# Boundary conditions for transition-metal dichalcogenide monolayers in the continuum model

Csaba G. Péterfalvi,* Andor Kormányos, and Guido Burkard

*Department of Physics, University of Konstanz, D-78464 Konstanz, Germany*

We derive the boundary conditions for MoS$_2$ and similar transition-metal dichalcogenide honeycomb (2H polytype) monolayers with the same type of $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian within the continuum model around the K points. In an effective two-band description, the electron-hole symmetry breaking quadratic terms are also taken into account. We model the effect of the edges with a linear edge constraint method that has been applied previously to graphene. Focusing mainly on zigzag edges, we find that different reconstruction geometries with different edge atoms can generally be described with one scalar parameter varying between 0 and $2\pi$. We analyze the edge states and their dispersion relation in MoS$_2$ in particular, and we find good agreement with the results of previous density functional theory calculations for various edge types.

PACS number(s): 73.20.At, 73.21.Hb, 73.22.−f, 73.22.Dj

## I. INTRODUCTION

Semiconducting transition metal dichalcogenides (TMDCs) have attracted much attention lately. This applies especially to their monolayers, which have a direct band gap (MoS$_2$, MoSe$_2$, MoTe$_2$, WS$_2$, and WSe$_2$) as opposed to their three-dimensional crystals with an indirect band gap. TMDCs are interesting for a number of applications ranging from flexible and transparent field-effect transistors [1–4] to logical [5] and optoelectronic devices [6–9]. These applications necessitate the knowledge of the electronic structure of finite flakes and nanoribbons, which in turn also relies on the accurate description of the edges [10]. Studying the edge states is also very important when charge carriers need to be injected into a layer of TMDC. The edges of TMDCs have already been studied experimentally [11–17] and theoretically [10,18–26] by several groups. Recently, using STM spectroscopy, experimental evidence of metallic edge states has been found in single-layer MoS$_2$ [17].

*Ab initio* calculations, like density functional theory (DFT) combined with tight-binding (TB) models [26] have allowed many detailed studies of TMDCs. Nevertheless, the description of mesoscopic edge-terminated systems still represents an outstanding challenge for theory. In such cases, a continuum model can offer an efficient way to describe the system. Here, we study the boundary conditions (BCs) and the low-energy in-gap edge states in TMDCs in general, and in MoS$_2$ in particular. Inspired by previous works on the analogous problem in graphene [27,28], we derive the general boundary conditions for the zigzag and armchair types of edges in monolayer TMDCs, see Fig. 1. The BCs are given in the form of two linear matrix equations, one for the wave function and one for its derivative normal to the edge. These equations can be expressed using an initially unknown matrix $M$, in which the number of free parameters is reduced using the symmetries of the system and the edge. Due to the continuum nature of our method, not all the parameters in $M$ can be determined exclusively by means of symmetry considerations. Some atomistic details of the boundary elude this procedure, and therefore, some parameters need to be fitted to results

obtained from other methods with atomistic resolution, such as DFT or density functional based tight-binding (DFTB) methods. We will demonstrate this fitting procedure through the example of zigzag edges in MoS$_2$.

To describe a finite system in MoS$_2$, we start with the bulk Hamiltonian for one spin state,

$$\mathbf{H} = \begin{pmatrix} \alpha k^2 + \varepsilon_{\mathrm{v}} & \gamma k_- & 0 & 0 \\ \gamma k_+ & \beta k^2 + \varepsilon_{\mathrm{c}} & 0 & 0 \\ 0 & 0 & \beta' k^2 + \varepsilon_{\mathrm{c}}' & \gamma k_- \\ 0 & 0 & \gamma k_+ & \alpha' k^2 + \varepsilon_{\mathrm{v}}' \end{pmatrix}. \quad (1)$$

In this form, the effective $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian [30] is cast into the valley isotropic representation [28], where the blocks of $\mathbf{H}$ correspond to the $K$ and $K'$ valleys, respectively. The basis functions, consisting of mainly the $d$ orbitals of the Mo atoms, are the wave functions of the valence band and the conduction band, as in Ref. [30]. The basis of the 4-spinor thus reads $\{\psi_{\mathrm{v}}, \psi_{\mathrm{c}}, -\psi_{\mathrm{c}}', \psi_{\mathrm{v}}'\}$. In the $K^{(\prime)}$ valley, the valence band maximum is $\varepsilon_{\mathrm{v}}^{(\prime)}$, the conduction band minimum is $\varepsilon_{\mathrm{c}}^{(\prime)}$, and we also take into account the quadratic terms with $\alpha^{(\prime)}$ and $\beta^{(\prime)}$, which break the electron-hole symmetry. We will refer to the nonequivalence of the two blocks of $\mathbf{H}$, which is due to the spin-orbit interaction (SOI), as *valley asymmetry*. We assume that $\gamma$ is the same in both valleys, and since we focus on the vicinity of the $K$ points, we neglect the trigonal warping and cubic terms [30]. The distance from the $K$ points is denoted by $k$, and $k_\pm = k_x \pm i k_y$. The eigenvalues of the block of $\mathbf{H}$ for the $K$ valley are

$$\varepsilon_\pm = \tfrac{1}{2}(\varepsilon_{\mathrm{v}} + \varepsilon_{\mathrm{c}} + k^2(\alpha + \beta))$$
$$\pm \sqrt{k^2 \gamma^2 + \tfrac{1}{4}(k^2(\alpha - \beta) - E_{\mathrm{g}})^2}, \quad (2)$$

while $\varepsilon_\pm'(k')$ in the $K'$ valley can be obtained with the respective parameters. The energy gap between the conduction- and valence-band edges is $E_{\mathrm{g}}^{(\prime)} = \varepsilon_{\mathrm{c}}^{(\prime)} - \varepsilon_{\mathrm{v}}^{(\prime)}$. The corresponding eigenstates are

$$\left(1, \frac{\varepsilon + E_{\mathrm{g}}/2 - \alpha k^2}{\gamma k_-}, 0, 0\right)^{\mathrm{T}} \quad \text{in the } K, \text{ and}$$

$$\left(0, 0, 1, \frac{\varepsilon - E_{\mathrm{g}}'/2 - \beta' k'^2}{\gamma k_-'}\right)^{\mathrm{T}} \quad \text{in the } K' \text{ valley}, \quad (3)$$

—————
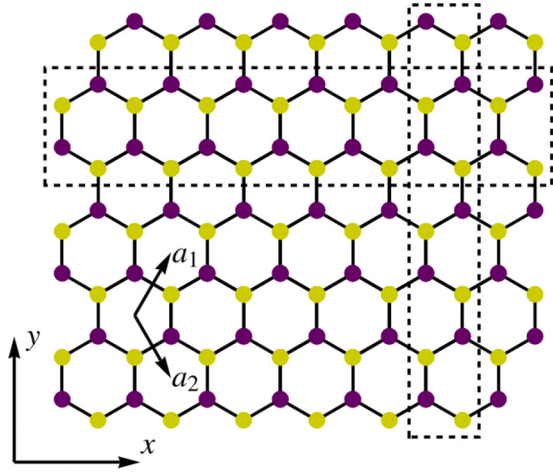*csaba.peterfalvi@uni-konstanz.de

FIG. 1. (Color online) The honeycomb lattice of transition-metal dichalcogenide monolayers. Rectangles with dashed lines show the unit cells of the armchair and zigzag nanoribbons. The width of the nanoribbons can be parametrized by the number $N_d$ of the A-B dimers in the corresponding honeycomb unit cell without regarding any adatoms attached to the possibly reconstructed edge. In this sense, the width of the framed armchair ribbon would be 14, and that of the framed zigzag ribbon would be 7 in this figure. The actual width of an armchair ribbon is $\frac{1}{2}N_d a$, and that of a zigzag ribbon is $\frac{\sqrt{3}}{2}N_d a$, where the lattice constant in MoS$_2$ is $a = |\boldsymbol{a_1}| = |\boldsymbol{a_2}| = 3.1565$ Å [29].

where T denotes transposition. At an edge, where only the parallel component of the wave vector $k_\parallel$ remains a good quantum number, the degeneracy is eightfold: twofold for the two valleys, twofold for the two solutions of the wave number at a given $\varepsilon = \varepsilon_+^{(\prime)}(k_1^{(\prime)}) = \varepsilon_-^{(\prime)}(k_2^{(\prime)})$ energy, and twofold for the $\pm$ sign of the wave vector's normal component $k_\perp$, see Fig. 2. For now, we neglect the electron spin, and we assume that the armchair and zigzag edges mix these eight states to satisfy the boundary conditions.

