$\hbar \sum_{\mathbf{K}} \frac{K_x}{K_\mu} K_x^2 K_y^2 K_z^2 F_{K,c}^* G_{K,v}, \tag{24}$$

which, for a cubic crystal, is equal to zero except for $\mu = x$. So

$$\langle c|\hat{p}_x|\mu_v\rangle = \delta_{x,\mu}\hbar \sum_{\mathbf{K}} K_x^2 K_y^2 K_z^2 F_{K,c}^* G_{K,v} \equiv \delta_{x,\mu}\mathcal{P}_{cv}. \tag{25}$$

The matrix elements of the $\hat{w}_{\mathbf{k}}$ operator between the threefold valence states $|\mu_v\rangle$ and the nondegenerate conduction state $|c\rangle$ then reduce to

$$\langle c|\hat{w}_{\mathbf{k}}|\mu_v\rangle = \langle c|\frac{\hbar}{m_0}(k_x\hat{p}_x + k_y\hat{p}_y + k_z\hat{p}_z)|\mu_v\rangle$$
$$= \frac{\hbar\mathcal{P}_{cv}}{m_0}k_\mu. \tag{26}$$

When the $|c\rangle$ state only is included in the $P_\perp$ projector appearing in Eq. (8), we get

$$\mathcal{W}_{\mu',\mu} = \langle \mu_v'|\hat{w}_{\mathbf{k}}\frac{|c\rangle\langle c|}{-E_g}\hat{w}_{\mathbf{k}}|\mu_v\rangle \tag{27}$$
$$= -\frac{\hbar^2|\mathcal{P}_{cv}|^2}{m_0^2 E_g}k_{\mu'}k_\mu \equiv -\frac{\hbar^2}{2m_0}\gamma_1\langle\mu_v'|\hat{\mathcal{B}}^{(1)}|\mu_v\rangle,$$

where γ_1 is a dimensionless parameter given by

$$\gamma_1 = \frac{2|\mathcal{P}_{cv}|^2}{m_0 E_g}. \tag{28}$$

The corresponding 3×3 matrix for $\hat{\mathcal{W}}$ then reads $-(\hbar^2\gamma_1/2m_0)\hat{\mathcal{B}}^{(1)}$, where $\hat{\mathcal{B}}^{(1)}$ is a symmetric real matrix

$$\hat{\mathcal{B}}_\mu^{(1)} = \begin{pmatrix} k_x^2 & \cdot & \cdot \\ k_x k_y & k_y^2 & \cdot \\ k_z k_x & k_y k_z & k_z^2 \end{pmatrix}. \tag{29}$$

Since $\det\hat{\mathcal{B}}^{(1)} = 0$, as easy to check, we readily see that the equation $\det\hat{\mathcal{D}} = 0$, for coupling to $|c\rangle$ only, has a solution $\tilde{\varepsilon}_{v,\mathbf{k}} = 0$, which corresponds to an unchanged positive mass m_0. More precisely, the solutions of $\det\hat{\mathcal{D}} = 0$ read as $\tilde{\varepsilon}_{v,\mathbf{k}} = (\hbar^2/2m_0)\gamma_1 e_1$ with e_1 solution of

$$0 = \begin{vmatrix} k_x^2 + e_1 & \cdot & \cdot \\ k_x k_y & k_y^2 + e_1 & \cdot \\ k_z k_x & k_y k_z & k_z^2 + e_1 \end{vmatrix} = e_1^2(e_1 + k^2). \tag{30}$$

So the coupling to $|c\rangle$ only leads to a partial splitting of the valence orbital states $|\mu_v\rangle$, with two unchanged branches still having a positive mass m_0 which corresponds to the $e_1 = 0$ degenerate solution, and one branch $e_1 = -k^2$ which gives

$$\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + (1 - \gamma_1)\frac{\hbar^2 k^2}{2m_0}. \tag{31}$$

This branch can have a negative effective mass for $\gamma_1 > 1$, that is, a coupling $\mathcal{P}_{cv}$ between $|\mu_v\rangle$ and $|c\rangle$ large enough to have $2|\mathcal{P}_{cv}|^2/m_0$ larger than the band gap E_g.

B. Coupling to the $|\mu_c\rangle$ subspace only

To better catch the role of the orbital state degeneracy, we now consider the coupling of the threefold valence states $|\mu_v\rangle$ to the threefold conduction states $|\mu_c\rangle$, these two sets of states having opposite parity.

Using the periodic wave functions given in Sec. III B, we find

$$\langle\mu_c'|\hat{p}_x|\mu_v\rangle = \sum_{\mathbf{K}}(K_{\mu'}G_{K,c}^*)\hbar K_x\left(\frac{K_x K_y K_z}{K_\mu}G_{K,v}\right), \tag{32}$$

which differs from zero for (μ', μ) equal to (y, z) or to (z, y). So, we end, since (x, y, z) play the same role, with

$$\langle y_c|\hat{p}_x|z_v\rangle = \langle z_c|\hat{p}_x|y_v\rangle = \hbar\sum_{\mathbf{K}}K_x^2 K_y^2 G_{K,c}^* G_{K,v} \equiv \mathcal{Q}_{cv}, \tag{33}$$

and similar results obtained from cyclic permutations.

