

VARIATIONAL FORMULATION FOR WANNIER FUNCTIONS  
WITH ENTANGLED BAND STRUCTURE\*ANIL DAMLE<sup>†</sup>, ANTOINE LEVITT<sup>‡</sup>, AND LIN LIN<sup>§</sup>

**Abstract.** Wannier functions provide a localized representation of spectral subspaces of periodic Hamiltonians and play an important role for interpreting and accelerating Hartree–Fock and Kohn–Sham density functional theory calculations in quantum physics and chemistry. For systems with isolated band structure, the existence of exponentially localized Wannier functions and numerical algorithms for finding them are well studied. In contrast, for systems with entangled band structure, Wannier functions must be generalized to span a subspace larger than the spectral subspace of interest to achieve favorable spatial locality. In this setting, little is known about the theoretical properties of these Wannier functions, and few algorithms can find them robustly. We develop a variational formulation to compute these generalized maximally localized Wannier functions. When paired with an initial guess based on the selected columns of the density matrix method, our method can robustly find Wannier functions for systems with entangled band structure. We formulate the problem as a constrained nonlinear optimization problem and show how the widely used disentanglement procedure can be interpreted as a splitting method to approximately solve this problem. We demonstrate the performance of our method using real materials including silicon, copper, and aluminum. To examine more precisely the localization properties of Wannier functions, we study the free electron gas in one and two dimensions, where we show that the maximally localized Wannier functions only decay algebraically. We also explain using a one dimensional example how to modify them to obtain superalgebraic decay.

**Key words.** Wannier function, localization, entangled band, metallic system, variational method, optimization, free electron gas

**AMS subject classifications.** 65Z05, 82D25, 65F30, 65K10

**DOI.** 10.1137/18M1167164

**1. Introduction.** Localized representations of electronic wavefunctions have a wide range of applications in quantum physics, chemistry, and materials science. In an effective single particle theory such as Hartree–Fock theory and Kohn–Sham density functional theory (KSDFT) [18, 22], the electronic wavefunctions are characterized by eigenfunctions of single particle Hamiltonian operators. These eigenfunctions generally have significant magnitude in large portions of the computational domain. However, the physically meaningful quantity is not each individual eigenfunction, but the subspace spanned by the collection of a set of eigenfunctions. This is often referred to as the Kohn–Sham subspace, and it is often possible to reduce the computational complexity of various methods by using an alternative, localized representation of the subspace.

---

\*Received by the editors January 26, 2018; accepted for publication (in revised form) October 29, 2018; published electronically January 30, 2019.

<http://www.siam.org/journals/mms/17-1/M116716.html>

**Funding:** The first author was supported by the National Science Foundation under grant DMS-1606277. The third author was partially supported by the National Science Foundation under grant DMS-1652330, the Department of Energy under grants DE-SC0017867 and DE-AC02-05CH11231, and the SciDAC project.

<sup>†</sup>Department of Computer Science, Cornell University, Ithaca, NY 14853 (damle@cornell.edu).

<sup>‡</sup>Inria Paris, F-75589 Paris Cedex 12, Université Paris-Est, CERMICS (ENPC), F-77455 Marne-la-Vallée, France (antoine.levitt@inria.fr).

<sup>§</sup>Department of Mathematics, University of California, Berkeley, Berkeley, CA 94720, and Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 (linlin@math.berkeley.edu).

Wannier functions provide one such localized representation of the Kohn–Sham subspace. They require significantly less memory to store and are the foundation of so-called linear scaling methods [21, 15, 4] for solving quantum problems. They can also be used to analyze chemical bonding in complex materials, interpolate the band structure of crystals, accelerate ground and excited state electronic structure calculations, and form reduced order models for strongly correlated many body systems [25].

Wannier functions are not uniquely determined and depend on a choice of gauge (a rotation among the occupied states), which strongly influences their localization. For periodic systems with an isolated band structure, the localization properties of Wannier functions have been studied extensively [20, 3, 30, 5, 34]. Interestingly, the existence of localized Wannier functions in this case is characterized by a topological invariant. For physical systems without a magnetic field (the Hamiltonian is invariant under time-reversal symmetry), this topological invariant is trivial, and it is known that there exists a gauge leading to Wannier functions with exponential decay [5, 33]. In this setting, efficient numerical algorithms have been developed to compute these exponentially localized functions [26, 19, 16, 11, 32, 9, 29, 6, 10]. In particular, the widely used maximally localized Wannier function (MLWF) procedure minimizes the variance (or “spread”) [13, 26] over all possible choices of gauge to obtain localized Wannier functions. In the insulating case, it is known that minimizers of this spread are exponentially localized [34].

The situation becomes significantly more challenging for systems with entangled band structure. Entangled band structure arises in metallic systems, but also in insulating systems when conduction bands or a selected range of bands are to be localized. A straightforward definition of Wannier functions requires the set of all Wannier functions to exactly span the selected spectral subspace. However, such Wannier functions are known to decay slowly in real space. Therefore, the definition of Wannier functions has been generalized to refer to functions spanning a subspace larger than, but containing, the given entangled spectral subspace, referred to as a “frozen window” [35]. This is useful, for instance, in band interpolation, where the additional Wannier functions give rise to extra bands that can simply be ignored. Finding such generalized Wannier functions numerically is considerably more complex. Furthermore, little is known theoretically about the localization properties of the constructed generalized Wannier functions. In order to be consistent with the terminology in the physics literature, we will refer to these generalized Wannier functions simply as Wannier functions, unless otherwise noted.

In this paper, we develop a variational formulation for finding Wannier functions in the entangled setting. We formulate the problem as a nonlinear constrained optimization problem. Practical Wannier function calculations indicate that such nonlinear optimization problems can have many local minima. Hence the solution can strongly depend on the initial guess, and the difficulty of constructing a good initial guess is often a significant impediment to finding Wannier functions in a robust fashion. In order to avoid being trapped at undesirable local minima, we use the recently developed selected columns of the density matrix (SCDM) methodology to construct the initial guess for our variational formulation. This strategy is applicable to both the isolated case [9] and the entangled case [8].

Our variational formulation can be obtained in several theoretically equivalent constructions. We find that one of these formulations yields the so-called partly occupied Wannier functions [36] and can be solved efficiently using standard numerical algorithms for minimization under orthogonality constraints. Our formulation also reveals that the widely used “disentanglement” procedure [35] can be viewed as a

splitting method for solving the constrained optimization problem, which only performs a single alternation step between the two pieces of the objective function, and therefore does not achieve a global minimum of the spread. We verify the performance of the variational formulation with real materials such as silicon, copper, and aluminum. In these examples, we find that the fully converged variational formulation consistently provides orbitals with a smaller spread than that from the disentanglement procedure and is more robust to the choice of initial guess. We also find that the difference between the variational formulation and the disentanglement procedure is often small when used for band structure interpolation.

The variational formulation allows us to study the decay properties of Wannier functions for metallic systems. We present the localization properties of generalized Wannier functions for the free electron gas in one and two dimensions (1D and 2D). We find that minimizers of the spread exhibit a weak algebraic decay, related to singularities that we identify in  $\mathbf{k}$  space. This slow decay is shown to be not a fundamental property of disentangled Wannier functions, but rather a consequence of the fact that minimizing the spread only imposes finite second moments (or square-integrable first derivatives in  $\mathbf{k}$  space). In particular we show in 1D how to modify the maximally localized Wannier functions to obtain superalgebraic decay.

The rest of the paper is organized as follows. We first introduce several background topics such as Bloch–Floquet theory, Wannier functions, and the SCDM methodology in section 2. We then present a variational formulation for Wannier functions in section 3, and discuss the relation between our variational formulation and existing methods. Numerical results for the free electron gas and real materials are given in sections 4 and 5, followed by conclusion and discussion. Some of the technical details related to the implementation of the variational formulation are given in Appendix A.

## 2. Preliminaries.

**2.1. Bloch–Floquet theory.** We first briefly review Bloch–Floquet theory for crystal structures. The *Bravais lattice* with lattice vectors  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3 \in \mathbb{R}^3$  is defined as

$$(2.1) \quad \mathbb{L} = \left\{ \mathbf{R} = \sum_{i=1}^3 n_i \mathbf{a}_i \mid n_1, n_2, n_3 \in \mathbb{Z} \right\},$$

and the lattice vectors define a unit cell in the Bravais lattice

$$(2.2) \quad \Gamma = \left\{ \mathbf{r} = \sum_{i=1}^3 c_i \mathbf{a}_i \mid -1/2 \leq c_1, c_2, c_3 < 1/2 \right\}.$$

The Bravais lattice induces a reciprocal lattice denoted  $\mathbb{L}^*$ , which is the support of the Fourier transform of  $\mathbb{L}$ -periodic functions. The lattice vectors of  $\mathbb{L}^*$  are denoted by  $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$  with  $\mathbf{b}_i \cdot \mathbf{a}_j = 2\pi\delta_{ij}$ . A unit cell of the reciprocal lattice is selected and called (with some abuse of language) the *Brillouin zone* and is defined as

$$(2.3) \quad \Gamma^* = \left\{ \mathbf{k} = \sum_{i=1}^3 c_i \mathbf{b}_i \mid -1/2 \leq c_1, c_2, c_3 < 1/2 \right\}.$$

For a potential  $V$  that is real-valued and  $\mathbb{L}$ -periodic, i.e.,

$$(2.4) \quad V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}) \quad \forall \mathbf{r} \in \mathbb{R}^3, \mathbf{R} \in \mathbb{L},$$

we consider the Schrödinger operator in  $\mathbb{R}^3$

$$\mathcal{H} = -\frac{1}{2}\Delta + V.$$

The Bloch–Floquet theory allows us to relabel the spectrum of  $\mathcal{H}$  using two indices  $(n, \mathbf{k})$ , where  $n \in \mathbb{N}$  is the band index, and  $\mathbf{k} \in \Gamma^*$  is the Brillouin zone index. The generalized (not square-integrable) eigenfunction  $\psi_{n,\mathbf{k}}(\mathbf{r})$  is known as a Bloch orbital and satisfies

$$\mathcal{H}\psi_{n,\mathbf{k}}(\mathbf{r}) = \varepsilon_{n,\mathbf{k}}\psi_{n,\mathbf{k}}(\mathbf{r}).$$

Importantly,  $\psi_{n,\mathbf{k}}$  can be decomposed as

$$(2.5) \quad \psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n,\mathbf{k}}(\mathbf{r}),$$

where  $u_{n,\mathbf{k}}(\mathbf{r})$  is a periodic function with respect to  $\mathbb{L}$ . Eigenpairs  $(\varepsilon_{n,\mathbf{k}}, u_{n,\mathbf{k}})$  can therefore be obtained by solving the eigenvalue problem

$$(2.6) \quad \mathcal{H}(\mathbf{k})u_{n,\mathbf{k}} = \varepsilon_{n,\mathbf{k}}u_{n,\mathbf{k}}, \quad n \in \mathbb{N}, \quad \mathbf{k} \in \Gamma^*,$$

where  $\mathcal{H}(\mathbf{k}) = \frac{1}{2}(-i\nabla + \mathbf{k})^2 + V(\mathbf{r})$ . For each  $\mathbf{k}$ , the eigenvalues  $\varepsilon_{n,\mathbf{k}}$  are ordered non-decreasingly, and  $\{\varepsilon_{n,\mathbf{k}}\}$  as a function of  $\mathbf{k}$  for a fixed  $n$  is called a *band*. The set of all eigenvalues is called the *band structure* of the crystal and characterizes the spectrum of the operator  $\mathcal{H}$ . If  $\min_{\mathbf{k} \in \Gamma^*} \varepsilon_{N+1,\mathbf{k}} > \max_{\mathbf{k} \in \Gamma^*} \varepsilon_{N,\mathbf{k}}$ , then the first  $N$  bands are *isolated*. This is, for instance, the case in the occupied bands of an insulator. When the gap condition does not hold, the band structure becomes entangled.<sup>1</sup> Entangled band structure appears not only in metallic systems, but also insulating systems when a Wannier representation of part of the conduction bands is required.