The Hamiltonian **H** is not specific to MoS$_2$, as it can also describe similar transition-metal dichalcogenide monolayers with honeycomb lattices, like MoSe$_2$, MoTe$_2$, WS$_2$, and WSe$_2$ [29]. The $M$ matrix method that we present in this paper is also as general as the Hamiltonian itself. Nevertheless, in the last section, we will demonstrate the use of our method through the example of MoS$_2$ [21].

## II. THE $M$ MATRIX METHOD

Compared to the calculation for graphene [27,28], an additional boundary condition is needed here for the first derivative of the wave function $\Psi'$, because of the quadratic terms in **H**. As we will see, the second condition can also be expressed by the same, constant $M$ matrix. Inspired by Ref. [27], we introduce hard wall boundary conditions for $\Psi$ and $\Psi'$ with the Hermitian matrix, $M_0$:

$$[\mathbf{H} + M_0 \delta(\boldsymbol{n} \cdot \boldsymbol{r})]\Psi(\boldsymbol{r}) = E\Psi(\boldsymbol{r}), \quad (4)$$

where $\delta(\boldsymbol{n} \cdot \boldsymbol{r})$ is the Dirac delta function modeling the confinement potential at the boundary, which crosses the origin here for simplicity, and the unit vector $\boldsymbol{n}$ is normal to the boundary and pointing outwards. To obtain information about
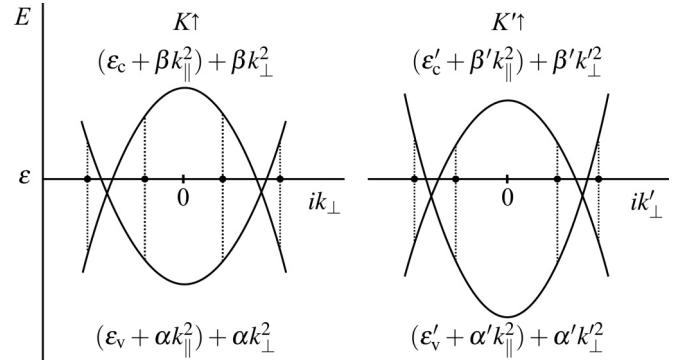


FIG. 2. Schematic representation of the eight degenerate states in $k$ space that are mixed by the armchair and zigzag edges. These in-gap states denoted by dots are characterized by the same spin, the same energy $\varepsilon$ and the same quantum number $k_\parallel$, where $k_\parallel$ is the component of $\boldsymbol{k}$ parallel to the edge. The dispersion curves are labeled with the corresponding functional forms, which are the diagonal elements of the Hamiltonian **H**. For solutions localized at the edge, the normal component $k_\perp$ is imaginary, which results in the opposite curvature of the valence and conduction bands. States from these two bands with the same wave vector compose the corresponding in-gap edge state. From the left to the right, the normal components of the wave vectors are labeled as $-ik_{\perp 1}, -ik_{\perp 2}, ik_{\perp 2}, ik_{\perp 1}, -ik'_{\perp 1}, -ik'_{\perp 2}, ik'_{\perp 2}$, and $ik'_{\perp 1}$.

the unknown matrix $M_0$, we first consider the infinitesimal integral of Eq. (4) through the boundary:

$$\lim_{\epsilon \to 0} \int_{-\epsilon}^{+\epsilon} [\mathbf{H}_\gamma + \mathbf{H}_{\alpha,\beta} + M_0 \delta(\boldsymbol{n} \cdot \boldsymbol{r})]\Psi(\boldsymbol{r})\,\boldsymbol{n} \cdot d\boldsymbol{r} = 0, \quad (5)$$

where the Hamiltonian is split into parts with zeroth-order, linear, and quadratic terms in $k$ as $\mathbf{H} = \mathbf{H}_{E_g} + \mathbf{H}_\gamma + \mathbf{H}_{\alpha,\beta}$ with indices referring to the parameters they depend on. The integrated zeroth order part and the right hand side of (4) vanish for any finite wave function. After carrying out the integral in the given limit, we arrive at the following equation for $\Psi = \Psi(\boldsymbol{r})|_{\boldsymbol{n} \cdot \boldsymbol{r} = 0^-}$ at the boundary:

$$\gamma \widetilde{J}_\gamma \Psi + A\Psi' + \tfrac{1}{2}M_0\Psi = 0, \quad (6)$$

where $\Psi' = \Psi'(\boldsymbol{r})|_{\boldsymbol{n} \cdot \boldsymbol{r} = 0^-}$ denotes the derivative normal to the boundary, the diagonal matrix $A = \text{diag}(\alpha, \beta, \beta', \alpha')$, and $\widetilde{J}_\gamma = \tau_0 \otimes \exp(i\frac{\pi}{2}\boldsymbol{n} \cdot \boldsymbol{\sigma})$. Here and throughout the paper, $\boldsymbol{s}$, $\boldsymbol{\tau}$, and $\boldsymbol{\sigma}$ are denoting vectors of the Pauli matrices acting in the spin space, valley space and the space of the basis functions $\{\psi_v, \psi_c\}$, with the zeroth component being the two-dimensional identity matrix. Note that the operator $\widetilde{J}_\gamma$ is unitary and anti-Hermitian, and so $\widetilde{J}_\gamma^2 = -1$. Moreover, $A$ combined with the derivative operator is anti-Hermitian. Therefore if we calculate the expectation value of these terms by multiplying them by $\Psi^\dagger$ from the left, the anti-Hermitian operators yield purely imaginary numbers, but the Hermitian matrix $M_0$ yields a real number, meaning that they have to be zero separately:

$$\gamma \Psi^\dagger \widetilde{J}_\gamma \Psi + \Psi^\dagger A\Psi' = 0 \quad \text{and} \quad (7)$$

$$\Psi^\dagger M_0 \Psi = 0. \quad (8)$$

A second consideration concerns the total current, which can be determined as

$$J = \frac{i}{\hbar}[\mathbf{H}, \mathbf{n}\cdot\mathbf{r}]$$

$$= \frac{1}{\hbar}\begin{pmatrix} -2i\alpha\mathbf{n}\cdot\mathbf{\nabla} & \gamma e^{-i\chi} & 0 & 0 \\ \gamma e^{i\chi} & -2i\beta\mathbf{n}\cdot\mathbf{\nabla} & 0 & 0 \\ 0 & 0 & -2i\beta'\mathbf{n}\cdot\mathbf{\nabla} & \gamma e^{-i\chi} \\ 0 & 0 & \gamma e^{i\chi} & -2i\alpha'\mathbf{n}\cdot\mathbf{\nabla} \end{pmatrix}$$

$$(9)$$

measured in the direction of $\mathbf{n} = (\cos\chi, \sin\chi)$, which is normal to the boundary, and $\mathbf{\nabla}$ stands for the gradient operator. We split $J$ in diagonal and off-diagonal parts:

$$J = J_\gamma + J_{\alpha,\beta} = -i\frac{\gamma}{\hbar}\tau_0 \otimes \exp\left(i\frac{\pi}{2}\mathbf{n}\cdot\mathbf{\sigma}\right) - i\frac{2}{\hbar}A(\mathbf{n}\cdot\mathbf{\nabla}),$$

$$(10)$$

where $J_\gamma = -i\frac{\gamma}{\hbar}\widetilde{J}_\gamma$. It is a natural requirement as a boundary condition that the expectation value of the normal current at the boundary needs to vanish:

$$\langle J\rangle = -i\frac{\gamma}{\hbar}\langle\widetilde{J}_\gamma\rangle - i\frac{2}{\hbar}\langle A(\mathbf{n}\cdot\mathbf{\nabla})\rangle = 0,$$

$$(11)$$

which can be translated into $\gamma\Psi^\dagger\widetilde{J}_\gamma\Psi + 2\Psi^\dagger A\Psi' = 0$ at the boundary. Comparing this to Eq. (7), we see immediately that

$$\Psi^\dagger\widetilde{J}_\gamma\Psi = 0 \quad \text{and} \quad (12)$$

$$\Psi^\dagger A\Psi' = 0 \;, \quad (13)$$

meaning that the expectation values of the diagonal and off-diagonal currents both vanish: $\langle J_{\alpha,\beta}\rangle = \langle J_\gamma\rangle = 0$.

