4 Wannier Functions and Construction of Model Hamiltonians

Jan Kuneš

Institute of Physics, Academy of Sciences
Praha, Czech Republic

Contents

1   Introduction

Although introduced already in 1937, Wannier functions might be enjoying their 'golden age' right now, thanks to the rapid growth of methods linking first-principles band structure calculations with model theories based on the second quantization formalism, such as LDA+DMFT. While using Wannier functions as a computational tool is the likely goal of the reader, there is another fascinating aspect of Wannier functions, namely the connection between the spatial localization of Wannier functions and the topological properties of the corresponding Bloch states. It is this topology that gave the name to the topological insulators.

In the first part of these notes we establish the connection between the smoothness and periodicity of Bloch functions as functions of the electron quasi-momenta and the exponential localization of the corresponding Wannier functions. As this is a rather mathematical topic we do not attempt a comprehensive presentation. Our goal is to point out the general ideas and concepts and to direct the reader to the original literature. In the second part of the notes we present the commonly used computational methods for construction of the Wannier functions and examples of their application.

1.1   Electron in periodic potential

The introduction of density functional theory [1] started the era of *ab initio* calculations of electronic structure for real materials. The Kohn-Sham equation has the form of Schrödinger equation for non-interacting electrons

$$-\nabla^2 \psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = \epsilon \psi(\mathbf{r}), \tag{1}$$

where we chose atomic units with $\hbar = 1$, $m_e = 1/2$. With periodic boundary conditions (at infinity) and the requirement that $\psi(\mathbf{r})$ is normalizable, the equation represents an eigenvalue problem for a Hermitian Hamiltonian $H = -\nabla^2 + V(\mathbf{r})$. If the crystal potential $V(\mathbf{r})$ possesses translational symmetry

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}), \tag{2}$$

where $\mathbf{R}$ is a vector of the crystal lattice, the Bloch theorem [2] allows a partial diagonalization. The eigenfunctions of H can be written in the form

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}), \tag{3}$$

where $u_{n,\mathbf{k}}(\mathbf{r})$ is an $\mathbf{r}$-periodic function, n is a discrete band index, and $\mathbf{k}$ is a continuous index, a vector from the first Brillouin zone. The corresponding eigenvalues $\epsilon_{n,\mathbf{k}}$ are continuous and periodic functions of $\mathbf{k}$ – they are said to form energy bands. The cell periodic functions $u_{n,\mathbf{k}}(\mathbf{r})$ obey the equation

$$(-i\nabla - \mathbf{k})^2 u_{n,\mathbf{k}}(\mathbf{r}) + V(\mathbf{r})u_{n,\mathbf{k}}(\mathbf{r}) = \epsilon_{n,\mathbf{k}} u_{n,\mathbf{k}}(\mathbf{r}), \tag{4}$$

with periodic boundary conditions in the unit cell. Besides mathematical elegance, the Bloch theorem is crucial for numerical applications since it replaces the Hamiltonian H with continuous spectrum, by a (continuous) set of Hamiltonians $H_{\mathbf{k}} = (-i\nabla - \mathbf{k})^2 + V(\mathbf{r})$ with discrete spectra, i.e., a problem which is numerically tractable with standard methods of linear algebra. One of the key characteristics of any material is the presence or absence of an energy gap separating the ground state from the excitations, which translates into presence or absence of a direct gap between the occupied valence bands and empty conduction bands in the case of non-interacting electrons in a periodic solid. On the other hand, for the existence and other properties of the Wannier functions the band filling is irrelevant. Also the distinction between direct and indirect gaps does not matter for the properties of Wannier functions. We say that the nth band is isolated if $\epsilon_{n-1,\mathbf{k}} < \epsilon_{n,\mathbf{k}} < \epsilon_{n+1,\mathbf{k}}$ for each $\mathbf{k}$, i.e., the nth band is separated by a finite, possibly indirect, gap from the bands below and above. Similarly a composite band is isolated if $\epsilon_{n_{\min}-1,\mathbf{k}} < \epsilon_{n_{\min},\mathbf{k}} \leq \epsilon_{n_{\max},\mathbf{k}} < \epsilon_{n_{\max}+1,\mathbf{k}}$. Bands that are not isolated will be called entangled.

1.2 Why localized basis?

The reasons why the description of materials in terms of localized orbitals is attractive are both conceptual and technical. By conceptual we mean those features that do not simplify computations, but provide better insights into the physics. For example, the chemist's language of chemical bonds is 'difficult' to understand for a physicist speaking the 'Bloch wave' language. The path from a set of isolated atoms with the *localized* atomic orbitals to a periodic solid with *extended* Bloch states may be difficult to grasp, as it involves qualitative differences such as localized versus extended or discrete versus continuous. The Wannier functions provide a natural extension of the concept of atomic orbitals into solids and thus bridge this gap. Another example is provided by the theory of dielectric polarization [3]. While in terms of Bloch states the polarization is expressed through a rather abstract concept of Berry phase, the formulation in terms of Wannier functions uses a simple notion of the center of mass of the corresponding charge distribution.

Many important physical properties and phenomena involve spatially localized objects such as impurities or defects in the crystal structure, screened electron-electron interaction, composite excitations such as excitons or polarons. The technical advantages of localized orbitals in the context of local electronic correlations are evident. As we typically consider only the short-range part of the electron-electron interaction explicitly, expressing the interaction in terms of localized orbitals reduces the number of non-zero terms considerably (for a more detailed discussion of the Hubbard model see section 5.5).

While atomic orbitals appear to provide a natural framework there is a problem. As put by Wannier [4] in the context of semiconductor physics: "It would no doubt be more satisfactory for insulating crystals, to discuss the Hamiltonian using atomic functions rather than Bloch functions. But this line of attack has been hampered by the fact that atomic functions are not orthogonal". Obviously, the orthogonality of a basis set is a great advantage, in particular when the formalism of second quantization is to be used.

2   Preliminaries

2.1   Orthogonal atomic orbitals

There are various ways to introduce the Wannier functions (WFs). Let us follow Wannier [4] for a moment and consider the following warm-up problem. Assume that we have a solid formed by atoms sitting on lattice sites $\mathbf{R}$ with a single valence orbital described by an atomic wave function $v(\mathbf{x})$, i.e., the energies of the other atomic orbitals are far away so that their effect can be neglected. The basis $\{v(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$ is in general non-orthogonal. How to orthogonalize it while preserving the explicit translational symmetry of the basis set? A straightforward Gram-Schmidt procedure does break the translational symmetry of the basis. Instead, the following construction may be used.