The matrix elements of the $\hat{w}_{\mathbf{k}}$ operator between the threefold valence states $|\mu_v\rangle$ and the threefold conduction states $|\mu_c'\rangle$ then read as

$$\langle\mu_c'|\hat{w}_{\mathbf{k}}|x_v\rangle = \langle\mu_c'|\frac{\hbar}{m_0}(k_x\hat{p}_x + k_y\hat{p}_y + k_z\hat{p}_z)|x_v\rangle$$
$$= \frac{\hbar\mathcal{Q}_{cv}}{m_0}(k_y\delta_{\mu',z} + k_z\delta_{\mu',y}). \tag{34}$$

When the $P_\perp$ projector appearing in Eq. (8) only contains the $|\mu_c\rangle$ states, we find

$$\mathcal{W}_{\mu',\mu} = \langle\mu_v'|\hat{w}_{\mathbf{k}}\sum_{\mu''}\frac{|\mu_c''\rangle\langle\mu_c''|}{-(E_g + \Delta_c)}\hat{w}_{\mathbf{k}}|\mu_v\rangle$$
$$\equiv -\frac{\hbar^2}{2m_0}\gamma_3\langle\mu_v'|\hat{\mathcal{B}}^{(3)}|\mu_v\rangle, \tag{35}$$

where the dimensionless parameter γ_3 is given by

$$\gamma_3 = \frac{2|\mathcal{Q}_{cv}|^2}{m_0(E_g + \Delta_c)}. \tag{36}$$

The corresponding 3×3 matrix for $\hat{\mathcal{W}}$ then reads $-(\hbar^2\gamma_3/2m_0)\hat{\mathcal{B}}^{(3)}$ where the $\hat{\mathcal{B}}^{(3)}$ operator in the $\mu = (x, y, z)$ basis also is a symmetric real matrix

$$\hat{\mathcal{B}}_\mu^{(3)} = \begin{pmatrix} k_y^2 + k_z^2 & \cdot & \cdot \\ k_x k_y & k_z^2 + k_x^2 & \cdot \\ k_z k_x & k_y k_z & k_x^2 + k_y^2 \end{pmatrix}. \tag{37}$$

The determinant of this matrix is equal to $4k_x^2 k_y^2 k_z^2$; so, for any $k_\mu = 0$, the equation $\det\hat{\mathcal{D}} = 0$ has a solution $\tilde{\varepsilon}_{v,\mathbf{k}} = 0$, that is, valence electrons still having a positive mass m_0. More precisely, the solutions of $\det\hat{\mathcal{D}} = 0$ read as $\tilde{\varepsilon}_{v,\mathbf{k}} = (\hbar^2/2m_0)\gamma_3 e_3$ with e_3 given by

$$0 = \begin{vmatrix} k_y^2 + k_z^2 + e_3 & \cdot & \cdot \\ k_x k_y & k_z^2 + k_x^2 + e_3 & \cdot \\ k_z k_x & k_y k_z & k_x^2 + k_y^2 + e_3 \end{vmatrix}$$
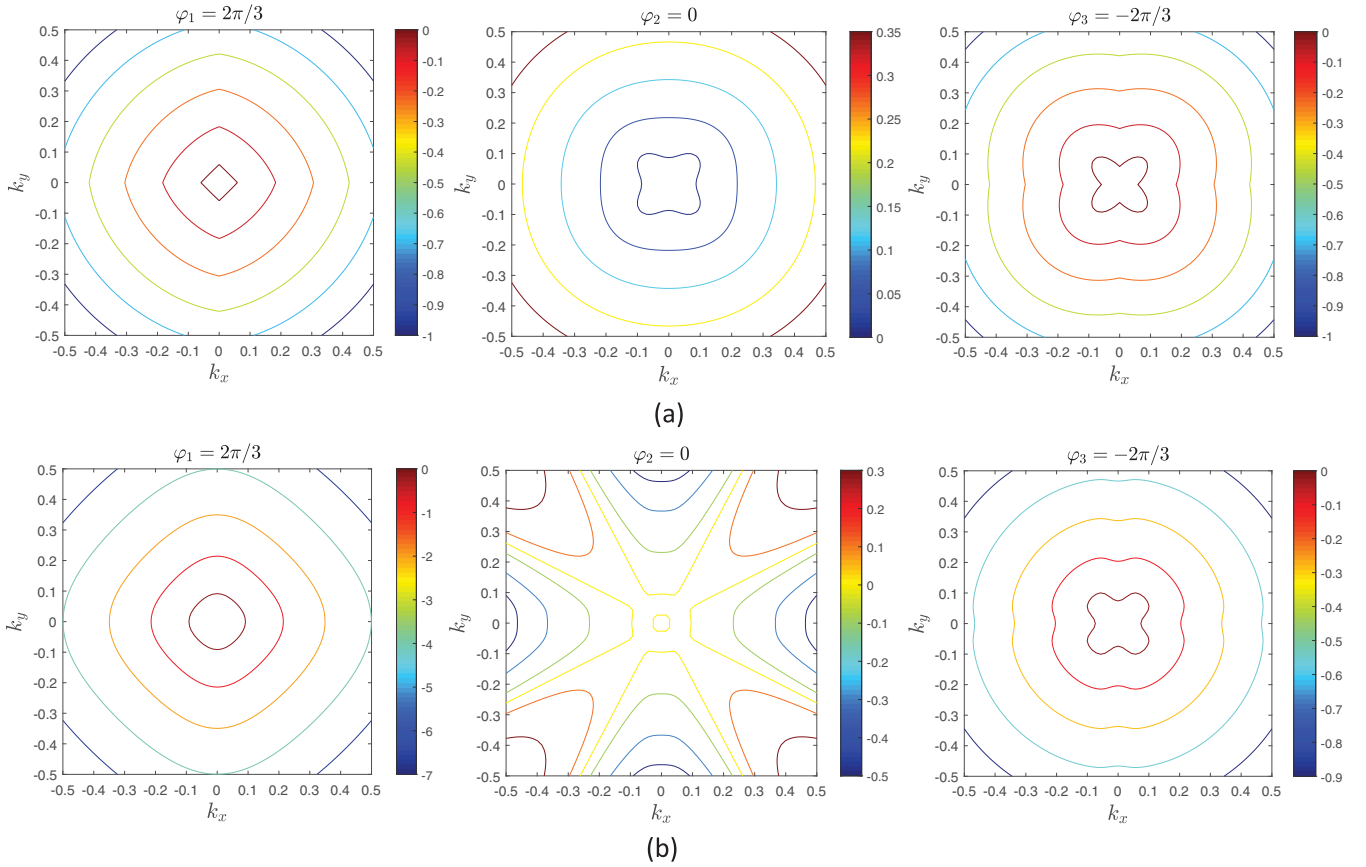$$= e_3(e_3 + k^2)^2 + 4k_x^2 k_y^2 k_z^2. \tag{38}$$