**2.2. Wannier functions.** For simplicity, we first consider systems with isolated first  $N$  bands—an assumption we will drop toward the end of this section. Rotating the set of functions  $\{\psi_{n,\mathbf{k}}\}$  by an arbitrary unitary matrix  $U(\mathbf{k}) \in \mathbb{C}^{N \times N}$ , we can define a new set of functions

$$(2.7) \quad \tilde{\psi}_{n,\mathbf{k}}(\mathbf{r}) = \sum_{m=1}^N \psi_{m,\mathbf{k}}(\mathbf{r})U_{m,n}(\mathbf{k}), \quad \mathbf{k} \in \Gamma^*.$$

A given set of such matrices  $\{U(\mathbf{k})\}_{\mathbf{k} \in \Gamma^*}$  is called a *gauge*.

For each  $\mathbf{k}$ , we consider the density matrix  $P(\mathbf{k})$ , which is the projector on the the eigenspace corresponding to the first  $N$  eigenvalues of  $H(\mathbf{k})$

$$(2.8) \quad P(\mathbf{k}) = \sum_{n=1}^N |\psi_{n,\mathbf{k}}\rangle\langle\psi_{n,\mathbf{k}}| = \sum_{n=1}^N |\tilde{\psi}_{n,\mathbf{k}}\rangle\langle\tilde{\psi}_{n,\mathbf{k}}|.$$

Importantly, for each  $\mathbf{k}$ , the density matrix  $P(\mathbf{k})$  is gauge-invariant. If  $\mathcal{C}$  is a contour in the complex plane enclosing the eigenvalues  $\varepsilon_{1,\mathbf{k}}, \dots, \varepsilon_{N,\mathbf{k}}$  (and only those), then the Cauchy integral formula yields an alternative representation of  $P(\mathbf{k})$

$$(2.9) \quad P(\mathbf{k}) = \frac{1}{2\pi i} \int_{\mathcal{C}} \frac{1}{\lambda - H(\mathbf{k})} d\lambda.$$

Since  $H(\mathbf{k})$  is analytic, it follows that so is  $P(\mathbf{k})$ .

<sup>1</sup>If  $\min_{\mathbf{k} \in \Gamma^*} \varepsilon_{N+1,\mathbf{k}} \leq \max_{\mathbf{k} \in \Gamma^*} \varepsilon_{N,\mathbf{k}}$ , but  $\varepsilon_{N+1,\mathbf{k}} > \varepsilon_{N,\mathbf{k}}$  is valid in a pointwise sense, the band structure can also be viewed as isolated. However, for simplicity we will treat such a case using the more general formulation for entangled band structure.

Given a choice of gauge, the Wannier functions are defined as [38]

$$(2.10) \quad w_{n,\mathbf{R}}(\mathbf{r}) = \frac{1}{|\Gamma^*|} \int_{\Gamma^*} \tilde{\psi}_{n,\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{R}} d\mathbf{k}, \quad \mathbf{r} \in \mathbb{R}^3, \mathbf{R} \in \mathbb{L},$$

where  $|\Gamma^*|$  is the volume of the first Brillouin zone. This represents a unitary transformation from the family  $(\psi_{n,\mathbf{k}})_{n=1,\dots,N, \mathbf{k} \in \Gamma^*}$  to  $(w_{n,\mathbf{R}})_{n=1,\dots,N, \mathbf{R} \in \mathbb{L}}$ . In particular, the Wannier functions  $w_{n,\mathbf{R}}$  are orthogonal to each other and span the same space as the range of the total density matrix  $\frac{1}{|\Gamma^*|} \int_{\Gamma^*} P(\mathbf{k}) d\mathbf{k}$ . They are also translation invariant:  $w_{n,\mathbf{R}}(\mathbf{r}) = w_n(\mathbf{r} - \mathbf{R})$ .

For insulating systems, in the absence of topological obstructions, there exists a gauge such that  $\tilde{\psi}_{n,\mathbf{k}}$  is analytic and  $\mathbb{L}^*$ -periodic in  $\mathbf{k}$ , implying that the Fourier transform of  $w_{n,\mathbf{R}}$  is analytic, and therefore that each Wannier function decays exponentially as  $|\mathbf{r}| \rightarrow \infty$  [3, 5, 33]. The Wannier localization problem is reduced to the problem of finding a gauge  $\{U(\mathbf{k})\}$  such that  $w_{n,\mathbf{0}}$  is localized, or, equivalently, that  $\tilde{\psi}_{n,\mathbf{k}}$  is smooth with respect to  $\mathbf{k}$ . This can be done by minimizing the “spread functional” [13, 26]

$$(2.11) \quad \Omega[\{U(\mathbf{k})\}] = \sum_{n=1}^N \int |w_{n,\mathbf{0}}(\mathbf{r})|^2 \mathbf{r}^2 d\mathbf{r} - \left| \int |w_{n,\mathbf{0}}(\mathbf{r})|^2 \mathbf{r} d\mathbf{r} \right|^2,$$

which corresponds to the sum of the variances of the Wannier functions. Here  $w_{n,\mathbf{0}}$  depends on  $U(\mathbf{k})$  through  $\tilde{\psi}_{n,\mathbf{k}}$  as in (2.10). This problem is usually solved by a minimization algorithm such as steepest descent or conjugate gradient with projections at each step to respect the constraints that  $U(\mathbf{k})$  must be unitary [27, 28].

For systems with entangled band structure, the density matrix  $P(\mathbf{k})$  as defined in (2.8) is no longer smooth with respect to  $\mathbf{k}$ . As a result, there is no choice of gauge  $U(\mathbf{k})$  that leads to a set of rotated Bloch orbitals that is smooth with respect to  $\mathbf{k}$ , and Wannier functions defined strictly according to (2.10) will then decay very slowly in real space [15]. In order to enhance the localization properties of Wannier functions, the definition of Wannier functions has been generalized so that the spectral subspace of interest is only a proper subspace of that spanned by Wannier functions [35]. In the physics literature, the spectral subspace is described by a “frozen window” along the energy spectrum, and the Wannier functions are linear combinations of orbitals from a larger set described by an “outer window.”

More specifically, we first fix a number of bands  $N_o$  that determines the outer window<sup>2</sup> and then proceed to look for Wannier functions built out of  $\psi_{1,\mathbf{k}}, \dots, \psi_{N_o,\mathbf{k}}$ . Next, for each  $\mathbf{k}$  point we fix a set of frozen bands  $\mathcal{N}_f(\mathbf{k}) \subset [N_o]$ , and let  $N_f(\mathbf{k}) = |\mathcal{N}_f(\mathbf{k})|$ .  $\mathcal{N}_f(\mathbf{k})$  are often defined as the bands within a fixed energy window that we will try to reproduce. Correspondingly, we define the frozen density matrix as

$$(2.12) \quad P_f(\mathbf{k}) = \sum_{n \in \mathcal{N}_f(\mathbf{k})} |\psi_{n,\mathbf{k}}\rangle \langle \psi_{n,\mathbf{k}}|,$$

which is the projection onto the states within the frozen energy window. Again, the frozen density matrix as defined in (2.12) is not smooth with respect to  $\mathbf{k}$ .

<sup>2</sup>To simplify the exposition we assume a constant number of bands in the outer window, but this can be relaxed to a variable number of bands  $N_o(\mathbf{k})$ . Practically, we simply choose  $N_o(\mathbf{k}) = N_o$  to be sufficiently large to contain all bands in and near the window of interest.

We now seek to construct a set of  $N_w$  Wannier functions that span the subspace defined by the range of  $\frac{1}{|\Gamma^*|} \int_{\Gamma^*} P_f(\mathbf{k}) d\mathbf{k}$ . We introduce the gauge matrices  $U(\mathbf{k}) \in \mathbb{C}^{N_o \times N_w}$  with orthonormal columns, with  $|N_f(\mathbf{k})| \leq N_w \leq N_o$ , such that

$$(2.13) \quad \tilde{\psi}_{n,\mathbf{k}}(\mathbf{r}) = \sum_{m=1}^{N_o} \psi_{m,\mathbf{k}}(\mathbf{r}) U_{m,n}(\mathbf{k}), \quad \mathbf{k} \in \Gamma^*, n = 1, \dots, N_w.$$

This may be equivalently expressed in matrix form as  $\tilde{\Psi}(\mathbf{k}) = \Psi(\mathbf{k})U(\mathbf{k})$ . This choice of gauge also induces a density matrix of rank  $N_w$  for each  $\mathbf{k}$  defined as

$$(2.14) \quad P_w(\mathbf{k}) = \sum_{n=1}^{N_w} |\tilde{\psi}_{n,\mathbf{k}}\rangle \langle \tilde{\psi}_{n,\mathbf{k}}| = \tilde{\Psi}(\mathbf{k}) \tilde{\Psi}^*(\mathbf{k}) = \Psi(\mathbf{k})U(\mathbf{k})U^*(\mathbf{k})\Psi^*(\mathbf{k}).$$

Note that unlike the case with isolated band structure where  $N_o = N_w$ , here  $N_w \neq N_o$  implies that  $U(\mathbf{k})U^*(\mathbf{k}) \neq I_{N_o}$ . Furthermore, since the set of orbitals in the frozen window is only a subset of all possible orbitals, in general the projectors  $P_w$  and  $P_f$  do not span the same space. In order to ensure that our Wannier functions span the same subspace as the subspace associated with the frozen window, we require that

$$(2.15) \quad P_w(\mathbf{k})P_f(\mathbf{k}) = P_f(\mathbf{k}) \quad \forall \mathbf{k} \in \Gamma^*.$$

**3. Variational formulation for Wannier functions with entangled band structure.** We now proceed to develop a variational formulation for Wannier functions. First, we illustrate how to encode the desired constraints when paired with the aforementioned spread functional. Subsequently, to facilitate a numerical solution of the optimization problem, we refine how the constraints are expressed. Last, we discuss the relation between the existing disentanglement procedure to our formulation and discuss how we construct an initial guess using the SCDM methodology.

**3.1. Formulating the optimization problem.** Without loss of generality, for the following discussion we assume that the frozen orbitals ( $\Psi_f$ ) are always ordered before the rest of the orbitals ( $\Psi_r$ ). In terms of the notation from the previous section, this means that for each  $\mathbf{k}$  the frozen orbitals are represented by the set  $\mathcal{N}_f(\mathbf{k}) = \{1, 2, \dots, N_f(\mathbf{k})\}$  with  $N_f(\mathbf{k})$  simply representing the number of frozen orbitals per  $\mathbf{k}$ -point. Now, we may partition the orbitals and the gauge using the following block form:

$$(3.1) \quad \Psi(\mathbf{k}) = [\Psi_f(\mathbf{k}) \quad \Psi_r(\mathbf{k})], \quad U(\mathbf{k}) = \begin{bmatrix} U_f(\mathbf{k}) \\ U_r(\mathbf{k}) \end{bmatrix}.$$

The matrices  $U_f(\mathbf{k})$  and  $U_r(\mathbf{k})$  are of size  $N_f(\mathbf{k}) \times N_w$  and  $(N_o - N_f(\mathbf{k})) \times N_w$ , encoding the weight assigned to the frozen subspace and its complement in the Wannier functions, respectively. The condition that the Wannier functions represent the frozen bands as in (2.15) can conveniently be expressed in terms of these matrices as follows.