First, we form the Bloch sum (discrete Fourier transform) of the atomic orbitals

$$\phi(\mathbf{k}, \mathbf{x}) = \frac{A(\mathbf{k})}{N^{1/2}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \, v(\mathbf{x} - \mathbf{R}), \tag{5}$$

where the summation runs over the lattice vectors $\mathbf{R}$. The normalization constant $A(\mathbf{k})$ is given by

$$\frac{1}{A(\mathbf{k})^2} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \int d\mathbf{x} \, v^*(\mathbf{x}) v(\mathbf{x} - \mathbf{R}). \tag{6}$$

It is easy to verify that the basis $\{\phi(\mathbf{k}, \mathbf{x})\}_{\mathbf{k}}$, indexed by the reciprocal vectors $\mathbf{k}$, is orthonormal. Next, we perform the inverse Fourier transform

$$w(\mathbf{R}, \mathbf{x}) = \frac{1}{N^{1/2}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} \phi(\mathbf{k}, \mathbf{x}) = \sum_{\mathbf{R}'} c(\mathbf{R} - \mathbf{R}') v(\mathbf{x} - \mathbf{R}'). \tag{7}$$

The new functions $w(\mathbf{R}, \mathbf{x})$ are combinations of the original atomic functions $v(\mathbf{x} - \mathbf{R}')$ centered at different lattice sites with coefficients $c(\mathbf{R})$ given by

$$c(\mathbf{R}) = \frac{1}{N} \sum_{\mathbf{k}} A(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{R}}. \tag{8}$$

Let us make a few observations about $w(\mathbf{R}, \mathbf{x})$. First, since the coefficients $c(\mathbf{R} - \mathbf{R}')$ in (7) depend only on the difference $\mathbf{R} - \mathbf{R}'$ we can write $w(\mathbf{R}, \mathbf{x}) \equiv w(\mathbf{x} - \mathbf{R})$, i.e. $w(\mathbf{R}, \mathbf{x})$ on different lattice sites are shifted images of the same functional form $w(\mathbf{x})$. Second, since (7) is a unitary transformation of an orthonormal basis $\{\phi(\mathbf{k}, \mathbf{x})\}_{\mathbf{k}}$ the basis $\{w(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$ is also orthonormal as can be verified by explicit calculation. In fact, we have performed forward and inverse Fourier transformation, but we did not get the original atomic functions. The 'magic' is in the normalization factor $A(\mathbf{k})$, which was introduced between the Fourier transforms. Let us look at the behavior of $A(\mathbf{k})$ in more detail. If the atomic functions $v(\mathbf{x} - \mathbf{R})$ do not overlap at all, expression (6) returns $A(\mathbf{k}) = 1$ and the inverse Fourier transform (7) recovers the original atomic functions. Once the overlap between $v(\mathbf{x} - \mathbf{R})$ on different sites is finite, the $\mathbf{k}$-dependence of $A(\mathbf{k})$ leads to admixture of atomic functions from neighboring sites in $w(\mathbf{x} - \mathbf{R})$. Generally, the larger the overlap of $v(\mathbf{x} - \mathbf{R})$'s the stronger the $\mathbf{k}$-dependence of

$A(\mathbf{k})$ and the less the $w(\mathbf{R} - \mathbf{x})$ resembles the starting atomic functions. For exponentially localized atomic functions the orthonormal $w(\mathbf{R} - \mathbf{x})$ are also exponentially localized [4].

Since the orbitals $\{v(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$ give rise to a single band we can say that we have constructed a localized, translationally invariant orthonormal basis $\{w(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$, which spans the Hilbert space of this band.

Typically, in numerical computations we do not not have an atomic basis for a given band to start with. Thus a question arises whether a similar construction is possible for an arbitrary band and how it can be realized if we know the Bloch eigenstates in some basis, as is usually the case.

2.2 Asymptotic behavior of Fourier coefficients

Localization, that is how fast a function decays away from its center of mass, is one of the central issues concerning Wannier orbitals. Since the Wannier functions are, vaguely speaking, Fourier transforms of the Bloch functions let us review some properties of the Fourier series [5]. We do not aspire to provide a full mathematical background of the complicated issue of localization. Our intention is to point out the intimate connection between smoothness of a function and the convergence (localization) of its Fourier series.

The Fourier coefficients of a function $f(x)$ of a real variable x integrable (L_1 integrable) on the interval $[0, 2\pi)$ are defined by

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} dx\, e^{-inx} f(x) \tag{9}$$

and the corresponding Fourier series reads

$$F(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx}. \tag{10}$$

Let us assume that $f(x)$ is periodic, $f(x) = f(x + 2\pi)$, and piecewise smooth up to the second derivative, i.e., there is a finite number of points α_i where the first derivative $f^{(1)}(x)$ is discontinuous and $f^{(1)}(x)$ and $f^{(2)}(x)$ are continuous and finite on every interval $[\alpha_i, \alpha_{i+1}]$. This is sufficient for the series (10) to converge uniformly towards $f(x)$ and for the Fourier coefficients to converge as $|a_n| < K/n^2$, where K is a finite constant. Differentiating (10) with respect to x and applying the above statement to higher derivatives of $f(x)$ it is easy to see that the smoother the function, the faster its Fourier coefficients decay. In general, if a discontinuity appears first in the kth-derivative, the Fourier coefficients decay as K/n^{k+1}. Therefore a necessary condition for an exponential decay $a_n \propto \exp(-Kn)$ is that $f(x)$ is smooth, i.e., has derivatives to an arbitrary order. However, this is not sufficient as is shown by an example at the end of this section.

To derive a sufficient condition for the exponential decay of a_n, let us assume that $f(x)$ can be analytically continued into the complex plane. Introducing a new variable $z = e^{ix}$, we map the real interval $[0, 2\pi)$ onto a unit circle and define a function $\tilde{f}(z)$ by $\tilde{f}(e^{ix}) = f(x)$. Now, let
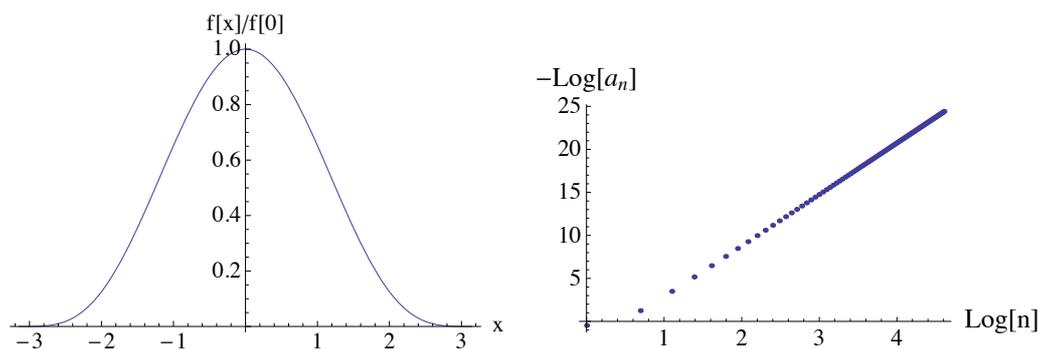
Fig. 1: *The graph of $x \to \left(1 - \frac{x}{\pi}\right)^4 \left(1 + \frac{x}{\pi}\right)^4$ (left) and its Fourier coefficients a_n in a log-log plot (right).*

us assume that $\tilde{f}(z)$ is analytic on a ring $r < 1 < R$ which implies that there exists a Laurent series

$$\tilde{f}(z) = \sum_{n=-\infty}^{\infty} b_n z^n, \tag{11}$$

where the coefficients b_n are given by an integral over a contour γ, which encircles zero and lies inside the ring $r < z < R$,

$$b_n = \frac{1}{2\pi i} \oint_\gamma \frac{\tilde{f}(z)dz}{z^{n+1}}. \tag{12}$$

Taking the unit circle for γ and comparing to (9) one sees that $a_n = b_n$. Let us define $\rho = \min(R, 1/r)$. The analyticity of $\tilde{f}(z)$ implies that the series (11) is absolutely convergent for any real x from the interval $1/\rho < x < \rho$, but diverges for either $x > \rho$, $x < 1/\rho$, or both. This means that $b_{|n|} x^{|n|} \to 0$ for $x < \rho$, but $b_{|n|} x^{|n|} \to \infty$ for $x > \rho$. This is the precise meaning of $b_{|n|} \sim \exp\left(-|n|h\right)$, where $h = \ln \rho$.

We close this section with a few examples of the convergence of Fourier series for functions defined on $[-\pi, \pi)$ and periodically repeated over the real axis. In Fig. 1 we start with $\left(1 - x/\pi\right)^4 \left(1 + x/\pi\right)^4$, which has a discontinuous 5th derivative at $x = \pm\pi$. Its Fourier coefficients exhibit a power law decay, which appears linear in the log-log plot, Fig. 1.

