

**Robust determination of maximally localized Wannier functions**Éric Cancès,<sup>1</sup> Antoine Levitt,<sup>2</sup> Gianluca Panati,<sup>3</sup> and Gabriel Stoltz<sup>1</sup><sup>1</sup>*Université Paris-Est, CERMICS (ENPC), Inria, F-77455 Marne-la-Vallée, France*<sup>2</sup>*Inria Paris, F-75589 Paris Cedex 12, Université Paris-Est, CERMICS (ENPC), F-77455 Marne-la-Vallée, France*<sup>3</sup>*Dipartimento di Matematica, La Sapienza Università di Roma, Roma, Italy*

(Received 27 July 2016; revised manuscript received 16 January 2017; published 6 February 2017)

We propose an algorithm to determine maximally localized Wannier functions (MLWFs). This algorithm, based on recent theoretical developments, does not require any physical input such as initial guesses for the Wannier functions, unlike popular schemes based on the projection method. We discuss how the projection method can fail on fine grids when the initial guesses are too far from MLWFs. We demonstrate that our algorithm is able to find localized Wannier functions through tests on two-dimensional systems, simplified models of semiconductors, and realistic DFT systems by interfacing with the WANNIER90 code. We also test our algorithm on the Haldane and Kane-Mele models to examine how it fails in the presence of topological obstructions.

DOI: [10.1103/PhysRevB.95.075114](https://doi.org/10.1103/PhysRevB.95.075114)**I. INTRODUCTION**

Wannier functions are a well-established tool in the solid-state physics community. In addition to providing intuition on chemical bonding, they are the theoretical and computational underpinning of many developments such as tight-binding approximations [1], interpolation of band structures [2], and the so-called modern theory of polarization or orbital magnetization [3,4].

Wannier functions are, however, not uniquely defined, and the choice of the phase of the (quasi-) Bloch functions can have a dramatic impact on their spatial localization. As for a single isolated Bloch band, it was realized already in the 1960s that in specific situations, appropriate choices of gauge lead to exponentially localized Wannier functions [5–7]. Progress was made in the 1980s by Nenciu [8] as well as Helffer and Sjöstrand [9], who proved in the general case the existence of a Bloch gauge yielding exponentially localized Wannier functions.

Whenever Bloch bands intersect each other, as it generically happens in three-dimensional (3D) crystals, Bloch functions cannot be smooth, and not even continuous, at the crossing points [excluding the exceptional case of one-dimensional (1D) systems]. In this situation, a multiband approach is mandatory. As early realized [5,6,10], in insulators and semiconductors it is convenient to consider all the Bloch bands below the Fermi level as a whole, and to replace Bloch functions with quasi-Bloch functions, namely, eigenfunctions of the projector  $P(\mathbf{k})$  on the occupied bands. Their preimages under the Bloch transform are called composite Wannier functions since they refer to a composite family of energy bands. For the sake of a simpler terminology, we will skip the adjective “composite” in the next sections.

The existence of an orthonormal basis of well-localized Wannier functions is equivalent to the existence of an orthonormal frame of quasi-Bloch functions, spanning  $\text{Ran } P(\mathbf{k})$ , which is both smooth and periodic as a function of  $\mathbf{k}$ . Such an existence problem has been a long-standing problem in theoretical solid-state physics. The solution for 1D systems was provided in [11,12], while a solution for 2D and 3D systems required the crucial use of geometric ideas and methods [13,14].

From a practical viewpoint, Marzari and Vanderbilt introduced an optimization procedure to minimize the spread of Wannier functions [15] and compute maximally localized Wannier functions (MLWFs). The corresponding localization functional is now known as the Marzari-Vanderbilt (MV) functional. The approach based on the minimization of the MV functional was developed before a full understanding of the theoretical criteria for localization was gained. It, however, gives very satisfactory results in many situations, and has become one of the standard tools of computational solid-state physics. It was conjectured in [15] that global minimizers of the MV functional are exponentially localized, a fact which was proved recently in [16], thereby providing a firm and mathematically consistent ground for the MV optimization.

It has, however, been observed that, in some situations, the minimization of the MV functional could fail because the algorithm remained trapped in “false local minima” presenting unphysical oscillations [15,17]. This issue is to be distinguished from physically relevant “real” local minima [18] and was associated with bad initial guesses obtained by the projection method and fine  $\mathbf{k}$ -point meshes. This problem has been considered recently in [19], where an alternative, more robust method is presented. This method is, however, still based on a projection and requires a physical input in the choice of basis functions, while the algorithm we present here does not require any input or parameter tuning. In that sense, it is similar to the recently proposed selected columns of the density matrix (SCDM) algorithm [20,21], which uses columns of the density matrix to provide localized Wannier functions. Compared to this method, our approach works directly in Bloch space, avoiding any representation of the density matrix in real space, and is readily implementable using the same input as standard MLWFs computation. It also does not depend on a potentially ill-conditioned matrix inversion.

We will show in this work that the issue of “false local minima” occurs when the initial guess corresponds to a Bloch gauge with vortexlike discontinuities, which prevent the convergence of the MV optimization algorithm on fine samplings of the Brillouin zone. To provide an admissible initial guess for the MV algorithm, we should really find a continuous Bloch gauge, a mathematically nontrivial task. This issue has been studied recently in [22,23], where the

authors develop constructive algorithms which complement the abstract existence results of [13,14]. Both methods are, however, not trivial to implement in practice, that of [22] requiring a perturbation argument that might yield to Bloch gauge with very sharp transitions, and that of [23] relying on cumbersome interpolation procedures.

In this work, we present an algorithm inspired by the theoretical works of [22,23] but more suited for actual implementation. This method fixes the gauge at the  $\Gamma$  point in the Brillouin zone and then incrementally on its neighboring points, until the whole Brillouin zone is covered. Certain pseudoperiodicity conditions have then to be satisfied, which we enforce by the method of obstruction matrices developed in [23]. This requires finding continuous and periodic logarithms of families of unitaries, a problem discussed in [22]. We propose an algorithm to solve this problem. Although we are unable to prove that this step always produces a continuous Bloch gauge, we conjecture it to be the case, and support this claim by numerical tests.

Our algorithms only provide a continuous Bloch gauge, which yields algebraically decaying Wannier functions. Therefore, we see this algorithm not as an end in itself but as a way to obtain good initial guesses for the MV localization procedure. We show by numerical examples on two- and three-dimensional systems that this two-step process always yields localized Wannier functions, which is not the case for state-of-the-art methods based on the projection method. We therefore advocate the use of our method when good initial guesses are not available. We, however, emphasize that our method only finds “real” local minima of the Marzari-Vanderbilt functional, which might or might not possess optimal spread (i.e., be a global minimum). Since they are devoid of any physical input, they might also not correspond to physically intuitive Wannier functions. Therefore, we do not see our method as a replacement of the traditional construction of Wannier functions for systems where physical insight is available, but rather as an automated procedure that will always produce exponentially localized Wannier functions in the case when intuition is lacking.

This paper is organized as follows. Section II introduces the notation and recalls some well-known facts on Wannier functions. We next present our algorithm in a one-dimensional case in Sec. III, then its extensions to the two- and three-dimensional settings in Secs. IV and V. The subtle issue of the appropriate choice of continuous and periodic logarithms is discussed in Sec. VI. We finally illustrate the approach with numerical results in Sec. VII, and present some perspectives in Sec. VIII.

## II. NOTATION AND STATEMENT OF THE PROBLEM

We consider a  $d$ -dimensional crystal. We denote by  $\mathcal{R} = \sum_{j=1}^d \mathbb{Z} \mathbf{a}_j$  the corresponding Bravais lattice, by  $Y$  a unit cell centered on the origin, and by  $\mathcal{R}^* = \sum_{j=1}^d \mathbb{Z} \mathbf{a}_j^*$  the reciprocal lattice, with  $\mathbf{a}_i \cdot \mathbf{a}_j^* = 2\pi \delta_{ij}$ . We choose the vectors  $\{\mathbf{a}_j^*\}$  as a basis for the reciprocal lattice. In these coordinates, the Brillouin zone is

$$\mathcal{B} = \left\{ \mathbf{k} \in \mathbb{R}^d : \mathbf{k} = \sum_{j=1}^d k_j \mathbf{a}_j^* \text{ for } k_j \in \left[ -\frac{1}{2}, \frac{1}{2} \right] \right\} \simeq \left[ -\frac{1}{2}, \frac{1}{2} \right]^d,$$

where the symbol  $\simeq$  refers to the fact that hereafter we will identify the vector  $\mathbf{k} = \sum_{j=1}^d k_j \mathbf{a}_j^*$  with its components  $(k_1, \dots, k_d)$  with respect to the reciprocal basis.

### A. Smooth families of projectors and their symmetries

The key input of our algorithm is a family  $P(\mathbf{k})$  of orthogonal projectors of rank  $J$ , which is smooth with respect to the wave vector  $\mathbf{k} = \sum_j k_j \mathbf{a}_j^* \in \mathbb{R}^d$ . Such projectors are obtained, in standard solid-state physics computations, by spectral projections of the effective one-body Schrödinger operator

$$H = -\frac{1}{2} \Delta + V_{\text{per}}, \quad (1)$$

where  $V_{\text{per}}$  is a real-valued  $\mathcal{R}$ -periodic potential. More precisely, introduce the Bloch orbitals  $\psi_{n,\mathbf{k}}$ , with periodic parts  $u_{n,\mathbf{k}}$ , so that

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = u_{n,\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{r}}.$$

The integer  $n$  labels the band index, while  $\mathbf{k} \in \mathcal{B}$  denotes the quasimomentum. For a fixed  $\mathbf{k} \in \mathcal{B}$ , the periodic functions  $(u_{n,\mathbf{k}})_{n \geq 1}$  form an orthogonal basis of  $L^2_{\text{per}}(Y)$  consisting of solutions of the following eigenvalue problem:

$$\begin{aligned} H(\mathbf{k})u_{n,\mathbf{k}} &= \varepsilon_{n,\mathbf{k}} u_{n,\mathbf{k}}, \\ H(\mathbf{k}) &= \frac{1}{2} (-i \nabla + \mathbf{k})^2 + V_{\text{per}}, \end{aligned}$$

$$\int_Y \overline{u_{n,\mathbf{k}}(\mathbf{r})} u_{m,\mathbf{k}}(\mathbf{r}) d\mathbf{r} = \delta_{nm},$$

where the  $\varepsilon_{n,\mathbf{k}}$  are labeled by  $n$  in increasing order. Our convention for the Bloch transform is

$$f_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R} \in \mathcal{R}} f(\mathbf{r} + \mathbf{R}) e^{-i\mathbf{k} \cdot (\mathbf{r} + \mathbf{R})}.$$

In order to construct Wannier functions from the Bloch orbitals, we identify a set  $\mathcal{I}$  of  $J$  Bloch bands in the energy spectrum. We assume that they are isolated in the sense that

$$\inf_{\mathbf{k} \in \mathcal{B}, n \in \mathcal{I}, m \notin \mathcal{I}} |\varepsilon_{n,\mathbf{k}} - \varepsilon_{m,\mathbf{k}}| > 0.$$

This gap condition is satisfied for instance by the occupied bands of an insulator or a semiconductor. It ensures that the spectral projector

$$P(\mathbf{k}) = \sum_{n \in \mathcal{I}} |u_{n,\mathbf{k}}\rangle \langle u_{n,\mathbf{k}}| \quad (2)$$

is a smooth (and even analytic) function of  $\mathbf{k}$  [12,16].

The underlying symmetries of the Hamiltonian  $H(\mathbf{k})$  translate into corresponding symmetries of the family of projectors  $P(\mathbf{k})$ . We first note that, because the potential  $V_{\text{per}}$  is real,  $P(\mathbf{k})$  satisfies the time-reversal property

$$P(-\mathbf{k}) = C P(\mathbf{k}) C, \quad (3)$$

where  $C$  is the antiunitary operator corresponding to complex conjugation, namely,

$$C u_{n,\mathbf{k}} = \overline{u_{n,\mathbf{k}}}.$$

We also note that, for any  $\mathbf{K} \in \mathcal{R}^*$ ,

$$P(\mathbf{k} + \mathbf{K}) = \tau_{\mathbf{K}} P(\mathbf{k}) \tau_{-\mathbf{K}}, \quad (4)$$

with the following unitary translation operators:

$$(\tau_{\mathbf{K}}f)(\mathbf{r}) = e^{-i\mathbf{K}\cdot\mathbf{r}}f(\mathbf{r}).$$

The convention we use here is slightly different from the one in [23]: our definition of  $\tau_{\mathbf{K}}$  ensures that  $\tau_{\mathbf{K}}u_{n,\mathbf{k}}$  is an eigenvector of  $H(\mathbf{k} + \mathbf{K})$  (so that  $\tau_{\mathbf{K}}$  is a translation by  $-\mathbf{K}$ ). The family of operators  $\tau_{\mathbf{K}}$  is a unitary group representation of  $\mathcal{R}^*$ , in the sense that

$$\tau_{\mathbf{K}+\mathbf{K}'} = \tau_{\mathbf{K}}\tau_{\mathbf{K}'}, \quad \text{and} \quad \tau_{-\mathbf{K}} = \tau_{\mathbf{K}}^*. \quad (5)$$

In view of the time-reversal symmetry (3) and the translation property (4), we can restrict ourselves to studying the reduced Brillouin zone

$$\begin{aligned} \mathcal{B}_{\text{red}} &= \left\{ \mathbf{k} = \sum_{j=1}^d k_j \mathbf{a}_j^* \text{ with } \begin{array}{l} k_1 \in [0, \frac{1}{2}] \\ k_j \in [-\frac{1}{2}, \frac{1}{2}] \end{array} \text{ for } j \geq 2 \right\} \\ &\simeq \left[0, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]^{d-1}, \end{aligned}$$

and map functions defined for  $\mathbf{k} \in \mathcal{B}_{\text{red}}$  to functions defined for any  $\mathbf{k} \in \mathbb{R}^d$  by reflection and translation.

*Remark 1.* Similarly to [23], we do not use the specific structure of the model (1), so that the approach presented here can be applied to various other periodic models of quantum mechanics, such as tight-binding models or relativistic models described by a Dirac operator. We only require a continuous family of projectors  $P(\mathbf{k})$  satisfying (3) and (4), for some unitary operators  $\tau_{\mathbf{K}}$  and an antiunitary operator  $C$  satisfying

$$\tau_{\mathbf{K}+\mathbf{K}'} = \tau_{\mathbf{K}}\tau_{\mathbf{K}'}, \quad \tau_{-\mathbf{K}} = \tau_{\mathbf{K}}^*, \quad C^2 = \text{Id}, \quad \text{and} \\ \tau_{\mathbf{K}}C = C\tau_{-\mathbf{K}}.$$

*Remark 2.* It is possible to further restrict the Brillouin zone under consideration when  $P(\mathbf{k})$  has more symmetries than the time-reversal and translation symmetries (3) and (4). The construction we present here does not take into account these possible additional symmetries, and therefore may produce Wannier functions that do not respect these symmetries. Our algorithm could be extended to work only on the irreducible Brillouin zone, but this is outside the scope of this paper.