**PROPOSITION 3.1.** *The following statements are equivalent:*

1.  $P_w(\mathbf{k})P_f(\mathbf{k}) = P_f(\mathbf{k})$ .
2.  $U_f(\mathbf{k})U_f^*(\mathbf{k}) = I_{N_f(\mathbf{k})}$ .
3.  $U_f(\mathbf{k})U_r^*(\mathbf{k}) = 0$  and  $U_f(\mathbf{k})$  has full row rank.
4.  $U(\mathbf{k}) = \begin{bmatrix} I_{N_f(\mathbf{k})} & 0 \\ 0 & Y(\mathbf{k}) \end{bmatrix} X(\mathbf{k})$ , where  $X(\mathbf{k})$  is a unitary matrix of size  $N_w \times N_w$ , and  $Y(\mathbf{k})$  is a matrix with orthogonal columns of size  $(N_o - N_f(\mathbf{k})) \times (N_w - N_f(\mathbf{k}))$ .

*Proof.* Since each  $\mathbf{k}$  point is treated independently, for simplicity we drop the  $\mathbf{k}$  dependence in the proof below.

1  $\Leftrightarrow$  2: From the definition of  $P_w, P_f$  we have

$$P_w P_f = \Psi \begin{bmatrix} U_f U_f^* & U_f U_r^* \\ U_r U_f^* & U_r U_r^* \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \Psi^* = \Psi \begin{bmatrix} U_f U_f^* & 0 \\ U_r U_f^* & 0 \end{bmatrix} \Psi^*,$$

$$P_f = \Psi \begin{bmatrix} I_{N_f} & 0 \\ 0 & 0 \end{bmatrix} \Psi^*.$$

The result follows since  $\Psi$  has orthogonal columns. Note that  $U_f U_f^* = I_{N_f}$  in conjunction with  $UU^*$  being a projector implies that  $U_r U_f^* = 0$ .

2  $\Leftrightarrow$  3: From the partition of  $U$  we have

$$UU^* = \begin{bmatrix} U_f U_f^* & U_f U_r^* \\ U_r U_f^* & U_r U_r^* \end{bmatrix}.$$

If  $U_f U_f^* = I_{N_f}$ , since  $UU^*$  is a projector, it follows that  $U_r U_f^* = 0$ . On the other hand, if  $U_r U_f^* = U_f U_r^* = 0$ , then the fact that  $UU^*$  is a projector implies that  $U_f U_f^*$  is a projector as well. Since it has full row rank, it must therefore be the identity matrix.

3  $\Leftrightarrow$  4: If 3 is true, then

$$P_U = UU^* = \begin{bmatrix} I_{N_f} & 0 \\ 0 & U_r U_r^* \end{bmatrix}.$$

Then  $U_r U_r^*$  is a projector with rank  $N_w - N_f$ , or equivalently

$$U_r U_r^* = YY^*$$

for some  $(N_o - N_f) \times (N_w - N_f)$  matrix  $Y$  with orthogonal columns. Then

$$U = P_U U = \begin{bmatrix} I_{N_f} & 0 \\ 0 & YY^* \end{bmatrix} U = \begin{bmatrix} I_{N_f} & 0 \\ 0 & Y \end{bmatrix} X,$$

where

$$X = \begin{bmatrix} I_{N_f} & 0 \\ 0 & Y^* \end{bmatrix} U,$$

and it can be readily verified that  $X$  is unitary. The reverse direction is obvious.  $\square$

Proposition 3.1 gives us various concise ways to impose the desired condition on the span of Wannier functions, and we may, for instance, consider using condition 2. Since the smoothness requirement for  $\tilde{\psi}_{n,\mathbf{k}}$  with respect to the Brillouin zone index  $\mathbf{k}$  can be realized by minimizing the spread functional (2.11), finding the desired smooth gauge  $U(\mathbf{k})$  can be recast as the following constrained optimization problem:

$$(3.2) \quad \begin{aligned} \inf_{\{U(\mathbf{k})\}} \quad & \Omega[\{U(\mathbf{k})\}] \\ \text{s.t.} \quad & U^*(\mathbf{k})U(\mathbf{k}) = I_{N_w}, \quad U_f(\mathbf{k})U_f^*(\mathbf{k}) = I_{N_f(\mathbf{k})}. \end{aligned}$$

The difficulty at this stage is that numerical optimization of (3.2) with respect to these constraints may not be easy. In particular, the set of matrices  $U$  satisfying  $U^*U = I_{N_w}$  and  $U_f U_f^* = I_{N_f}$  does not necessarily possess a smooth manifold structure. This complicates the application of standard methods for the minimization of functions over orthogonality constraints.

On the other hand, the condition 4 in Proposition 3.1 represents  $U$  in a factorized form, hereinafter referred to as the  $(X, Y)$  representation. This representation of the matrix  $U(\mathbf{k})$  gives rise to Wannier functions composed of the  $N_f(\mathbf{k})$  functions in the frozen window, and another set of  $N_w - N_f(\mathbf{k})$  functions, encoded by the matrix  $Y(\mathbf{k})$ . This  $Y$  encapsulates all the necessary information about the projector  $P_w = \tilde{\Psi} \tilde{\Psi}^*$ . The unitary  $X(\mathbf{k})$  matrix mixes these  $N_w$  Wannier functions among themselves to produce a smooth gauge. In the  $(X, Y)$  representation, the variational formulation (3.2) can be written as

$$(3.3) \quad \begin{aligned} & \inf_{\{X(\mathbf{k}), Y(\mathbf{k})\}} \Omega[\{U(\mathbf{k})\}], \\ & \text{s.t. } U(\mathbf{k}) = \begin{bmatrix} I_{N_f(\mathbf{k})} & 0 \\ 0 & Y(\mathbf{k}) \end{bmatrix} X(\mathbf{k}), \\ & X^*(\mathbf{k})X(\mathbf{k}) = I_{N_w}, \\ & Y^*(\mathbf{k})Y(\mathbf{k}) = I_{N_w - N_f(\mathbf{k})}. \end{aligned}$$

This optimization problem is equivalent to the “partly occupied Wannier functions” [36]. This also directly generalizes the maximally localized Wannier functions procedure [26] by Marzari and Vanderbilt for the isolated case.

The  $(X, Y)$  representation is a redundant representation, and a given  $U$  can be reproduced by many pairs  $(X, Y)$ . However, in contrast to the formulation (3.2), the constraint in (3.3) defines a Riemannian manifold where  $X(\mathbf{k})$  and  $Y(\mathbf{k})$  are independent matrices with orthogonality constraints. This allows us to use standard algorithms for the minimization of differentiable functions on Riemannian manifolds to solve the problem. We refer to Appendix A for the details of the computation of the gradient of the objective function  $\Omega$ .

**3.2. Implementation.** For our implementation, we modified the Julia [2] library `Optim.jl` for unconstrained optimization to accommodate constraints represented by Riemannian manifolds [12, 1]. Our modifications have been integrated into that library and are available online.<sup>3</sup> For the numerical tests that follow we used the limited-memory BFGS algorithm [31] with Hager–Zhang line search [17], which gave the best performance compared to other readily available algorithms (steepest descent, conjugate gradient, BFGS) and line searches (fixed step, backtracking).

In order to generate the initial guess for numerical optimization, we need to convert a given matrix  $U$  to a pair  $(X, Y)$  that parametrizes it. It will also be useful to consider matrices  $U$  that do not satisfy the constraints  $U^*U = I_{N_w}$  and  $U_f U_f^* = I_{N_f}$  exactly but only approximately. This will allow us to project  $U$  to the admissible set that satisfies these constraints.

To find a pair  $(X, Y)$  that represents a given  $U$ , we first choose  $Y$  to minimize the error on the projector  $UU^*$  measured by the Frobenius norm via

$$(3.4) \quad \inf_{Y^*Y=I} \left\| UU^* - \begin{bmatrix} I_{N_f} & 0 \\ 0 & YY^* \end{bmatrix} \right\|_F^2.$$

A solution to this problem can be computed using the eigenvalue decomposition

$$(3.5) \quad U_r U_r^* = V S V^*.$$

When  $U$  satisfies the constraints,  $U_r U_r^*$  is a projector of rank  $N_w - N_f$ , and we can choose  $Y$  to be the columns of  $V$  corresponding to the  $N_w - N_f$  nonzero eigenvalues

<sup>3</sup><https://github.com/JuliaNLSSolvers/Optim.jl>.

of  $S$ . When  $U$  does not satisfy the constraints, we pick  $Y$  as the columns of  $V$  corresponding to the largest  $N_w - N_f$  eigenvalues of  $S$ .

Once  $Y$  is computed, we can find  $X$  that minimizes the error on  $U$ :

$$(3.6) \quad \inf_{X^* X = I} \left\| U - \begin{bmatrix} I_{N_f} & 0 \\ 0 & Y \end{bmatrix} X \right\|_F^2.$$

When  $U$  satisfies the constraints, the solution of this problem is simply

$$X = \begin{bmatrix} I_{N_f} & 0 \\ 0 & Y^* \end{bmatrix} U.$$

Otherwise, the solution can again be obtained via the singular value decomposition

$$(3.7) \quad \begin{bmatrix} I_{N_f} & 0 \\ 0 & Y^* \end{bmatrix} U = \tilde{V}_l \tilde{S} \tilde{V}_r^*$$

and setting  $X = \tilde{V}_l \tilde{V}_r^*$ . This step is also called the Löwdin orthogonalization procedure [24].

**3.3. Relation to disentanglement.** Our variational formulation also gives us a concise way to understand the “disentanglement” procedure of Souza, Marzari, and Vanderbilt [35], in which the spread functional is split into two parts

$$(3.8) \quad \Omega[\{U(\mathbf{k})\}] = \Omega_I[\{U(\mathbf{k})\}] + \tilde{\Omega}[\{U(\mathbf{k})\}].$$

Here  $\Omega_I$  is called the gauge-invariant part (depending on  $P_w$ , and hence only on  $YY^*$ ), and  $\tilde{\Omega}$  is called the gauge dependent part (depending on  $X$ ). Instead of optimizing (3.3) directly, [35] proposes to use a two-step procedure. The first one optimizes the gauge-invariant part only:

$$(3.9) \quad \inf_{\{Y(\mathbf{k})\}} \Omega_I[\{U(\mathbf{k})\}].$$

This is numerically expedient as  $\Omega_I$  only depends on  $YY^*$ . In fact, it is analogous to minimization problems in electronic structure (for instance, the Hartree–Fock model), where one minimizes the energy, which only depends on the spectral subspace, over all possible orthogonal orbitals. The authors in [35] accordingly obtain a nonlinear eigenvalue problem as the first-order optimality conditions, which they solve using a damped self-consistent field (SCF) iteration.

Once  $\{Y(\mathbf{k})\}$  is obtained, it is fixed and so is the projector  $P_w$ . A second minimization problem

$$(3.10) \quad \inf_{\{X(\mathbf{k})\}} \tilde{\Omega}[\{U(\mathbf{k})\}]$$

is then solved with respect to the gauge matrix  $X(\mathbf{k})$ . This optimization problem is of the same nature as the one for an isolated set of bands.

The total spread from the above two-step procedure is necessarily larger or equal to the global minimum of (3.3). Interestingly, although the optimal spread can be substantially lower than that found by the two-step disentanglement procedure, numerical experiments show that the quality of Wannier interpolation, measured, for instance, by the quality of band structure interpolation, is often similar in both cases.

**3.4. Selected column of the density matrix.** While the primary purpose of this work is to introduce and analyze a variational formulation of Wannier functions, both the objective function and the constraints are nonlinear, and hence there may exist multiple local minima. It is practically important to seed such methods with a good initial guess. Here, we summarize the recently developed unified methodology for Wannier localization of entangled band structure [8] based on the SCDM methodology [9]. Importantly, this method is direct and robust—no initial guess is required and it will generate valid output—and thus may be reliably used to generate an initial guess.