Function $\left(1 + \cos(x)\right)\sqrt{1.1 + \cos(x)}$ provides an example of an analytic function with an infinite Fourier series. The Fourier coefficients decay exponentially as shown in Fig. 2

Finally we study the Fourier series of a 'bump' function $\exp\left(-1/[1 - (x/\pi)^2]\right)$, which is smooth, but not analytic. Its Fourier coefficients exhibit $|a_n| \sim |\pi n|^{-3/4} exp(-\sqrt{|\pi n|})$ asymptotic behavior as can be shown by the saddle point method [6]. The decay of a_n is obviously faster than a power law, but slower than exponential. The actual Fourier coefficients obtained by numerical integration are shown in Fig. 3.
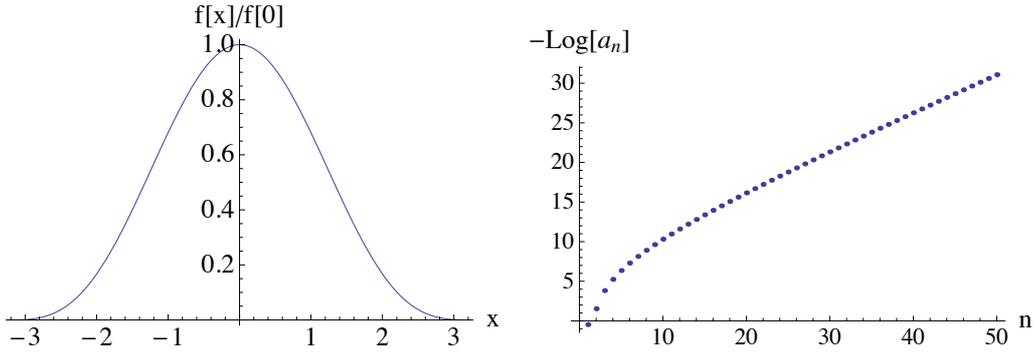
Fig. 2: *The graph of $x \to \left(1 + \cos(x)\right)\sqrt{1.1 + \cos(x)}$ (left) and its Fourier coefficients a_n in a log-lin plot (right).*
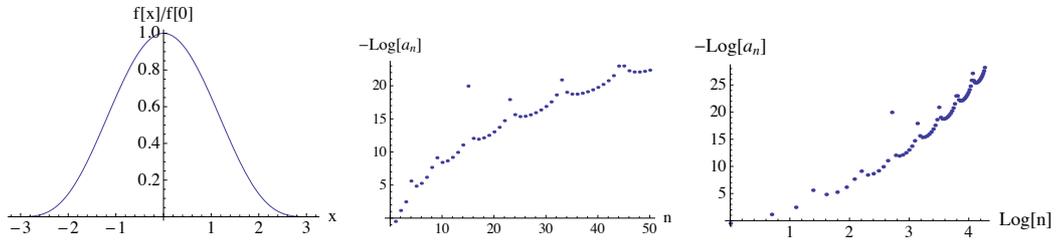
Fig. 3: *The graph of $x \to \exp\left(-\frac{1}{1-(x/\pi)^2}\right)$ (left) and its Fourier coefficients a_n in a log-lin (center) and log-log (right) plot.*

3 Wannier functions

3.1 Basic definitions

The Wannier function (WF) of an isolated band is defined as

$$w_n(\mathbf{r} - \mathbf{R}) = \frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} d\mathbf{k}\, e^{-i\mathbf{k}\cdot\mathbf{R}} \psi_{n,\mathbf{k}}(\mathbf{r}), \tag{13}$$

where V is the unit cell volume, d is the dimension, $\psi_{n,\mathbf{k}}(\mathbf{r})$ are the Bloch functions (3) corresponding to the nth band. The integration runs over the first Brillouin zone. The inverse transformation reads

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} w_n(\mathbf{r} - \mathbf{R}). \tag{14}$$

The transformation properties of the Bloch functions under lattice translation (3) ensure the mutual orthogonality of WFs centered in different unit cells. The definition (13) is not unique since the overall phase associated with the Bloch function is arbitrary. For reasons that become clear later we will consider only situations where $\psi_{n,\mathbf{k}}(\mathbf{r})$ is a smooth function of $\mathbf{k}$. A particular choice of the phases will be called a gauge and the transformation

$$\psi_{n,\mathbf{k}}(\mathbf{r}) \to e^{i\phi(\mathbf{k})} \psi_{n,\mathbf{k}}(\mathbf{r}), \tag{15}$$

between different gauges, where $\phi(\mathbf{k})$ is an analytic real-valued function, will be called a gauge transformation.

The generalization to the case of isolated composite bands is straightforward.

$$\mathbf{w}(\mathbf{r} - \mathbf{R}) = \frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} d\mathbf{k}\, \mathbf{U}(\mathbf{k})\boldsymbol{\psi}_{\mathbf{k}}(\mathbf{r}), \qquad (16)$$

where $\mathbf{w}(\mathbf{r})$ and $\boldsymbol{\psi}_{\mathbf{k}}(\mathbf{r})$ are vectors

$$\mathbf{w} = \begin{pmatrix} w_{n_{\min}} \\ \vdots \\ w_{n_{\max}} \end{pmatrix}, \quad \boldsymbol{\psi}_{\mathbf{k}} = \begin{pmatrix} \psi_{n_{\min},\mathbf{k}} \\ \vdots \\ \psi_{n_{\max},\mathbf{k}} \end{pmatrix} \qquad (17)$$

in the band index spanning the composite band, and $\mathbf{U}(\mathbf{k})$ is a unitary matrix acting on the band index. We define a quasi-inverse transformation

$$\tilde{\psi}_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \mathbf{w}(\mathbf{r} - \mathbf{R}), \qquad (18)$$

where the quasi Bloch states $\tilde{\psi}_{\mathbf{k}}(\mathbf{r})$ have the transformation property (3), but are not the eigenstates of the Hamiltonian. The gauge transformation (15) generalizes to a set of unitary transformations $\mathbf{U}(\mathbf{k})$ analytic and periodic in $\mathbf{k}$.

The construction of Wannier functions poses several non-trivial problems.

Existence. Given an isolated band, is it possible to span the corresponding Hilbert space with a basis of exponentially localized Wannier functions? As discussed below an equivalent question reads: Is it possible to find a gauge such that the resulting Bloch states are periodic and analytic functions of $\mathbf{k}$? The requirement of periodicity has topological implications and it turns out that the existence or non-existence of exponentially localized Wannier function is a topological characteristic of a given band.

Multiple bands. How do the arguments about existence of the exponentially localized WFs generalize to the case of isolated composite bands? Mathematically the difference between single and multiple bands lies in the Abelian (commutative multiplication of numbers) or non-Abelian (non-commutative multiplication of matrices) character of the respective gauge transformations. As a results, the technique used originally for a single band cannot be simply generalized to a composite band. For example, the results for composite bands depend on the dimension of the lattice.

Entanglement. Often it is not possible to find an isolated (composite) band spanning the desired part of the Hilbert space. Nevertheless, we would like to have a localized WFs basis representing the bands of interest, if only approximately. What is the optimal procedure to obtain WFs with such property?

Uniqueness. Assuming we can construct exponentially localized WFs, what is the meaning of the remaining gauge freedom? Under which additional conditions are the WFs defined uniquely and what are their physical consequences?