## B. Orthonormal frames

The fundamental element to construct well-localized Wannier functions and the output of our algorithm is a Bloch frame depending smoothly on the wave vector  $\mathbf{k}$ . By definition, a Bloch frame (or, shortly, a frame) is a mapping from  $\mathbf{k} \in \mathbb{R}^d$  to an orthonormal basis  $u_{\mathbf{k}} = (u_{1,\mathbf{k}}, \dots, u_{J,\mathbf{k}})$  of  $\text{Ran}[P(\mathbf{k})]$ , such that each component satisfies the pseudoperiodicity property

$$u_{n,\mathbf{k}+\mathbf{K}} = \tau_{\mathbf{K}}u_{n,\mathbf{k}} \quad \forall (\mathbf{k}, \mathbf{K}) \in \mathbb{R}^d \times \mathcal{R}^*. \quad (6)$$

The latter condition expresses the fact that such functions are compatible with the symmetries of the family of projectors given by (4).

We emphasize that, despite the similarity in the notation, we do not assume that the functions  $u_{n,\mathbf{k}}$  are (the periodic part of) Bloch functions. It is only assumed that

$$P(\mathbf{k})u_{n,\mathbf{k}} = u_{n,\mathbf{k}},$$

while  $H(\mathbf{k})u_{n,\mathbf{k}} = \varepsilon_{n,\mathbf{k}}u_{n,\mathbf{k}}$  does not hold true in general. Such functions are sometimes called quasi-Bloch functions in the literature.

As already noticed [12], the existence of a smooth Bloch frame is not trivial in view of the competition between the smoothness of  $u$  and the property (6), which encodes a global topological constraint. For instance, it is known that in some models with broken time-reversal symmetry (e.g., in the presence of a magnetic field or in a Chern insulator) there cannot exist any such continuous frame, due to a topological obstruction [13,24–26].

For two given frames  $u$  and  $v$ , we write  $u_{\mathbf{k}}^*v_{\mathbf{k}}$  for the  $J \times J$  matrix with entries

$$(u_{\mathbf{k}}^*v_{\mathbf{k}})_{nm} = \int_Y \overline{u_{n,\mathbf{k}}(x)}v_{m,\mathbf{k}}(x)dx.$$

For a  $J \times J$  matrix  $U(\mathbf{k})$ , we write

$$[u_{\mathbf{k}}U(\mathbf{k})]_n = \sum_{m \in \mathcal{I}} u_{m,\mathbf{k}}U(\mathbf{k})_{mn}.$$

Note that, if  $u$  is a frame, then so is  $uU$  when the matrices  $U(\mathbf{k})$  are unitary for all values of  $\mathbf{k} \in \mathbb{R}^d$ . Similarly, for an operator  $A$ ,  $Au_{\mathbf{k}}$  is obtained by applying  $A$  to all components  $u_{j,\mathbf{k}}$  of  $u_{\mathbf{k}}$  independently. These conventions allow for a uniform notation whether  $u_{j,\mathbf{k}}$  are functions or coefficients in a basis set of size  $N$  (in which case  $u_{\mathbf{k}}$  is a  $N \times J$  matrix, and the previous definitions are simply restatements of matrix multiplication rules).

## C. Well-localized Wannier functions

(Composite) Wannier functions are defined, for a given frame  $u$ , by

$$\forall \mathbf{r} \in \mathbb{R}^d, \quad w_n(\mathbf{r}) = \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} u_{n,\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k}, \quad (7)$$

and the translations of these functions:

$$\begin{aligned} \forall (\mathbf{r}, \mathbf{R}) \in \mathbb{R}^d \times \mathbb{R}, \quad w_{n,\mathbf{R}}(\mathbf{r}) &:= w_n(\mathbf{r} - \mathbf{R}) \\ &= \frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} u_{n,\mathbf{k}}(\mathbf{r}) e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R})} d\mathbf{k}. \end{aligned}$$

The Wannier functions  $\{w_{n,\mathbf{R}}\}_{n \in \mathcal{I}, \mathbf{R} \in \mathcal{R}}$  form a complete orthonormal basis of the subspace of  $L^2(\mathbb{R}^d)$  associated with the chosen  $J$  bands.

The localization properties of the Wannier functions are determined by the regularity of the frame  $u$ . This can be seen from the fact that, because  $u$  satisfies the pseudoperiodicity property (6), an integration by parts gives

$$\frac{1}{|\mathcal{B}|} \int_{\mathcal{B}} \left( \frac{\partial}{\partial k_i} u_{n,\mathbf{k}}(\mathbf{r}) \right) e^{i\mathbf{k}\cdot\mathbf{r}} d\mathbf{k} = -ir_i w_n(\mathbf{r}).$$

Therefore, a  $C^\infty$  frame yields Wannier functions that decay faster than any polynomial, and an analytic frame yields exponentially localized Wannier functions (see Sec. 2 of [16] for a more precise statement involving the Sobolev regularity of  $u_{n,\mathbf{k}}$ ). In practice, the frame straightforwardly coming out of the simulations, computed on the reduced Brillouin zone, is usually not smooth because of the arbitrary phases of the  $u_{n,\mathbf{k}}$  and of the possible crossings between the energy levels  $\varepsilon_{n,\mathbf{k}}$ . Therefore, some correction through a family of unitary

matrices  $U(\mathbf{k})$  has to be applied to the frame in order to transform it into a smooth frame  $uU$ . Although it is easy to fix phases to impose local smoothness, global smoothness is a more complicated issue because the frame, extended by reflection and translation to the full space  $\mathbb{R}^d$ , might not be continuous at the boundary of the reduced Brillouin zone. Certain compatibility or gluing conditions should be satisfied in order for this extension to be smooth.

Our problem is to compute real-valued localized Wannier functions, which, after Bloch transform, is equivalent to the following.

*Problem 3.* Find a smooth frame  $u$  such that

$$\forall \mathbf{k} \in \mathbb{R}^d, \quad u_{-\mathbf{k}} = Cu_{\mathbf{k}}. \quad (8)$$

The additional property (8) ensures that the Wannier functions are real valued.

We discuss in the next section an algorithm for constructing continuous frames. We do not attempt to impose a higher degree of regularity, as done in [22,23] via abstract procedures, but consider a more pragmatic approach where a subsequent smoothing of the continuous frame is obtained with the MV algorithm. Our hope is that the continuous frame we obtain at the end of our procedure is a sufficiently good initial guess for the minimization of the MV functional in order to actually converge towards a minimizer.

### III. ALGORITHM IN THE ONE-DIMENSIONAL CASE

We present here a practical algorithm to solve Problem 3 in dimension 1. It can be seen as a discretization of well-known procedures to construct smooth frames [23], dating back to at least [12], and discussed in [15] Sec. IV C 1. We, however, carefully present this simple case since, as in [22], it is the building block of the algorithm in higher dimensions.

We start from a given frame  $u$  on the reduced Brillouin zone  $\mathcal{B}_{\text{red}} = [0, \frac{1}{2}]$ , which we then extend to the whole space  $\mathbb{R}$  by the relations (6) and (8). More precisely, we first extend  $u$  to  $k \in (-\frac{1}{2}, 0)$  by  $u_k = Cu_{-k}$ , and then to any  $k \in \mathbb{R}$  by  $u_k = \tau_{-K}u_{k+K}$  where  $K \in \mathbb{Z}$  is such that  $k + K \in (-\frac{1}{2}, \frac{1}{2}]$ . This procedure yields a globally continuous frame if and only if the frame on  $\mathcal{B}_{\text{red}}$  is continuous and satisfies the following compatibility conditions at 0 and  $\frac{1}{2}$ :

$$u_0 = Cu_0, \quad (P1)$$

$$u_{1/2} = \tau_1 Cu_{1/2}. \quad (P2)$$

The algorithm we propose consists of four steps: choosing a starting frame at 0 satisfying (P1), propagating this to  $[0, \frac{1}{2}]$ , enforcing (P2) at  $\frac{1}{2}$ , and propagating this fix back to  $[0, \frac{1}{2}]$ .

*Step 1: Fix 0.* The first step is to choose a frame  $u_0$  satisfying (P1). Since  $H(0)$  is real valued, this can be done by choosing an orthonormal set of real-valued eigenfunctions of  $H(0)$ .

*Step 2: Propagate from  $k = 0$  to  $\frac{1}{2}$ .* Evolving the eigenvector at  $k = 0$  to  $\frac{1}{2}$  can be done in various ways, for instance by the Sz-Nagy intertwining unitary as done in [22]. Here, we describe a natural way of performing this operation on a mesh of  $k$  points, using a Löwdin orthogonalization procedure. Assume that a mesh  $0 = k_0 < k_1 < \dots < k_N = \frac{1}{2}$  of  $\mathcal{B}_{\text{red}} =$

$[0, \frac{1}{2}]$  is given. We construct iteratively

$$\begin{aligned} \tilde{u}_{k_i} &= P(k_i)u_{k_{i-1}}, \\ u_{k_i} &= \tilde{u}_{k_i}(\tilde{u}_{k_i}^* \tilde{u}_{k_i})^{-1/2}. \end{aligned} \quad (9)$$

Since  $P$  is continuous, provided the mesh spacing is sufficiently small, the overlap matrix  $\tilde{u}_{k_i}^* \tilde{u}_{k_i}$  has its eigenvalues bounded away from 0, so that  $u_{k_i}$  is a well-defined orthonormal basis of  $\text{Ran}[P(k_i)]$ .

*Remark 4.* When the mesh spacing tends to zero, it can be shown that  $u$  converges to the solution of the following ordinary differential equation (ODE), with initial condition  $u_0$ :

$$\frac{du_k}{dk} = \frac{dP(k)}{dk}u_k - \frac{1}{2}u_k \left( \left[ \frac{dP(k)}{dk}u_k \right]^* u_k + u_k^* \left[ \frac{dP(k)}{dk}u_k \right] \right). \quad (10)$$

In particular, the function  $k \mapsto u_k$  is smooth.

*Step 3: Enforcing (P2) at  $k = \frac{1}{2}$ .* The previous step yields a frame that is continuous, satisfies the compatibility condition (P1), but not (P2). Note, however, that  $\text{Ran}[P(\frac{1}{2})] = \text{Ran}[\tau_1 P(-\frac{1}{2})] = \text{Ran}[\tau_1 C P(\frac{1}{2})]$ , so that  $u_{1/2}$  and  $\tau_1 Cu_{1/2}$  are both orthonormal bases of the same space. There exists therefore a unitary matrix  $U_{\text{obs}}$ , which we call ‘‘obstruction matrix’’ following [23], such that

$$u_{1/2}U_{\text{obs}} = \tau_1 Cu_{1/2}. \quad (11)$$

This matrix can be explicitly computed as  $U_{\text{obs}} = u_{1/2}^*(\tau_1 Cu_{1/2})$ . A simple computation also shows that  $U_{\text{obs}}^T = U_{\text{obs}}$ :

$$\begin{aligned} [U_{\text{obs}}]_{nm} &= \langle u_{n,1/2}, \tau_1 Cu_{m,1/2} \rangle = \langle C\tau_1 Cu_{m,1/2}, Cu_{n,1/2} \rangle \\ &= \langle \tau_{-1}u_{m,1/2}, Cu_{n,1/2} \rangle \\ &= \langle u_{m,1/2}, \tau_1 Cu_{n,1/2} \rangle = [U_{\text{obs}}]_{mn}. \end{aligned}$$

Therefore,  $\overline{U_{\text{obs}}} = U_{\text{obs}}^* = U_{\text{obs}}^{-1}$ .

We now look for a matrix  $U_{\text{heal}}$  such that  $u_{1/2}U_{\text{heal}}$  satisfies the compatibility condition (P2). Applying  $\tau_1 C$  and using (11), we get

$$\begin{aligned} \tau_1 C(u_{1/2}U_{\text{heal}}) &= (\tau_1 Cu_{1/2})\overline{U_{\text{heal}}} \\ &= u_{1/2}(U_{\text{obs}}\overline{U_{\text{heal}}}) \\ &= (u_{1/2}U_{\text{heal}})(U_{\text{heal}}^* U_{\text{obs}} \overline{U_{\text{heal}}}). \end{aligned}$$

We see that  $u_{1/2}U_{\text{heal}}$  satisfies the compatibility condition (P2) if and only if  $U_{\text{heal}}^* U_{\text{obs}} \overline{U_{\text{heal}}} = \text{Id}_J$ , i.e.,  $U_{\text{heal}} U_{\text{heal}}^T = U_{\text{obs}}$ . Since  $(U_{\text{obs}}^{1/2})^T = (U_{\text{obs}}^T)^{1/2} = U_{\text{obs}}^{1/2}$ , we can choose

$$U_{\text{heal}} = U_{\text{obs}}^{1/2}.$$

With this choice,  $u_{1/2}U_{\text{heal}}$  satisfies (P2). We define the fractional power of a normal matrix in the usual way, by applying the exponent to its eigenvalues. For matrices with complex eigenvalues  $z$  such as  $U_{\text{obs}}$ , defining  $z^\alpha = e^{\alpha \ln z}$  requires choosing a branch cut in the complex plane for the logarithm, which we fix by using the principal determination, i.e., for all  $z \in \mathbb{C}^*$ ,  $\text{Im}(\ln z) \in (-\pi, \pi]$ .

*Step 4: Correct the frame on  $\mathcal{B}_{\text{red}}$ .* The last step is to globally define a new frame  $u'$  on  $\mathcal{B}_{\text{red}}$  by interpolation:

$$u'_k = u_k U_{\text{heal}}^{2k} = u_k U_{\text{obs}}^k.$$

Note that  $u'_0 = u_0 = Cu_0 = Cu'_0$ , and so  $u'$  satisfies (P1). In the limit of infinitely small mesh spacing, the mapping  $k \mapsto u_k$  is continuous, and so is  $k \mapsto U_{\text{obs}}^k$ , which implies that  $k \mapsto u'_k$  is continuous on  $\mathcal{B}_{\text{red}}$ .

The compatibility conditions finally ensure that  $u_k$  can be extended to a continuous frame for all  $k \in \mathbb{R}$ .

*Remark 5.* We have only imposed continuity in our construction, but we actually get a smooth frame. This is because the frame we constructed is the same as the one we would obtain by defining  $u$  on  $\mathbb{R}$  via the propagation (9), then  $u'_k = u_k U_{\text{obs}}^k$ . The latter frame is smooth because both  $k \mapsto u_k$  and  $k \mapsto U_{\text{obs}}^k$  are.

#### IV. ALGORITHM IN THE TWO-DIMENSIONAL CASE

As in the one-dimensional case, we first construct a frame on the reduced Brillouin zone

$$\mathcal{B}_{\text{red}} \simeq \left[0, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right],$$

and then extend it by symmetry to the whole space using (6) and (8). This yields a globally continuous frame if and only if the frame on the reduced Brillouin zone is continuous and the following compatibility conditions are satisfied:

$$u_{(0,k_2)} = Cu_{(0,-k_2)}, \quad (\text{E1})$$

$$u_{(k_1,1/2)} = \tau_{(0,1)}u_{(k_1,-1/2)}, \quad (\text{E2})$$

$$u_{(1/2,k_2)} = \tau_{(1,0)}Cu_{(1/2,-k_2)}. \quad (\text{E3})$$

Note that these conditions are now edge conditions, relating two points of the same edge [for (E1) and (E3)] or two points on opposite edges [for (E2)], in contrast to the point conditions of the one-dimensional case. These edge conditions imply compatibility conditions at the special points  $(0,0)$ ,  $(0, \pm \frac{1}{2})$ ,  $(\frac{1}{2}, \pm \frac{1}{2})$ , and  $(\frac{1}{2}, 0)$ , which are the fixed points of the symmetry group. All these conditions can be visualized on Fig. 1.