The SCDM method for entangled band structure first constructs a quasi-density matrix

$$(3.11) \quad f(H(\mathbf{k})) = \sum_{n=1}^{N_o} f(\varepsilon_{n,\mathbf{k}}) |\psi_{n,\mathbf{k}}\rangle \langle \psi_{n,\mathbf{k}}|.$$

For insulating systems  $f$  would be 1 on the occupied bands and 0 otherwise, yielding the projector  $P$  as before. For entangled band structure, however, the function  $f(\cdot)$  is chosen to be large on the bands of interest and decays rapidly, but smoothly, away from them [8]. The SCDM algorithm constructs a gauge by selecting a common set of columns of the  $\mathbf{k}$ -dependent (quasi)-density matrix  $f(H(\mathbf{k}))$ . In practice, it is often sufficient to select these columns based on an “anchor” point denoted  $\mathbf{k}_0$ —generically chosen to be the so-called Gamma-point  $(0, 0, 0)^T$ .

We now briefly outline the SCDM method and refer the reader to [8] for more details. Let  $\Psi_{\mathbf{k}} \in \mathbb{C}^{N_g \times N_o}$  be the matrix with orthogonal columns that represents  $\{\psi_{n,\mathbf{k}}(\mathbf{r})\}$  on a discrete grid in the unit cell, and  $\mathcal{E}(\mathbf{k}) = \text{diag}\{\{\varepsilon_{n,\mathbf{k}}\}_{n=1}^{N_o}\}$  be a diagonal matrix encoding the corresponding eigenvalues. SCDM identifies  $N_w$  columns of  $f(H(\mathbf{k}))$  based on the leading  $N_w$  columns of the permutation matrix  $\Pi$ , computed via the QR factorization with column pivoting (QRCP) procedure

$$(3.12) \quad \Psi_{\mathbf{k}_0}^* \Pi = QR.$$

This set of columns is denoted by  $\mathcal{C} = \{\mathbf{r}_n\}_{n=1}^{N_w} \subset \Gamma$ . Now, for each  $\mathbf{k}$ -point define  $\Xi(\mathbf{k}) \in \mathbb{C}^{N_o \times N_w}$  as

$$(3.13) \quad \Xi_{m,n}(\mathbf{k}) = f(\varepsilon_{m,\mathbf{k}}) \psi_{m,\mathbf{k}}^*(\mathbf{r}_n).$$

It is expected that the selected columns of the quasi-density matrix

$$(3.14) \quad \phi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{m=1}^{N_o} \psi_{m,\mathbf{k}}(\mathbf{r}) \Xi_{m,n}(\mathbf{k}), \quad 1 \leq n \leq N_w,$$

for  $n = 1, \dots, N_w$  is smooth with respect to  $\mathbf{k}$ . Therefore, if the singular values of  $\Xi(\mathbf{k})$  are uniformly bounded away from 0 in the Brillouin zone,  $U(\mathbf{k})$  constructed via Löwdin orthogonalization [24] of  $\Xi(\mathbf{k})$  has orthogonal columns. Substituting  $\Xi(\mathbf{k})$  by  $U(\mathbf{k})$  in (3.14), we yield a set of orthogonal orbitals that are smooth with respect to  $\mathbf{k}$ .

In this framework the frozen bands are not represented exactly, so prior to use in our optimization procedure we must convert from  $U$  to a pair  $(X, Y)$  via our aforementioned scheme. Since the projector on the frozen set varies discontinuously with  $\mathbf{k}$ , this procedure does not produce a continuous gauge. However, if the function  $f$  is chosen appropriately, it should be close to one. This is further corroborated by

the quality of band interpolation; despite the substantially larger spread, we achieve in the numerical results section using the SCDM initial guess without explicitly freezing any bands.

**4. Free electron gas.** Investigating the decay properties of Wannier functions numerically for real materials is difficult because of the very large number of  $\mathbf{k}$  points needed to observe the asymptotic decay rate.<sup>4</sup> Therefore, prior to embarking upon experiments with real materials we consider a model problem: the free electron gas. This model, i.e.,  $V = 0$ , is explicitly solvable and poses a very interesting benchmark for disentanglement, even in 1D or 2D. Here, we define the lattice as  $\mathbb{L} = 2\pi\mathbb{Z}^d$ , so that  $\mathbb{L}^* = \mathbb{Z}^d$  and we let the Brillouin zone be the set  $[0, 1]^d$ .

In this case the eigenfunctions of the operator  $H(\mathbf{k}) = (-i\nabla + \mathbf{k})^2$  are given by

$$v_{\mathbf{K}}(\mathbf{r}) = \frac{1}{\sqrt{|\Gamma|}} e^{i\mathbf{K}\cdot\mathbf{r}}$$

for  $\mathbf{K}$  in the reciprocal lattice  $\mathbb{Z}^d$ . The corresponding eigenvalues are  $\varepsilon_{\mathbf{K},\mathbf{k}} = |\mathbf{K} + \mathbf{k}|^2$ . Notice that since  $\varepsilon_{\mathbf{K},\mathbf{k}}$  is the squared distance from  $\mathbf{k}$  to  $-\mathbf{K}$ , the dispersion relation is the set of squared distances of  $\mathbf{k}$  to the points of the reciprocal lattice  $\mathbb{Z}^d$ .

We order the eigenfunctions in nondecreasing order based on their associated eigenvalues. We let  $u_{n,\mathbf{k}} = v_{\mathbf{K}_n}$ , where  $\varepsilon_{\mathbf{K}_n,\mathbf{k}}$  is the  $n$ th eigenvalue of  $H(\mathbf{k})$ , this choice being arbitrary in the presence of degeneracies. The matrix elements  $M_{mn}^{\mathbf{k},\mathbf{b}} = \langle u_{m,\mathbf{k}}, u_{n,\mathbf{k}+\mathbf{b}} \rangle$  of overlap between neighboring  $\mathbf{k}$  points used in the optimization process (see Appendix) then assume a particularly simple expression:  $M_{mn}^{\mathbf{k},\mathbf{b}} = 1$  if  $u_{m,\mathbf{k}}$  and  $u_{n,\mathbf{k}+\mathbf{b}}$  are associated with the same  $\mathbf{K}$ , and  $M_{mn}^{\mathbf{k},\mathbf{b}} = 0$  otherwise. In particular, this matrix differs from the identity (it is a permutation matrix) near eigenvalue crossings, where  $\varepsilon_{\mathbf{K},\mathbf{k}} = \varepsilon_{\mathbf{K}',\mathbf{k}}$  with  $\mathbf{K}' \neq \mathbf{K}$ .

The free electron gas also makes it particularly easy to compute the Wannier functions via their Fourier transforms. For a Wannier function given in  $\mathbf{k}$ -space by

$$\tilde{\psi}_{n\mathbf{k}}(\mathbf{r}) = \sum_{n \in \mathbb{N}} e^{i\mathbf{k}\cdot\mathbf{r}} v_{\mathbf{K}_m}(\mathbf{r}) U_{mn}(\mathbf{k}),$$

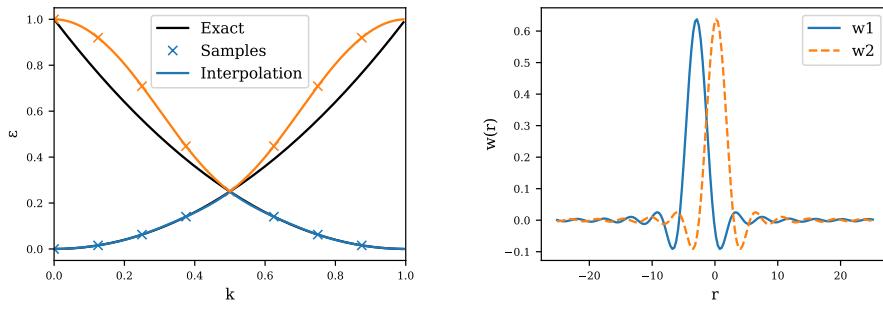
it holds that

$$w_{n,\mathbf{0}}(\mathbf{r}) = \int_{[0,1]^d} \tilde{\psi}_{n\mathbf{k}}(\mathbf{r}) d\mathbf{k} = \frac{1}{(2\pi)^d} \int_{[0,1]^d} \sum_{m \in \mathbb{N}} U_{mn}(\mathbf{k}) e^{i(\mathbf{k} + \mathbf{K}_m) \cdot \mathbf{r}},$$

from which it follows that  $U_{mn}(\mathbf{k})$  is simply the Fourier transform of  $w_{n,\mathbf{0}}$  at frequency  $\xi = \mathbf{K}_m + \mathbf{k}$ .

Notably, the free electron gas possesses a large number of symmetries and consequently it has a number of properties that are not expected for generic systems. For instance, eigenvalue crossings are numerous (of co-dimension 1, i.e., points in 1D, lines in 2D, and planes in 3D), while they are expected from the von Neumann–Wigner theorem [37] to be rare. This theorem predicts that, in the absence of particular symmetries, the crossing of eigenvalues of a Hermitian matrix are a phenomenon of codimension 3. This is believed to be true (although unproved in many cases) for “generic” Schrödinger operators [23]: for a generic  $V$ , the band structure is only expected to show crossings in codimension 3 (isolated points in 3D, and no crossings in lower dimensions). The free electron gas, possessing all the symmetries a Schrödinger operator can have, may thus be considered a worst case system for disentanglement.

<sup>4</sup>For instance, in [39] the authors report a fast (consistent with exponential) decay up to grid sizes of  $15 \times 15 \times 15$ , although the convergence appears to slow down after that.



(a) Exact band plot, samples on the  $k$ -space grid, and Wannier interpolation.

(b) Wannier functions in real space.

FIG. 1. *Optimized Wannier functions obtained on a  $k$ -space grid of size  $N = 8$ .*

#### 4.1. The one-dimensional case.

**4.1.1. One frozen band, two Wannier functions.** The dispersion relation of the one-dimensional free electron gas results in a crossing between the first and second eigenvalues at half-integers, and between the second and third eigenvalues at integers (see Figure 1(a)). Because of the crossing between the first and second band, any single Wannier function representing the first band has a maximal decay of  $1/r$ . We therefore consider the problem of finding two Wannier functions representing the first band ( $N_w = 2, N_f = 1$ ). We do not set an outer window but discretize the band structure using  $2L + 1$  Fourier modes with  $L = 10$ . As we will see, the Wannier functions we find are compactly supported in Fourier space, and therefore do not depend on the choice of  $L \geq 2$ .

For this simple system, initializing the gauge  $U(\mathbf{k})$  randomly our optimization algorithm converges robustly to the same set of Wannier functions, up to a change of sign and a shift by a lattice vector. It would therefore seem that the global minimizer of the spread is unique (up to the invariant degrees of freedom described above) and real. The SCDM algorithm with the settings  $\mu = 0$  and  $\sigma = 2$  yields Wannier function that are visually indistinguishable from the optimized Wannier functions and have a very similar spread (2.51 for the SCDM algorithm as compared to 2.44 for the optimized Wannier functions).

In Figure 1 we observe that the first band (which is frozen) is exactly reproduced. Since the Wannier functions are localized, the Wannier interpolation is very good, and in particular has no trouble reproducing the crossing. Furthermore, the optimized Wannier functions are symmetric. (They are real, and the second one is a translate of the first by half a lattice vector.)