Implementation. How to compute the Wannier functions numerically? The theory of Wannier functions operates with the concepts of analytic continuation or topology with respect to $\mathbf{k}$,

which are inherently connected to the notion of continuity. How to perform practical calculations in the computer, which deal inevitably with discrete quantities?

In the following we will discuss these issues in some detail.

3.2 Existence of exponentially localized Wannier functions – single band

In the following we will sketch the conditions and prove of the existence of exponentially localized WFs for a single isolated band.

Let us assume that a given isolated band has eigenvalues $\{\epsilon_{\mathbf{k}}\}$ and the corresponding Bloch states $\{|\psi_{\mathbf{k}}\rangle\}$. The key ingredients for exponential localization are analyticity and periodicity as a function of $\mathbf{k}$ (see 2.2). The existence of exponentially localized WFs is equivalent to the possibility of finding a set of Bloch states such that $\langle\mathbf{r}|\psi_{\mathbf{k}}\rangle$ is a periodic and analytic function of $\mathbf{k}$ throughout the Brillouin zone. To appreciate the question of localization one has to realize that the eigenvalue problem does not specify $\langle\mathbf{r}|\psi_{\mathbf{k}}\rangle$ uniquely, but only up to an arbitrary multiplicative phase factor, which may differ from $\mathbf{k}$-point to $\mathbf{k}$-point. While the exponentially localized WFs exist under rather general conditions (see below), it is not true that they exist for every Hermitian periodic Hamiltonian.

The first partial answer to the question of existence of exponentially localized WFs in **1D** was given by Kohn [7] for non-relativistic electrons in a periodic potential with inversion symmetry. He proved the existence of exponentially localized WFs as well as their uniqueness upon the requirement of reality and definite parity. Kohn studied the solutions of the initial-value problem for the Schrödinger equation

$$-\frac{d^2}{dx^2}\phi_{1,2}(E,x) + V(x)\phi_{1,2}(E,x) = E\phi_{1,2}(E,x) \tag{19}$$

as a function of the energy E (the indices 1 and 2 stand for the two linearly independent solutions corresponding to different initial conditions). The solutions obtained this way are analytic functions of E. Kohn showed that Bloch functions, defined by the periodicity condition $\psi(E, x+a) = \lambda\psi(E,x)$, can be build as $\psi(E,x) = \alpha(E)\phi_1(E,x) + \beta(E)\phi_2(E,x)$ in such a way that analyticity as a function of E and in turn as a function of k is preserved. It turns out that while analyticity of $\langle x|\psi_k\rangle$ can be satisfied relatively easily, the requirement of periodicity on top of it, is a rather stringent condition.

A straightforward generalization of Kohn's global analysis, based on the initial value problem to the Schrödinger equation, to **higher dimensions** is not possible. To analyze the dimensions two and higher des Cloizeaux [8] introduced the band projection operator $P(\mathbf{k})$

$$P(\mathbf{k}) = |\psi_{\mathbf{k}}\rangle\langle\psi_{\mathbf{k}}| \tag{20}$$

and proved the analyticity in $\mathbf{k}$ of its $\mathbf{r}$-elements $\langle\mathbf{r}|P(\mathbf{k})|\mathbf{r}'\rangle$ as a function of $\mathbf{k}$ for an isolated as well as composite band. While $\mathbf{k}$-analyticity of Bloch states $\langle\mathbf{r}|\psi_{\mathbf{k}}\rangle$ implies the analyticity of $\langle\mathbf{r}|P(\mathbf{k})|\mathbf{r}'\rangle$, the opposite is not true. To illustrate the problem we follow Ref. [9] and try to
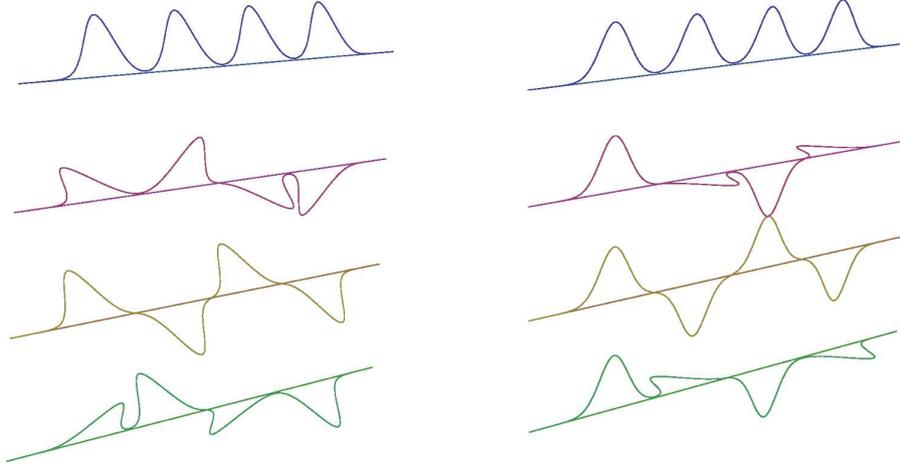
Fig. 4: *Cartoon depiction of the projection method. The Bloch functions of different* **k** *with random overall phases (left) and the same functions with the aligned phase on the first site (right).*

construct the **k**-analytic Bloch state $|\psi_{\mathbf{k}}\rangle$ as

$$\langle\mathbf{r}|\psi_{\mathbf{k}}\rangle = G^{-1/2}(\mathbf{k})\langle\mathbf{r}|P(\mathbf{k})|M_0\rangle, \tag{21}$$

$$G(\mathbf{k}) = \langle M_0|P(\mathbf{k})|M_0\rangle, \tag{22}$$

where $|M_0\rangle$ is an arbitrary trial orbital. The idea is to let the Bloch functions on the lattice site specified by the trial orbital $|M_0\rangle$ add up with the same sign (constructive interference) and to interfere destructively on the other lattice sites as depicted in Fig. 4. The Bloch function $|\psi_{\mathbf{k}}\rangle$ defined by (21) is analytic as long as the non-negative $G(\mathbf{k})$ is different from 0. Thus points $G(\mathbf{k}_0) = 0$ need special attention. As $G(\mathbf{k})$ is analytic, $G(\mathbf{k} - \mathbf{k}_0) \propto (\mathbf{k} - \mathbf{k}_0)^{2p}$ must be a quadratic (or even higher) form. Here again the 1D is distinct, as only in 1D a square root $G^{1/2}(\mathbf{k})$ analytic at $\mathbf{k}_0$ is guaranteed to exist, but can possibly change sign. Assuming time-reversal symmetry implies that zeros of $G(\mathbf{k})$ appear in pairs $\mathbf{k}_0$ and $-\mathbf{k}_0$ due to Kramers degeneracy (the cases $G(0) = 0$ and $G(\pi) = 0$ may be excluded by the choice of $|M_0\rangle$). Thus in a 1D time-reversal symmetric system a periodic and analytic $G^{1/2}(\mathbf{k})$ exists. It can be shown that in this case the zeros of $G^{1/2}(\mathbf{k})$ are canceled out by the corresponding zeros in $\langle\mathbf{r}|P(\mathbf{k}_0)|M_0\rangle$ and $\langle\mathbf{r}|\psi_{\mathbf{k}}\rangle$ (21) is periodic and analytic.

Similar argument cannot be used in higher dimensions since singularities of $G^{1/2}(\mathbf{k})$ at zeros of $G(\mathbf{k})$ cannot be removed. The present technique allows to prove that exponentially localized Wannier functions exist in time-reversal invariant systems with inversion symmetry or systems that can be adiabatically connected (i.e., keeping the band isolated along the adiabatic path) to a system with localized WFs [9]. On the other hand, in 2D quantum Hall systems (i.e., no time-reversal symmetry) one can prove that a zero of $G(\mathbf{k})$ exists for any choice of $|M_0\rangle$ [10] and that exponentially localized WFs do not exist for bands carrying the Hall current. Using a different approach based on the band projection operator Nenciu proved the existence of exponentially

localized Wannier functions for an isolated band in systems with time reversal symmetry in arbitrary dimension [11].