(a)

(b)

FIG. 3. (a) Contours of the valence electron energies $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}}$ (in the unit of $\hbar^2/2m_0$) as a function of (k_x, k_y) for $k_z = 0.05$. They correspond to the three solutions given in Eq. (43) for coupling between $|\mu_v\rangle$ and $|\mu_c\rangle$ only. In each subfigure, the color of each contour curve represents the energy value given by the color bar on the right. The solution for $\varphi_2 = 0$ corresponds to a positive mass, as can be deduced from the fact that the $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}}$ energy increases from 0 with increasing k. The other two solutions, $\varphi_1 = 2\pi/3$ and $\varphi_3 = -2\pi/3$, correspond to negative masses, as seen from the fact that the $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}}$ energy decreases from 0 with increasing k. (b) Same for Eq. (46) that correspond to coupling between $|\mu_v\rangle$ and both $|c\rangle$ and $|\mu_c\rangle$ states of the conduction band. The solution for $\varphi_2 = 0$ corresponds to a positive mass along $k_x = \pm k_y$ and a negative mass along $k_x = 0$ or $k_y = 0$; the change occurs along the yellow contour lines which correspond to $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}} = 0$, that is, the inverse of the effective mass equal to zero. The other two solutions produce negative masses, as $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}}$ decreases from 0 with increasing k. We have taken $\gamma_1 = 16.7$ and $\gamma_3 = 3.5$ as for GaAs [32].

For $k_x k_y k_z = 0$, this gives

$$\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + \frac{\hbar^2}{2m_0}k^2 \quad \text{nondegenerate,} \tag{39}$$

$$\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + \frac{\hbar^2}{2m_0}(1 - \gamma_3)k^2 \quad \text{twofold.} \tag{40}$$

So the coupling to $|\mu_c\rangle$ only does not yet provide valence electrons with an effective mass negative for whatever $\mathbf{k}$.

It is possible to analytically solve Eq. (38) for $k_x k_y k_z \neq 0$ by using the Cardano's trick [31]: The three solutions of $x^3 + 3ax - b = 0$ are

$$x^{(n)} = \sum_{\tau=\pm 1} e^{i\tau\varphi_n} \left[\frac{b + \tau\sqrt{b^2 + 4a^3}}{2} \right]^{1/3} \tag{41}$$

for $\varphi_n = (0, \pm 2\pi/3)$. To use it, we first introduce $\Delta_3 = e_3 + 2k^2/3$. From Eq. (38), we find that the resulting cubic polynomial equation for Δ_3 has no quadratic term

$$0 = \Delta_3^3 - \frac{k^4}{3}\Delta_3 - \frac{2k^6}{27} + 4k_x^2 k_y^2 k_z^2. \tag{42}$$

Since $k_x^2 k_y^2 k_z^2 \leqslant (k^2/3)^3$, Eq. (41) gives the three solutions of the above equation as

$$\Delta_3^{(n)} = \sum_{\tau=\pm 1} e^{i\tau\varphi_n} \left[\left(\frac{k^2}{3}\right)^3 - 2P_\mathbf{k} + 2i\tau\sqrt{P_\mathbf{k}\left(\frac{k^2}{3}\right)^3 - P_\mathbf{k}^2} \right]^{1/3}, \tag{43}$$

with $P_\mathbf{k} = k_x^2 k_y^2 k_z^2$. The shapes of the resulting energy contours for $\varepsilon_{v,\mathbf{k}}$ are shown in Fig. 3(a).