We take care of these conditions in a sequential manner: we first find a frame on  $\{0\} \times [-\frac{1}{2}, \frac{1}{2}]$  that satisfies (E1), extend this to a frame satisfying (E2), then enforce the condition (E3) on the right edge.

*Step 1: Constructing a frame satisfying (E1) and (E2).* We first use the one-dimensional construction to obtain a frame  $u_{(0,k_2)}$  with  $-\frac{1}{2} \leq k_2 \leq \frac{1}{2}$ , satisfying (E1) as well as

$$u_{(0,1/2)} = \tau_{(0,1)}u_{(0,-1/2)}. \quad (12)$$

Then, for every  $k_2 \in [-\frac{1}{2}, \frac{1}{2}]$ , we propagate from  $(0,k_2)$  to  $(\frac{1}{2}, k_2)$  using the propagation procedure described in (9) in the one-dimensional case. This defines a frame on  $\mathcal{B}_{\text{red}}$ .

The compatibility condition (12) at  $k_1 = 0$  is propagated for any  $k_1 \in [0, \frac{1}{2}]$ , which implies that (E2) is satisfied. Let us prove this when starting from  $k_1 = 0$  and going to  $k_1 = \Delta k$ . We define  $\tilde{u}_{(\Delta k, k_2)} = P(\Delta k, k_2)u_{(0, k_2)}$ . Then, using (12) and then (4),

$$\begin{aligned} \tilde{u}_{(\Delta k, 1/2)} &= P(\Delta k, 1/2)\tau_{(0,1)}u_{(0, -1/2)} \\ &= \tau_{(0,1)}P(\Delta k, -1/2)u_{(0, -1/2)} = \tau_{(0,1)}\tilde{u}_{(\Delta k, -1/2)}, \end{aligned}$$

so that (E2) follows upon normalizing the frames as  $u_{\mathbf{k}} = \tilde{u}_{\mathbf{k}}(\tilde{u}_{\mathbf{k}}^* \tilde{u}_{\mathbf{k}})^{-1/2}$ . However, (E3) is not satisfied in general.

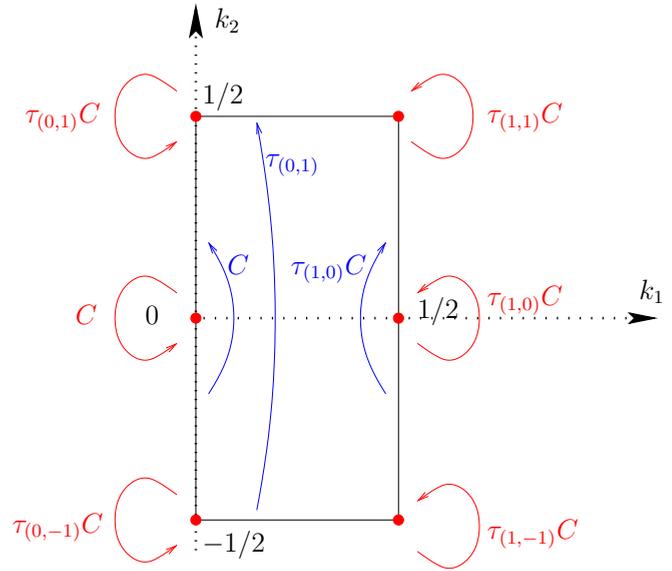


FIG. 1. Symmetries in the two-dimensional reduced Brillouin zone. The blue arrows correspond to the edge conditions (E1), (E2), and (E3), and the red arrows to the point conditions implied by the edge conditions.

*Step 2: Enforcing the compatibility conditions (E3) at the corners  $(\frac{1}{2}, \pm \frac{1}{2})$ .* Before satisfying (E3) on the right edge, we enforce the condition at the corner  $(\frac{1}{2}, \frac{1}{2})$ , where the condition (E3) implies

$$u_{(1/2,1/2)} = \tau_{(1,1)}Cu_{(1/2,1/2)}. \quad (13)$$

This is a point compatibility condition analogous to the one encountered in the one-dimensional case. To enforce it, we consider the unitary obstruction matrix

$$U_{\text{obs}} = u_{(1/2,1/2)}^* \tau_{(1,1)}Cu_{(1/2,1/2)},$$

and define the modified frame

$$u'_{\mathbf{k}} = u_{\mathbf{k}} U_{\text{obs}}^{k_1}. \quad (14)$$

The modified frame  $u'$  satisfies (13) by construction. Note also that the transformation (14) preserves the compatibility conditions (E2) and (E1). Moreover, (E2) and (13) imply that  $u'$  also satisfies the corner condition at  $(\frac{1}{2}, -\frac{1}{2})$ :

$$u'_{(1/2,-1/2)} = \tau_{(1,-1)}Cu'_{(1/2,-1/2)}.$$

In order to simplify the notation, we drop the prime in the remainder of the algorithm, and simply denote by  $u$  the modified frame obtained at the end of this step.

*Step 3: Enforcing (E3) on the right edge.* We still need to modify  $u$  to satisfy (E3) for any value of  $k_2 \in [-\frac{1}{2}, \frac{1}{2}]$ . We define to this end the following family of obstruction matrices:

$$U_{\text{obs}}(k_2) = u_{(1/2,k_2)}^* \tau_{(1,0)}Cu_{(1/2,-k_2)}.$$

These matrices satisfy by construction

$$U_{\text{obs}}(-k_2) = U_{\text{obs}}(k_2)^T, \quad U_{\text{obs}}(1/2) = U_{\text{obs}}(-1/2) = \text{Id}_J.$$

At this point, we would like to define a new frame by

$$u'_{(k_1, k_2)} = u_{(k_1, k_2)} U_{\text{obs}}(k_2)^{k_1}.$$

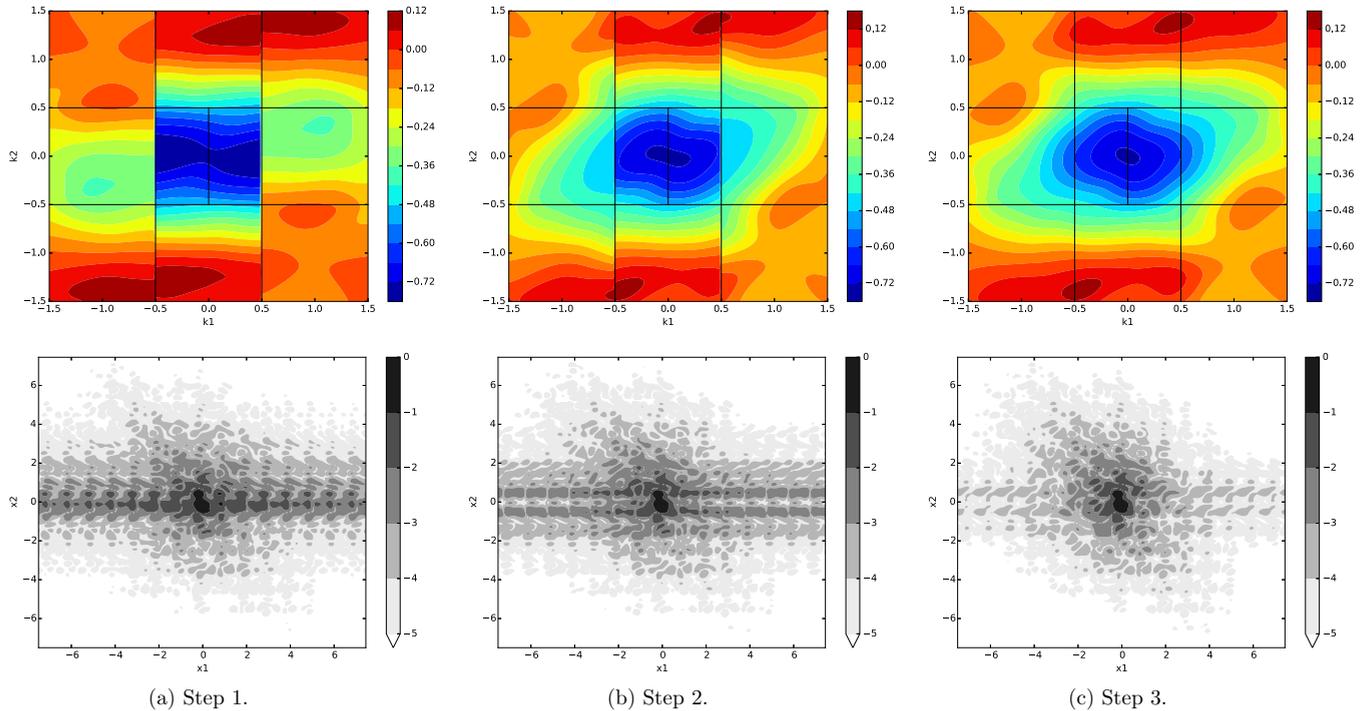


FIG. 2. Real part of the average of the first component  $\int_Y \text{Re}[u_{1,\mathbf{k}}(r)]dr$ , of a Bloch frame corresponding to  $J = 5$  Bloch bands, as a function of  $\mathbf{k} \in [-\frac{3}{2}, \frac{3}{2}]^2$  (top) and associated Wannier function  $w_1$  in logarithmic scale (bottom). The real part of the average of  $u_{1,\mathbf{k}}$  is used as a proxy to visualize the regularity of the frame  $u_{\mathbf{k}}$ . After step 1, the frame satisfies the compatibility conditions related to vertical translations, and the associated Wannier function is localized vertically, but not horizontally (left panel). After step 2, the frame satisfies all the compatibility conditions at the corners  $(\frac{1}{2} + n_1, \frac{1}{2} + n_2)$  (with  $n_1, n_2 \in \mathbb{Z}$ ), but not on the vertical edges between these points. The associated Wannier function is still delocalized horizontally (middle panel). After step 3, the frame satisfies all the compatibility conditions and is globally continuous, although not smooth. The associated Wannier function is localized, but exhibits a slow algebraic horizontal decay (right panel). See Sec. VII for the details of the model used.

The frame  $u'$  satisfies all the compatibility conditions (E1)–(E3), is continuous with respect to  $k_1$ , but may fail to be continuous with respect to  $k_2$ , because the eigenvalues of  $U_{\text{obs}}$  may pass through the branch cut of the logarithm in the negative real axis. We are therefore faced with the problem of the continuous periodic logarithm: find a matrix-valued function  $k_2 \mapsto L(k_2)$ , with  $L(k_2)$  Hermitian, which is continuous and satisfies the following conditions for all  $k_2 \in [-\frac{1}{2}, \frac{1}{2}]$ :

$$\begin{aligned} \exp[iL(k_2)] &= U_{\text{obs}}(k_2), \quad L(-k_2) = L(k_2)^T, \\ L(-1/2) &= L(1/2) = 0_J. \end{aligned}$$

If such a function exists, the frame

$$u'_{\mathbf{k}} = u_{\mathbf{k}} \exp(ik_1 L(k_2))$$

is continuous and satisfies the compatibility conditions (E1)–(E3). We delay the discussion of this problem to Sec. VI.

The behavior of the algorithm on a simple example can be visualized in Fig. 2.

*Remark 6.* The algorithm produces a continuous frame. However, in contrast to the 1D case, it does not provide any additional regularity. Because the 1D construction guarantees a smooth frame, there are no discontinuities of derivatives in the vertical direction, i.e., at the points  $(k_1, \frac{1}{2} + \mathbb{Z})$ . However, the condition for the first derivative to be continuous at a point

$(\frac{1}{2}, k_2)$  is

$$\left. \frac{\partial u'}{\partial k_1} \right|_{(1/2, k_2)} = -\tau_{(1,0)} C \left. \frac{\partial u'}{\partial k_1} \right|_{(1/2, -k_2)}$$

which is in general incompatible with (E3). This yields discontinuous derivatives at  $(\frac{1}{2} + \mathbb{Z}, k_2)$ .

This means that the Wannier function will only decay algebraically in the  $\mathbf{a}_1$  direction, as can be seen in the right panel of Fig. 2. This is, however, easily fixed by the application of the MV minimization algorithm, as will be shown on Fig. 5 in Sec. VII. We note that the MV algorithm is essentially equivalent to other forms of smoothing, such as for instance the one employed in [22]: the MV algorithm can be seen as the gradient flow of a Dirichlet-type energy functional, i.e., a heat flow, whose solution can be written as a convolution similar to the one used in [22].

### V. ALGORITHM IN THE THREE-DIMENSIONAL CASE

We use the same induction to pass from the two-dimensional to the three-dimensional case as the one to pass from the one-dimensional to the two-dimensional case. We define a frame on the reduced Brillouin zone

$$\mathcal{B}_{\text{red}} \simeq \left[0, \frac{1}{2}\right] \times \left[-\frac{1}{2}, \frac{1}{2}\right]^2,$$

and extend it to the full reciprocal space by symmetry. The compatibility conditions to be satisfied by the frame  $u$  are now

$$u_{(0,k_2,k_3)} = Cu_{(0,-k_2,-k_3)}, \quad (\text{F1})$$

$$u_{(k_1,\pm 1/2,k_3)} = \tau_{(0,\pm 1,0)}u_{(k_1,\mp 1/2,k_3)}, \quad (\text{F2})$$

$$u_{(k_1,k_2,\pm 1/2)} = \tau_{(0,0,\pm 1)}u_{(k_1,k_2,\mp 1/2)}, \quad (\text{F3})$$

$$u_{(1/2,k_2,k_3)} = \tau_{(1,0,0)}Cu_{(1/2,-k_2,-k_3)}. \quad (\text{F4})$$

*Step 1: Constructing a frame satisfying (F1)–(F3).* Our first step is to use the two-dimensional algorithm to build a frame  $u_{(0,k_2,k_3)}$  on the face  $k_1 = 0$ . This frame satisfies (F1), and additionally

$$u_{(0,\pm 1/2,k_3)} = \tau_{(0,\pm 1,0)}u_{(0,\mp 1/2,k_3)},$$

$$u_{(0,k_2,\pm 1/2)} = \tau_{(0,0,\pm 1)}u_{(0,k_2,\mp 1/2)}.$$

For any  $(k_2,k_3) \in [-\frac{1}{2}, \frac{1}{2}]^2$ , we propagate from  $(0,k_2,k_3)$  to  $(\frac{1}{2},k_2,k_3)$  using (9). This defines a continuous frame on  $\mathcal{B}_{\text{red}}$  which satisfies (F1)–(F3), but not (F4).

*Step 2: Enforcing (F4) on the edges of the face  $k_1 = \frac{1}{2}$ .* As in the two-dimensional case, before enforcing (F4) on the whole face  $k_1 = \frac{1}{2}$ , we enforce it on the four edges of its boundary. In order to do this, we first fix the corner  $(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})$ , for which the compatibility condition is

$$u_{(1/2,1/2,1/2)} = \tau_{(1,1,1)}Cu_{(1/2,1/2,1/2)}. \quad (15)$$

We define the obstruction matrix

$$U_{\text{obs}} = u_{(1/2,1/2,1/2)}^* \tau_{(1,1,1)} Cu_{(1/2,1/2,1/2)},$$

and introduce

$$u'_{\mathbf{k}} = u_{\mathbf{k}} U_{\text{obs}}^{k_1}.$$

By construction,  $u'$  satisfies (15). The compatibility conditions (F1)–(F3) are also still valid for the modified frame  $u'$ . In addition, in view of (15) and (F2) and (F3), the frame  $u'$  satisfies the compatibility conditions at the other corners  $(\frac{1}{2}, \frac{1}{2}, -\frac{1}{2})$ ,  $(\frac{1}{2}, -\frac{1}{2}, \frac{1}{2})$ , and  $(\frac{1}{2}, -\frac{1}{2}, -\frac{1}{2})$ .