3.3 Existence of exponentially localized Wannier functions – composite bands

The arguments from the previous section cannot be simply generalized to the case of composite bands. It is nevertheless possible to prove that time-reversal invariance of the Hamiltonian is sufficient for exponentially localized WFs to exist for spinless fermions in 2D and 3D [12]. The proof relies on the concepts of connection, curvature, and Chern numbers from differential geometry. The basic idea is that Brillouin zone provides a base manifold and the inner product in the Hilbert space of the cell-periodic functions $u_{n,\mathbf{k}}$ supplies a sense of parallelism between nearby $\mathbf{k}$-points, so called (Berry) connection

$$A^\alpha_{mn}(\mathbf{k}) = \left(u_{n,\mathbf{k}} \left| \frac{\partial}{\partial k_\alpha} u_{m,\mathbf{k}} \right. \right) , \tag{23}$$

where the scalar product $(\cdot|\cdot)$ denotes an $\mathbf{r}$-integration over the unit cell. Although the connection is not gauge-invariant it allows for the calculation of gauge invariant quantities by standard techniques of differential geometry. Following [12, 13], we define the trace of the curvature corresponding to $A^\alpha_{mn}(\mathbf{k})$ by

$$B^{\alpha\beta}(\mathbf{k}) = \mathrm{tr} \left(\frac{\partial \mathbf{A}^\beta}{\partial k_\alpha} - \frac{\partial \mathbf{A}^\alpha}{\partial k_\beta} - [\mathbf{A}^\alpha, \mathbf{A}^\beta] \right) \tag{24}$$

$$= 2\,\mathrm{Im} \sum_{n=n_{\min}}^{n_{\max}} \left(\frac{\partial}{\partial k_\alpha} u_{n,\mathbf{k}} \left| \frac{\partial}{\partial k_\beta} u_{n,\mathbf{k}} \right. \right) \tag{25}$$

$$= \mathrm{Tr} \left(P(\mathbf{k}) \left[\frac{\partial}{\partial k_\alpha} P(\mathbf{k}), \frac{\partial}{\partial k_\alpha} P(\mathbf{k}) \right] \right) . \tag{26}$$

Here the first line is the definition of $B^{\alpha\beta}(\mathbf{k})$ in terms of the connection $A^\alpha_{mn}(\mathbf{k})$, with the trace and the commutator acting on the band indices. The second line is $B^{\alpha\beta}(\mathbf{k})$ for the connection (23). The third line shows explicitly the gauge invariance and analyticity by expressing $B^{\alpha\beta}(\mathbf{k})$ in terms of the generalized band projection operator $P(\mathbf{k})$, with the trace and commutator understood in the operator sense. The generalized band projection operator (20)

$$P(\mathbf{k}) = \sum_{n=n_{\min}}^{n_{\max}} |\psi_{n,\mathbf{k}}\rangle\langle\psi_{n,\mathbf{k}}| \tag{27}$$

is an analytic function of $\mathbf{k}$ [8]. Panati [14] showed that the analytic and periodic quasi Bloch states exist only when all Chern numbers associated with the curvature $B^{\alpha\beta}(\mathbf{k})$ are zero. Modern examples of systems with non-trivial band topology, which prohibits construction of exponentially localized WFs, are topological insulators [15]. We close this section by pointing out that the topological characteristics are the property of a given band or composite bands. The topological characteristics change when other bands are included in the composite bands and thus the construction of localized Wannier functions may become possible.

3.4 Uniqueness

Analytic functions are very rigid objects in the sense that, once the function is known on an arbitrarily small open interval Δk^d (where $\mathbf{k}$ is viewed as a complex variable!), it is uniquely defined everywhere. Nevertheless it is clear that having an exponentially localized WF any gauge transformation (15) also produces exponentially localized WF.

Uniqueness of the WF center. Let us define the center of WF as $\langle w|\mathbf{r}|w\rangle$ (the center of mass). We will show that for an isolated band the WF center does not depend on the gauge.

$$\langle w|\mathbf{r}|w\rangle = \int d\mathbf{r} \left(\frac{V}{(2\pi)^d}\right)^2 \int_{\mathrm{BZ}} d\mathbf{k}d\mathbf{k}' \, \psi_{\mathbf{k}}^*(\mathbf{r}) \, \mathbf{r} \, \psi_{\mathbf{k}'}(\mathbf{r})$$
$$= -i\frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} d\mathbf{k} \, \langle u_{\mathbf{k}}|\nabla_{\mathbf{k}} u_{\mathbf{k}}\rangle, \tag{28}$$

where we have used $\mathbf{r}\psi_{\mathbf{k}}(\mathbf{r}) = i\nabla_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{r}) - ie^{-i\mathbf{k}\cdot\mathbf{r}}\nabla_{\mathbf{k}}u_{\mathbf{k}}(\mathbf{r})$ and the fact the $\psi_{\mathbf{k}}(\mathbf{r})$ is a periodic function of $\mathbf{k}$. Now, performing the gauge transformation (15) $\tilde{\psi} = e^{i\phi}\psi$ we get

$$\langle \tilde{w}|\mathbf{r}|\tilde{w}\rangle = \langle w|\mathbf{r}|w\rangle + \frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} d\mathbf{k} \, \nabla_{\mathbf{k}}\phi(\mathbf{k})$$
$$= \langle w|\mathbf{r}|w\rangle + \mathbf{R}, \tag{29}$$

where $\mathbf{R}$ is a lattice vector. We have used the fact that $\phi(\mathbf{k})$ is smooth and periodic modulo 2π. Therefore, we can conclude that the position of the WF center is unique modulo a lattice vector $\mathbf{R}$. This uncertainty is not surprising if we keep in mind that it is the grid of WFs periodically repeated over the whole lattice which represents the energy band and that none of the lattice points has a special meaning. In case of composite bands it is the center of mass of all Wannier functions associated with a given lattice point $\mathbf{R}$ which is unique up to a lattice translation.

Reality of Wannier function. It was found empirically that in systems with real Hamiltonian, e.g., without spin-orbit coupling and external magnetic field, the exponentially localized WFs obtained with the maximum-localization method [13] (discussed in section 4.1) are real, up to a trivial overall phase. It was conjectured [13] that real localized Wannier functions can be constructed in such systems. A simple criterion for existence of real localized WFs is presented in Ref. [12].

Symmetry constraints. In some cases Wannier functions may be uniquely defined by symmetry requirements. We present a simple example here. Let us assume that we have an isolated band and a real Wannier function $w(\mathbf{r})$ of definite parity (e.g., even). We show that such a WF is unique.

$$w(\mathbf{r}) = \frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} \psi_{\mathbf{k}}(\mathbf{r}) \tag{30}$$

$$w^*(\mathbf{r}) = w(\mathbf{r}), \quad w(-\mathbf{r}) = w(\mathbf{r}). \tag{31}$$

Let us consider another WF obtained by a gauge transformation

$$
\begin{aligned}
\tilde{w}(\mathbf{r}) &= \frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} e^{i\phi(\mathbf{k})} \psi_{\mathbf{k}}(\mathbf{r}), \\
&= \frac{V}{(2\pi)^d} \sum_{\mathbf{R}} \left(\int_{\mathrm{BZ}} e^{i\phi(\mathbf{k})} e^{i\mathbf{k}\cdot\mathbf{R}} \right) w(\mathbf{r}-\mathbf{R}).
\end{aligned}
\tag{32}
$$

Requiring reality and evenness of $\tilde{w}(\mathbf{r})$ leads to a constraint on the gauge transformation $e^{i\phi(\mathbf{k})} = e^{-i\phi(\mathbf{k})}$ and thus $\phi(\mathbf{k}) = N\pi$ (with integer N), which means that $w(\mathbf{r})$ is unique up to a trivial change of sign.