We assume now that the boundary problem has at least one solution. From Eq. (12), it follows that we could only have four linearly independent solutions in the four-dimensional Hilbert space of $\Psi$, if $\widetilde{J}_\gamma$ was 0, which is not the case. Therefore the number of the linearly independent solutions for $\Psi$ must be less than four. One could argue that $\Psi'$ should be treated as four additional independent variables adding four more dimensions to the Hilbert space, opening up the possibility to solutions that are linearly independent in the subspace of $\Psi'$, but not in the subspace of $\Psi$. This option can be dismissed if we regard Eq. (6), which fully determines $\Psi'$ for a given $\Psi$, because $A$ is invertible. This means that the subspace of the solutions in $\Psi$ will determine the corresponding subspace in $\Psi'$ with the same dimensionality.

We now move to the basis spanned by the solutions to $\Psi$, and we denote the size of this subspace by $d$, which is $1 \leqslant d \leqslant 3$. The first $d$ coordinates of $\Psi$ can thus be chosen arbitrarily and the last $4-d$ have to be zero. For Eq. (12), this means that the top left $d\times d$ block of $\widetilde{J}_\gamma$ must be 0 in this basis. At the same time, the last $4-d$ columns of $M_0$ can be chosen freely. We then define the matrix $M_1$ in the following way:

$$M_1 = \begin{pmatrix} 0 & -Y^\dagger \\ -Y & 0 \end{pmatrix}, \quad (14)$$

where $Y$ is a $(4-d)\times d$-sized matrix containing the matrix elements of the bottom left block of $\widetilde{J}_\gamma$ (row indices from

$d+1$ to 4 and column indices from 1 to $d$). The top left $d\times d$-sized block contains zeros, and the same holds for the bottom right $(4-d)\times(4-d)$-sized block. Since $\widetilde{J}_\gamma$ is unitary and anti-Hermitian, it is easy to prove that it consists of the same blocks as $M_1$, except for the sign change of the bottom left block. It follows from here that

$$\widetilde{J}_\gamma\Psi + M_1\Psi = 0 \quad (15)$$

for every solution of the boundary problem, and $M_1$ is a Hermitian matrix. Eq. (15) multiplied by $\gamma$ and subtracted from Eq. (6) leads us to

$$A\Psi' + M_2\Psi = 0, \quad (16)$$

where we introduced $M_2 = M_0/2 - \gamma M_1$, which must be also Hermitian, given that both $M_0$ and $M_1$ have this property.

Using that $\widetilde{J}_\gamma^2 = -1$, Eq. (15) can be rewritten as $\widetilde{J}_\gamma M_1\Psi = \Psi$, and this boundary condition becomes

$$M\Psi = \Psi \quad (17)$$

with the definition $M \equiv \widetilde{J}_\gamma M_1$. It follows from the similar block structure of $\widetilde{J}_\gamma$ and $M_1$, and from $Y^\dagger Y = 1$, that $M$ is a diagonal matrix, the first $d$ elements of which are 1 along the diagonal, and the last $4-d$ are $-1$. From here, it is straightforward to see that $\{\widetilde{J}_\gamma, M\} = 0$, $M$ is Hermitian and unitary, and $M_1$ must be unitary too. From $\{\widetilde{J}_\gamma, M\} = 0$, the properties of $\widetilde{J}_\gamma$ and the basis invariance of the trace, it also follows that $M$ is traceless, therefore it has two eigenvalues of 1 and two of $-1$, making $d = 2$, and all the blocks in (14) $2\times 2$. Another way to show this can be found in the basis where $\widetilde{J}_\gamma$ is a real antisymmetric matrix, and $M_1$ is consequently a real symmetric matrix. The trace of the product of two such matrices is always 0. At this point, we see that $M$ has two eigenvectors with eigenvalues of 1, and two other eigenvectors with eigenvalues of $-1$. Note that the basis consisting of these eigenvectors is not fully specified as long as these pairs of eigenvectors can be rotated in their respective two-dimensional subspaces. We can specify the basis by fixing $Y$. For example, if $Y = \sigma_0$, then $\widetilde{J}_\gamma = -i\tau_2 \otimes \sigma_0$ and $M_1 = -\tau_1 \otimes \sigma_0$. Independently from the specific choice of $Y$, we can call this basis the eigenbasis of $M$, because $M = \tau_3 \otimes \sigma_0$ is diagonal. The subspace of $M$ with eigenvalues $\pm 1$ is mapped by $\widetilde{J}_\gamma$ into the subspace with eigenvalues $\mp 1$, which is a basis-invariant feature of course.

From Eqs. (13) and (16), we obtain that $\Psi^\dagger M_2\Psi = 0$, which means that the top left $2\times 2$ block of the Hermitian matrix $M_2$ must be 0 in the eigenbasis of $M$. In this basis, one can easily show that $(M+1)M_2(M+1) = 0$, which also holds however in every other basis. From here, we obtain $(M+1)M_2\Psi = 0$, which combines with (16) to the boundary condition for the derivative of the 4-spinor:

$$(M+1)A\Psi' = 0. \quad (18)$$

Note that if $A = 0$, then (18) is trivial, and we formally return to the solution for gapped graphene with the only boundary condition of Eq. (17) to be solved with four linearly independent functions at the edge.

The two boundary conditions for $\Psi$ and $\Psi'$, Eqs. (17) and (18) represent the first main results of this work. Both equations are expressed with the same $M$ matrix, which is

Hermitian, unitary, and satisfies $\{\widetilde{J}_\gamma, M\} = 0$. These are the same conditions that led to Eq. (2.8) of Ref. [28]:

$$M = \sin\lambda\, \tau_0 \otimes (\boldsymbol{\xi} \cdot \boldsymbol{\sigma}) + \cos\lambda\, (\boldsymbol{\nu} \cdot \boldsymbol{\tau}) \otimes (\boldsymbol{\mu} \cdot \boldsymbol{\sigma}), \quad (19)$$

where $\boldsymbol{\xi}$, $\boldsymbol{\nu}$, and $\boldsymbol{\mu}$ are three-dimensional unit vectors, such that $\boldsymbol{\xi}$ and $\boldsymbol{\mu}$ are orthogonal to each other and also to $\boldsymbol{n}_b$, which is in turn normal to the boundary, and pointing outwards. The vector $\boldsymbol{\nu}$ can be parametrized by the polar angle $\varphi$ and the azimuthal angle $\phi$, and the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\xi}$ with a single polar angle $\theta$ and $\theta + \pi/2$. The optional minus sign in the latter, can be merged with the sign of the angle $\lambda$. Assuming that no external magnetic field is present, we now invoke time reversal symmetry. In order to explicitly take into account the spin degree of freedom, we extend the Hilbert space. The Hamiltonian **H** must be extended with the same two blocks along the diagonal, but this time with $\varepsilon_c'$, $\varepsilon_v'$, $\alpha'$, and $\beta'$ parametrizing the $K$ valley, and $\varepsilon_c$, $\varepsilon_v$, $\alpha$, and $\beta$ the $K'$ valley. The $2 \times 2$ blocks of **H** can now be labeled as

$$\begin{bmatrix} K\uparrow & & & \\ & K'\uparrow & & \\ & & K\downarrow & \\ & & & K'\downarrow \end{bmatrix}. \quad (20)$$

This $8 \times 8$ Hamiltonian is commuting with the time-reversal operator

$$T = -s_x \otimes \tau_y \otimes \sigma_y \mathcal{C}, \quad (21)$$

where $\mathcal{C}$ is the operator for complex conjugation. Upon time reversal, the real components of the $\boldsymbol{k}$ vector change sign, thus also changing the valley, and the electron spin flips. The edges on the other hand, do not couple different electron spin states, so in the basis defined at (20), two blocks of $M$ appear along the diagonal possibly mixing states in different valleys but only with the same spin. Allowing that the parameter angles defined at (19) are different for the two spin states, we require that $\mathrm{diag}(M(\lambda,\varphi,\phi,\theta), M(\lambda',\varphi',\phi',\theta'))$ commutes with $T$, where diag denotes a (block-)diagonal matrix with the arguments along the diagonal. This can be solved with $\lambda' = -\lambda$, and $\varphi' = \varphi$, $\phi' = \phi$, and $\theta' = \theta$, and unless $\alpha + \beta = \alpha' + \beta' = 0$, electron-hole symmetry or valley symmetry cannot be used to further reduce the number of the four free parameters of $M$, unlike in Ref. [28].