However, closer inspection reveals that the optimized Wannier functions are not exponentially localized, and in fact decay as  $1/r^2$  for large  $r$  (Figure 2(b)). The origin of this slow decay of the Wannier functions is the kink that appears at  $\xi = \pm 3/2$  in their Fourier transform in Figure 2(a), which also shows that the Wannier functions are compactly supported on  $[-3/2, 3/2]$  in Fourier space. This is because only the first three bands are occupied by the Wannier functions, as can be seen via inspection of the  $U$  matrix. Since band 3 crosses with band 4 at  $k = 1/2$  (corresponding to basis functions  $K = 1$  and  $K = -2$ ), the continuity of  $\tilde{\psi}_{nk}$  with respect to  $k$  means that  $U_{31}(1/2) = U_{32}(1/2) = 0$ . But the first derivative is not zero at  $1/2$ , which in turn creates the kink in Fourier space at  $K = 1, k = 1/2$  and  $K = -2, k = 1/2$ , that is,  $\xi = \pm 3/2$ .

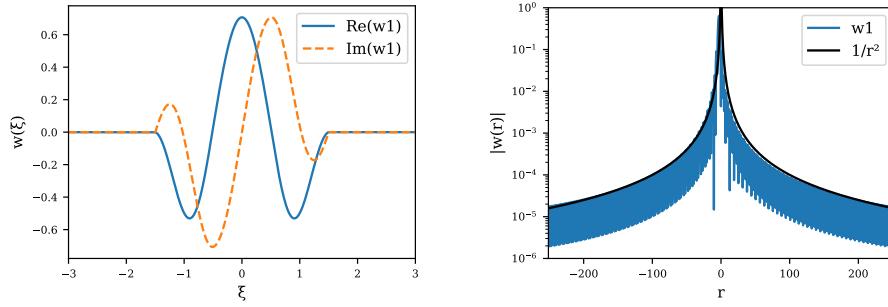
(a)  $w_1$  in Fourier space: a clear kink is visible at  $\xi = \pm 3/2$ .(b)  $w_1$  in real space, showing excellent agreement with a decay rate  $r^{-2}$ .

FIG. 2. Optimized Wannier functions obtained on a  $\mathbf{k}$ -space grid of size  $N = 80$ . (Only  $w_1$  is represented for clarity.) The kink in Fourier space at  $\xi = \pm 3/2$  translates to a  $r^{-2}$  decay. The maximum reconstruction error on the first band is  $3 \times 10^{-5}$ .

The fact that the optimized Wannier functions are only weakly localized is surprising at first glance, because it was proven in [34] that, for isolated bands, maximally-localized Wannier functions are exponentially localized. What is different in this case? The crucial point in the analysis of [34] is that the spread is similar to a Dirichlet energy, or a  $H^1$  norm, in  $\mathbf{k}$  space. Then the  $\tilde{\psi}_{nk}$  are shown to satisfy an elliptic equation in  $\mathbf{k}$  space, which by a bootstrap argument implies their analyticity. Here, this argument breaks down because the constraint  $P_w(k)P_f(k) = P_f(k)$  is discontinuous at crossings. This creates an effective “boundary condition” for the gauge  $U(k)$  at crossings that destroys the regularity. A simple analogy is that eigenfunctions of the Laplacian on  $[0, 1]$  (critical points of the Dirichlet energy) are smooth on  $\mathbb{R}$  when periodic boundary conditions are imposed, but generically produce kinks at 0 and 1 when Dirichlet boundary conditions are imposed. Here also, the effective boundary condition  $U_{31}(1/2) = U_{32}(1/2) = 0$  produces a kink.

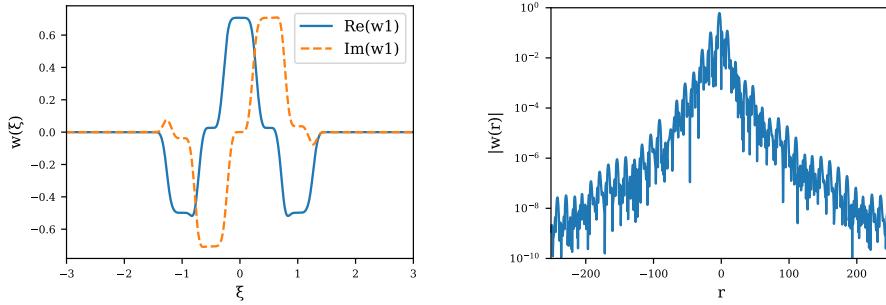
To remedy this, we show on this particular example how to build Wannier functions that are of class  $C^\infty$  in Fourier space (and therefore decay faster than any inverse polynomial in real space). In order to do so, we define the function  $\alpha(x) = e^{-1/x}$  for  $x \geq 0$  and  $\alpha(x) = 0$  for  $x < 0$ . This function is  $C^\infty$  and identically zero for  $x \leq 0$ . The function  $f(x) = \frac{\alpha(x)}{\alpha(x) + \alpha(1-x)}$  is therefore  $C^\infty$  on  $\mathbb{R}$ , equal to 0 for  $x \leq 0$ , and equal to 1 for  $x \geq 1$ .

Given  $U(0^+), U(1/2^\pm), U(1^-)$  obtained from the optimized Wannier functions, we construct a new gauge

$$\tilde{U}(k) = \begin{cases} (1 - f(2k))U(0^+) + f(2k)U(1/2^-) & \text{if } 0 < k \leq 1/2, \\ (1 - f(2k-1))U(1/2^+) + f(2k-1)U(1^-) & \text{if } 1/2 < k < 1. \end{cases}$$

This produces a new set of  $\tilde{\psi}_{nk}$  that are smooth with respect to  $\mathbf{k}$ , but not orthogonal and do not span the frozen bands. We now impose those conditions in the same way we do for the SCDM procedure as described in section 3.4. In this specific example, this produces a smooth gauge, as illustrated in Figure 3.

The Wannier functions obtained in this way display more rugged variations in Fourier space (a consequence of the use of the function  $f$  above), and accordingly the decay is slower for small values of  $x$ . However, because the gauge is smooth, the Wannier functions decay much faster for large values. Since the gauge is of



(a) Smoothed Wannier function in Fourier space: although the variations are more rugged than Figure 2, it is a  $C^\infty$  function.

(b) Smoothed Wannier function in real space: the asymptotic decay is much faster than in Figure 2.

FIG. 3. *Smoothed Wannier functions obtained on a  $\mathbf{k}$ -space grid of size  $N = 80$  (see main text for details, only the first Wannier function represented for clarity). The maximum reconstruction error on the first band is  $2 \times 10^{-9}$ .*

class  $C^\infty$  but not analytic, the Wannier functions decay faster than any polynomial, but not exponentially. Since the error of Wannier interpolation on the first band is determined by the interaction of the Wannier functions on the supercell with their periodic images, the faster asymptotic decay leads to a better reconstruction of the first band in this case. Numerical tests indicate that the cross-over point (above which the reconstruction error on the first band with the smoothing procedure is smaller than that with the optimized Wannier functions) is around  $N = 12$ .

**4.1.2. General case in 1D.** In the one-dimensional free electron gas, crossings only happen at half-integers, and between two bands at the same time. Accordingly, whenever  $N_w \geq N_f + 1$  we can construct optimized Wannier functions. Numerical results indicate that they are very similar to the ones above: their support in Fourier space increases with  $N_w$  and  $N_f$  but remains compact, they are translates of each other, and they decay as  $1/r^2$ . When  $N_w = N_f$  (that is, we are treating a metal as if it was an insulator), the gauge is discontinuous, and the corresponding Wannier functions decay as  $1/r$ .

**4.2. The two-dimensional case.** In 2D, the first four bands of the free electron gas are degenerate at  $\mathbf{k} = (1/2, 1/2)$ , corresponding to the wave vectors

$$\mathbf{K} = (0, 0), (-1, 0), (0, -1), (-1, -1).$$

This means that  $P_w(1/2, 1/2)$  must span the four-dimensional subspace corresponding to those four wave vectors. Therefore, freezing the first band can only produce localized Wannier functions when  $N_w \geq 4$ . Accordingly, we consider the case  $N_f = 1$  and  $N_w = 4$ .

Similarly to the one-dimensional case, the optimized Wannier functions are real and differ from each other only by a change of origin. However, they are not compactly supported in Fourier space and instead have decaying components on arbitrarily large wave vectors  $\mathbf{K}$  (see Figure 4(a)). Their support in Fourier space is found to display a checkerboard pattern, for which we do not have a simple explanation.

The asymptotic decay rate of a Wannier function is determined by the smoothness of its Fourier transform. In this case, the first derivative is discontinuous along edges, which corresponds to a decay of the Wannier functions in the  $x$  and  $y$  directions as  $1/r^2$ .

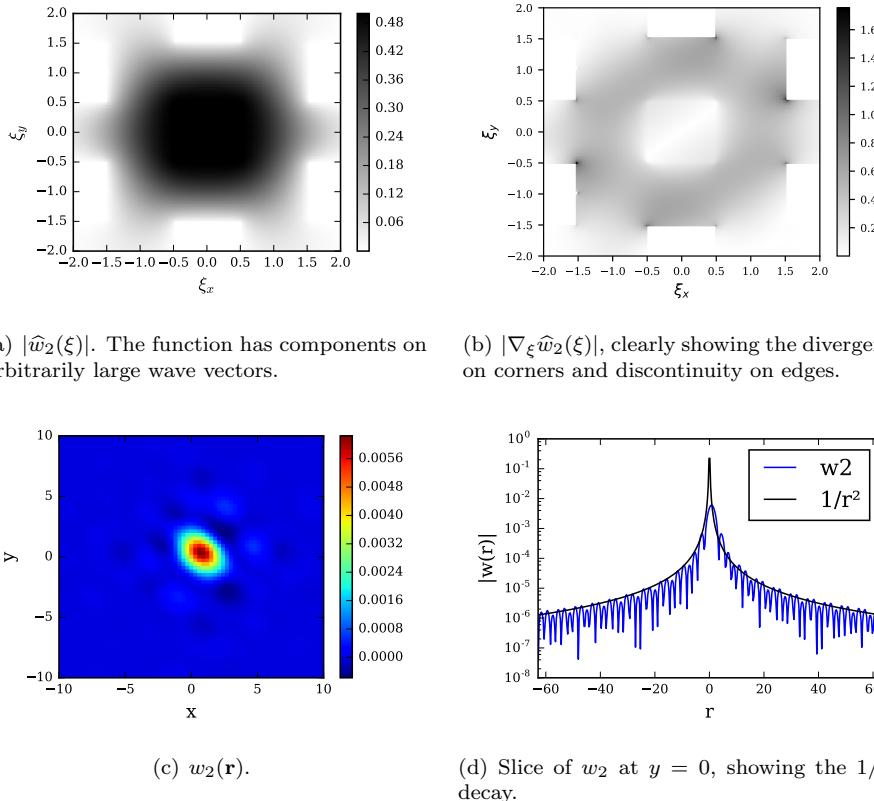


FIG. 4. Second of four optimized Wannier functions for the free electron gas in 2D with one frozen band, on a  $40 \times 40$   $\mathbf{k}$ -point grid. The other three Wannier functions are similar to  $w_2$ .

**4.3. Discussion.** These results show that the maximally localized Wannier functions only decay algebraically in general. Although numerical results are harder to obtain for real materials, we expect our analysis to carry through for very fine  $\mathbf{k}$ -point meshes: an eigenvalue crossing at a particular  $\mathbf{k}$  point acts as a constraint on the gauge, which must at that point be able to span the crossing eigenspace for the gauge to be continuous. Minimizing the second moments of the Wannier functions yields a gauge with a square-integrable but discontinuous first derivative at the crossing points, resulting in a weak asymptotic localization.