While it is possible that reality and symmetry properties define exponentially localized WFs uniquely also for composite bands for cases with simple symmetry, we are not aware of rigorous proofs of such theorems. In general, the exponential localization only fixes the long-range behavior of the WFs, while the short range form depends on the particular computational method. For a discussion of point group aspects of construction of localized orbitals see Ref. [16].

4   Numerical methods for construction of Wannier functions

In the following we will describe two methods that are commonly used in connection with present band structure codes. We will also describe an approach for construction of Wannier functions in case of entangled bands, i.e., a situation where composite bands cannot be isolated from the rest of the spectrum. The first method is the Marzari and Vanderbilt maximally-localized WF construction [13], which consists in finding a gauge in which the second moment of the density distribution is minimized. The second method, proposed by Ku [17], is a realization of the projection construction (21). A common feature of these methods is that they start from a known set of the Bloch eigenstates of the Hamiltonian and search for the unitary transformation (16) according to some criteria. These two approaches are commonly used for construction of first principles tight-binding models that are used in the context of many-electron calculations, in particular LDA+DMFT.

4.1   Maximally localized Wannier functions

The maximal-localization (MALOC) method aims to minimize the spread functional

$$
\Omega = \sum_{n} \langle r^2 \rangle_n - \langle \mathbf{r} \rangle_n^2,
\tag{33}
$$

where n is the WF index and $\langle \mathbf{r} \rangle_n$ is the position of the center of nth WF given by the $\mathbf{r}$-matrix element (28), while $\langle r^2 \rangle_n$ is the corresponding r^2 matrix element. The reader is invited to check that Ω depends neither on the choice of the particular lattice site $\mathbf{R}$ associated with WF nor on the choice of the coordinate origin. Marzari and Vanderbilt proceed by splitting Ω into gauge

independent Ω_I and gauge dependent $\tilde{\Omega}$ parts

$$\Omega = \Omega_I + \tilde{\Omega} \tag{34}$$

$$\Omega_I = \sum_n \langle r^2 \rangle_n - \sum_{\mathbf{R},n,m} |\langle \mathbf{R}, m|\mathbf{r}|0, n\rangle|^2 \tag{35}$$

$$\tilde{\Omega} = \sum_n \sum_{\mathbf{R}, m \neq 0, m} |\langle \mathbf{R}, m|\mathbf{r}|0, n\rangle|^2. \tag{36}$$

Expressed in terms of the Berry connection (23), $\tilde{\Omega}$ can be written as a sum of two terms

$$\tilde{\Omega} = \sum_n \frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} d\mathbf{k} \, |\mathbf{A}_{nn}(\mathbf{k}) - \langle \mathbf{r} \rangle_n|^2 - \sum_{n \neq m} \frac{V}{(2\pi)^d} \int_{\mathrm{BZ}} d\mathbf{k} \, |\mathbf{A}_{mn}(\mathbf{k})|^2, \tag{37}$$

where $\mathbf{A}_{mn}(\mathbf{k})$ is a d-dimensional vector (d=1,2,3) and $|\cdot|^2$ is the corresponding norm. The task of constructing MALOC WFs can be formulated as follows: given some initial gauge we are looking for a gauge transformation which minimizes $\tilde{\Omega}$. Definition (23) of the Berry connection is not useful for numerical calculations and has to be replaced with a discrete approximation, which reduces to (23) and (37) in the continuum limit. Among the different possibilities Marzari and Vanderbilt choose

$$M_{mn}^{(\mathbf{k},\mathbf{b})} = (u_{m,\mathbf{k}}|u_{n,\mathbf{k}+\mathbf{b}}) \tag{38}$$

and

$$\tilde{\Omega} = \frac{1}{N} \sum_{\mathbf{k},\mathbf{b}} w_\mathbf{b} \sum_n \left(-\operatorname{Im} \ln M_{nn}^{(\mathbf{k},\mathbf{b})} - \mathbf{b} \cdot \langle \mathbf{r} \rangle_n \right)^2 + \frac{1}{N} \sum_{\mathbf{k},\mathbf{b}} w_\mathbf{b} \sum_{n \neq m} \left| M_{mn}^{(\mathbf{k},\mathbf{b})} \right|^2. \tag{39}$$

Here we assume to have a uniform $\mathbf{k}$-mesh throughout the Brillouin zone. For each $\mathbf{k}$-point a set of vectors $\mathbf{b}$ defines connections to its neighbors. Moreover, we assume periodicity, which means that a connection is defined between points of the opposite sides of the Brillouin zone. The quadrature of (37) associates a weight $w_\mathbf{b}$ with each $\mathbf{b}$-vector [13]. The same discretization for $\langle \mathbf{r} \rangle_n$ leads to the expression

$$\langle \mathbf{r} \rangle_n = -\frac{1}{N} \sum_{\mathbf{k},\mathbf{b}} w_\mathbf{b} \mathbf{b} \operatorname{Im} \ln M_{nn}^{(\mathbf{k},\mathbf{b})}. \tag{40}$$

Using the logarithm in the discrete approximation preserves the gauge invariance of $\langle \mathbf{r} \rangle_n$. It also ensures the proper shift of the WF centers by a lattice vector upon the gauge transformation $e^{i\mathbf{k}\cdot\mathbf{R}}$. However, the discretization is not without problems. In particular, for $M_{nn}^{(\mathbf{k},\mathbf{b})}$ to provide an approximation of the continuum limit, we shall require that $1 - M_{nn}^{(\mathbf{k},\mathbf{b})} \to 0$ as $\mathbf{b} \to 0$. This is typically not the case in numerical calculations, where phases of the eigenvectors are more or less random, no matter how small $\mathbf{b}$. In other words, the gauge transformation between the computer output and a smooth $\mathbf{k}$-dependence of $|u_{n,\mathbf{k}}\rangle$, which yields localized WFs, corresponds to an erratic function of $\mathbf{k}$. It is nearly impossible to find such a transformation by local minimization methods. Another way to see the problem is to realize that $M_{nn}^{(\mathbf{k},\mathbf{b})}$ is complex and thus its logarithm is a multivalued function. A consistent choice of the branch leading to localized WFs becomes numerically impossible. Using more $\mathbf{k}$-points makes the problem even worse as it does not remove the random phase, but only makes the phase fluctuations sharper.

The way out consists in preprocessing the Bloch states so that the phase differences between nearby $\mathbf{k}$-points are small. This can be achieved by defining a set of trial orbitals $|M_n\rangle$, similar to (21). Projecting these onto the Bloch states

$$|\Phi_n(\mathbf{k})\rangle = \sum_m |\psi_{m,\mathbf{k}}\rangle \langle \psi_{m,\mathbf{k}}|M_n\rangle \tag{41}$$

$$S_{mn}(\mathbf{k}) = \langle \Phi_m(\mathbf{k})|\Phi_n(\mathbf{k})\rangle \tag{42}$$

and performing symmetric orthogonalization

$$|\tilde{\psi}_{n,\mathbf{k}}\rangle = \sum_m \left[S^{-1/2}(\mathbf{k}) \right]_{mn} |\psi_{m,\mathbf{k}}\rangle \tag{43}$$

we get the desired result. Now, when constructing $M_{mn}^{(\mathbf{k},\mathbf{b})}$ from $|\tilde{\psi}_{n,\mathbf{k}}\rangle$ the phase of $M_{mn}^{(\mathbf{k},\mathbf{b})}$ is small and we can choose $|\operatorname{Im} \ln M_{nn}^{(\mathbf{k},\mathbf{b})}| < \pi$. The functional $\tilde{\Omega}$ can then be minimized by a sequence of unitary transformations using the method of steepest-descent. For details we refer the reader to Ref. [13].