There is another symmetry of **H** that swaps the valleys and flips the spin, but does not change the sign of the momentum relative to the $K$ points. This is represented by the operator

$$Q = -s_x \otimes \tau_y \otimes \sigma_y \times$$
$$\times \mathrm{diag}(-k_+, k_-, -k_+, k_-, -k_+, k_-, -k_+, k_-)/k. \quad (22)$$

This symmetry corresponds to a $\pi$ rotation with respect to the $y$ axis at an armchair edge (see Fig. 1) and results in spin degeneracy there, in contrast to zigzag edges, where $Q$ represents no symmetry and we observe spin splitting. Unfortunately, this symmetry operation cannot be used to gain further information on $M$ by requiring $[M, Q] = 0$, because of the momentum-operators in $Q$ acting on edge states, which are superpositions of the bulk solutions with different $k_\perp$.

Apart from the mass term, and the valley-dependent quadratic terms $\alpha$ and $\beta$, there are more differences between
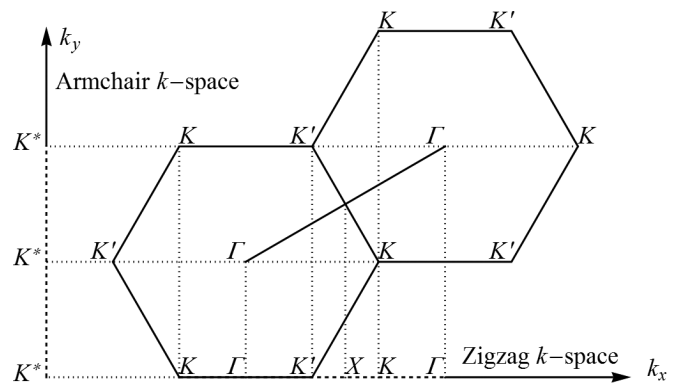


FIG. 3. Projections to the reciprocal spaces of zigzag and armchair nanoribbons. In armchair ribbons the valleys are fully mixed and all the $K$ points are equivalent, while in zigzag ribbons, the $K$ and $K'$ points remain independent, the valley mixing is zero.

graphene and the transition-metal dichalcogenide monolayers which we focus on here. While in graphene the basis functions are the orbitals in the two sublattices, the basis we are using here consists of the extended wave functions of the valence and conduction bands. This makes it less straightforward to see how the angles $\lambda$, $\varphi$, $\phi$, and $\theta$ relate to the geometry of a given edge. We will discuss here the two special cases of zigzag and armchair edges (Figs. 1 and 3), with a strong emphasis on the zigzag edge for being energetically favorable[11,14] and more relevant from the applications' point of view. From Fig. 3, it is clear that an armchair edge mixes the $K$ and $K'$ valleys, while these remain nonequivalent in the 1D Brillouin zone of a zigzag edge. This suggests that the polar angle $\varphi$ responsible for the valley mixing must be 0 or $\pi$ for zigzag, and from symmetry considerations, we also deduce that $\varphi = \pi/2$ causing the maximal mixing of the valleys, must belong to the armchair orientation, just like in graphene.[1] This also means that $\phi$ drops out of Eqs. (17) and (18) for zigzag, leaving only $\lambda$ and $\theta$ to characterize the various zigzag edges. As it turns out, $\phi$ matters for armchair edges, but only if edge states from different edges can interfere. For a single armchair edge, it is also $\lambda$ and $\theta$ that determine the edge states and their dispersion relation. Without studying the atomistic details of an edge, based on symmetry considerations alone the values of $\lambda$ and $\theta$ cannot be inferred within the continuum model. However, fitting to *ab initio* band structures is a straightforward way to relate the different edge geometries to different values of these parameters.

With $\varphi = 0$ for zigzag, $M$ can be simplified to

$$M_{\mathrm{zz}} = \begin{pmatrix} \cos(\vartheta^+) & \sin(\vartheta^+) & 0 & 0 \\ \sin(\vartheta^+) & -\cos(\vartheta^+) & 0 & 0 \\ 0 & 0 & -\cos(\vartheta^-) & -\sin(\vartheta^-) \\ 0 & 0 & -\sin(\vartheta^-) & \cos(\vartheta^-) \end{pmatrix},$$
$$(23)$$

where $\vartheta^\pm = \theta \pm \lambda$. The $\varphi = \pi$ case is already included if we shift $\vartheta$ by $\pi$. Since the zigzag edge does not mix the valleys, $\vartheta^+$

---

[1]Note that a finite $\lambda$ parameter in $M$ does not allow complete mixing of the valleys at an armchair edge. Only at vanishing valley asymmetry ($\lambda = 0$) can the valley mixing be complete with $\varphi_{\mathrm{armchair}} = \pi/2$.

and $\vartheta^-$ can be fitted separately. As we know from above, the same $M_{zz}$ also works for the opposite spin, but with $-\lambda$. Note that $\lambda$ is expected to be small, for it is finite only because of the valley asymmetry. If all the parameters in **H** were the same in both valleys, then **H** would commute with $-\tau_y \otimes \sigma_y \mathcal{C}$ (the $4 \times 4$ off-diagonal block of $T$), which in turn has to commute also with $M_{zz}$, resulting in $\lambda = 0$ as in graphene (see Eq. (2.10) of Ref. [28]).

## III. SINGLE EDGES AND NANORIBBONS

### A. General discussion

Here we apply the $M$ matrix method to various kinds of edges. Since we are interested in the in-gap edge states, the normal component of the wave vector $k_\perp$ will always be imaginary. Out of the eight edge solutions of the eigenvalue problem, we can only work with the four that are decaying away from the edge of a half-infinite system and we discard the diverging solutions. This leads to four unknown coefficients and eight equations defined by (17) and (18). However, the rank of each of the $4 \times 4$ matrices defining this linear system of equations is 2, yielding in total four equations to solve for the four coefficients. The determinant of this $4 \times 4$ matrix is zero for the nontrivial solutions; this delivers the dispersion relation of the edge states, and the null space of the matrix provides us with the coefficients.

In a nanoribbon, the dispersion relation can be calculated in the same way, but with eight states resulting in a determinant of an $8 \times 8$ matrix. Since the edge states are concentrated in the vicinity of the edges, in a few nanometer wide nanoribbon, they become independent. Further increments of the width do not affect the bands of the edge states any more, and the bands of the single edges independently add up to form the band structure of the nanoribbon. It is important to note that two edges, characterized by $\lambda_{1,2}$ and $\theta_{1,2}$ (or by $\vartheta^\pm_{1,2}$ in zigzag), can be combined to form a nanoribbon, if we attribute $-\lambda$ and $-\theta$ ($-\vartheta^\pm$) to the edge bordering the ribbon on the opposite side relative to the one that was used to obtain these angles. For example, in the case of an armchair ribbon with equivalent edges, $\lambda, \theta, -\lambda$, and $-\theta$ has to enter the above mentioned $8 \times 8$ matrix to give rise to the correct dispersion relation originating from the combined effect of the two identical edges.

This method can be generalized to flakes with more complex shapes and mixed types of edges, like equilateral triangles, rectangles, hexagons, etc. Each edge has its respective $\theta$ and $\lambda$, and by studying the symmetries of the system of equations posed by the boundary conditions, it must be determined how to couple these to form the boundaries of the same flake. In the case of armchair edges, the relative distance and orientation of the edges also matter if the corresponding states are overlapping, meaning that the knowledge of the angles $\phi$ is also required for the accurate description. As we see at the end of this section through a simple example in graphene, this can be determined from an appropriate TB model.

### B. Application to MoS$_2$

To demonstrate the use of our method, we calculate the edge states in MoS$_2$. To this end, we need the material parameters appearing in the Hamiltonian. Because of the mirror symmetry

of the lattice to the $x$-$y$ plane, **H** is block-diagonal in the $z$ component of the real spin and takes the form of (20) with the valley-dependent sets of parameters $\{\alpha, \beta, \gamma, \varepsilon_v, \varepsilon_c\}$ and $\{\alpha', \beta', \gamma, \varepsilon_v', \varepsilon_c'\}$. We assume that no external field or special edge geometry breaks this symmetry of the bulk, and therefore the full Hamiltonian including $M$ also preserves this symmetry. Our continuum model, which describes the edge states as linear combinations of the bulk states using only the basis of the valence and the conduction bands, will not be able to describe all the in-gap states. For example, states that are almost entirely localized on the terminating row of sulfur atoms (see, e.g., Fig. 5 of Ref. [10]) cannot be expanded in the basis we use. The bulk wave functions of this basis require that the parameters appearing in **H** to be extracted from a bulk DFT study. The values of these bulk parameters are taken from Ref. [29], where $\gamma$ is reported to be spin-independent: $\alpha = -2.16$ eV Å$^2$, $\alpha' = -2.62$ eV Å$^2$, $\beta = 4.09$ eV Å$^2$, $\beta' = 4.35$ eV Å$^2$, $\gamma = 2.76$ eV Å, $E_g = \varepsilon_c - \varepsilon_v = 1.821$ eV, $E_g' = \varepsilon_c' - \varepsilon_v' = 1.67$ eV, $\varepsilon_c - \varepsilon_c' = 3$ meV, and $\varepsilon_v' - \varepsilon_v = 148$ meV. Note that the contribution of the free Hamiltonian $\hbar^2/(2m_e) = 3.81$ eV Å$^2$ is included in the $\alpha, \alpha', \beta$ and $\beta'$ parameters, where $m_e$ is the free electron mass.