Our findings are to be contrasted with the recent theoretical result of [7], which proves under generic hypotheses that there exists almost exponentially localized Wannier functions. This simply means that, unlike in the case of insulators, for entangled band structures minimizing the second moments is not an asymptotically optimal strategy.

To get a faster asymptotic decay, one could minimize higher moments. This corresponds to minimizing integrals of higher derivatives, which have to be approximated by more complex stencils, and require the computation of additional overlaps between the  $u_{n\mathbf{k}}$  than simply the nearest neighbors. This becomes numerically expensive and complex to implement. Another possibility is to perform local “smoothing surgeries” similar to the one we demonstrated in 1D. However, this is likely to be useful only for very fine  $\mathbf{k}$ -point grids.

**5. Real materials.** Having studied the decay properties of Wannier functions for the free electron gas, we now consider the performance of our variational formulation for several real materials. This includes valence and conduction bands of silicon (semiconductor), conduction bands of copper (metal), and valence bands of aluminum (metal). We always start with the aforementioned SCDM based initial guess. We compare the result obtained from the variational formulation to that obtained from the disentanglement procedure in `Wannier90`, as well as the result obtained directly from the SCDM initial guess without further refinement.

In these experiments, the choice of the parameters of the SCDM procedure are chosen to yield good baseline band structure interpolation. However, they are not “optimized” to minimize band structure interpolation error. These experiments often show how the two optimization methods are comparable, though in some situations we are able to find Wannier functions with smaller spreads using our variational method even given the same initial guess. One interesting point that we will see play out throughout our examples is that the value of the spread and band interpolation quality may not be directly connected, i.e., Wannier functions with considerably different values of spread can yield a qualitatively similar interpolation error.

All of the  $\mathbf{k}$ -point calculations and reference band structure calculations were performed with `Quantum ESPRESSO` [14]. The SCDM initial guess was constructed using the code available online.<sup>5</sup> Our new variational formulation was implemented in the Julia language and is available online.<sup>6</sup>

**5.1. Silicon.** Here, we compute the lowest 16 bands of silicon on an  $8 \times 8 \times 8$   $\mathbf{k}$ -point grid and then proceed to compute eight Wannier functions. This includes the four valence bands and four additional low lying conduction bands. For the SCDM procedure we use  $\mu = 11.0$  and  $\sigma = 2.0$  with  $f$  corresponding to “entangled case 1” in [8]—a complementary error function. For `Wannier90` and our method, we freeze bands below 12 eV and set the outer window maximum at infinity. For `Wannier90` the prescribed convergence criteria of  $1 \times 10^{-10}$  on the spread was reached after 225 disentanglement iterations and 95 spread reduction iterations, and for our method after 149 iterations.

Figure 5 shows the band structure interpolation using the three methods. We see that for the four valence bands all three methods perform very well. Furthermore, while there are differences in the interpolation of the conduction bands, no one method clearly outperforms the others. As expected, if we consider the total spread (see Table 1) of the final localized orbitals, our variational formulation yields the most compact orbitals.

In Table 2 we report the per orbital spread for each method and observe that to the number of significant digits reported all the orbitals found by our method have the same spread (they do not vary until the fourth decimal place). In contrast, `Wannier90` seems to converge to two distinct sets of orbitals with slightly different spreads. In Figure 6, we illustrate the differences between the orbitals found with our variational method and those found via `Wannier90`. We observe that the orbitals obtained from our variational formulation resemble more closely  $sp^3$  hybridized orbitals centering around each Si atom, as indicated from chemical intuition. Interestingly, by using the output of our variational method as input to `Wannier90` we are able to force `Wannier90` to converge to the same point as our method. Unfortunately, it is difficult

<sup>5</sup><https://github.com/asdamle/SCDM>.

<sup>6</sup><https://github.com/antoine-levitt/wannier>.

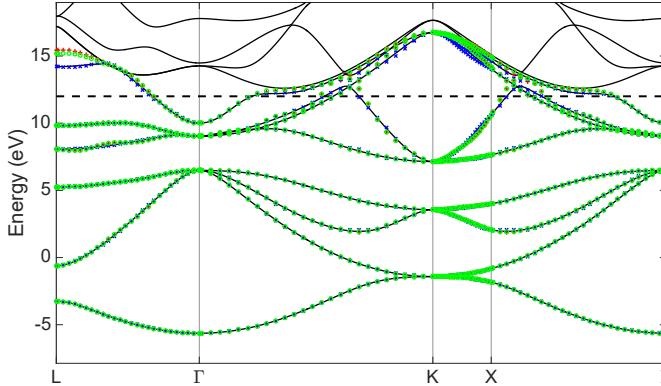


FIG. 5. Wannier interpolation of silicon with 8  $k$ -points using (blue Xs) SCDM, (green circles) our variational formulation, and (red +s) Wannier90 compared with a (black line) reference calculation. The frozen window is the region below the dotted black line.

TABLE 1  
Silicon spread and valence band error comparison.

	Final spread ( $\text{\AA}^2$ )	max error (eV)	RMSE (eV)
Variational	25.177	0.069	0.021
Wannier90	27.00	0.083	0.023
SCDM	45.206	0.112	0.029

TABLE 2  
Spreads of the eight individual Wannier functions for silicon.

	Orbital spread ( $\text{\AA}^2$ )							
	Variational	3.15	3.15	3.15	3.15	3.15	3.15	3.15
Wannier90	3.16	3.16	3.16	3.16	3.59	3.59	3.59	3.59
SCDM	4.93	4.93	4.93	4.93	6.37	6.37	6.37	6.37

to pinpoint a specific cause for the apparent convergence of Wannier90 to a worse local minimum in this setting.

**5.2. Aluminum.** We now repeat the same experiments as before, but for the valence bands of the aluminum system with an  $8 \times 8 \times 8$   $k$ -point grid. Specifically, we start with six bands and seek four Wannier functions. For the SCDM procedure we use  $\mu = 8.42$  and  $\sigma = 4.0$  with  $f$  corresponding to “entangled case 1” in [8]—a complementary error function. Once again, Wannier90 was run until convergence at  $10^{-10}$  for both the disentanglement and spread minimization, and reached that threshold after 2,437 and 91 iterations. Our method converged to a spread reduction tolerance of  $10^{-10}$  after 138 iterations. As we observe in Figure 7, all three methods once again perform well—particularly below the Fermi energy. The final spreads for each of the three methods are reported in Table 3 along with the spreads of each orbital. While both Wannier90 and our method improve upon the spread of the SCDM initial guess, we do find a slightly smaller spread with our optimization procedure. Here, bands below 11.6 eV were frozen in both Wannier90 and our method, and the outer window was set to  $\infty$ .

We also consider the convergence of the band interpolation error in this setting, looking at both maximum error and RMSE. As before, for both our method and Wannier90 we freeze bands below 11.6 eV and set the outer window to infinity. For



FIG. 6. Example orbitals from our variational method (left) and *Wannier90* (right). Here the red and gray isosurfaces are plotted at values  $\pm 0.5$  for both normalized orbitals and the yellow spheres indicate the locations of the silicon atoms. All of the orbitals we find with our variational method seem to clearly have  $sp^3$  hybrid character, as seen in the left figure. In contrast, as illustrated on the right, some of the orbitals found by *Wannier90* with larger spread do not share this behavior as clearly.

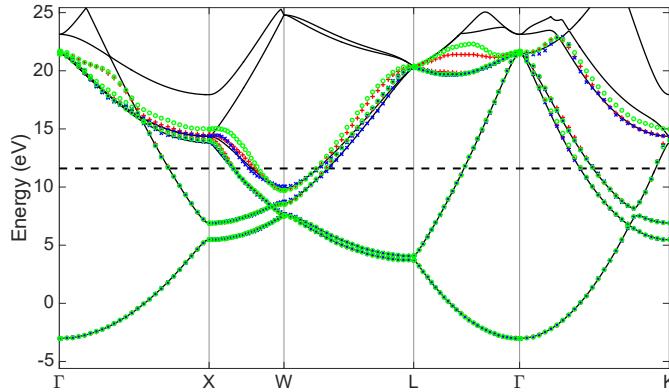


FIG. 7. *Wannier* interpolation of aluminum with 8  $\mathbf{k}$ -points per direction using (blue Xs) SCDM, (green circles) our variational formulation, and (red +s) *Wannier90* compared with a (black line) reference calculation. The frozen window is the region below the dotted black line.

TABLE 3  
Spreads of the four individual Wannier functions for aluminum and the final spread.

	Orbital spread ( $\text{\AA}^2$ )				Final spread ( $\text{\AA}^2$ )
Variational	2.01	2.02	2.02	2.03	8.07
Wannier90	2.09	2.1	2.11	2.11	8.41
SCDM	3.44	4.02	4.21	4.21	15.89

all three methods we then measure the error of band interpolation at or below the Fermi energy (8.42 eV). In all cases, we capped *Wannier90* at 5,000 disentanglement and spread reduction iterations and considered it converged at a tolerance of  $10^{-10}$ . Similarly, we considered our method converged if the objective function changed by

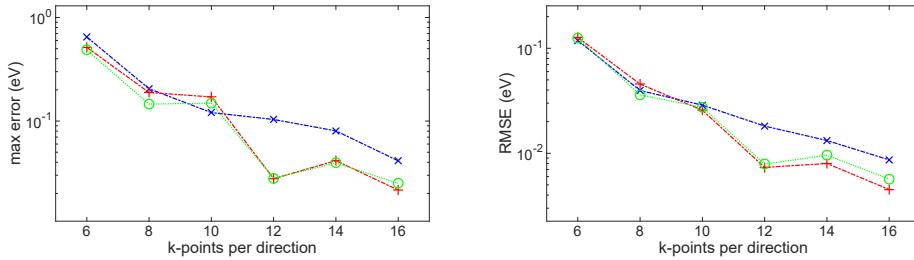


FIG. 8. *Max and RMSE in band interpolation for aluminum as computed using (blue Xs) SCDM, (green circles) our variational formulation, and (red +s) Wannier90.*

less than  $10^{-10}$  between successive iterates, or we reached 5,000 iterations. Typically our method took 100–250 iterations to converge, whereas the Wannier90 disentanglement would hit the iteration cap and spread reduction took 90–400 iterations to converge. Figure 8 shows broadly similar behavior for all three methods, though generally the two optimization methods do noticeably improve on the SCDM initial guess as more  $\mathbf{k}$ -points are used. We expect that asymptotically the optimization based methods should perform better. However, given the relatively small number of grid points per direction and the complexity of the band structure, it seems we are still in the preasymptotic regime.

**5.3. Copper.** Last, we illustrate the behavior of our method when used to interpolate seven conduction bands of copper around the Fermi energy. The main interest here is to interpolate the bands with  $d$ -orbital characters. Such bands may intersect with other bands not of interest within the energy window, and it is thus not easy to simply look at band interpolation error within the energy window. Rather, we place emphasis on the qualitative behavior of the interpolation and the orbital spread yielded by our variational formulation. Interestingly, in this case we observed a high sensitivity of Wannier90 to the SCDM initial guess based on the parameters (corresponding to “entangled case 2” in [8]—a Gaussian) and a contrasting robustness of our variational method. In all cases we used a frozen window of 13.5 to 17 eV and no outer window for both Wannier90 and our variational method.

When fixing the parameter  $\mu = 15.5$  and varying  $\sigma$  from 3.0 to 6.0, both SCDM and our variational method robustly generated good band interpolation.<sup>7</sup> However, for a range of  $\sigma$  Wannier90 with disentanglement failed to reach convergence. To sweep over several values of  $\sigma$  we limit Wannier90 to 5,000 iterations each for the disentanglement and spread minimization, and we limit our variational method to 1,000 iterations.