4.2 Projection method

It is possible to use only the initialization step described in the previous section to construct Wannier functions. This approach was adopted by Ku *et al.* [17] and used subsequently by others [18, 19]. It was successfully applied to investigation of oxides and similar materials for construction of atom-centered WFs. The advantage of this approach is that WFs retain the symmetry of the trial orbitals. The second advantage is that the construction is simple and guaranteed to converge as long as the overlaps with the trial functions are non-zero and localized WFs exist. This makes the method attractive for iterative procedures, where accidental freezing in a wrong local minima might spoil the calculation. This is typically the case when constructing tight-binding models, where we have a good idea of the shape and symmetry of the desired Wannier orbitals and want to get the quantitative information such as hopping parameters. On the other hand, when it comes to molecular orbitals or low-symmetry situations, where constructing good trial orbitals might be complicated, the MALOC method is preferable. This also applies to situations where the positions of Wannier centers are not fixed by symmetry and need to be calculated accurately, e.g., in ferroelectric materials.

4.3 Entangled bands

In some situations the number of constructed WFs is smaller than the number of Bloch bands. Typically this happens when no isolated band or composite bands can be found. In such a case it may also happen that the number of Bloch states per $\mathbf{k}$-point $N_{\mathbf{k}}$ varies although it must always be at least equal to N, the number of desired WFs. The technique to construct WFs in such situations is known as disentanglement and was introduced by Souza *et al.* [20]. The idea is to find an N-dimensional subspace $\mathcal{S}(\mathbf{k})$ of the $N_{\mathbf{k}}$-dimensional space spanned by the Bloch states and $\mathbf{k}$ such that the overlap between $\mathcal{S}(\mathbf{k})$ and $\mathcal{S}(\mathbf{k} + \mathbf{b})$ at neighboring $\mathbf{k}$-points is maximized
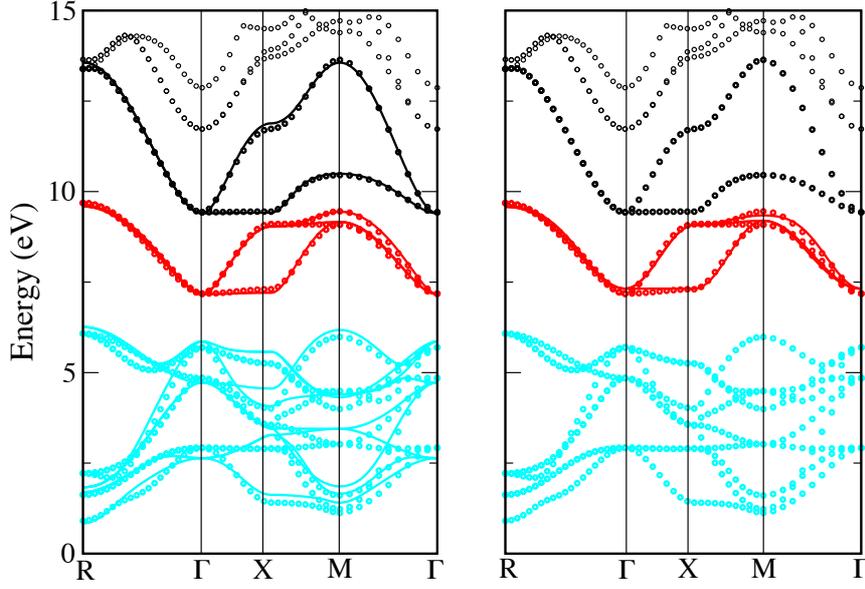
Fig. 5: *LDA band structure of SrVO$_3$ (circles) with dominant contributions marked by color: O-p (blue), V-d-t_{2g} (red), and V-d-e_g (black). Tight-binding bands obtained from WFs for (i) [left panel] all V-d and O-p bands (ii) [right panel] only the t_{2g} band, with spatial cut-offs as described in the text (the Fermi level is at 8.16 eV).*

in some sense. Souza *et al.* showed that this is equivalent to minimizing the Ω_I functional. The WF construction then proceeds in two steps. First, iterative minimization of Ω_I with respect to unitary transformations at each **k**-point. Second, the above MALOC procedure preformed using $\mathcal{S}(\mathbf{k})$'s from the first step. The details and examples of application can be found in [20].

The disentanglement procedure can also be realized with the projection technique [17]. In this case the subspace $\mathcal{S}(\mathbf{k})$ is chosen at each **k**-point such that its overlap with a set of trial orbitals is maximal.

We have presented the two currently most popular methods for construction of Wannier functions in periodic solids. Examples of other methods can be found in Refs. [21,22].

5   Examples of applications

In the following we present several examples that demonstrate applications of Wannier functions. This is a rather small sample of possible applications and focuses on problems connected with the construction of tight-binding models. For other important applications of WFs such as investigation of dielectric polarization the reader is referred to literature [3]. The examples presented below use the MALOC procedure implemented in the wannier90 package [23] starting from the electronic structures obtained with the wien2k code [24] and preprocessed with wien2wannier [25].
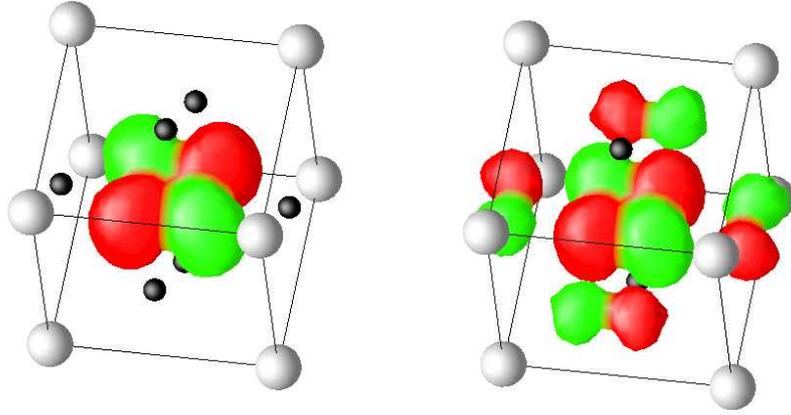
Fig. 6: *The xy Wannier orbital plotted as an isosurface of the charge density $|w(\mathbf{r})|^2$ and colored by the sign of $w(\mathbf{r})$. The left panel corresponds to the large energy window (i), the right panel to the small energy window (ii) for the same isovalue.*

5.1 Wide versus narrow energy window: $SrVO_3$

One of the basic parameters defining the WFs is the choice of the Hilbert space they should span. In practice this means the choice of bands or an energy window to be covered. In general, the larger the energy window the more localized WFs can be constructed. Taken to extreme, using the full spectrum of the Hamiltonian (up to infinite energy) we can build a basis out of Dirac delta functions. We will demonstrate the effect of the energy window on the localization of WFs for the example of a transition metal oxide by comparing two possible choices of the energy window (i) $3d$+O-$2p$ bands, (ii) $3d$ bands only.