In what follows, we fit and compare our results to the DFTB study of Erdogan *et al.* for nanoribbons [21]. Although SOI was not considered, the dispersion curves of sufficiently wide (32.8 Å) nanoribbons with several different kinds of edges were calculated, which offers the opportunity to fit $\vartheta$ edge-by-edge assuming that the two sides are independent. Such DFT calculations are typically done with nanoribbons and not with single edges because of the periodic boundary condition. However, we can consider the edge bands independent in nanoribbons over a few nm wide, because exponentially decaying edge states on either sides become negligible at the opposite edge and they do not interfere. Here, we first adopt the assumption that the DFTB band structure is actually spin-resolved and belongs to the spin-up states in the $K'$ valley, and then we repeat the calculation with the assumption that the same bands belong to spin-down states in the same valley. With this fitting to spin-independent bands, we aim to demonstrate the use of our method, rather than providing quantitatively accurate results. Accuracy could be obtained from fitting to nonmagnetic, spin-resolved bands, but to our best knowledge, no such study has been published to this day.

The fitted spin-up bands are plotted in Fig. 4, together with the original DFTB bands and nanoribbon structures from Ref. [21].[2] To fix the energy offset of the edge bands, we match the center of the bulk gap to the center of the nanoribbon energy gap by a rigid energy shift for the whole band structure. Then we adjust the energy of our edge band at the $K'$ point to that of the DFT-calculated band by tuning the only free parameter $\vartheta$ for each single edge band separately. The fitted $\vartheta$ values are unique apart from the $2\pi$ periodicity, and these

---

[2]The spin-down bands can be fitted equally well: $\vartheta_1^+ = 1.762$, $\vartheta_2^+ = \pi + 1.817$ for the antisymmetric edges, $\vartheta_1^+ = 1.737$, $\vartheta_2^+ = \pi + 1.886$ for the S$_2$-stripe edges, $\vartheta_1^+ = 1.758$, $\vartheta_2^+ = \pi + 1.860$ for the S-dimer edges, and $\vartheta_1^+ = 1.747$ for the S-half edges.

(a) Antisymmetric edges

$\vartheta_1^- = 1.795, \vartheta_2^- = \pi + 1.762.$

(b) $S_2$-stripe edges

$\vartheta_1^- = 1.768, \vartheta_2^- = \pi + 1.840.$

(c) S-dimer edges

$\vartheta_1^- = 1.791, \vartheta_2^- = \pi + 1.810.$
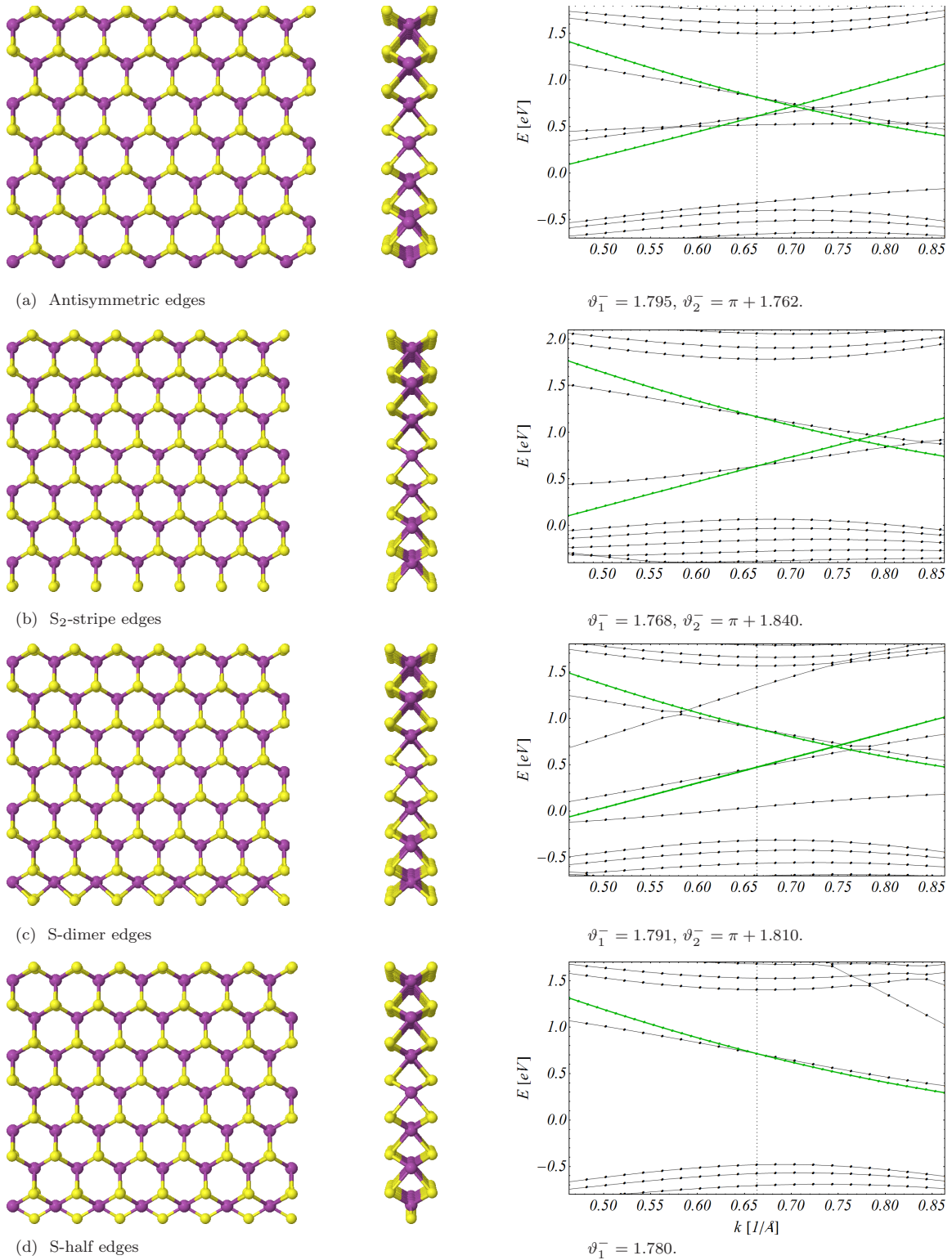
(d) S-half edges

$\vartheta_1^- = 1.780.$

FIG. 4. (Color online) On the left, various 12 dimers (32.8 Å) wide zigzag nanoribbons from Ref. [21], with the corresponding band structures on the right. The edge state bands from the continuum model for spin-up states are denoted by thick, green (gray) lines, and those from the DFTB calculation from Ref. [21] are denoted by thin, black lines. The dotted lines show the $K'$ point, and the wave number is measured from the $\Gamma$ point in these figures. The thick green (gray) curves correspond to the states at the two edges fitted with $\vartheta_1^-$ and $\vartheta_2^-$, where $\varphi_1 = \varphi_2 = 0$. The dots represent the data points the curves were fitted to.
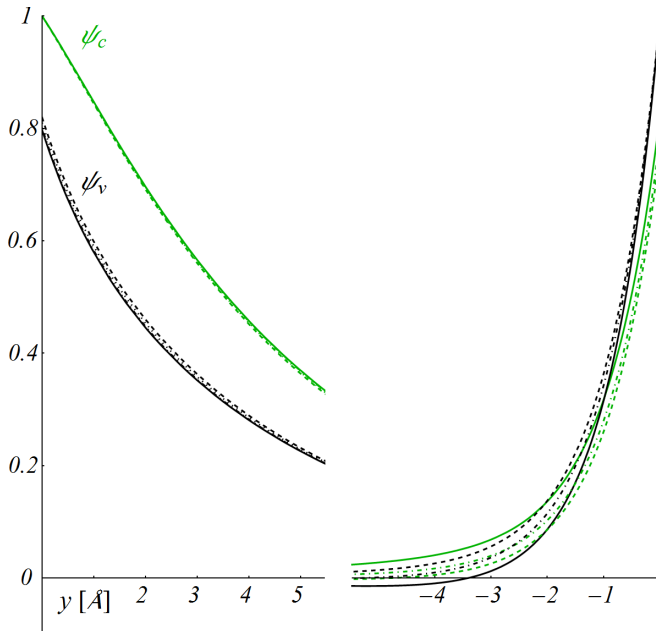
FIG. 5. (Color online) Coefficients of $\psi_v$ and $\psi_c$, i.e., the wave function at $k_x = 0$ in the $K\downarrow$ and $K'\uparrow$ valleys at the zigzag edges, within the width of two dimers $\sqrt{3}a$. Solid lines describe the antisymmetric edges [Fig. 4(a)], dashed lines the S$_2$-stripe edges [Fig. 4(b)] and dashed-dotted lines the S-dimer edges [Fig. 4(c)]. Dotted lines appearing only on the left describe the S-half ribbon, which does not support edge states on the right [Fig. 4(d)]. The wave function is calculated for independent edges with the $\vartheta$ values from Fig. 4. Note that all the components of $\Psi$ are real, and the maximum of $\psi_c$ ($\psi_v$) on the left (right) is always normalized to 1. For the corresponding energies, see the band structures in Fig. 4 at the $K'$ point.