*Remark 1.* For  $\sigma = 5.0$ , where we observed particularly bad performance of Wannier90 (see below), we let it run for 100,000 iterations of both the disentanglement and spread. While the disentanglement procedure converged after roughly 20,000 iterations, the spread minimization failed to converge even after 100,000 iterations. While this does not guarantee that the local minimum that Wannier90 may eventually find is poor and could simply be the optimization algorithm behaving poorly, we feel this is a reasonable comparison to make even without Wannier90

<sup>7</sup>We observe reduced accuracy of the interpolation near the edge of the frozen window. This seems to be a consequence of the fact that the constraints are only enforced in the frozen window coupled with the presence of a band just outside the window.

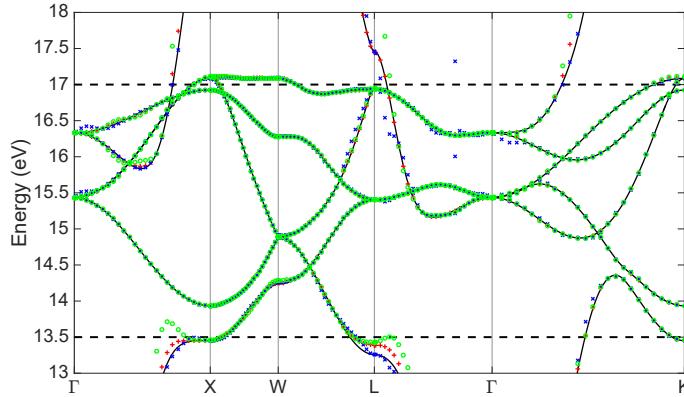


FIG. 9. Using an SCDM initial guess with  $\mu = 15.5$  and  $\sigma = 4.0$ , Wannier interpolation of copper with 10  $k$ -points using (blue Xs) SCDM, (green circles) our variational formulation, and (red +s) Wannier90 compared with a (black line) reference calculation. The frozen window is the region between the dotted black line.

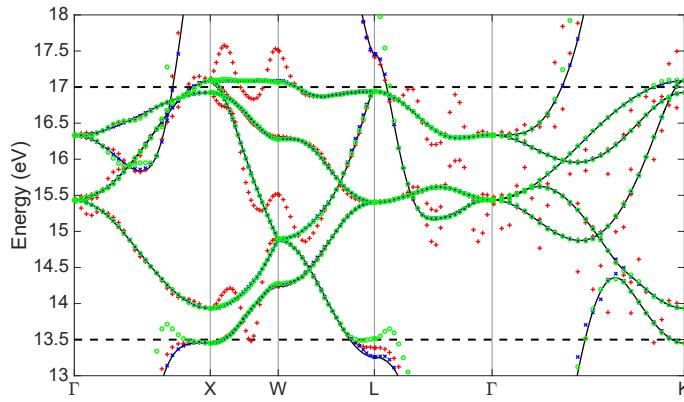


FIG. 10. Using an SCDM initial guess with  $\mu = 15.5$  and  $\sigma = 5.0$ , Wannier interpolation of Copper with 10  $k$ -points using (blue Xs) SCDM, (green circles) our variational formulation, and (red +s) Wannier90 compared with a (black line) reference calculation. The frozen window is the region between the dotted black line.

determining that it has converged. Interestingly, passing the output of Wannier90 in this setting to our variational method we were able to converge to a good gauge.

Figures 9 and 10 show the band interpolation of the three methods in the case where  $\sigma = 4.0$  and  $\sigma = 5.0$ . We also report the individual spreads of the orbitals in Tables 4 and 5. We observe that in both cases the SCDM initial guess and our optimized solution yield good band interpolation within the frozen window. In contrast, Wannier90 does not find a good local optima in the latter case and this results in poor interpolation quality. We further investigate this behavior in Table 6, where we report the total final spread of the methods as we vary  $\sigma$ . We see that for two of the parameter values Wannier90 failed to find a local optimum that is close to what our variational method finds.

TABLE 4

*Spreads of the seven individual Wannier functions for copper with the  $\sigma = 4.0$  initial guess.*

	Orbital spread ( $\text{\AA}^2$ )						
Variational	0.24	0.50	0.50	0.51	0.56	0.56	1.30
Wannier90	0.41	0.43	0.43	0.48	0.54	0.56	1.41
SCDM	1.49	2.34	2.46	3.06	3.13	3.68	7.50

TABLE 5

*Spreads of the seven individual Wannier functions for copper with the  $\sigma = 5.0$  initial guess.*

	Orbital spread ( $\text{\AA}^2$ )						
Variational	0.21	0.47	0.51	0.52	0.53	0.62	1.38
Wannier90	0.47	0.53	0.54	0.60	0.71	2.78	2.83
SCDM	1.21	1.47	1.56	1.78	1.97	2.53	8.87

TABLE 6

*Comparison of the spreads ( $\text{\AA}^2$ ) for copper as the  $\sigma$  parameter of the SCDM initial guess is varied.*

$\sigma$	3.0	3.5	4.0	4.5	5.0	5.5	6.0
SCDM	34.22	26.9	23.65	20.83	19.39	18.7	17.3
Our method	4.23	4.17	4.18	4.23	4.23	4.18	4.18
Wannier90	4.27	4.26	4.26	8.48	8.46	4.26	4.26

**6. Conclusion and discussion.** We have developed a variational formulation that, paired with a specific initialization strategy, is able to robustly construct Wannier functions for systems with entangled band structure. Importantly, the definition of Wannier functions must be generalized, allowing them to lie in a subspace that contains, but is larger than, the spectral subspace of interest. While this condition adds extra constraints to our optimization problem, and they can be phrased in many theoretically equivalent ways, we find one that is particularly convenient. This results in a formulation that matches that of partly occupied Wannier functions [36] and allows us to view the widely used disentanglement procedure as an alternating minimization algorithm—albeit one that only takes a single alternation step. As the underlying problem is nonconvex, our choice of initialization strategy via the SCDM methodology is key. As demonstrated with several real materials, our method is robust and effective at finding localized functions and enabling good quality band interpolation. Our variational formulation is versatile and can be modified relatively easily to accommodate additional constraints, such as symmetry constraints, for certain types of real materials. It would also be interesting to study the behavior of localization properties of generalized Wannier functions for systems with nontrivial topological characters.

We also study the free electron gas, providing interesting insights into the further theoretical study of the localization properties of generalized Wannier functions. We find that the minimization of the second moments of the Wannier functions only imposes a relatively weak algebraic decay. Our analysis suggests that, for real three-dimensional materials, the disentangled Wannier functions decay asymptotically slowly as well. Further localization is possible, but the method we present here is likely to only be useful for very fine  $\mathbf{k}$ -point grids. The computation of Wannier functions that are localized in both the preasymptotic and asymptotic regime remains an interesting open question.

**Appendix A. Gradient of the objective function.** If  $F$  is a function from a complex Hilbert space to  $\mathbb{R}$ , we recall that its gradient (sometimes written  $\partial F/\partial z^*$ ) is defined as the (unique) vector  $g$  such that

$$F(z + h) = F(z) + \operatorname{Re} \langle g, h \rangle + O(h^2).$$

All the derivatives and gradients below are in this sense. Note that this is different from the notion of derivative of a  $\mathbb{C} \rightarrow \mathbb{C}$  function, which is not relevant here. (Non-trivial complex-to-real functions are not complex differentiable.) The advantage of this definition is that it allows a straightforward translation of first-order (but not second-order) optimization algorithms in complex variables.

We use the numerical setup of [26]. Recall that the  $\mathbf{k}$ -point grid is discretized, with a total of  $N_{\mathbf{k}}$  points, and on each  $\mathbf{k}$  in that grid, we are given a  $N_g \times N_o$  matrix  $u_{\mathbf{k}}$  for each  $\mathbf{k}$ -point representing the  $N_o$  orbitals in the outer window discretized on a space of dimension  $N_g$ . The vectors  $\{\mathbf{b}\}$  are displacements from one  $\mathbf{k}$ -point to a set of neighbors, and  $w_{\mathbf{b}}$  are weights chosen to satisfy

$$\sum_{\mathbf{b}} w_{\mathbf{b}} \mathbf{b} \mathbf{b}^T = I_3,$$

so that the gradient of a function  $f(\mathbf{k})$  can be approximated by

$$\nabla f(\mathbf{k}) \approx \sum_{\mathbf{b}} w_{\mathbf{b}} (f(\mathbf{k} + \mathbf{b}) - f(\mathbf{k})) \mathbf{b}.$$

We look for a set of  $N_o \times N_w$  matrices  $\{U_{\mathbf{k}}\}$  with orthogonal columns, which define Wannier functions by (2.13). Let

$$M_{\mathbf{k}, \mathbf{b}, m, n} = \langle u_{m, \mathbf{k}}, u_{n, \mathbf{k} + \mathbf{b}} \rangle$$

be the  $N_o \times N_o$  overlap matrix between the bands, which is an input of the algorithm. Then the  $N_w \times N_w$  overlap matrix between the Wannier functions defined by  $U(\mathbf{k})$  is

$$N_{\mathbf{k}, \mathbf{b}} = U_{\mathbf{k}}^* M_{\mathbf{k}, \mathbf{b}} U_{\mathbf{k} + \mathbf{b}}.$$

The Marzari–Vanderbilt spread functional is given by

$$\begin{aligned} \Omega &= \sum_n \langle |\mathbf{r}|^2 \rangle_n - |\langle \mathbf{r} \rangle_n|^2, \text{ where} \\ \langle \mathbf{r}^2 \rangle_n &= \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}, \mathbf{b}} w_{\mathbf{b}} (1 - |N_{\mathbf{k} \mathbf{b} n n}|^2 + (\operatorname{Im} \ln N_{\mathbf{k} \mathbf{b} n n})^2) \\ \langle \mathbf{r} \rangle_n &= -\frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}, \mathbf{b}} w_{\mathbf{b}} \operatorname{Im} \ln N_{\mathbf{k} \mathbf{b} n n} \mathbf{b} \end{aligned}$$

[26, (11), (31), and (32)].

We need to compute  $\Omega(U + \Delta U)$  to first-order in  $\Delta U$ , from which we will identify  $\nabla \Omega$  by  $\Omega(U + \Delta U) - \Omega(U) = \operatorname{Re} \sum_{\mathbf{k}} \operatorname{Tr}((\nabla \Omega)_{\mathbf{k}}^* (\Delta U)_{\mathbf{k}}) + O(\Delta U^2)$ .