C. Coupling to both $|c\rangle$ and $|\mu_c\rangle$ states

As now shown, the coupling to the $|c\rangle$ and $|\mu_c\rangle$ states not only decreases the number of $\mathbf{k}$ momenta at which the valence electron effective mass stays equal to m_0, but this coupling also generates a warping dependence of the dispersion relation in $S_\mathbf{k} = k_x^2 k_y^2 + k_y^2 k_z^2 + k_z^2 k_x^2$.

For $P_\perp = |c\rangle\langle c| + \sum_\mu |\mu_c\rangle\langle\mu_c|$, we find

$$\mathcal{W}_{\mu',\mu} = -\frac{\hbar^2}{2m_0}\left[\gamma_1 \mathcal{B}_{\mu',\mu}^{(1)} + \gamma_3 \mathcal{B}_{\mu',\mu}^{(3)}\right] \equiv -\frac{\hbar^2}{2m_0}\langle\mu'_v|\hat{\mathcal{B}}|\mu_v\rangle. \tag{44}$$

Using the $\hat{\mathcal{B}}_\mu^{(1)}$ and $\hat{\mathcal{B}}_\mu^{(3)}$ matrices calculated above, we can write $\hat{\mathcal{B}}_\mu$ as $\gamma_3 k^2 \mathcal{I} + \hat{\mathcal{C}}_\mu$, where $\mathcal{I}$ is the 3×3 unit matrix while the matrix $\hat{\mathcal{C}}_\mu$ is given by

$$\hat{\mathcal{C}}_\mu = \begin{pmatrix} k_x^2 \gamma_- & \cdot & \cdot \\ k_x k_y \gamma_+ & k_y^2 \gamma_- & \cdot \\ k_z k_x \gamma_+ & k_y k_z \gamma_+ & k_z^2 \gamma_- \end{pmatrix}, \qquad (45)$$

with $\gamma_\pm = \gamma_1 \pm \gamma_3$.

By writing $\tilde{\varepsilon}_{v,\mathbf{k}}$ as $(\hbar^2/2m_0)e$ with $e = e' - \gamma_3 k^2$, we find that the valence electron eigenvalues resulting from the couplings to the $|c\rangle$ and $|\mu_c\rangle$ conduction states follow from

$$0 = e'^3 + (\gamma_1 - \gamma_3)k^2 e'^2 - 4\gamma_1 \gamma_3 S_\mathbf{k} e' + 4\gamma_3^2 (3\gamma_1 + \gamma_3) P_\mathbf{k}, \tag{46}$$

with $(S_\mathbf{k}, P_\mathbf{k})$ defined in Eq. (1). The above equation demonstrates a warping behavior through the $S_\mathbf{k}$ term; note that both the γ_1 and γ_3 couplings to the $|c\rangle$ and $|\mu_c\rangle$ conduction states are necessary to bring the warping into the problem. Moreover, we see that when these two couplings exist, that is, when $\gamma_1 \gamma_3 \neq 0$, the above equation has a solution $e' = \gamma_3 k^2$, that is, $e = 0$, for

$$k_x = 0, \quad k_y = \pm k_z, \tag{47}$$

and their cyclic permutations. So, for some k values, the valence electron effective mass still is positive and equal to m_0.

For arbitrary k's, we can obtain the three solutions of Eq. (46) by again using the Cardano's trick. The shapes of the energy contours for $\varepsilon_{v,\mathbf{k}}$ are shown in Fig. 3(b).

All this shows that including the couplings to the $|c\rangle$ state and the $|\mu_c\rangle$ states of the conduction band reduces the number of $\mathbf{k}$ momenta at which the valence electron effective mass stays equal to m_0. However, this mixing of orbital symmetries is not enough to produce a negative curvature for whatever $\mathbf{k}$. We are going to show that an additional mixing, that comes from the spin-orbit interaction, is necessary to endow a valence electron with a negative effective mass whatever its $\mathbf{k}$ momentum.

D. $\hat{\mathcal{B}}^{(1)}$ and $\hat{\mathcal{B}}^{(3)}$ operators in the $(1, 0, -1)$ basis

To derive the effective masses of valence electrons in the presence of spin, we have to turn from the (x, y, z) crystal basis, for which calculations are easy to perform with the help of cyclic permutations, to the $(1, 0, -1)$ basis for which such cyclic symmetry is broken by the spin quantization axis.

By inserting the $\hat{\mathcal{B}}_\mu^{(1)}$ matrix elements in the (x, y, z) basis, given in Eq. (29), into the general link given in Eq. (22) between matrices in different bases, we obtain the $\hat{\mathcal{B}}^{(1)}$ operator in the $(1, 0, -1)$ basis as

$$\hat{\mathcal{B}}_\eta^{(1)} = \begin{pmatrix} |k_1|^2 & \cdot & \cdot \\ k_0^* k_1 & |k_0|^2 & \cdot \\ k_{-1}^* k_1 & k_{-1}^* k_0 & |k_{-1}|^2 \end{pmatrix}, \qquad (48)$$

with $k_{\pm 1} = (\mp i k_x + k_y)/\sqrt{2}$ and $k_0 = i k_z$. We can check that $\det \hat{\mathcal{B}}_\eta^{(1)} = 0$ as expected, since the determinant is invariant under a basis change. We can also recover the eigenvalue equation for e_1 given in Eq. (30) but in a far heavier way, due to the broken symmetry of the $(1, 0, -1)$ basis along z.