were used for the plots of Fig. 5 as well.[3] It is apparent from Fig. 4 that all four band structures share an in-gap state with a negative slope. This suggests that this band belongs to the same edge that all the four nanoribbons have in common, namely the purely sulfur terminated zigzag edge (the top edges of the ribbons in Fig. 4). To this band, we can fit the angle $\vartheta_1^-$ with almost the same value in every cases. The edges on the other side are described by $\vartheta_2^-$, and in this case, we can only fit to one band without ambiguity even though in two cases [Figs. 4(a) and 4(c)], other in-gap bands appear in DFT, which are missing from the continuum model. Note that the edge with one S atom per Mo supports no edge state at all, see Fig. 4(d).

The deviations from the DFTB band structure can be explained with the following factors: (i) in the DFTB calculation, the spin-orbit coupling was neglected; (ii) the input parameters for our calculation ($\alpha$, $\beta$, $\gamma$, and $E_g$) were extracted from a bulk DFT calculation [29] with different exchange-correlation potential, etc.; finally, (iii) we used an effective, two-band

Hamiltonian neglecting the trigonal warping and the cubic terms [30]. Regarding all these differences, the agreement between the band structures is reasonably good. The most significant deviation manifests itself in the larger steepness of the bands that have lower energy at the $K'$ point. Note, however, that this band is moderately steeper also in Refs. [31] relative to our reference [21], and the derivative of our bands agrees especially well with the tight binding results in Fig. 9 from Ref. [26], even if we need to use slightly different $\vartheta$ angles to fit our bands to theirs.

The corresponding components of the wave function can be seen in Fig. 5, where the real-valued coefficients of $\psi_v$ and $\psi_c$ are plotted along the dimension normal to the various kinds of zigzag edges in Fig. 4. Here, the left edge always has the same, purely sulfur terminated structure, whereas the right edge has different reconstruction in the four cases. Since the S-half ribbon has no edge states on the right [see Fig. 4(d)], the dotted lines only appear on the left. The differences in the wave function on the right manifest in slightly different $\vartheta$ values for the otherwise identical left edges (see Fig. 4), and here, this prevents the $\psi_v$ curves and also the $\psi_c$ curves from exact overlap. This is due to the small, yet finite contribution of the electrons from the opposite edge 32.8 Å away. On the left, the overall predominance of the conduction band can be seen, but on the right, the contribution of the two bands is roughly the same. Accordingly, the left edge state will always have the higher energy, see Fig. 4. Using Eq. (13), we can verify a direct relation between these curves, namely that $\psi_v \alpha' \psi_v' + \psi_c \beta' \psi_c' = 0$ holds at both edges. The ratio of the $\alpha'$ and $\beta'$ parameters explains the relative height and steepness of the $\psi_v$ and $\psi_c$ curves at the two ends.

Note that in graphene, there are basically only two kinds of zigzag edges, and even they are related by symmetry. One edge is always terminated by atoms of sublattice A, and the other one by atoms of sublattice B. In our case, these atoms are different elements, we lose inversion symmetry, and also the fact that molybdenum atoms form bonds with six out-of-plane sulfur atoms increases the number of possible edge terminations. In the case of the antisymmetric edges [Fig. 4(a)], the purely S and purely Mo terminated edges correspond to the solid curves in Fig. 5 on the left and on the right respectively. This S–Mo antisymmetry constitutes the major difference between the wave functions on the two sides. The dangling bonds of the Mo atoms offer further possibility for various reconstructions of the edge with additional S atoms thus modifying the picture to some lesser extent on the right, but it does not change qualitatively. As we see, the obtained values for the $\vartheta^-$ angles on the right ($\vartheta_2^-$) and on the left ($\vartheta_1^-$) are also forming two distinct groups, emphasizing that the sublattice asymmetry is more important than the effect of the various edge reconstructions on the Mo-terminated side. This is also underlined regarding the energy of these states: the maximal energy difference at $k_x = 0$ between the two sides is 531 meV and the slopes of the bands differ in sign. The maximal energy difference between the various reconstructions on the right relative to the middle of the gap is 351 meV, and the slopes of the bands are almost the same. And finally, the maximal difference on the left due to edge states on the right reaching the left edge is only 26 meV. Remember that in the context of the continuum model, the different values of $\vartheta_1^-$ can be

_____

[3]At the $K'$ point, $+0.01$ change in $\vartheta_1^\pm$ means approximately 10 meV increase in energy and the same change in $\vartheta_2^\pm$ results in roughly 45 meV decrease in energy. Nevertheless, all the angles $\vartheta_{1,2}^\pm$ were fitted with an accuracy of 0.001.

understood to be due to the small, yet finite contribution of the states from the opposite edges at a distance of 32.8 Å. Otherwise, one kind of edge is characterized by one pair of angles $\vartheta^{\pm}$, which in the case of the purely S terminated left edge would be around 1.78. The precise determination of $\vartheta_1^{\pm}$ would necessitate fitting to *ab initio* calculations with somewhat wider ribbons, or the simultaneous fitting of $\vartheta_1^{\pm}$ and $\vartheta_2^{\pm}$ in a ribbon explicitly considering the interference of the edge states. This accuracy, however, would be justified only if we were fitting to a spin-resolved band structure accounting for the spin-orbit interaction as well.

There is one more detail to zigzag MoS$_2$ nanoribbons, which may be potentially important. Namely, experiments have shown evidence of weak ferromagnetism in MoS$_2$ nanostructures [12,13,15]. It was suggested that this can be explained by recent calculations on zigzag MoS$_2$ monolayer nanoribbons, which indicate that the ground state of these nanoribbons can be ferromagnetic, depending on the tensile strain [22,23], and also on the edge passivation (e.g., H saturation) [19,20,24]. The consequent energy splitting of the edge states, the *magnetization splitting,* depends on the average magnetic field felt by the electron in a given state. This is mainly determined by relation of the decay length of the state to the magnetic profile of the edge. Around the $K$ points, it is difficult to separate the effect of the SOI and that of the edge magnetization (EM). However, by studying the edge bands in the relevant literature on DFT band structures [18,20,24], we can see that the combined effect of the two is more or less a rigid shift along the energy axis for the two edge bands that can be described by our model. This shift depends on the band, but there seems to be negligible dependence on $k$ within one band, and so the change in the associated group velocities at the $K$ points are also negligibly small. Therefore, we assume that around the $K$ points, where the effective Hamiltonian **H** is accurate, the effect of EM on the one-dimensional edge states can also be accounted for by a simple energy shift of the bands. To our best knowledge, the band structure of the nonmagnetic state has not been published yet for such edges, although calculations were made, e.g., for the total energy difference of the magnetic and nonmagnetic states (see Fig. 6 of Ref. [24]). On average, this reaches only tens of meV, but some bands are affected by the EM more than others. This input would be needed to perform a proper fitting for $\vartheta_{1,2}^{\pm}$, after which, a simple energy-shift of the bands could restore the DFT band structures in a good approximation. Figure 6 shows the bands at the purely sulfur terminated zigzag edge in the nonmagnetic state, with the assumption that $\vartheta^{-} = \vartheta^{+} = 1.795$. The splitting at $k = 0$ is 38 meV. The order of magnitude of this energy gap can be estimated from the spin splitting of the states $\psi_{\mathrm{v}}$ and $\psi_{\mathrm{c}}$, and the ratio of their contribution to the given edge state (left part of Fig. 5). This splitting is significantly smaller than the values over 300 meV reported in the literature [18,24]. This could suggest that the EM is more important than the SOI for these states, provided that the above equality of the $\vartheta$ angles holds. We find, however, that the gap changes by 1 meV for a change of 0.001 in the value of $\lambda$ around $\theta \approx 1.795$, and, for example, $\lambda = 0.2$ would already result in a splitting of 343 meV due only to SOI in good agreement with the DFT calculations. Nevertheless, to know exactly the relative strength of these two effects, together with the precise
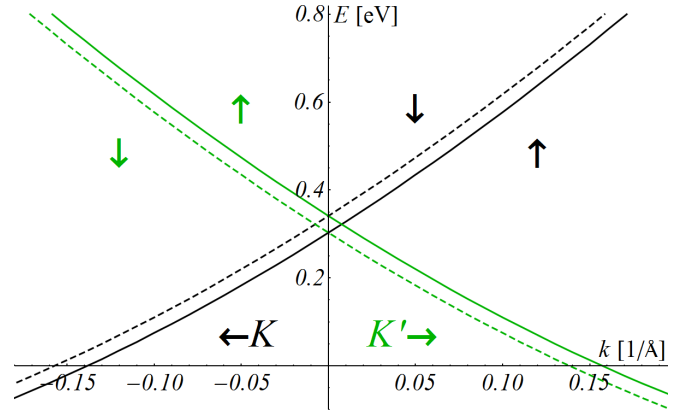