We next enforce the condition (F4) on the edge  $(k_1,k_2) = (\frac{1}{2}, \frac{1}{2})$ , namely,

$$u_{(1/2,1/2,k_3)} = \tau_{(1,1,0)}Cu_{(1/2,1/2,-k_3)}.$$

The corresponding unitary obstruction matrix is

$$U_{\text{obs}}(k_3) = u_{(1/2,1/2,k_3)}^* \tau_{(1,1,0)} Cu_{(1/2,1/2,-k_3)}.$$

Note that  $U_{\text{obs}}(-k_3) = U_{\text{obs}}(k_3)^T$  and  $U_{\text{obs}}(\pm \frac{1}{2}) = \text{Id}_J$ . As in the two-dimensional case, provided we can solve the logarithm problem, we find a Hermitian logarithm  $L(k_3)$  satisfying  $L(-k_3) = L(k_3)^T$ ,  $L(\pm \frac{1}{2}) = 0_J$ , and  $e^{iL(k_3)} = U_{\text{obs}}(k_3)$ , and modify the frame as

$$u''_{\mathbf{k}} = u'_{\mathbf{k}} \exp(ik_1 L(k_3)).$$

The modified frame  $u''$  then satisfies (F4) on the two edges  $(k_1,k_2) = (\frac{1}{2}, \pm \frac{1}{2})$ .

We repeat this construction on the edge  $(k_1,k_3) = (\frac{1}{2}, \frac{1}{2})$ , and introduce appropriate Hermitian matrices  $L(k_2)$  such that

$$u'''_{\mathbf{k}} = u''_{\mathbf{k}} \exp(ik_1 L(k_2))$$

satisfies (F4) on the two edges  $(k_1,k_3) = (\frac{1}{2}, \pm \frac{1}{2})$ .

To simplify the notation, we drop the primes in the remainder of the algorithm, and simply denote by  $u$  the modified frame obtained at the end of this step.

*Step 3: Enforcing (F4) on the whole face  $k_1 = \frac{1}{2}$ .* The compatibility condition (F4) is, on the whole face  $k_1 = \frac{1}{2}$ :

$$u_{(1/2,k_2,k_3)} = \tau_{(1,0,0)}Cu_{(1/2,-k_2,-k_3)}.$$

We therefore introduce the family  $U_{\text{obs}}(k_2,k_3)$  of unitary obstruction matrices

$$U_{\text{obs}}(k_2,k_3) = u_{(1/2,k_2,k_3)}^* \tau_{(1,0,0)} Cu_{(1/2,-k_2,-k_3)}.$$

Because of the previous step,

$$\begin{aligned} U_{\text{obs}}(1/2,k_3) &= U_{\text{obs}}(-1/2,k_3) = U_{\text{obs}}(k_2,1/2) \\ &= U_{\text{obs}}(k_2, -1/2) = \text{Id}_J. \end{aligned}$$

As in the two-dimensional case, we would like to define a new frame satisfying all the compatibility conditions as

$$u'_{\mathbf{k}} = u_{\mathbf{k}} U_{\text{obs}}(k_2,k_3)^{k_1}.$$

We again are faced with the logarithm problem, which we discuss in Sec. VI.

*Remark 7.* As a mathematical curiosity, this construction immediately extends to higher dimensions, again provided the logarithm problem can be solved. By induction, we first build a frame on  $\{0\} \times [-\frac{1}{2}, \frac{1}{2}]^{d-1}$ , then propagate it to  $[0, \frac{1}{2}] \times [-\frac{1}{2}, \frac{1}{2}]^{d-1}$ . We next fix the obstruction on the boundary of  $\{\frac{1}{2}\} \times [-\frac{1}{2}, \frac{1}{2}]^{d-1}$ , and conclude by finding a continuous logarithm of the obstruction matrix on  $\{\frac{1}{2}\} \times [-\frac{1}{2}, \frac{1}{2}]^{d-1}$  to correct the frame.

## VI. LOGARITHM PROBLEM

We come back in this section to the so-called logarithm problem, summarized in Problem 8 below.

An element  $\mathbf{k} \in \mathbb{R}^d$  is denoted  $(k_1, \mathbf{k}')$  in this section. We also introduce the set of edges in dimension 3 (vertices in dimension 2)

$$\begin{aligned} \mathcal{E} = \left\{ \mathbf{k}' = (k_2, \dots, k_d) \in \left[ -\frac{1}{2}, \frac{1}{2} \right]^{d-1} \right. \\ \left. \times \left| k_i \in \left\{ -\frac{1}{2}, \frac{1}{2} \right\} \text{ for some } 2 \leq i \leq d \right. \right\}. \end{aligned}$$

With this notation, we can summarize the logarithm problem as follows.

*Problem 8.* Given a continuous mapping  $U$  from  $[-\frac{1}{2}, \frac{1}{2}]^{d-1}$  to the set of unitary matrices, such that

$$\forall \mathbf{k}' \in \mathcal{E}, \quad U(\mathbf{k}') = \text{Id}_J, \quad (16)$$

and

$$\forall \mathbf{k}' \in [-1/2, 1/2]^{d-1}, \quad U(-\mathbf{k}') = U(\mathbf{k}')^T \quad (17)$$

find a continuous mapping  $L$  from  $[-\frac{1}{2}, \frac{1}{2}]^{d-1}$  to the set of Hermitian matrices, such that

$$\forall \mathbf{k}' \in \mathcal{E}, \quad L(\mathbf{k}') = 0_J,$$

and, for all  $\mathbf{k}' \in [-\frac{1}{2}, \frac{1}{2}]^{d-1}$ ,

$$e^{iL(\mathbf{k}')} = U(\mathbf{k}'), \quad L(-\mathbf{k}') = L(\mathbf{k}')^T. \quad (18)$$

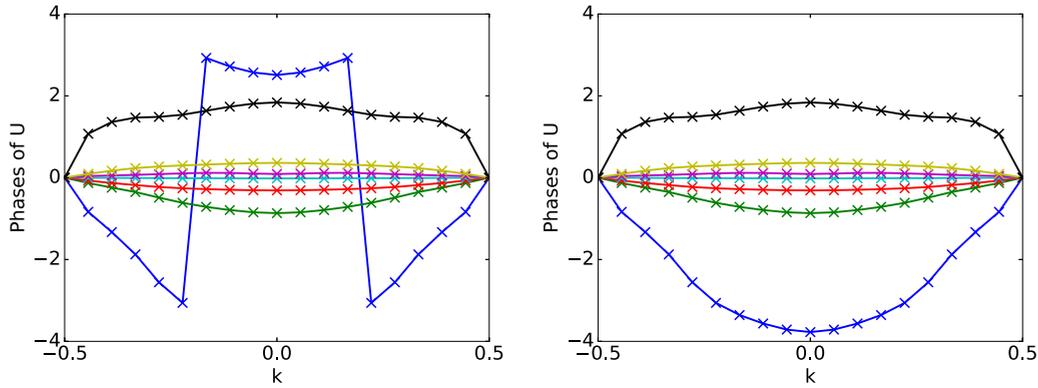


FIG. 3. Eigenvalues of  $L$ , with a naive determination (left) and with our algorithm (right). This example was obtained by a variation of the potential presented in Sec. VII with  $J = 7$ .

This is the problem as used by our construction in the previous two sections: (16) is satisfied because we fixed the obstruction matrix on the edges, while (17) stems from the structure of the conditions (E3) and (F4). Let us already note that this problem cannot be solved in general (see the counterexample in [22], which we recall in Sec. VIC). We describe an algorithm that solves the problem provided that the type of eigenvalue collisions present in this counterexample does not occur. In our numerical tests, we did not encounter such difficulties and were always able to solve the logarithm problem. We refer to Sec. VIC for a discussion of this issue.

### A. Two-dimensional case

A logarithm of  $U$  can be found by starting at  $k_2 = -\frac{1}{2}$ , and by following the eigenvalues of  $U$  to ensure the continuity of  $L$ . This procedure can be seen as a discrete counterpart of the phase following procedure used in [22, Lemma 2.13] for a single band.

Assume that  $[-\frac{1}{2}, 0]$  is partitioned as  $-\frac{1}{2} = k_{2,0} < k_{2,1} < \dots < k_{2,N} = 0$ . We set  $L(k_{2,0}) = 0$ , and iteratively determine  $L(k_{2,j})$  as a logarithm of  $U(k_{2,j})$ , taking into account the phase information in  $L(k_{2,j-1})$ . To this end, we diagonalize  $U(k_{2,j})$  as  $U(k_{2,j}) = W(k_{2,j})D(k_{2,j})W(k_{2,j})^*$ , where  $W(k_{2,j})$  is unitary and  $D(k_{2,j})$  a diagonal matrix whose entries have modulus 1. We then set  $L(k_{2,j}) = W(k_{2,j})E(k_{2,j})W(k_{2,j})^*$ , where  $E(k_{2,j})$  is the diagonal matrix given by

$$\begin{aligned}
 iE(k_{2,j})_n &= \ln[D(k_{2,j})_n] + 2\pi i p_{j,n}, \\
 p_{j,n} &= \operatorname{argmin}_{p \in \mathbb{Z}} \operatorname{dist}(\ln[D(k_{2,j})_n] + 2i\pi p, E(k_{2,j-1})),
 \end{aligned}
 \tag{19}$$

where  $\operatorname{dist}[x, E(k_{2,j-1})]$  is the distance of the imaginary number  $x$  to the set of diagonal elements of  $E(k_{2,j-1})$ . The matrix-valued function  $L$  is finally extended from  $[-\frac{1}{2}, 0]$  to  $[-\frac{1}{2}, \frac{1}{2}]$  by the relation  $L(-k_2) = L(k_2)^T$ . The continuity at  $k_2 = 0$  is ensured because  $L(0)^T = L(0)$  [see (17)].

Figure 3 shows the algorithm in action on one example. With a naive phase determination, i.e.,  $E(k_{2,j})_n = \ln[D(k_{2,j})_n]$ , the matrix-valued function  $L$  is not continuous when the eigenvalues of  $U$  cross  $-1$ , and therefore the eigenvalues of  $L$  cross  $\pm\pi$  (left panel). This is because the complex logarithm is only continuous outside its branch cut,

chosen here on the negative real axis. When that happens in our procedure, the integer  $p$  jumps to accommodate this, and the eigenvalues of  $L$  evolve smoothly (right panel).

If the eigenvalues of  $U$  never collide for values of  $k_2$  in the open interval  $(-\frac{1}{2}, \frac{1}{2})$ , then  $L$  is continuous in the limit of infinitesimal mesh spacing. This is a consequence of the fact that the minimizer in (19) is always uniquely determined, and the phase following procedure ensures that the eigenvalues are determined continuously when they cross the value  $-1$  (see also the continuity result stated in [22, Lemma 2.13]). The continuity of the frame is preserved even if a crossing happens, as long as the integer  $p_n$  is the same for all the colliding eigenvalues. On the other hand, when two eigenvalues with different values of  $p_n$  collide,  $L$  is discontinuous. In the example of Fig. 3, this would correspond to a collision of the top and bottom eigenvalues. In this case, our algorithm fails. We, however, never observed such a situation in all the tests we have performed. We discuss this in more details in Sec. VIC.

### B. Three-dimensional case

The three-dimensional case is handled similarly to the two-dimensional case, in slices: for every  $k_2$ , we find a continuous logarithm of  $k_3 \mapsto U(k_2, k_3)$ , which we call  $L(k_2, k_3)$ . As long as there are no crossings between eigenvalues with different values of  $p_n$ , this  $L$  is continuous with respect to both  $k_2$  and  $k_3$ , which solves the problem.

### C. Eigenvalue collisions

The prototypical case of a problematic eigenvalue collision is exemplified by the following situation (see [22, Example 2.25]): let  $J = 2$  and, for  $k \in \mathbb{R}$ , define

$$U(k) = - \begin{pmatrix} \cos(2\pi k) & -\sin(2\pi k) \\ \sin(2\pi k) & \cos(2\pi k) \end{pmatrix}. \tag{20}$$

The eigenvalues of this matrix are (see Fig. 4)

$$\lambda^\pm(k) = e^{\pm 2i\pi(k+1/2)}.$$

If the matrix-valued function  $L$  is continuous at  $k = 0$ , the phases of its eigenvalues must, respectively, vary by amounts of  $2\pi$  and  $-2\pi$  when evolving from  $k = -\frac{1}{2}$  to  $\frac{1}{2}$ . Therefore,  $L(1/2)$  must have eigenvalues  $\pm 2\pi$  and cannot be equal to

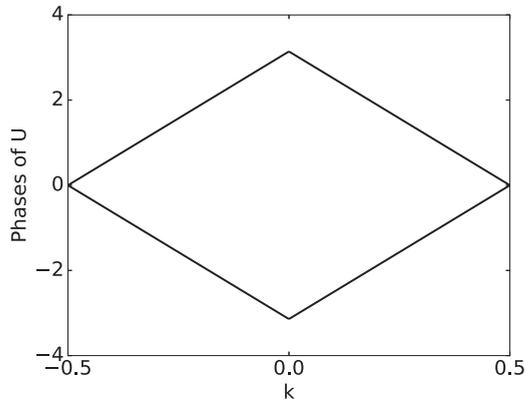


FIG. 4. Imaginary part of  $\ln[\lambda^\pm(k)]$ , the principal logarithm of the eigenvalues of (20).

0. This shows that no matrix logarithm can simultaneously be continuous and satisfy  $L(-\frac{1}{2}) = L(\frac{1}{2})$ .

In [22], the authors avoid this problem by relaxing Problem 8 to the following weaker problem: Find a sequence of Hermitian, continuous matrices  $L_1, \dots, L_N$  for  $\mathbf{k}' \in [-\frac{1}{2}, \frac{1}{2}]^{d-1}$ , equal to  $0_J$  on  $\mathcal{E}$ , and such that

$$e^{-iL_N(\mathbf{k}')/2} \dots e^{-iL_1(\mathbf{k}')/2} U(\mathbf{k}') e^{-iL_1(\mathbf{k}')/2} \dots e^{-iL_N(\mathbf{k}')/2} = \text{Id}_J, \quad (21)$$

$$\forall i \in \{1, \dots, N\}, \quad L_i(-\mathbf{k}') = L_i(\mathbf{k}')^T. \quad (22)$$

This is in fact sufficient for our purposes. Coming back to the last step of the algorithm in the three-dimensional case, assume that, instead of finding  $L$  satisfying (18) and setting

$$u'_\mathbf{k} = u_\mathbf{k} e^{ik_1 L(k_2, k_3)},$$

we find a family of functions  $(L_i)_{i=1, \dots, N}$  with values in the space of Hermitian matrices, which in addition satisfy (21). Then, we can fix the frame in  $N$  steps by defining

$$u'_\mathbf{k} = u_\mathbf{k} e^{ik_1 L_1(k_2, k_3)} \dots e^{ik_N L_N(k_2, k_3)}.$$

The corrected frame  $u'$  is continuous and satisfies all the required compatibility conditions.