In the same way, changing from (x, y, z) to $(1, 0, -1)$ basis gives the $\hat{\mathcal{B}}^{(3)}$ operator in the $(1, 0, -1)$ basis as

$$\hat{\mathcal{B}}_\eta^{(3)} = \begin{pmatrix} |k_1|^2 + |k_0|^2 & \cdot & \cdot \\ k_0^* k_1 & 2|k_1|^2 & \cdot \\ -k_1^{*2} & k_0 k_1 & |k_{-1}|^2 + |k_0|^2 \end{pmatrix}. \qquad (49)$$

It still is rather easy to check that the determinant of this matrix is equal to $4P_\mathbf{k}$, but the derivation of Eq. (38) for e_3 takes more time.

It can be of interest to note that the off-diagonal terms of the $\hat{\mathcal{B}}^{(1)}$ and $\hat{\mathcal{B}}^{(3)}$ operators are identical when written in the (x, y, z) basis, while only two of them are identical when written in the $(1, 0, -1)$ basis. Difference between $\mathcal{B}_{-1,1}^{(1)}$ and $\mathcal{B}_{-1,1}^{(3)}$ comes from the broken symmetry along the spin quantization axis which produces different diagonal terms, $(-\mathcal{B}_{x,x}^{(1)} + \mathcal{B}_{y,y}^{(1)})/2$ and $(-\mathcal{B}_{x,x}^{(3)} + \mathcal{B}_{y,y}^{(3)})/2$, in the $(1, -1)$ couplings [see Eq. (22)].

V. SPIN-ORBIT EFFECT ON VALENCE EFFECTIVE MASSES

A. From the $|\eta\rangle \otimes |\pm 1/2\rangle$ to the $|\tilde{\jmath}\rangle$ states

The first step to handle the effect of spin-orbit coupling on the valence states is to go from the $|\eta\rangle \otimes |\pm 1/2\rangle$ states to their linear combinations $|\tilde{\jmath}\rangle$ which are eigenstate of the spin-orbit interaction. In the following, we consider that the spin-orbit splitting Δ_{so} is large enough and the $\mathbf{k}$ momentum small enough to possibly restrict the spin-orbit eigenstates to the fourfold states $|\tilde{\jmath} = 3/2\rangle$ given in Eqs. (13) and (14), in spite of the fact that these states are coupled to the $|\tilde{\jmath} = 1/2\rangle$ states which have the same parity, through their common coupling to the conduction states $|c\rangle$ or $|\mu_c\rangle$.

For $\hat{A}$ operator that does not act on spin, like the $\hat{w}_\mathbf{k}$ operator, the parts of the $|\tilde{\jmath}\rangle$ states that are coupled must have the same spin. Accordingly, we get, using the $\hat{A}_\eta$ matrix elements defined in Eq. (19),

$$A_{\frac{3}{2},\frac{3}{2}} = \langle 3/2|\hat{A}|3/2\rangle = \langle 1/2| \otimes \langle 1|\hat{A}|1\rangle \otimes |1/2\rangle = A_{1,1}. \tag{50}$$

In the same way

$$A_{\frac{1}{2},\frac{3}{2}} = \frac{\langle -1/2| \otimes \langle 1| + \sqrt{2}\langle 1/2| \otimes \langle 0|}{\sqrt{3}} \hat{A}|1\rangle \otimes |1/2\rangle$$
$$= \sqrt{\frac{2}{3}} A_{0,1}, \tag{51}$$

and so on.

The $\hat{A}$ operator in the $(3/2, 1/2, -1/2, -3/2)$ basis then appears as

$$\hat{A}_{\tilde{\jmath}} = \begin{pmatrix} A_{1,1} & \cdot & \cdot & \cdot \\ \sqrt{\frac{2}{3}} A_{0,1} & \frac{A_{1,1} + 2A_{0,0}}{3} & \cdot & \cdot \\ \frac{1}{\sqrt{3}} A_{-1,1} & \sqrt{2} \frac{A_{0,1} + A_{-1,0}}{3} & \frac{A_{1,1} + 2A_{0,0}}{3} & \cdot \\ 0 & \frac{1}{\sqrt{3}} A_{-1,1} & \sqrt{\frac{2}{3}} A_{-1,0} & A_{-1,-1} \end{pmatrix}. \tag{52}$$

Note that for real $A_{\mu',\mu}$'s, as in the case of present interest,

$$A_{0,1} + A_{-1,0} = \frac{-A_{z,x} - iA_{z,y}}{\sqrt{2}} + \frac{A_{x,z} + iA_{y,z}}{\sqrt{2}} = 0. \quad (53)$$

So all elements in the second diagonal are equal to zero.