FIG. 6. (Color online) Band structure of the purely sulfur terminated zigzag edge in MoS$_2$ (top edges in Fig. 4), where electrons are propagating in opposite directions in the $K$ and $K'$ valleys. The bands are calculated for the nonmagnetic state, that is the spin splitting is only due to the bulk SOI, where the assumption $\vartheta^{-} = \vartheta^{+} = 1.795$ is used. Down arrows and dashed lines denote down-spin states, up-arrows and solid lines up-spin states. The left-propagating states in the $K$ valley are in black, and the right-propagating states in the $K'$ valley are in green (gray).

values of $\vartheta^{\pm}$, a fit to a nonmagnetic band structure would be needed.

Note that this large spin splitting combined with the fact that $K$ and $K'$ electrons are propagating in opposite directions could be used to generate spin- and valley-polarized edge currents at room temperature and above [32]. The EM can further enhance the spin splitting for one of the valleys. In a first approximation, the magnetization is expected to shift the $K\uparrow$ band up (or down) by the same amount as the $K'\downarrow$ band shifts down (or up). A similar statement holds for the other two bands as well, but the magnetization splitting of these will be slightly different due to the valley asymmetry.

In Ref. [31], the edge states of MoS$_2$ nanoribbons are studied in relation to the width-dependent electrostatic polarization, with the help of first-principles DFT calculations. According to this study, the wider the ribbon is, the more these polarization charges are screened, but in the case of incomplete screening in narrower ribbons, a transverse electric field is induced by the polarization charges. As a result, a potential energy difference appears between the edges, and the edge bands are pushed upward at one side of the ribbon, and pushed downward at the other side. This phenomenon is independent of the spin state, but otherwise, in our model, it can be accounted for in a similar fashion as EM, with a rigid shift of the bands.

Recently, several works [25,32] used a "gapped-graphene" type $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian to model the bulk electronic structure of monolayer MoS$_2$ in the vicinity of the $K$ ($K'$) points. While such analogy to graphene can be useful to understand bulk properties, the analogy breaks down at the edges. To highlight the difference to gapped graphene, it is instructive to consider a zigzag edge in the limit of vanishing or negligible $\alpha$ and $\beta$ parameters and equal energy gaps for the two valleys and spins. Here we assume that the restored electron-hole

symmetry of the bulk is not broken by the edge. From this, it follows that $\theta = \vartheta = \pi/2$, and we obtain edge states with Dirac-like, linear dispersion $\varepsilon = \pm\gamma k_\parallel$ for bands belonging to the $K$ and $K'$ valleys respectively. The normal component $k_\perp = iE_g/(2\gamma)$ depends only on material parameters. These results are independent from the edge geometry, that is all zigzag edges are equivalent in the model as long as $\alpha = \beta = 0$. Differences between the edges can be taken into account only if finite values for $\alpha$ and $\beta$ are allowed. However, why can we not find these in-gap states in gapped graphene [33]? To answer this, we recall that the basis of graphene is a sublattice basis, in which the edge inherently breaks the electron-hole symmetry, since all the terminating sites belong to one particular sublattice, and all have the same on-site energy, which is also the energy of the flat bands (see Fig. 1 of Ref. [33]). In MoS$_2$, the basis functions are extended wave functions of the valence and conduction bands, and the terminating sites do not belong to either of these. Therefore it is conceivable that none of the two basis functions is given preference by the edge, thus not causing any symmetry breaking. Based on these findings, we expect the presence of edge states showing the preserved electron-hole symmetry of the gapped system.

In the case of an armchair ribbon, provided there is no asymmetric edge reconstruction, we have the same kind of edges on both sides: $\lambda_1 = \lambda_2$ and $\theta_1 = \theta_2$. We have two different linear combinations of states localized to the opposite edges, and these combinations are split in energy, but they are degenerate in spin. In narrow ribbons, where the edge states do not vanish completely at the opposite edges, the difference $\Delta\phi = \phi_2 - \phi_1$ affects the band structure, but the angles $\phi_{1,2}$ individually do not. In wide ribbons, where the opposite edges do not interfere, the angles $\phi$ have no effect on the dispersion, the splitting vanishes, and we end up with fourfold degenerate edge states: twice for the spin and twice for the two linear combinations, where the latter ones can eventually be split into states belonging to their respective edges only. Here, the angles $\phi$ turn out to be arbitrary phase differences between the valleys, as in $\{\psi_v, \psi_c, -\psi'_c \exp(i\phi), \psi'_v \exp(i\phi)\}$, and this phase difference does not change upon time reversal.

It is apparent from Fig. 3(b) of Ref. [21] that the band structure of the edge states does not change qualitatively with the width in the given range, and the valley-mixing effect of $\Delta\phi$ can be neglected. Because of the size of our basis, we cannot describe all the bands, but only one of the degenerate flat bands around the $K^*$ point (Fig. 3), denoted by $\Gamma$ in Fig. 3(b) of Ref. [21]. For small values of $\lambda$, almost in the full range of $\theta$, we see a flat band indeed that could be matched with any of the two between 0 and 1 eV. Without further input, however, e.g., regarding the composition of these states, we cannot tell which one to fit to.

Note however that in graphene nanoribbons, we can also describe extended edge-to-edge bulk states with the $M$ matrix method, where $\Delta\phi$ also has an effect on the dispersion of these states. This can be used to determine its value eg. by means of a TB model, which enables the analytical determination of band structure, like it has been done in Ref. [34]. For the spectrum at the $K$ point, these authors found $\varepsilon_r = t|1 + 2\cos\frac{r\pi}{N_d+1}|$, where

$t$ is the transfer integral between nearest-neighbor carbon sites, and $r \in \{1, \ldots, N_d\}$ is an integer indexing the bands. For the lowest, positive energy $\varepsilon_{r_0}$, $|\frac{r\pi}{N_d+1} - \frac{2\pi}{3}|$ is minimal with $r = r_0$. It can also be shown easily that $\Delta\phi = \pm 2W\varepsilon_{r_0}/\gamma$ holds, where $W = \frac{1}{2}N_d a$, and $\gamma = \frac{\sqrt{3}}{2}ta$. Combining all this, we obtain that $\Delta\phi = \pm\frac{2}{3}\pi$ if $N_d$ modulo 3 is 0 or 1, and $\Delta\phi = 0$ if $N_d$ modulo 3 equals 2. The sign of $\Delta\phi$ is irrelevant for armchair nanoribbons because of their symmetric edges, this is why we cannot find a definite sign from the energy spacing. Nevertheless, it is sensible to think of this effect as if every additional dimer added $2\pi/3$ to $\phi_2$ counting from the narrowest ribbon with $N_d = 3$, $\phi_1 = 0$, and $\phi_2 = 2\pi/3$. As for the in-gap states in TMDC armchair nanoribbons, we speculate that $\Delta\phi$ may be determined by the width of the ribbon as well. The main arguments are that the individual angles $\phi_1$ and $\phi_2$ do not affect the in-gap bands, but only their difference, and the effect of $\Delta\phi$ vanishes with the vanishing overlap of opposite lying edge states in an infinitely wide ribbon. These findings suggest that the relative geometry of the edges matter, as in the case of graphene, where $\Delta\phi$ is directly proportional to the ribbon width $W$. The determination of the exact relation between $\Delta\phi$ and $W$ in TMDCs is an open problem, which could probably be solved by a TB calculation [26,35–39].