In [22], the authors use the relaxed problem (21) in two steps. Their proof relies crucially on the analyticity properties of the eigenvalues and eigenvectors, and is tied to the one-dimensional case  $d - 1 = 1$ , i.e.,  $d = 2$ . Moreover, it does not provide a practical way of computing the Hermitian matrices  $L_i$ . A natural extension to their method to dimension  $d = 3$  is to slightly perturb  $U(\mathbf{k})$  to eliminate any problematic eigenvalue collision, and replace eigenvalue crossings with avoided crossings, which the authors allude to in [22, Remark 1.10]. For the two- and three-dimensional cases, where it is necessary to avoid crossings in one- and two-dimensional settings, respectively, generic perturbations will turn the crossings into avoided crossings because the space of unitary matrices with repeated eigenvalues is of codimension 3. This argument can be made precise using the transversality theorem [27], but the construction is technical and we refrain from doing so here. This allows us to find a continuous logarithm  $L_1(\mathbf{k})$  of a perturbation  $\tilde{U}(\mathbf{k})$  of  $U(\mathbf{k})$  by the procedure above. Then,

$$\widehat{U}(\mathbf{k}) = e^{-iL_1(\mathbf{k})/2} U(\mathbf{k}) e^{-iL_1(\mathbf{k})/2}$$

is very close to  $\text{Id}_J$ , and therefore its eigenvalues never cross  $-1$ . There is therefore no obstruction in finding a logarithm  $L_2(\mathbf{k})$  of  $\widehat{U}(\mathbf{k})$ . Finally,

$$e^{-iL_2(\mathbf{k})/2} e^{-iL_1(\mathbf{k})/2} U(\mathbf{k}) e^{-iL_1(\mathbf{k})/2} e^{-iL_2(\mathbf{k})/2} = \text{Id}_J.$$

The problem with this approach is that it is not clear how to best implement in practice the perturbation argument because small gaps in  $\tilde{U}(\mathbf{k})$  result in eigenvectors that are continuous but have very large variations, yielding large derivatives for  $L_1(\mathbf{k})$ . Therefore, although we could in principle design a scheme to treat such eigenvalue collisions, it is likely to be unstable on coarse meshes, and would require some parameters to be fine tuned.

We do not discuss such extensions here since, as explained in Sec. VII, we have never encountered any problematic eigenvalue collisions in our numerical tests on two- and three-dimensional systems. We do not know whether there is a topological reason forbidding such crossings, or whether crossings only occur on subspaces of codimension 3 and are therefore generically absent in two- and three-dimensional situations, where the obstruction matrices depend on 1 and 2 parameters, respectively.

## VII. NUMERICAL RESULTS

This section presents numerical tests of the method proposed here, as well as a comparison with the projection method. We first detail our methodology. We then test the algorithms on two-dimensional toy models, three-dimensional semiconductors with effective potentials, and density functional theory (DFT) computations. Results on topological insulators are presented in an appendix.

### A. Spatial discretization

We present in this section numerical experiments illustrating our algorithm. We solve the periodic Schrödinger equation (1) using a Galerkin basis consisting of all plane waves

$$e_{\mathbf{K}}(\mathbf{r}) = e^{i\mathbf{K} \cdot \mathbf{r}},$$

for  $\mathbf{K} \in \mathcal{R}^*$  such that  $|\mathbf{K}|^2 \leq E_c$ , for some fixed energy cutoff  $E_c$ . Note that the basis we consider does not depend on the value of  $\mathbf{k} \in \mathcal{B}_{\text{red}}$ , in contrast to most plane-wave codes, which use the  $\mathbf{k}$ -dependent cutoff condition  $|\mathbf{k} + \mathbf{K}|^2 \leq E_c$ . This is done for convenience in our rudimentary implementation. This has a number of undesirable properties. For instance, the equality

$$H(\mathbf{k} + \mathbf{K}) = \tau_{\mathbf{K}} H(\mathbf{k}) \tau_{-\mathbf{K}}$$

is only approximately valid, in the regime where  $\mathbf{k}$  and  $\mathbf{K}$  are not too large. In addition,  $\tau_{\mathbf{K}}$  is not unitary when restricted to the vector space spanned by the Galerkin basis. These properties are, however, recovered in the limit when  $E_c \rightarrow +\infty$ . In our tests, we take  $E_c$  large enough to avoid these problems.

### B. Two-dimensional case

As a first test, we use an artificial potential with randomly chosen coefficients, with unit cell  $Y = [-\frac{1}{2}, \frac{1}{2}]^2$ . More

precisely,

$$V(\mathbf{r}) = \sum_{(j,k) \in \mathbb{Z}^2} \widehat{V}_{j,k} e^{-2i\pi(j,k) \cdot \mathbf{r}} + \text{c.c.}, \quad (23)$$

where the only nonzero coefficients  $\widehat{V}_{j,k}$  were selected randomly. The results we present here correspond to the following set of coefficients:

$$\begin{aligned} \widehat{V}_{(0,0)} &= 0, \quad \widehat{V}_{(0,1)} = 5.50 + 10.37i, \quad \widehat{V}_{(0,2)} \\ &= -12.50 + 11.52i, \quad \widehat{V}_{(1,-1)} = 11.70 + 10.21i, \\ \widehat{V}_{(1,-1)} &= 11.70 + 10.21i, \quad \widehat{V}_{(1,0)} = -10.19 + 10.04i, \\ \widehat{V}_{(1,1)} &= -7.85 + 1.97i, \quad \widehat{V}_{(2,0)} = -2.02 + 4.61i. \end{aligned}$$

With this potential, the first five bands are isolated from the higher-energy ones, so we choose  $J = 5$ .

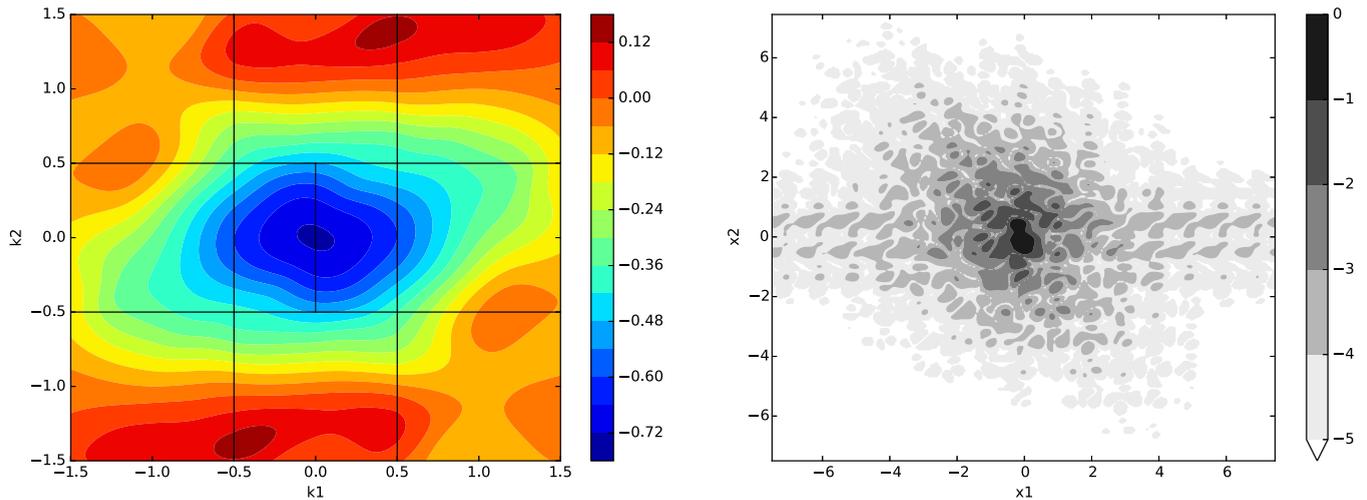
We note that this system is entirely unphysical, and was chosen to be a challenging test for our algorithm. On this system, it is hard to get good initialization for the projection

method because the centers and shapes of the Wannier functions are *a priori* unknown.

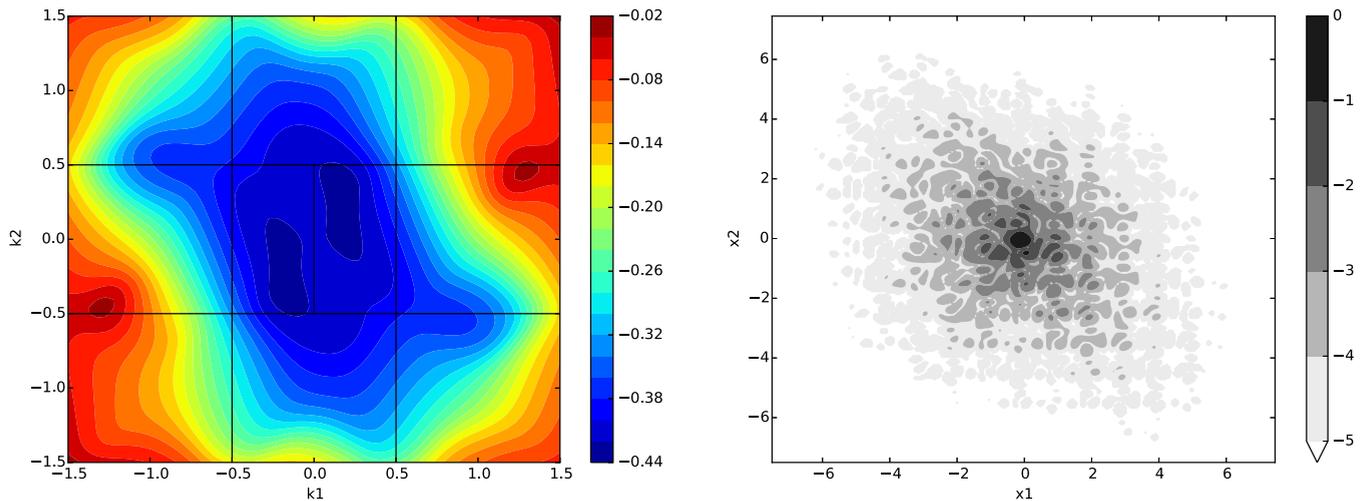
### 1. Results obtained with the proposed algorithm

The result of our algorithm can be seen in Fig. 5, where we represent the first component  $u_{1,\mathbf{k}}$  of our Bloch frame as a function of  $\mathbf{k}$ , using the same methodology as in Fig. 2. We also plot the associated Wannier function  $w_1$  to check its spatial decay. On this very simple example, there is no ambiguity in finding the logarithm on the right edge because the eigenvalues of the obstruction matrix never cross the value  $-1$ , as can be seen in Fig. 6. We therefore obtain a continuous frame, which leads to localized Wannier functions. However, there are slight discontinuities in the first derivative at the edges  $k_1 = \frac{1}{2} + p$  for  $p \in \mathbb{Z}$ . This translates into a slow spatial decay of the Wannier function in the  $x_1$  direction.

To smooth the discontinuities of  $\partial_{k_1} u_{n,\mathbf{k}}$ , we use the MV procedure. The output of this minimization is given in Fig. 5(b).



(a) With our algorithm, before MV minimization.



(b) With our algorithm, after MV minimization.

FIG. 5. Result of our algorithm for the model specified by (23). As in Fig. 2, we plot the average real part of the first component  $u_{1,\mathbf{k}}$  of the Bloch frame with  $J = 5$  as a function of  $\mathbf{k}$  (left) and corresponding Wannier function (right).

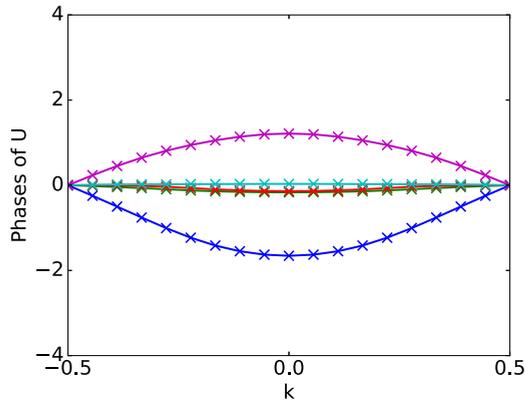


FIG. 6. Phase of the obstruction matrix.

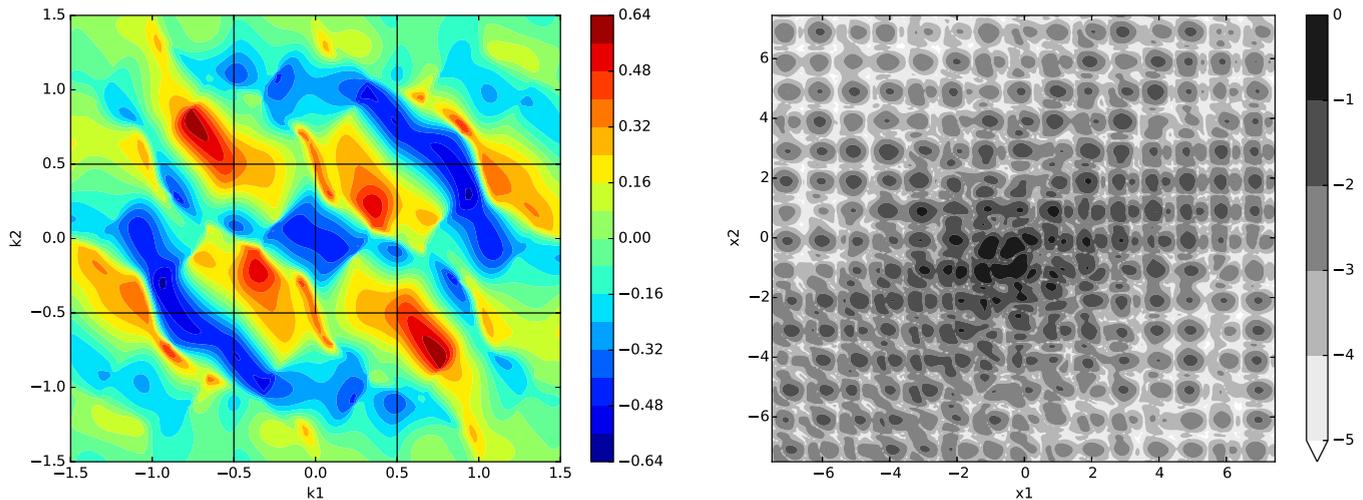
On this simple example, the main features of the frame do not change, but some extra regularization occurs. This removes the spatial tail of the Wannier function in the  $x_1$  direction.

We checked that the Wannier function we obtain presents an exponential decay, provided the cutoff energy  $E_c$  and the number of iterations of the MV algorithm are sufficiently large.