B. Coupling to $|c\rangle \otimes |\pm 1/2\rangle$ states only

From the $\hat{\mathcal{B}}_\eta^{(1)}$ matrix in the 3×3 orbital subspace $(1, 0, -1)$, given in Eq. (48), we can derive its expression in the 4×4 subspace $(3/2, 1/2, -1/2, -3/2)$, by using Eq. (52). We find

$$\hat{\mathcal{B}}_{\tilde{j}}^{(1)} = \begin{pmatrix} |k_1|^2 & \cdot & \cdot & \cdot \\ \sqrt{\frac{2}{3}}k_0^* k_1 & \frac{|k_1|^2 + 2|k_0|^2}{3} & \cdot & \cdot \\ \frac{1}{\sqrt{3}}k_1^2 & 0 & \frac{|k_1|^2 + 2|k_0|^2}{3} & \cdot \\ 0 & \frac{1}{\sqrt{3}}k_1^2 & \sqrt{\frac{2}{3}}k_0 k_1 & |k_1|^2 \end{pmatrix}. \quad (54)$$

The determinant of this matrix is equal to zero; so there still is a solution with a positive valence electron mass m_0. More precisely, the eigenvalue equation obtained from this matrix reads

$$0 = e_1^2 \left(e_1^2 + 2k^2/3 \right)^2. \quad (55)$$

It has a twofold solution $e_1 = 0$, that leads to $\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + \hbar^2 k^2 / 2m_0$ and a twofold solution $e_1 = -2k^2/3$, that leads to

$$\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + \frac{\hbar^2 k^2}{2m_0}\left(1 - \frac{2}{3}\gamma_1\right). \quad (56)$$

This shows that the coupling to the nondegenerate conduction level splits the upper fourfold valence states resulting from spin-orbit interaction, into two states still having a positive effective mass equal to m_0 and two states with an effective mass possibly negative for a coupling γ_1 large enough. The effect of the spin-orbit interaction slightly reduces this possibility through a prefactor increase from $(1 - \gamma_1)$ to $(1 - 2\gamma_1/3)$.

C. Coupling to $|\mu_c\rangle \otimes |\pm 1/2\rangle$ only

For a spin-orbit splitting small compared to the band gap, we can neglect its effect on the threefold orbital states $|\mu_c\rangle$ of the conduction band, and consider that the six states $|\mu_c\rangle \otimes |\pm 1/2\rangle$ have the same energy $(E_g + \Delta_c)$.

From the $\hat{\mathcal{B}}_\eta^{(3)}$ matrix in the 3×3 orbital subspace $(1, 0, -1)$, given in Eq. (49), we can derive its expression in the 4×4 subspace $(3/2, 1/2, -1/2, -3/2)$, by using Eq. (52). We find

$$\hat{\mathcal{B}}_{\tilde{j}}^{(3)} = \begin{pmatrix} |k_1|^2 + |k_0|^2 & \cdot & \cdot & \cdot \\ \sqrt{\frac{2}{3}}k_0^* k_1 & \frac{5|k_1|^2 + |k_0|^2}{3} & \cdot & \cdot \\ -\frac{1}{\sqrt{3}}k_1^{*2} & 0 & \frac{5|k_1|^2 + |k_0|^2}{3} & \cdot \\ 0 & -\frac{1}{\sqrt{3}}k_1^{*2} & \sqrt{\frac{2}{3}}k_0 k_1 & |k_1|^2 + |k_0|^2 \end{pmatrix}. \quad (57)$$

This matrix has obvious similarities with $\hat{\mathcal{B}}_{\tilde{j}}^{(1)}$, but some crucial differences. These differences bring to $\hat{\mathcal{B}}_{\tilde{j}}^{(3)}$ a determinant $k^8/9$ which differs from zero for *all* directions of the electron $\mathbf{k}$ momentum. As a result, there is no more solution $\tilde{\varepsilon}_{v,\mathbf{k}} = 0$:
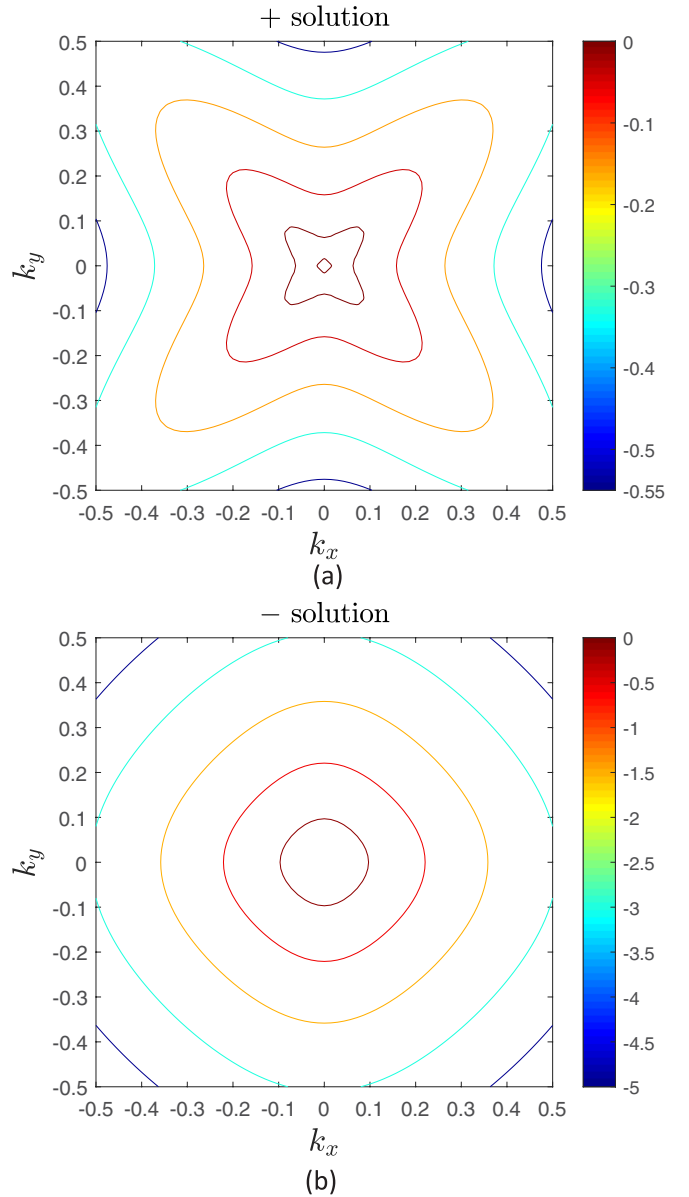