## IV. CONCLUSIONS

We derived the general boundary conditions for zigzag and armchair edges of transition-metal dichalcogenide monolayers within a continuum model. We used an effective, two-band, $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian around the $K$ point, where the electron-hole symmetry breaking quadratic terms were taken into account. We modeled the effect of the edges with the $M$ matrix method, and after reducing the number of free parameters by symmetry considerations, we found that different edge geometries can be described with two or three scalar parameters varying between 0 and $2\pi$. We discussed the case of single edges and interacting edges of nanoribbons, and we fitted the free parameters to *ab initio* band structures. Focusing mainly on zigzag edges, we analyzed the edge states and their dispersion relation in MoS$_2$ in particular, and we demonstrated a good agreement with DFT based tight-binding calculations.

Note that our method is not restricted to in-gap states, but it can be applied to low-energy bulk states as well, as long as they are close to the conduction band or the valence band edge. Therefore our method can be used to study bulk band states in edge-terminated nanostructures, such as quantum dots and nanoribbons [40]. This also opens up the possibility to fit the $\vartheta_{1,2}^\pm$ angles for the two edges using these extended bulk states. The advantage of this would be that the bulk bands are not much affected by the EM in wider ribbons, but the disadvantage is that the bulk states connect the edges and a simultaneous fitting for $\vartheta_1^\pm$ and $\vartheta_2^\pm$ is needed. The ribbons cannot be too wide either so that the quantized bulk bands are well-separated in energy. Alternatively, as demonstrated in our work, fitting to noninteracting in-gap edge states is the simplest way to obtain $\vartheta_{1,2}^\pm$, but for accuracy, nonmagnetic band structures need to be calculated first.

[1] G.-H. Lee, Y.-J. Yu, X. Cui, N. Petrone, C.-H. Lee, M. S. Choi, D.-Y. Lee, C. Lee, W. J. Yoo, K. Watanabe, T. Taniguchi, C. Nuckolls, P. Kim, and J. Hone, ACS Nano **7**, 7931 (2013).

[2] H.-Y. Chang, S. Yang, J. Lee, L. Tao, W.-S. Hwang, D. Jena, N. Lu, and D. Akinwande, ACS Nano **7**, 5446 (2013).

[3] S. Das, R. Gulotty, A. V. Sumant, and A. Roelofs, Nano Lett. **14**, 2861 (2014).

[4] J. Kang, W. Liu, and K. Banerjee, Appl. Phys. Lett. **104**, 093106 (2014).

[5] B. Radisavljevic, M. B. Whitwick, and A. Kis, ACS Nano **5**, 9934 (2011).

[6] Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, and M. S. Strano, Nat. Nano **7**, 699 (2012).

[7] L. Britnell, R. M. Ribeiro, A. Eckmann, R. Jalil, B. D. Belle, A. Mishchenko, Y.-J. Kim, R. V. Gorbachev, T. Georgiou, S. V. Morozov, A. N. Grigorenko, A. K. Geim, C. Casiraghi, A. H. C. Neto, and K. S. Novoselov, Science **340**, 1311 (2013).

[8] B. W. H. Baugher, H. O. H. Churchill, Y. Yang, and P. Jarillo-Herrero, Nat. Nano **9**, 262 (2014).

[9] A. Pospischil, M. M. Furchi, and T. Mueller, Nat. Nano **9**, 257 (2014).

[10] M. V. Bollinger, K. W. Jacobsen, and J. K. Nørskov, Phys. Rev. B **67**, 085410 (2003).

[11] J. V. Lauritsen, J. Kibsgaard, S. Helveg, H. Topsøe, B. S. Clausen, E. Lægsgaard, and F. Besenbacher, Nat. Nanotechnol. **2**, 53 (2007).

[12] J. Zhang, J. M. Soon, K. P. Loh, J. Yin, J. Ding, M. B. Sullivian, and P. Wu, Nano Lett. **7**, 2370 (2007).

[13] S. Tongay, S. S. Varnoosfaderani, B. R. Appleton, J. Wu, and A. F. Hebard, Appl. Phys. Lett. **101**, 123105 (2012).

[14] A. M. van der Zande, P. Y. Huang, D. A. Chenet, T. C. Berkelbach, Y. You, G.-H. Lee, T. F. Heinz, D. R. Reichman, D. A. Muller, and J. C. Hone, Nat. Mater. **12**, 554 (2013).

[15] D. Gao, M. Si, J. Li, J. Zhang, Z. Zhang, Z. Yang, and D. Xue, Nanoscale Res. Lett. **8**, 129 (2013).

[16] X. Yin, Z. Ye, D. A. Chenet, Y. Ye, K. O'Brien, J. C. Hone, and X. Zhang, Science **344**, 488 (2014).

[17] C. Zhang, A. Johnson, C.-L. Hsu, L.-J. Li, and C.-K. Shih, Nano Lett. **14**, 2443 (2014).

[18] Y. Li, Z. Zhou, S. Zhang, and Z. Chen, J. Am. Chem. Soc. **130**, 16739 (2008).

[19] A. Vojvodic, B. Hinnemann, and J. K. Nørskov, Phys. Rev. B **80**, 125416 (2009).

[20] C. Ataca, H. Şahin, E. Aktürk, and S. Ciraci, J. Phys. Chem. C **115**, 3934 (2011).

[21] E. Erdogan, I. H. Popov, A. N. Enyashin, and G. Seifert, Eur. Phys. J. B **85**, 1 (2012).

[22] P. Lu, X. Wu, W. Guo, and X. C. Zeng, Phys. Chem. Chem. Phys. **14**, 13035 (2012).

[23] H. Pan and Y.-W. Zhang, J. Phys. Chem. C **116**, 11752 (2012).

[24] H. Pan and Y.-W. Zhang, J. Mater. Chem. **22**, 7280 (2012).

[25] J. Klinovaja and D. Loss, Phys. Rev. B **88**, 075404 (2013).

[26] G.-B. Liu, W.-Y. Shan, Y. Yao, W. Yao, and D. Xiao, Phys. Rev. B **88**, 085433 (2013).

[27] E. McCann and V. I. Fal'ko, J. Phys. Condens. Matter **16**, 2371 (2004).

[28] A. R. Akhmerov and C. W. J. Beenakker, Phys. Rev. B **77**, 085423 (2008).

[29] A. Kormányos, G. Burkard, M. Gmitra, J. Fabian, V. Zólyomi, N. D. Drummond, and V. Fal'ko, 2D Mater. **2**, 022001 (2015).

[30] A. Kormányos, V. Zólyomi, N. D. Drummond, P. Rakyta, G. Burkard, and V. I. Fal'ko, Phys. Rev. B **88**, 045416 (2013).

[31] M. Gibertini and N. Marzari, Nano Lett. **15**, 6229 (2015).

[32] D. Xiao, G.-B. Liu, W. Feng, X. Xu, and W. Yao, Phys. Rev. Lett. **108**, 196802 (2012).

[33] W. Yao, S. A. Yang, and Q. Niu, Phys. Rev. Lett. **102**, 096801 (2009).

[34] K. Wakabayashi, K.-i. Sasaki, T. Nakanishi, and T. Enoki, Sci. Technol. Adv. Mater. **11**, 054504 (2010).

[35] E. Cappelluti, R. Roldán, J. A. Silva-Guillén, P. Ordejón, and F. Guinea, Phys. Rev. B **88**, 075409 (2013).

[36] H. Rostami, A. G. Moghaddam, and R. Asgari, Phys. Rev. B **88**, 085440 (2013).

[37] R.-L. Chu, G.-B. Liu, W. Yao, X. Xu, D. Xiao, and C. Zhang, Phys. Rev. B **89**, 155317 (2014).

[38] H. Rostami and R. Asgari, Phys. Rev. B **91**, 075433 (2015).

[39] S. Fang, R. K. Defo, S. Shirodkar, S. Lieu, G. Tritsaris, and E. Kaxiras, Phys. Rev. B **92**, 205108 (2015).

[40] L. Brey and H. A. Fertig, Phys. Rev. B **73**, 235411 (2006).