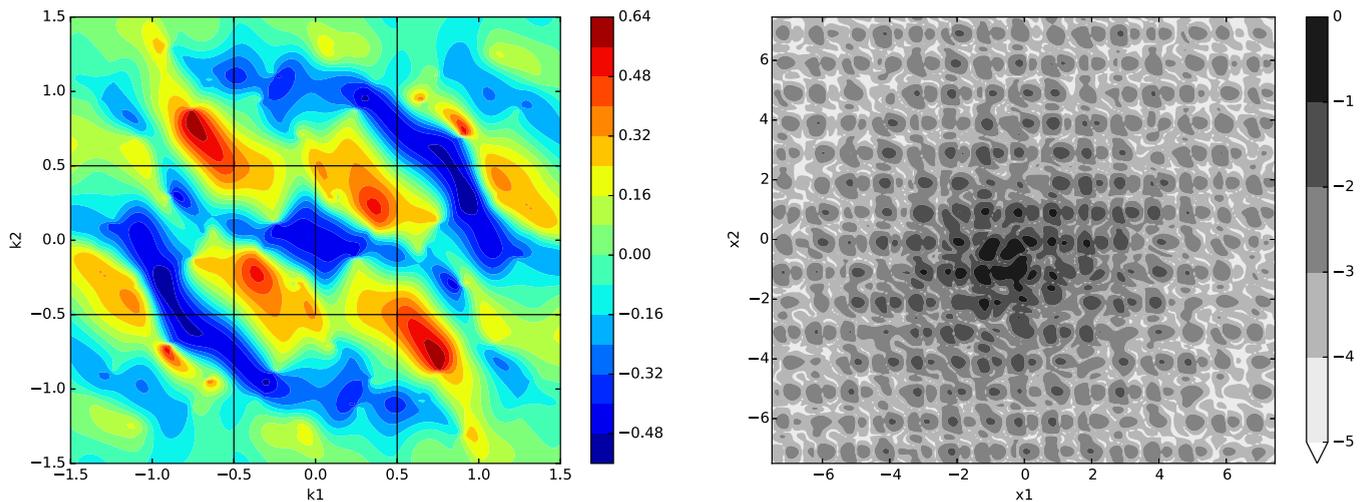
### 2. Comparison with the projection method

We next compare the results of our algorithm to those obtained from the projection method commonly used to initialize the MV algorithm [15]. The initial guesses for the projection method are  $J$  Gaussian functions centered at random positions. The width of these Gaussian functions is chosen to match approximately the size of the localized Wannier functions we obtained with our method. Figure 7 represents the Bloch frame obtained with one example of such a projection, before and after running the MV minimization.

Although the frame is smooth in most regions, and obeys by construction the symmetry properties, we can see isolated points of discontinuity, which visually manifest themselves as rapid changes of color. Moreover, the MV minimization



(a) With the projection method, before MV minimization.



(b) With the projection method, after MV minimization.

FIG. 7. Result of the projection method with an initial guess corresponding to Gaussian functions centered at random positions, for the same model and methodology as in Fig. 5.

procedure fails to further localize the initial guess, a situation the authors of [15] refer to as “false local minima.”

The behavior around the isolated points of singularity can be understood as follows. We denote by  $v_{\mathbf{k}}$  the periodic parts of the Bloch transform of the trial Wannier functions used as an initial guess. When  $J = 1$ , according to the projection method, the corresponding frame is given by

$$u_{\mathbf{k}} = \frac{P(\mathbf{k})v_{\mathbf{k}}}{\|P(\mathbf{k})v_{\mathbf{k}}\|} = \frac{\langle w_{\mathbf{k}}, v_{\mathbf{k}} \rangle w_{\mathbf{k}}}{|\langle w_{\mathbf{k}}, v_{\mathbf{k}} \rangle|} = e^{i\theta(\mathbf{k})} w_{\mathbf{k}},$$

where  $w_{\mathbf{k}}$  is any vector in the one-dimensional vector space  $\text{Ran } P(\mathbf{k})$ , which can be chosen locally continuous, and  $\theta(\mathbf{k})$  is the phase of  $\langle w_{\mathbf{k}}, v_{\mathbf{k}} \rangle$ . This phase is well defined and continuous as long as  $\langle w_{\mathbf{k}}, v_{\mathbf{k}} \rangle$  is nonzero. The condition  $\langle w_{\mathbf{k}}, v_{\mathbf{k}} \rangle = 0$  is a set of two real equations, which generically admits point solutions in dimension 2 and solutions on a one-dimensional manifold in dimension 3. The frame  $u$  is singular at these values. The analysis when  $J > 1$  is more involved but yields the same conclusions: the frame is singular when the overlap matrix  $[P(\mathbf{k})v_{\mathbf{k}}]^*(P(\mathbf{k})v_{\mathbf{k}})$  is not invertible, which generically is a set of two equations. By formal analogy with similar phenomena in other application fields, we call such singular sets *vortices*. Vortices are points in dimension 2 and lines in dimension 3. A more quantitative information about the behavior of frames around vortices can be obtained by introducing the notion of eigenspace vorticity [28], which generalizes the pseudospin winding number appearing in the literature on graphene [29].

As can be seen in Fig. 7(b), the MV minimization fails to resolve these vortices: on a finite mesh, these pathological singularities appear as local minima of the functional. This is because the MV algorithm was designed with a continuous input in mind. The continuity ensures that, when the mesh is fine enough, the matrix elements  $M_{mn,\mathbf{k},\mathbf{b}}$ , defined in (26) and used crucially in the algorithm are close to 1. This allows to define their logarithm unambiguously. In the presence of vortices, this quantity is not close to 1, even on fine meshes, and the algorithm is not well defined.

In order to confirm that the undesirable behavior of the MV algorithm arises from the existence of initial vortices, we consider the case of a single band. The dominant contribution to the MV functional is given by [16]

$$\tilde{\Omega}[u] = \int_B \int_Y |\nabla_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r})|^2 d\mathbf{r} d\mathbf{k}. \quad (24)$$

Consider a model vortex at  $\mathbf{k} = 0$  (without loss of generality), and let  $\mathbf{k} = (\rho \cos \theta, \rho \sin \theta)$ . We consider the model vortex  $u_{\mathbf{k}} = e^{i\theta} u_0$  in a neighborhood of 0. Then  $\nabla_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r})$  has a  $1/\rho$  singularity at 0. Therefore, on a  $N \times N$  grid, we can expect that the MV functional (24) diverges like  $\ln(N)$ . In dimension 3, where vortices are lines,  $\nabla_{\mathbf{k}} u_{\mathbf{k}}$  also presents a  $1/\rho$  divergence, where  $\rho$  is the distance from  $\mathbf{k}$  to the line of singularities. This also yields a logarithmic divergence of the functional. The reasoning is similar for  $J > 1$ .

Such a divergence is observed in our simulations: for the simple case presented here, the values of the MV functional at convergence of the algorithm on a  $20 \times 20$ ,  $40 \times 40$ ,  $80 \times 80$ , and  $160 \times 160$  grid, respectively, are 3.7, 5.7, 7.9, and 9.6. This is consistent with a logarithmic divergence. However, since the divergence is relatively mild (logarithmic only), when the mesh

is sufficiently coarse, these vortices are found not to impact the algorithm, which converges to nonsingular minima. For instance, the MV algorithm on the example considered here converges to a smooth minimum on a  $7 \times 7$  grid in  $\mathbf{k}$  space, but stalls without removing the vortices on  $10 \times 10$  grids and finer.

This explains the observation in [15, Sec. IV D 2], where the MV minimization is reported to sometimes fail for random initial guesses. The authors state that “this problem (false local minima) is not associated with the presence of a large number of bands, but instead with the use of fine  $\mathbf{k}$ -point meshes,” consistent with our observation that the energy contribution of a vortex diverges logarithmically with the grid spacing. They also “never observed the system to become trapped in a false local minimum when starting from reasonable trial projection functions.” In the situation we consider here (a random potential), “reasonable trial projection functions” are hard to devise in advance. We have found vortices to occur for a large class of initial inputs, although of course not for ones close to real maximally localized Wannier functions, where the overlap matrix  $[P(\mathbf{k})v_{\mathbf{k}}]^*[P(\mathbf{k})v_{\mathbf{k}}]$  has its eigenvalues bounded away from zero.

By contrast, our algorithm always succeeded in constructing a good frame, even on more complicated systems where the eigenvalues of the obstruction matrix cross the value  $-1$ , as in Fig. 3. In order to investigate whether the type of problematic eigenvalue collisions of the obstruction matrix described in Sec. VI was possible, we played with the Fourier coefficients  $\widehat{V}$ , but were not able to see any. In fact, the only way we could make the eigenvalues of the obstruction matrix collide is by closing the gap  $\inf_{\mathbf{k} \in B} [\varepsilon_{J+1}(\mathbf{k}) - \varepsilon_J(\mathbf{k})]$ , which violates the hypotheses of our problem (we only consider isolated bands). We do not know if it is fundamentally impossible for eigenvalues of the obstruction matrix to cross at a finite gap (for topological reasons, for instance), or if it is simply an exceptional situation which we failed to encounter in our tests.

### C. Three-dimensional case

The previous two-dimensional example with a random potential was artificial and not representative of real systems. To test our algorithm on more realistic cases, we use the simple effective one-body potentials of [30] in a zinc-blende structure, discretized on a plane-wave basis. These pseudopotentials were chosen to obtain representative band structures at minimal cost. In this framework, the potential  $V$  reads as

$$V(\mathbf{r}) = \sum_{\mathbf{K} \in \mathcal{R}^*} \widehat{V}(\mathbf{K}) e^{-i\mathbf{K} \cdot \mathbf{r}},$$

$$\widehat{V}(\mathbf{K}) = V_S(|\mathbf{K}|) \cos(\mathbf{K} \cdot \boldsymbol{\tau}) + i V_A(|\mathbf{K}|) \sin(\mathbf{K} \cdot \boldsymbol{\tau}),$$

where  $\boldsymbol{\tau} = a(1, 1, 1)/8$  and  $a$  is the lattice constant of the zinc-blende structure. The form factors  $V_S$  and  $V_A$ , as well as the parameters  $a$  for various compounds, are tabulated in [30].

We first test our algorithm on silicon. We study the first four bands, which are isolated from the others. As in the 2D case, our algorithm was able to produce a continuous frame, as seen in Fig. 8, where we plot again the average real part of the first component of the Bloch frame, on the cut plane  $k_2 = 0$ .

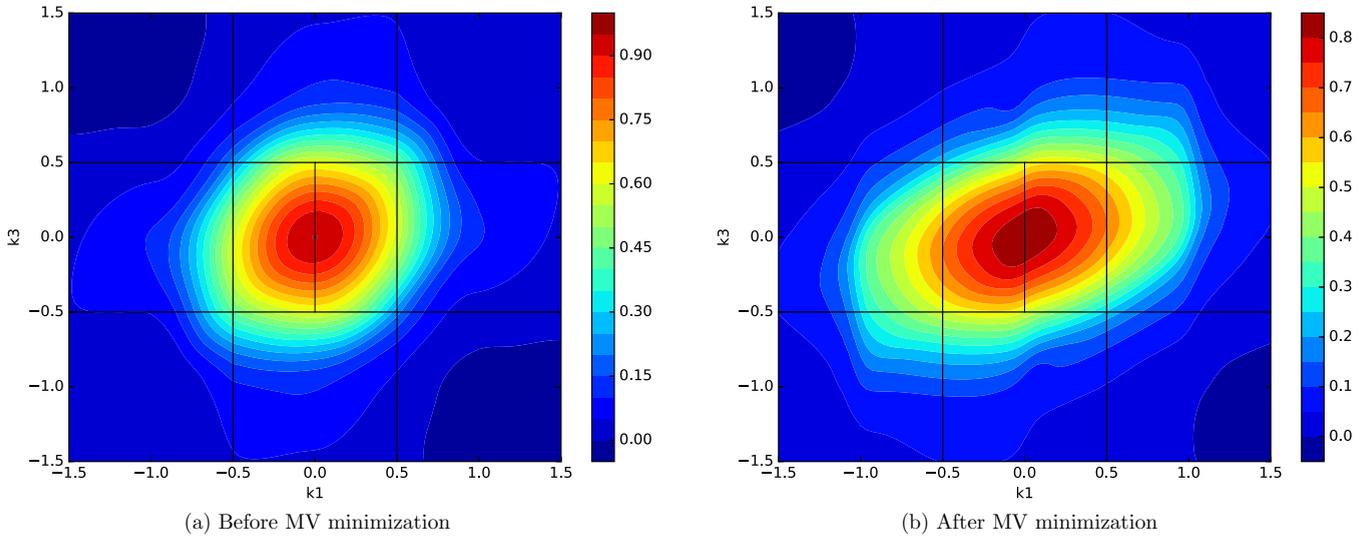


FIG. 8. Average real part of the first component of the Bloch frame obtained by our method on the cut plane  $k_2 = 0$ , before (left) and after (right) the MV algorithm, on a  $24 \times 24 \times 24$  grid.

Figure 9 represents the  $J = 4$  different phases of the eigenvalues of the obstruction matrix  $U_{\text{obs}}(k_2, k_3)$  in the three stages of the algorithm: fixing the corners, the edges, and the face. Before fixing the corners, there is a conical intersection in the eigenvalues at  $(0, 0)$ . This is a nongeneric case, presumably due to eigenvalue degeneracy at  $\Gamma$  in the model, itself related

to a particular symmetry of the potential. However, this intersection is harmless, as it connects eigenvalues with the same phase determination. After fixing the corners, the conical intersection moves to the corners, and the phase of all the eigenvalues has a maximum of about 0.6, well below  $\pi$ . Fixing the edges is not necessary in this case.

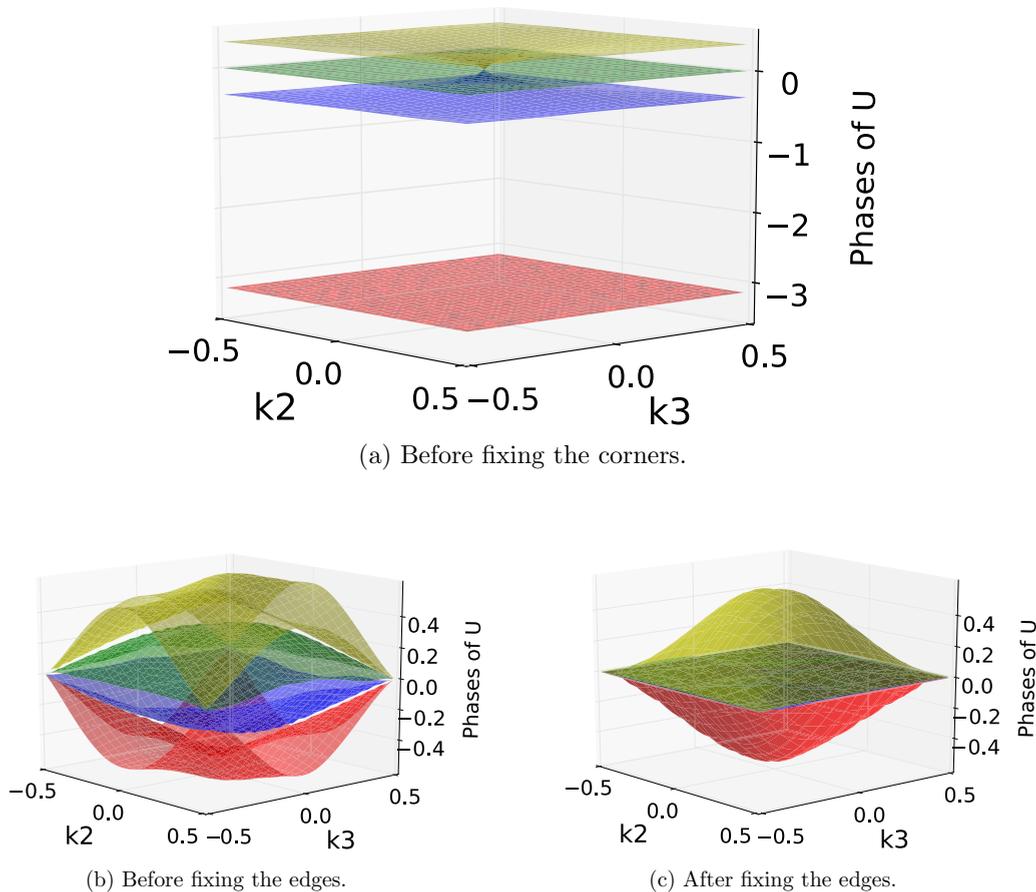


FIG. 9. Phase of the obstruction matrix for the silicon example.