FIG. 4. Energy contours of $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}}$ (in the unit of $\hbar^2/2m_0$) corresponding to (a) the $+$ solution and (b) the $-$ solution given in Eq. (63). Parameters are the same as in Fig. 3 for GaAs. The two solutions correspond to a decrease of $\varepsilon_{v,\mathbf{k}} - \varepsilon_{v,\mathbf{0}}$ from 0 for whatever $\mathbf{k}$, that is, a negative effective mass for all $\mathbf{k}$ directions.

All the curvatures in the valence dispersion relation now differ from the one of a free electron.

By writing $\tilde{\varepsilon}_{v,\mathbf{k}}$ as $(\hbar^2/2m_0)\gamma_3 e_3$, we find, from the determinant of the above matrix, the equation for e_3 as

$$0 = \left(e_3^2 + 4k^2 e_3/3 + k^4/3\right)^2 = (e_3 + k^2)^2 (e_3 + k^2/3)^2. \quad (58)$$

The resulting energies for the fourfold $|\tilde{j} = 3/2\rangle$ states then are

$$\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + \frac{\hbar^2}{2m_0}(1 - \gamma_3)k^2 \quad \text{twofold}, \quad (59)$$

$$\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + \frac{\hbar^2}{2m_0}\left(1 - \frac{\gamma_3}{3}\right)k^2 \quad \text{twofold}. \quad (60)$$

These valence electrons all have a negative effective mass for a large enough coupling γ_3 between the $|\mu_v\rangle$ and $|\mu_c\rangle$ states.

This shows that in order to have a negative valence mass whatever the direction of the $\mathbf{k}$ momentum, it is necessary to mix the $|\mu_v\rangle$ valence orbital states not only through their coupling to the threefold orbital states $|\mu_c\rangle$ of the conduction band but also between themselves through the spin states via the spin-orbit interaction.

D. Coupling to $|c\rangle \otimes |\pm 1/2\rangle$ and $|\mu_c\rangle \otimes |\pm 1/2\rangle$ states

As in the absence of spin, we write the energy change as $\tilde{\varepsilon}_{v,\mathbf{k}} = (\hbar^2/2m_0)e$ with $e = (e' - \gamma_3 k^2)$. We find that e' follows from the cancellation of the 4×4 determinant given by

$$0 = \begin{vmatrix} |k_1|^2\gamma_- + e' & \cdot & \cdot & \cdot \\ \sqrt{\tfrac{2}{3}}k_0^*k_1\gamma_+ & \frac{|k_1|^2+2|k_0|^2}{3}\gamma_- + e' & \cdot & \cdot \\ \frac{\gamma_1 k_1^2 - \gamma_3 k_1^{*2}}{\sqrt{3}} & 0 & \frac{|k_1|^2+2|k_0|^2}{3}\gamma_- + e' & \cdot \\ 0 & \frac{\gamma_1 k_1^2 - \gamma_3 k_1^{*2}}{\sqrt{3}} & \sqrt{\tfrac{2}{3}}k_0k_1\gamma_+ & |k_1|^2\gamma_- + e' \end{vmatrix}. \tag{61}$$

The resulting equation reads

$$0 = \left[e'^2 + \tfrac{2}{3}(\gamma_1 - \gamma_3)k^2 e' - \tfrac{4}{3}\gamma_1\gamma_3 S_{\mathbf{k}} \right]^2. \tag{62}$$

From its solutions, we find that the fourfold states $|\tilde{j} = 3/2\rangle$ split as two sets of twofold states with energies

$$\varepsilon_{v,\mathbf{k}} = \varepsilon_{v,\mathbf{0}} + \frac{\hbar^2}{2m_0}\left[\left(1 - \frac{\gamma_1 + 2\gamma_3}{3}\right)k^2 \pm \frac{1}{3}\sqrt{(\gamma_1 - \gamma_3)^2 k^4 + 12\gamma_1\gamma_3 S_{\mathbf{k}}} \right]. \tag{63}$$

These results were first derived by Dresselhaus, Kip, and Kittel [33]. The shapes of the energy contours for $\varepsilon_{v,\mathbf{k}}$ are shown in Fig. 4.

It is easy to check that the above equation leads to the energies obtained in Eq. (56) when $\gamma_3 = 0$ and the energies obtained in Eqs. (59) and (60) when $\gamma_1 = 0$. These results also confirm our previous observation that in order to have a negative valence electron mass whatever $\mathbf{k}$, it is necessary to mix the $|\mu_c\rangle$ valence states with the spin states through the spin-orbit interaction, and to couple these valence states to the threefold conduction states $|\mu_c\rangle$ through the γ_3 parameter.