We begin with  $\sum_n \langle |\mathbf{r}|^2 \rangle_n$  and consider the following quantity:

$$I = \sum_{\mathbf{k}\mathbf{b}n} w_{\mathbf{b}} f(N_{\mathbf{k}\mathbf{b}nn}),$$

where  $f : \mathbb{C} \rightarrow \mathbb{R}$ . Then, using the fact that the set of vectors  $\mathbf{b}$  is symmetric (contains  $\mathbf{b}$  as well as  $-\mathbf{b}$ ) and that  $w_{-\mathbf{b}} = w_{\mathbf{b}}$ , we get

$$\begin{aligned} \Delta I &= \operatorname{Re} \sum_{\mathbf{k}\mathbf{b}n} w_{\mathbf{b}} f'(N_{\mathbf{k}\mathbf{b}nn})^* \Delta N_{\mathbf{k}\mathbf{b}nn} \\ &= \operatorname{Re} \sum_{\mathbf{k}\mathbf{b}n} w_{\mathbf{b}} f'(N_{\mathbf{k}\mathbf{b}nn})^* (\Delta U_{\mathbf{k}}^* M_{\mathbf{k}\mathbf{b}} U_{\mathbf{k}+\mathbf{b}})_{nn} + f'(N_{\mathbf{k}\mathbf{b}nn}) (\Delta U_{\mathbf{k}+\mathbf{b}}^* M_{\mathbf{k}\mathbf{b}}^* U_{\mathbf{k}})_{nn} \\ &= \operatorname{Re} \sum_{\mathbf{k}\mathbf{b}n} w_{\mathbf{b}} f'(N_{\mathbf{k}\mathbf{b}nn})^* (\Delta U_{\mathbf{k}}^* M_{\mathbf{k}\mathbf{b}} U_{\mathbf{k}+\mathbf{b}})_{nn} + f'(N_{\mathbf{k}-\mathbf{b},\mathbf{b},nn}) (\Delta U_{\mathbf{k}}^* M_{\mathbf{k}-\mathbf{b},\mathbf{b}}^* U_{\mathbf{k}-\mathbf{b}})_{nn} \\ &= \operatorname{Re} \sum_{\mathbf{k}\mathbf{b}n} w_{\mathbf{b}} f'(N_{\mathbf{k}\mathbf{b}nn})^* (\Delta U_{\mathbf{k}}^* M_{\mathbf{k}\mathbf{b}} U_{\mathbf{k}+\mathbf{b}})_{nn} + f'(N_{\mathbf{k},-\mathbf{b},nn}^*) (\Delta U_{\mathbf{k}}^* M_{\mathbf{k},-\mathbf{b}} U_{\mathbf{k}-\mathbf{b}})_{nn} \\ &= \operatorname{Re} \sum_{\mathbf{k}\mathbf{b}n} w_{\mathbf{b}} (f'(N_{\mathbf{k}\mathbf{b}nn})^* + f'(N_{\mathbf{k}\mathbf{b}nn}^*)) (\Delta U_{\mathbf{k}}^* M_{\mathbf{k}\mathbf{b}} U_{\mathbf{k}+\mathbf{b}})_{nn}, \end{aligned}$$

and the gradient is therefore

$$(\nabla I)_{\mathbf{k}mn} = \sum_{\mathbf{b}} w_{\mathbf{b}} (f'(N_{\mathbf{k}\mathbf{b}nn})^* + f'(N_{\mathbf{k}\mathbf{b}nn}^*)) (M_{\mathbf{k}\mathbf{b}} U_{\mathbf{k}+\mathbf{b}})_{mn}.$$

It can be checked using similar arguments that, when

$$I = \sum_{\mathbf{k}\mathbf{b}n} w_{\mathbf{b}} \mathbf{b} g(N_{\mathbf{k}\mathbf{b}nn}),$$

then

$$(\nabla I)_{\mathbf{k}mn} = - \sum_{\mathbf{b}} w_{\mathbf{b}} (g'(N_{\mathbf{k}\mathbf{b}nn})^* + g'(N_{\mathbf{k}\mathbf{b}nn}^*)) (M_{\mathbf{k}\mathbf{b}} U_{\mathbf{k}+\mathbf{b}})_{mn} \mathbf{b}.$$

Applying these formulas with

$$\begin{aligned} f(z) &= 1 - |z|^2 + (\operatorname{Im} \ln z)^2, & f'(z) &= -2z + 2 \frac{i \operatorname{Im} \ln z}{z^*}, \\ g(z) &= -\operatorname{Im} \ln z, & g'(z) &= -\frac{i}{z^*} \end{aligned}$$

we get

$$(\nabla \Omega)_{\mathbf{k}mn} = \frac{4}{N_{\mathbf{k}}} \sum_{\mathbf{b}} \left( -N_{\mathbf{k}\mathbf{b}nn}^* - i \frac{\operatorname{Im} \ln N_{\mathbf{k}\mathbf{b}nn} + \langle \mathbf{r} \rangle_n \cdot \mathbf{b}}{N_{\mathbf{k}\mathbf{b}nn}} \right) (M_{\mathbf{k}\mathbf{b}} U_{\mathbf{k}+\mathbf{b}})_{mn}.$$

Note that this is the *unconstrained* gradient of  $\Omega$  with respect to  $U$ . Using the chain rule with  $U = \begin{bmatrix} I_{N_f} & 0 \\ 0 & Y \end{bmatrix} X$  one can easily derive the gradient of  $\Omega$  with respect to  $(X, Y)$ . Then we simply have to minimize with respect to  $(X, Y)$  subject to the orthogonality constraints for  $X$  and  $Y$  using standard methods [12, 1].

## REFERENCES

- [1] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2009.
- [2] J. BEZANSON, A. EDELMAN, S. KARPINSKI, AND V. S., *Julia: A fresh approach to numerical computing*, SIAM Rev., 59 (2017), pp. 65–98.
- [3] E. I. BLOUNT, *Formalisms of band theory*, Solid State Phys., 13 (1962), pp. 305–373.
- [4] D. R. BOWLER AND T. MIYAZAKI,  *$O(N)$  methods in electronic structure calculations*, Rep. Progr. Phys., 75 (2012), 036503.
- [5] C. BROUDER, G. PANATI, M. CALANDRA, C. MOUROUGANE, AND N. MARZARI, *Exponential localization of Wannier functions in insulators*, Phys. Rev. Lett., 98 (2007), 046402.
- [6] E. CANCÈS, A. LEVITT, G. PANATI, AND G. STOLTZ, *Robust determination of maximally-localized Wannier functions*, Phys. Rev. B, 95 (2017), 075114.
- [7] H. CORNEAN, D. GONTIER, A. LEVITT, AND D. MONACO, *Localised Wannier Functions in Metallic Systems*, preprint, arXiv:1712.07954, 2017.
- [8] A. DAMLE AND L. LIN, *Disentanglement via entanglement: A unified method for Wannier localization*, Multiscale Model. Simul., 16 (2018), pp. 1392–1410.
- [9] 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 (2015), pp. 1463–1469.
- [10] A. DAMLE, L. LIN, AND L. YING, *SCDM-K: Localized orbitals for solids via selected columns of the density matrix*, J. Comput. Phys., 334 (2017), pp. 1–15.
- [11] W. E, T. LI, AND J. LU, *Localized bases of eigensubspaces and operator compression*, Proc. Natl. Acad. Sci. USA, 107 (2010), pp. 1273–1278.
- [12] A. EDELMAN, T. ARIAS, AND S. SMITH, *The geometry of algorithms with orthogonality constraints*, SIAM J. Matrix Anal. Appl., 20 (1998), pp. 303–353.
- [13] J. M. FOSTER AND S. F. BOYS, *Canonical configurational interaction procedure*, Rev. Modern Phys., 32 (1960), p. 300.
- [14] P. GIANNONI, S. BARONI, N. BONINI, M. CALANDRA, R. CAR, C. CAVAZZONI, D. CERESOLI, G. L. CHIAROTTI, M. COCCIONI, I. DABO, A. D. CORSO, S. DE GIRONCOLI, S. FABRIS, G. FRATESI, R. GEBAUER, U. GERSTMANN, C. GOUGOUSSIS, A. KOKALJ, M. LAZZERI, L. MARTIN-SAMOS, N. MARZARI, F. MAURI, R. MAZZARELLO, S. PAOLINI, A. PASQUARELLO, L. PAULATTO, C. SBRACCIA, S. SCANDOLO, G. SCLAUZERO, A. P. SEITSONEN, A. SMOGUNOV, P. UMARI, AND R. M. WENTZCOVITCH, *Quantum Espresso: A modular and open-source software project for quantum simulations of materials*, J. Phys. Condens. Matter, 21 (2009), pp. 395502–395520.
- [15] S. GOEDECKER, *Linear scaling electronic structure methods*, Rev. Modern Phys., 71 (1999), pp. 1085–1123.
- [16] F. GYGI, *Compact representations of Kohn–Sham invariant subspaces*, Phys. Rev. Lett., 102 (2009), 166406.
- [17] W. W. HAGER AND H. ZHANG, *A new conjugate gradient method with guaranteed descent and an efficient line search*, SIAM J. Optim., 16 (2005), pp. 170–192.
- [18] P. HOHENBERG AND W. KOHN, *Inhomogeneous electron gas*, Phys. Rev., 136 (1964), pp. B864–B871.
- [19] E. KOCH AND S. GOEDECKER, *Locality properties and Wannier functions for interacting systems*, Solid State Commun., 119 (2001), 105.
- [20] W. KOHN, *Analytic properties of Bloch waves and Wannier functions*, Phys. Rev., 115 (1959), 809.
- [21] W. KOHN, *Density functional and density matrix method scaling linearly with the number of atoms*, Phys. Rev. Lett., 76 (1996), pp. 3168–3171.
- [22] W. KOHN AND L. SHAM, *Self-consistent equations including exchange and correlation effects*, Phys. Rev., 140 (1965), pp. A1133–A1138.
- [23] P. KUCHMINT, *An overview of periodic elliptic operators*, Bull. Amer. Math. Soc., 53 (2016), pp. 343–414.
- [24] P.-O. LÖWDIN, *On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals*, J. Chem. Phys., 18 (1950), pp. 365–375.
- [25] N. MARZARI, A. A. MOSTOFI, J. R. YATES, I. SOUZA, AND D. VANDERBILT, *Maximally localized Wannier functions: Theory and applications*, Rev. Modern Phys., 84 (2012), pp. 1419–1475.
- [26] N. MARZARI AND D. VANDERBILT, *Maximally localized generalized Wannier functions for composite energy bands*, Phys. Rev. B, 56 (1997), 12847.
- [27] A. MOSTOFI, J. YATES, Y. LEE, I. SOUZA, D. VANDERBILT, AND N. MARZARI, *Wannier90: A tool for obtaining maximally-localised Wannier functions*, Comput. Phys. Commun., 178 (2008), pp. 685–699.

- [28] A. MOSTOFI, J. YATES, G. PIZZI, Y. 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 (2014), pp. 2309–2310.
- [29] 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 (2015), 165134.
- [30] G. NENCIU, *Dynamics of band electrons in electric and magnetic fields: Rigorous justification of the effective Hamiltonians*, Rev. Modern Phys., 63 (1991), pp. 91–127, <https://doi.org/10.1103/RevModPhys.63.91>.
- [31] J. NOCEDAL AND S. WRIGHT, *Numerical Optimization*, Springer, New York, 2006.
- [32] V. OZOLIŅŠ, R. LAI, R. CAFLISCH, AND S. OSHER, *Compressed modes for variational problems in mathematics and physics*, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 18368–18373.
- [33] G. PANATI, *Triviality of Bloch and Bloch-Dirac bundles*, Ann. Henri Poincaré, 8 (2007), pp. 995–1011.
- [34] G. PANATI AND A. PISANTE, *Bloch bundles, Marzari-Vanderbilt functional and maximally localized Wannier functions*, Comm. Math. Phys., 322 (2013), pp. 835–875.
- [35] I. SOUZA, N. MARZARI, AND D. VANDERBILT, *Maximally localized Wannier functions for entangled energy bands*, Phys. Rev. B, 65 (2001), 035109.
- [36] K. S. THYGESEN, L. B. HANSEN, AND K. W. JACOBSEN, *Partly occupied Wannier functions*, Phys. Rev. Lett., 94 (2005), 026405, <https://doi.org/10.1103/PhysRevLett.94.026405>.
- [37] J. VON NEUMANN AND E. WIGNER, *Über merkwürdige diskrete eigenwerte*, in The Collected Works of Eugene Paul Wigner, Springer, New York, 1993, pp. 291–293.
- [38] G. H. WANNIER, *The structure of electronic excitation levels in insulating crystals*, Phys. Rev., 52 (1937), 191.
- [39] J. R. YATES, X. WANG, D. VANDERBILT, AND I. SOUZA, *Spectral and Fermi surface properties from Wannier interpolation*, Phys. Rev. B, 75 (2007), 195121.