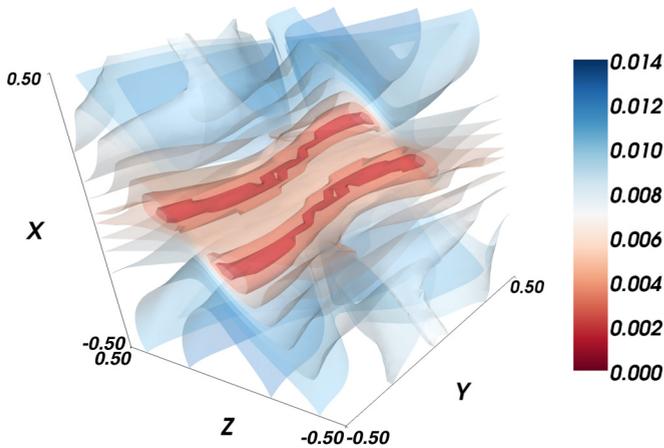
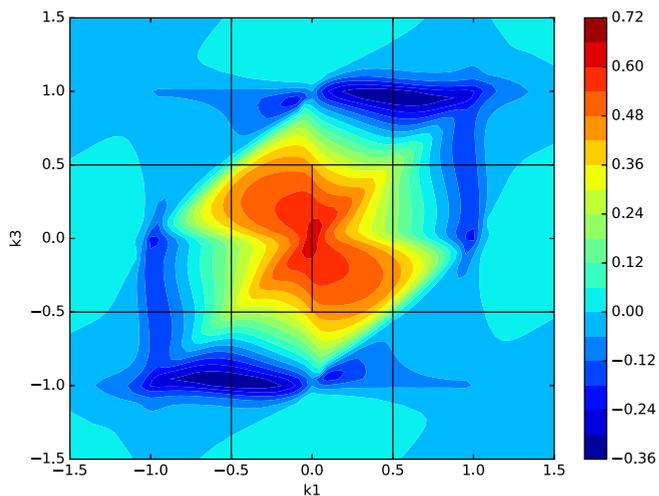


FIG. 10. Smallest eigenvalue of the overlap matrix (25). The lowest isosurface corresponds to the region where this matrix is nearly indefinite, and forms a vortex line.

To compare our method to the projection method, as in the 2D case, we use as initial Wannier functions Gaussians centered on random points. The projection method yields a continuous frame as long as the overlap matrix

$$O(\mathbf{k}) = [P(\mathbf{k})v_{\mathbf{k}}]^*[P(\mathbf{k})v_{\mathbf{k}}] \quad (25)$$

is positive definite for all  $\mathbf{k} \in \mathcal{B}$ . As can be seen in Fig. 10, this matrix becomes indefinite on vortex lines. This yields discontinuous frames, as illustrated in Fig. 11(a). In our tests with random positions of the Gaussians, this occurred about half the time. As in the 2D case, the MV algorithm is unable to remove these vortices when the sampling of the Brillouin zone is fine enough. In this example, the MV algorithm was able to remove the vortices on grids of size up to  $20 \times 20 \times 20$ , but not on finer grids. As expected, when the Wannier centers are selected appropriately, the MV algorithm converges to localized Wannier functions.



(a) Before MV minimization

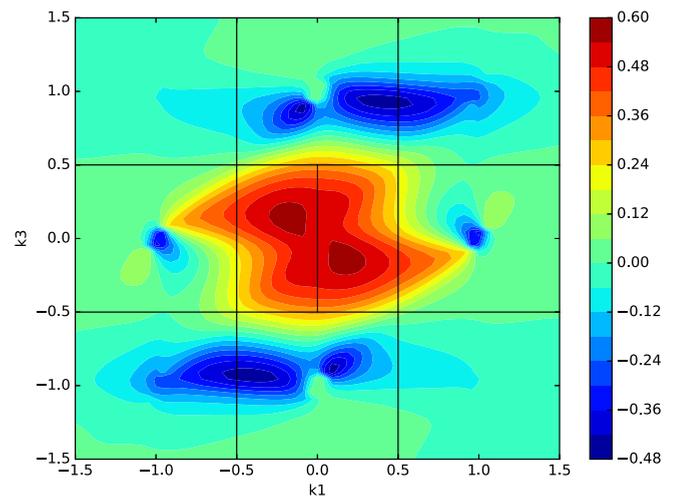
Silicon represents an easy test case because it possesses a relatively large direct gap between the valence and conduction bands. This makes the projector on the valence bands a smooth function of  $\mathbf{k}$ , which in turn produces well-localized Wannier functions. We also tried our algorithm on more complicated semiconductors such as indium arsenide, still using the pseudopotentials of [30]. Indium arsenide has a very small direct gap, resulting in sharp variations in the Brillouin zone near the band edges which require a fine sampling of the Brillouin zone to resolve accurately the band structure. Although the obstruction phases of Fig. 9 are more rugged in this case, their amplitude at the last step is still less than about 0.6, well below  $\pi$ , and our algorithm has no problem distinguishing the bands to fix the phases.

#### D. Interface with Wannier90 and tests on DFT systems

We have implemented our algorithm in a way that is compatible with the standard code WANNIER90 [17,31]. To that end, we note that our method, while presented here with frames, can also be implemented in the WANNIER90paradigm where one computes a fixed set of electronic orbitals  $u_{n\mathbf{k}}$  for each  $\mathbf{k}$  point, and then finds a set  $U_{\mathbf{k}}$  of unitaries from which the final frame is constructed as  $u'_{\mathbf{k}} = u_{\mathbf{k}}U_{\mathbf{k}}$ . The advantage of this approach for the MV algorithm is that only the low-dimensional unknowns  $U_{\mathbf{k}}$  have to be optimized, and that the only inputs from electronic structure codes are the overlap matrices

$$M_{nm,\mathbf{k},\mathbf{b}} = \langle u_{n\mathbf{k}+\mathbf{b}}, u_{m\mathbf{k}} \rangle, \quad (26)$$

where  $\mathbf{b}$  runs over nearest neighbors of 0 on the  $\mathbf{k}$ -space mesh. This separates the computation of the  $u_{n,\mathbf{k}}$  from the computation of Wannier functions, and facilitates the creation of independent libraries. Similarly to the traditional MV minimization, in our algorithm we only need as input the overlap matrices  $M_{nm,\mathbf{k},\mathbf{b}}$  (.mmn file) to compute the final frame  $u'_{\mathbf{k}} = u_{\mathbf{k}}U_{\mathbf{k}}$  and output the unitary matrix  $U_{\mathbf{k}}$  at each  $\mathbf{k}$  point (.amn file). Because this format does not explicitly



(b) After MV minimization

FIG. 11. Average real part of the first component of the Bloch frame obtained by the projection method on the cut plane  $k_2 = 0$ , before (left) and after (right) the MV algorithm, on a  $24 \times 24 \times 24$  grid. The MV algorithm is able to smooth out the general features of the frame, but not to remove the vortices.

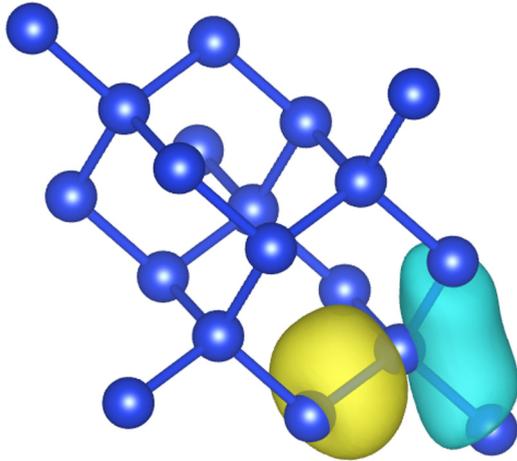


FIG. 12. One of the four Wannier functions produced by our algorithm for silicon. The positive and negative isovalues of the Wannier function are plotted, at a level equal to 20% of the maximal absolute value of the function. The Wannier functions obtained by our algorithm are localized, and appear to be close to combinations of the four bond-centered maximally localized Wannier functions.

account for the time-reversal symmetry, it is more convenient to use a modified version of our algorithm (see Appendix), which does not use the time-reversal symmetry. The resulting code is available at <https://github.com/antoine-levitt/wannier> and can readily be inserted in any workflow using WANNIER90.

We have tested our algorithm on bulk silicon computed by the code QUANTUM ESPRESSO [32] using the PBE exchange-correlation functional. Our conclusions are the same as those of the previous section using an effective potential approach. The phases of the obstruction matrix are similar to those of Fig. 9, and our algorithm produces a continuous frame. The corresponding Wannier functions are localized, as can be seen on Fig. 12.

These Wannier functions are not maximally localized, but form a good initial guess for the MV algorithm. By

contrast, initializing the algorithm with randomly centered *s*-type orbitals (as is the default in WANNIER90 in the absence of a specific prescription) might yield an algorithm that can converge to a “false local minimum,” especially on fine meshes, as can be seen in Fig. 13. However, when the initial guess is good enough, the convergence is satisfactory, and even faster than using our algorithm to produce an initial guess.

The computational time of our algorithm is dominated by the cost of a few operations on  $J \times J$  matrices at each  $\mathbf{k}$  point, and is therefore comparable to that of one iteration of the MV algorithm. By contrast, the projection method requires computations on the plane-wave grid, and is therefore much more costly. Therefore, the total time for the computation of MLWFs is lower with our algorithm, even when it requires more iterations to converge than the projection method with good initial guesses.

### VIII. CONCLUSION AND PERSPECTIVES

We proposed an algorithm to obtain well-localized Wannier functions without any initial guess or free parameters, and presented numerical results showing its correctness, as well as its superiority compared to the projection method on very fine meshes and when good initial guesses are not available. The flipside to this is that our method, which does not utilize the physics of the system, might not yield optimal Wannier functions, but only local minima. We anticipate our method to be useful for systems where no physical intuition is available, and where well-localized Wannier functions, even if not optimal or physically sensible, can be used for example for Wannier interpolation [2].

Our algorithm can fail when eigenvalues of the obstruction matrix corresponding to different number of turns  $p$  collide. We have never found this to be the case in practice. However, it is unclear whether such a phenomenon is impossible because of topological reasons or if it is simply rare and we have never found it in our tests. A better understanding of this issue is

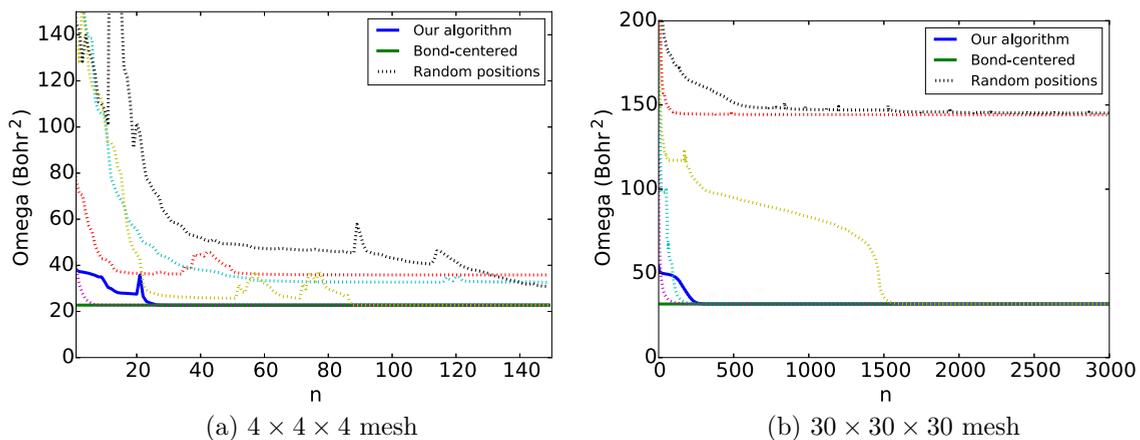


FIG. 13. Spread as a function of the number of iterations, for a coarse and a fine  $\mathbf{k}$ -point mesh. We use as initial guess (1) the one produced by our algorithm; (2) the projection method with bond-centered *s*-type orbitals; (3) the projection method with *s*-type orbitals with random centers (five different realizations). On the left panel, the black and light blue curves eventually converged to the global minimum after about 500 iterations, but the red one converged to a local minimum to machine precision. The red and black curves on the right panel seem to have converged to a local minimum, but the gradient is nonzero and the energy keeps decreasing, although very slowly; after 100 000 iterations, the energy was still above 136 bohrs<sup>2</sup>.

a worthwhile direction of research. Our algorithm could be adapted to tackle such collisions, but would be significantly more complicated, introduce free parameters, and be less robust for coarse  $\mathbf{k}$ -point meshes, which is why we refrain from doing so in this work.

As we demonstrated, a singular input to the MV algorithm may or may not yield a physically relevant answer, depending on the size of the mesh: singularities may not be seen on a coarse mesh, while fine meshes emphasize the divergent contribution of the vortices to the functional. In this case, the MV algorithm stalls and is unable to converge to a “true” local minimum. The behavior of the MV algorithm in this case would be interesting to study from a numerical analysis point of view. Another important question is whether any continuous frame will yield exponentially localized Wannier functions when used as initial guess to the MV algorithm.

On the numerical side, our algorithm, publicly available at <https://github.com/antoine-levitt/wannier>, can readily be interfaced in standard workflows using the WANNIER90 [17,31] code. Our construction preserves the time-reversal symmetry, but not any other symmetries the crystal might possess, unlike the construction in [33]. The extension of our method to additional symmetries is also an interesting direction for future work.

**ACKNOWLEDGMENT**

We wish to thank the anonymous referees for useful suggestions and comments.

**APPENDIX: SYSTEMS WITHOUT TIME-REVERSAL SYMMETRY AND TOPOLOGICAL INSULATORS**

Our algorithm was introduced in the framework of time-reversal-symmetric systems, and more precisely of a symmetry of *bosonic* type, which is one in which the antiunitary time-reversal operator  $C$  squares to 1. This is the case where we know theoretically that there exist localized Wannier functions. Although a detailed study is outside the scope of this paper, it is interesting to explore what happens when this assumption fails, and in particular investigate the case of topological insulators. Our algorithm can be adapted very simply to systems without time-reversal symmetry (or with a *fermionic* one, that squares

to  $-1$ ), and we explain it here for one- and two-dimensional systems (the three-dimensional case is a simple extension).

In 1D, we first start with a frame at  $\Gamma$ , which we do not impose to be real. Then, we propagate it using (9) to the segments  $[0, \frac{1}{2}]$  and  $[-\frac{1}{2}, 0]$ . At this point, generically  $u_{1/2} \neq \tau_1 u_{-1/2}$ : we compute the unitary obstruction matrix  $U_{\text{obs}} = u_{1/2}^* (\tau_1 u_{-1/2})$ , and set  $u'_k = u_k U_{\text{obs}}^k$ , which is now a continuous frame.