VI. CONCLUSION

It is commonly accepted that a periodic ion potential brings to semiconductor electrons a mass that drastically turns negative close to a band maximum. Yet, the precise physics for why electrons in a threefold valence orbital state all acquire a negative effective mass for whatever their momentum direction, is quite tricky. In this work, we consider various scenarios of couplings between threefold valence orbital states and conduction orbital states with different degeneracies that are close to the band gap, as well as coupling to spin states through spin-orbit interaction. Our analysis, based on a $\mathbf{k} \cdot \mathbf{p}$ approach, proves that while the coupling to conduction levels is the primary reason for driving the mass of the valence electrons toward a negative value, a further symmetry breaking of the valence orbital degeneracy along the spin quantization axis as a result of the spin-orbit interaction, is necessary to ensure that *all* valence electrons acquire a negative mass *whatever* the $\mathbf{k}$ momentum direction. Since spin-orbit interaction is a relativistic effect, the negative effective mass of semiconductor valence electrons thus constitutes an unexpected signature of quantum relativity, unrevealed until now.

[1] L. Esaki and R. Tsu, IBM J. Rev. Dev. **14**, 61 (1970).

[2] F. Capasso, Science **235**, 172 (1987).

[3] F. Bloch, Z. Phys. **52**, 555 (1928).

[4] C. Kittel, *Introduction to Solid State Physics*, 7th ed. (Wiley, New York, 1996).

[5] N. W. Ashcroft and N. D. Mermin, *Solid State Physics* (Holt, Rinehart & Winston, New York, 1976).

[6] L. M. Falicov, *Group Theory and Its Physical Applications* (University of Chicago Press, Chicago, 1966).

[7] G. L. Bir and G. E. Pikus, *Symmetry and Strain-Induced Effects in Semiconductors* (Wiley, New York, 1974).

[8] E. Ivchenko and G. E. Pikus, *Superlattices and Other Heterostructures*, Springer Series in Solid-State Sciences Vol. 110 (Springer, Berlin, 1997).

[9] P. Vogl, H. P. Hjalmarson, and J. D. Dow, J. Phys. Chem. Solids **44**, 365 (1983).

[10] S. Sapra, N. Shanthi, and D. D. Sarma, Phys. Rev. B **66**, 205202 (2002).

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

[12] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[13] N. Argaman and G. Makov, Am. J. Phys. **68**, 69 (2000).

[14] C. Gmachl, F. Capasso, D. L. Sivco, and A. Y. Cho, Rep. Prog. Phys. **64**, 1533 (2001).

[15] L. Sang, M. Liao, and M. Sumiya, Sensors **13**, 10482 (2013).

[16] G. Bastard, *Wave Mechanics Applied to Semiconductor Heterostructures* (Wiley, New York, 1988).

[17] H. Haug and S. W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 5th ed. (World Scientific, Singapore, 2009).

[18] P. Y. Yu and M. Cardona, *Fundamentals of Semiconductors*, 3rd ed. (Springer, Berlin, 2005).

[19] M. Combescot and S.-Y. Shiau, *Excitons and Cooper Pairs* (Oxford University Press, Oxford, 2015).

[20] R. de L. Kronig and W. G. Penney, Proc. R. Soc. Lond. A **130**, 499 (1931).

[21] M. G. Silveirinha and N. Engheta, Phys. Rev. B **86**, 161104(R) (2012).

[22] R. Chang, S. Potnis, R. Ramos, C. Zhuang, M. Hallaji, A. Hayat, F. Duque-Gomez, J. E. Sipe, and A. M. Steinberg, Phys. Rev. Lett. **112**, 170404 (2014).

[23] M. A. Khamehchi, K. Hossain, M. E. Mossman, Y. Zhang, T. Busch, Michael McNeil Forbes, and P. Engels, Phys. Rev. Lett. **118**, 155301 (2017).

[24] L. H. Thomas, Nature (Lond.) **117**, 514 (1926).

[25] M. Combescot, S.-Y. Shiau, and V. Voliotis, Phys. Rev. B **99**, 245202 (2019).

[26] C. Cohen-Tannoudji, B. Diu, and F. Laloe, *Quantum Mechanics* (Wiley, New York, 1977).

[27] L. C. Lew, Yan Voon, and M. Willatzen, *The* **k** · **p** *Method: Electronic Properties of Semiconductors* (Springer, Berlin, 2009).

[28] N. Cavassilas, F. Aniel, K. Boujdaria, and G. Fishman, Phys. Rev. B **64**, 115207 (2001).

[29] G. Fishman, *Semiconducteurs: Les Bases de la Théorie* **k** · **p** (Ecole Polytechnique, Paris, 2010).

[30] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics: Non-Relativistic Theory*, Vol. 3, 3rd ed. (Pergamon Press, London, 1977).

[31] R. W. D. Nickalls, Math. Gazette **77**, 354 (1993).

[32] M. Cardona, N. E. Christensen, and G. Fasol, Phys. Rev. B **38**, 1806 (1988); Y.-S. Kim, M. Marsman, G. Kresse, F. Tran, and P. Blaha, *ibid.* **82**, 205212 (2010).

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