In 2D, we apply the same construction as in the 1D case to build a continuous frame on the segment from  $(0, -\frac{1}{2})$  to  $(0, \frac{1}{2})$  that satisfies  $u_{(0,1/2)} = \tau_2 u_{(0,-1/2)}$ . Then, we propagate it horizontally and obtain a frame  $u_{(k_1,k_2)}$  on  $[-\frac{1}{2}, \frac{1}{2}]^2$ . Generically,  $u_{(1/2,k_2)} \neq \tau_1 u_{(-1/2,k_2)}$ , which we now fix in two steps. We define the top edge obstruction matrix  $U_{\text{obs,top}} = u_{(1/2,1/2)}^* (\tau_1 u_{(-1/2,1/2)})$ , set  $u'_{(k_1,k_2)} = u_{(k_1,k_2)} U_{\text{obs,top}}^{k_1}$ , and drop the primes for simplicity. We now define the obstruction matrix  $U_{\text{obs}}(k) = u_{(1/2,k)}^* (\tau_1 u_{(-1/2,k)})$ , which, like in the time-reversal-symmetric case, satisfies  $U_{\text{obs}}(-\frac{1}{2}) = U_{\text{obs}}(\frac{1}{2}) = \text{Id}_J$ . Provided we can find a continuous logarithm  $L(k)$  of  $U_{\text{obs}}(k)$ , we set  $u'_{(k_1,k_2)} = u_{(k_1,k_2)} e^{ik_1 L(k_2)}$  and obtain a continuous frame.

Note that, as this algorithm always produces a continuous frame if the logarithm problem is solved, it follows that the logarithm problem cannot be solved for materials with nonzero Chern numbers, where we know that there cannot exist a continuous frame [14]. We will now show how exactly the existence of a logarithm fails on a Chern insulator.

**Chern insulators: The Haldane model**

We test our algorithm on the prototype of Chern insulators, the two-band Haldane model. We use the same notation as in the original paper [25]: the parameters  $t_1$  and  $t_2$  are the nearest- and next-nearest-neighbors hopping terms,  $\phi$  is the phase, which breaks the time-reversal symmetry, and  $M$  is the onsite energy, which breaks the inversion symmetry. We used as parameters  $t_1 = t_2 = 1$  and  $M = 0.1$ , and consider the first band ( $J = 1$ ). Note that since the Haldane model only has one band, it is not a good test case to see the collision of eigenvalues we are concerned about in Sec. VI, but rather illustrates how our algorithm fails on a topological insulator with a nonzero Chern number, as any algorithm must, since continuous frames cannot exist in this case [13].

The Haldane model is a time-reversal-symmetric insulator for  $\phi = 0$ . When  $\phi$  is increased and the time-reversal symmetry

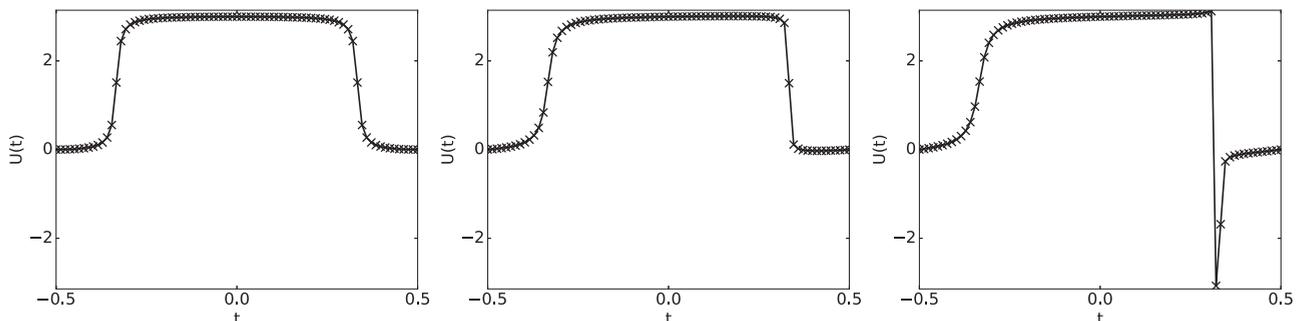


FIG. 14. Phase of  $U_{\text{obs}}(k_2)$  as a function of  $k_2$  for the Haldane model with  $\phi_c \approx 0.019$ , for  $\phi = 0$  (left),  $\phi = \phi_c - 0.005$  (middle), and  $\phi = \phi_c + 0.005$  (right). As  $\phi$  increases, the time-reversal symmetry breaks down and, at  $\phi_c$ , the gap closes and  $U$  jumps abruptly from  $-1$  to  $1$ . When  $\phi_c$  is increased into the Chern insulator regime, the map  $U$  has degree 1, and no choice of branch cut can make its phase continuous and periodic.

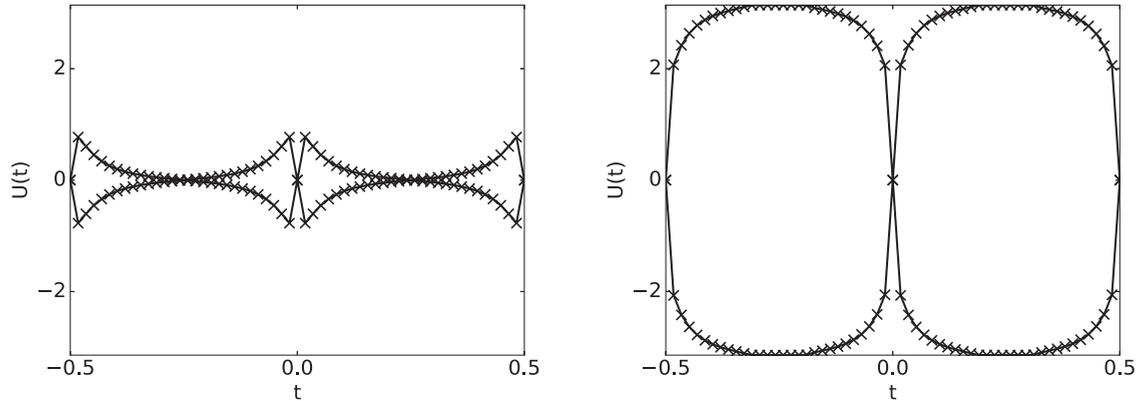


FIG. 15. Phase of  $U_{\text{obs}}(k_2)$  as a function of  $k_2$  for the Kane-Mele model, for  $\lambda_{v,c} = 3\sqrt{3}$ , with  $\lambda_v = \lambda_{v,c} + 0.02$  (left) and  $\lambda_{v,c} = \lambda_{v,c} - 0.02$  (right). As  $\lambda_{v,c}$  decreases, the system transitions from an even to an odd  $\mathbb{Z}_2$  invariant, and eigenvalues with different numbers of turns collide.

is broken, the band gap decreases until it closes at  $\phi_c = \arcsin[M/(3\sqrt{3}t_2)] \approx 0.019$ . After that point, it is a Chern insulator (with Chern number 1).

As can be seen in Fig. 14, when  $\phi$  is small, the phase of the obstruction matrix (here, a single number) evolves smoothly and symmetrically with respect to the origin:  $U_{\text{obs}}$  evolves in the upper half-circle  $|U_{\text{obs}}| = 1$ ,  $\text{Im } U_{\text{obs}} \geq 0$ , and goes from 1 to  $-1$  and back again, taking the same path in reverse. When  $\phi$  increases, the transition from  $-1$  back to 1 gets sharper, until it is discontinuous at the critical threshold  $\phi_c$ . When  $\phi$  increases again, this discontinuity is resolved, but this time  $U_{\text{obs}}$  goes through the lower half-circle on its way back. The net result is that a logarithm of  $U_{\text{obs}}(k_2)$  will pick up a phase factor of  $2\pi$  when going from  $k_2 = -\frac{1}{2}$  to  $\frac{1}{2}$ , and cannot therefore be continuous and periodic.

### $\mathbb{Z}_2$ topological insulators: The Kane-Mele model

We now turn to the case of  $\mathbb{Z}_2$  topological insulators [34].  $\mathbb{Z}_2$  topological insulators are characterized by a *fermionic*

time-reversal symmetry, squaring to  $-1$  instead of 1 as the one considered in this paper, and possess a  $\mathbb{Z}_2$  topological invariant. Systems with an odd invariant have a topological obstruction to the construction of frames respecting the time-reversal symmetry, but no obstruction to the construction of nonsymmetric frames [35].

We test our algorithm on the Kane-Mele model [36]. With the same notation as in [36], we choose  $a = 1$ ,  $t = 1$ ,  $\lambda_R = 0$ ,  $\lambda_{SO} = 1$ . For this choice of parameters, the system is in a regular insulator phase for  $\lambda_v > \lambda_{v,c} = 3\sqrt{3}$ , and in a quantum spin Hall phase for  $\lambda_v < \lambda_{v,c}$ .

We show in Fig. 15 that this transition introduces a collision of eigenvalues with different numbers of turns. The correspondence between the eigenvalue collisions and the  $\mathbb{Z}_2$  invariant was recently proved in [37, Proposition 5.7]. Let us stress that this collision occurs because of the nontrivial topological states of systems with *fermionic* time-reversal symmetry. For systems with a *bosonic* form of time reversal (the ones we consider in this paper), we found no such crossings.

- 
- [1] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, Maximally localized Wannier functions: Theory and applications, *Rev. Mod. Phys.* **84**, 1419 (2012).
  - [2] J. Yates, X. Wang, D. Vanderbilt, and I. Souza, Spectral and Fermi surface properties from Wannier interpolation, *Phys. Rev. B* **75**, 195121 (2007).
  - [3] R. Resta and D. Vanderbilt, Theory of polarization: a modern approach, *Physics of Ferroelectrics* (Springer, Berlin, 2007), pp. 31–68.
  - [4] N. A. Spaldin, A beginner's guide to the modern theory of polarization, *J. Solid State Chem.* **195**, 2 (2012).
  - [5] J. des Cloizeaux, Analytical properties of n-dimensional energy bands and Wannier functions, *Phys. Rev.* **135**, A698 (1964).
  - [6] J. des Cloizeaux, Energy bands and projection operators in a crystal: Analytic and asymptotic properties, *Phys. Rev.* **135**, A685 (1964).
  - [7] W. Kohn, Analytic properties of Bloch waves and Wannier functions, *Phys. Rev.* **115**, 809 (1959).
  - [8] G. Nenciu, Existence of the exponentially localised Wannier functions, *Commun. Math. Phys.* **91**, 81 (1983).
  - [9] B. Helffer and J. Sjöstrand, Équation de Schrödinger avec champ magnétique et équation de Harper, *Lect. Notes Phys.* **345**, 118 (1989).
  - [10] E. I. Blount, Formalism of band theory, in *Solid State Physics*, edited by F. Seitz and D. Turnbull (Academic, New York, 1962), pp. 305–373.
  - [11] A. Nenciu and G. Nenciu, Dynamics of Bloch electrons in external electric fields. ii. the existence of Stark-Wannier ladder resonances, *J. Phys. A: Math. Gen.* **15**, 3313 (1982).
  - [12] G. Nenciu, Dynamics of band electrons in electric and magnetic fields: Rigorous justification of the effective hamiltonians, *Rev. Mod. Phys.* **63**, 91 (1991).
  - [13] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, Exponential Localization of Wannier Functions in Insulators, *Phys. Rev. Lett.* **98**, 046402 (2007).
  - [14] G. Panati, Triviality of Bloch and Bloch-Dirac bundles, *Ann. Henri Poincaré* **8**, 995 (2007).

- [15] N. Marzari and D. Vanderbilt, Maximally localized generalized Wannier functions for composite energy bands, *Phys. Rev. B* **56**, 12847 (1997).
- [16] G. Panati and A. Pisante, Bloch bundles, Marzari-Vanderbilt functional and maximally localized Wannier functions, *Commun. Math. Phys.* **322**, 835 (2013).
- [17] A. Mostofi, J. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, wannier90: A tool for obtaining maximally-localised Wannier functions, *Comput. Phys. Commun.* **178**, 685 (2008).
- [18] F. Corsetti, On the properties of point defects in silicon nanostructures from ab initio calculations, Ph.D. thesis, Imperial College London, 2012.
- [19] J. I. Mustafa, S. Coh, M. L. Cohen, and S. G. Louie, Automated construction of maximally localized Wannier functions: Optimized projection functions method, *Phys. Rev. B* **92**, 165134 (2015).
- [20] A. Damle, L. Lin, and L. Ying, Compressed representation of Kohn–Sham orbitals via selected columns of the density matrix, *J. Chem. Theory Comput.* **11**, 1463 (2015).
- [21] A. Damle, L. Lin, and L. Ying, SCDM-k: Localized orbitals for solids via selected columns of the density matrix, [arXiv:1507.03354](https://arxiv.org/abs/1507.03354).
- [22] H. D. Cornean, I. Herbst, and G. Nenciu, On the construction of composite Wannier functions, *Ann. H. Poincaré* **17**, 3361 (2016).
- [23] D. Fiorenza, D. Monaco, and G. Panati, Construction of real-valued localized composite Wannier functions for insulators, *Ann. Henri Poincaré* **17**, 63 (2016).
- [24] B. A. Dubrovin and S. P. Novikov, Ground states of a two-dimensional electron in a periodic magnetic field, *Zh. Eksp. Teor. Fiz.* **79**, 1006 (1980) [*Sov. Phys. JETP* **52**, 511 (1980)].
- [25] F. D. M. Haldane, Model for a Quantum Hall Effect without Landau Levels: Condensed-matter Realization of the “Parity Anomaly”, *Phys. Rev. Lett.* **61**, 2015 (1988).
- [26] D. Monaco, G. Panati, A. Pisante, and S. Teufel (unpublished).
- [27] V. Guillemin and A. Pollack, *Differential Topology*, Vol. 370 (American Mathematical Society, Providence, RI, 2010).
- [28] D. Monaco and G. Panati, Topological invariants of eigenvalue intersections and decrease of Wannier functions in graphene, *J. Stat. Phys.* **155**, 1027 (2014).
- [29] C.-H. Park and N. Marzari, Berry phase and pseudospin winding number in bilayer graphene, *Phys. Rev. B* **84**, 205440 (2011).
- [30] M. L. Cohen and T. K. Bergstresser, Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures, *Phys. Rev.* **141**, 789 (1966).
- [31] A. Mostofi, J. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, An updated version of Wannier90: A tool for obtaining maximally-localised Wannier functions, *Comput. Phys. Commun.* **185**, 2309 (2014).
- [32] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. Chiarotti, M. Cococcioni, I. Dabo *et al.*, Quantum Espresso: a modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [33] R. Sakuma, Symmetry-adapted Wannier functions in the maximal localization procedure, *Phys. Rev. B* **87**, 235109 (2013).
- [34] M. Hasan and C. Kane, Colloquium: topological insulators, *Rev. Mod. Phys.* **82**, 3045 (2010).
- [35] D. Fiorenza, D. Monaco, and G. Panati,  $\mathbb{Z}_2$  invariants of topological insulators as geometric obstructions, *Commun. Math. Phys.* **343**, 1115 (2016).
- [36] C. J. Kane and E. J. Mele,  $\mathbb{Z}_2$  Topological Order and the Quantum Spin Hall Effect, *Phys. Rev. Lett.* **95**, 146802 (2005).
- [37] H. D. Cornean, D. Monaco, and S. Teufel, Wannier functions and  $\mathbb{Z}_2$  invariants in time-reversal symmetric topological insulators, *Rev. Math. Phys.* **29**, 1730001 (2017).