$SrVO_3$ has a rather simple band structure (see Fig. 5) consisting of isolated groups of bands derived from O-p, V-d-t_{2g} and V-d-e_g orbitals. For choice (i) we use all V-d and O-p bands, while for (ii) we select only the V-d bands of t_{2g} symmetry. The corresponding xy WF are shown in Fig. 6. The choice of the smaller energy window leads to the xy orbital having substantial weight on its oxygen neighbors. This is easy to understand. In terms of atomic-like V and O orbitals, the V-t_{2g} band is anti-bonding, while the O-$2p$ band is bonding. When we choose to build WFs from the anti-bonding band only, we cannot recover the atomic orbitals. The O-$2p$ atomic character contained in the anti-bonding band cannot disappear. It is preserved in the more extended shape of WFs. Note, that the anti-bonding character is apparent in the WF having a node (changes sign) between the O and V sites. Using a larger energy window (i) we include both bonding and anti-bonding bands. This allows us to construct WFs which closely resemble atomic V and O functions. There is still some V-xy weight on the O sites to ensure orthogonality of the WFs, but it is substantially less than in case (ii) where the O tails of the xy functions are enforced by hybridization.

The spatial extent of the Wannier orbitals is also reflected in the hopping integrals. With the small energy window (ii) the t_{2g} bands are well described when at least nn- and nnn-hoppings
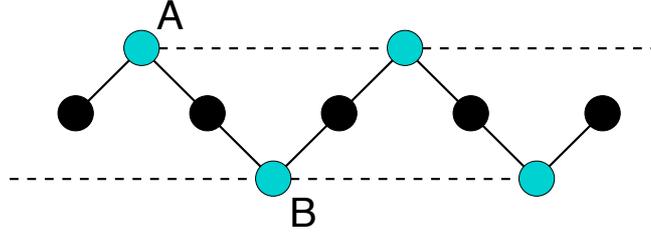
Fig. 7: *A simple zig-zag chain with two types of atoms (black and blue). The lines indicate possible hoppings. (including longer-range hoppings does not break the A-B symmetry).*
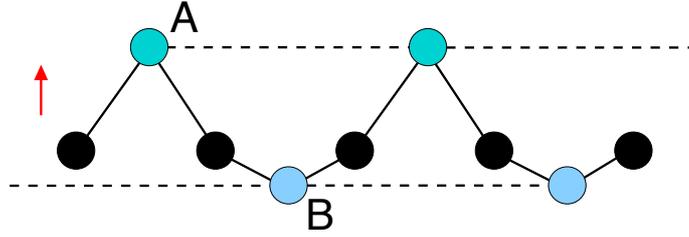
Fig. 8: *The zig-zag chain in perpendicular electric field, which breaks the A-B symmetry as indicated by a uniform shift of blue atoms.*

are considered (see Fig. 5), which in the $\{xy, yz, zx\}$ basis read

$$t_{100}[\text{meV}] = \begin{pmatrix} -268 & 0 & 0 \\ 0 & -30 & 0 \\ 0 & 0 & -268 \end{pmatrix}, \qquad t_{101}[\text{meV}] = \begin{pmatrix} 7 & 10 & 0 \\ 10 & 7 & 0 \\ 0 & 0 & -93 \end{pmatrix}. \qquad (44)$$

The longest nnn t_{2g}-t_{2g} hopping corresponds to a length of of 5.4 Å. Using the more localized orbitals (i) we can achieve similar accuracy (see Figure 5) by considering only V-V nn-hopping and V-O nnn-hopping, which translate into a direct spatial cut-off of only 4.3 Å. The obvious price to be paid are larger matrices (14×14 for (i) versus 3×3 in case (ii)).

5.2   Unfolding band structures: LaOFeAs

Wannier functions provide a tool for an unbiased and quantitatively accurate construction of tight binding models. In some situations it is possible that the tight-binding Hamiltonian has a symmetry higher than the original band structure. The recently much studied LaOFeAs is a good example. However, before we discuss this material let us consider the simpler case of a zig-zag chain shown in Fig. 7. Geometrically, the sites A and B are not related by a translation. However, when we consider the symmetry of the graph representing the hoppings (full and dashed lines are shown as an example) only the black and the blue sites are distinct. Another way to see the equivalence between A and B sites is to realize that they are connected by a 'gauge' transformation which changes the sense of 'up' and 'down'. Only when we add a field which breaks the up-down symmetry (e.g. uniform electric field perpendicular to chain) the
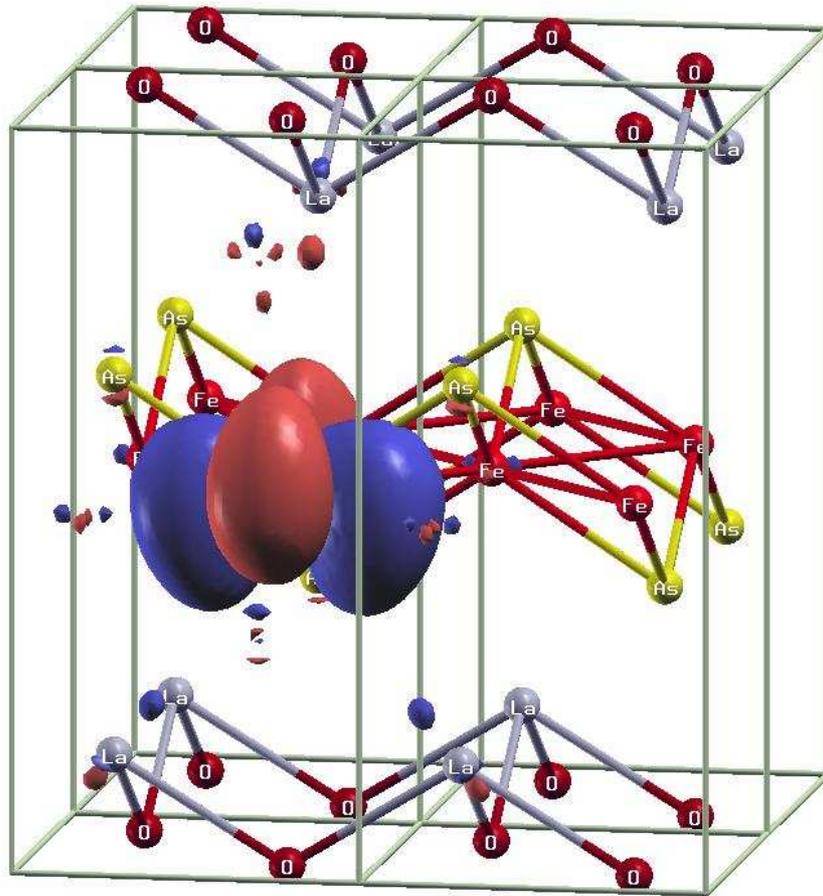
Fig. 9: *The Fe x^2-y^2 Wannier orbital for the $d-p$ model of LaOFeAs. The deviation from the atomic x^2-y^2 orbital reflecting the geometry of Fe-As plane is clearly visible.*

A-B equivalence is broken (see Fig. 8). The Fe-As plane in LaOFeAs with As atom alternating above and below the Fe plane resembles the zig-zag chain.

As bandstructure codes use the geometrical symmetry of the crystal structure, the electronic structure calculation is performed in a unit cell with two Fe atoms and the corresponding Brillouin zone. Unfolding the bandstructure amounts to doubling the Brillouin zone and halving the number of bands. A straightforward way to unfold the bandstructure is to construct the WFs and the corresponding tight-binding Hamiltonian. After a simple geometrical transformation the intra-cell hoppings can be converted into inter-cell ones, to obtain a tight-binding Hamiltonian for the smaller unit cell with a single Fe atom. The construction of the corresponding unfolded bandstructure is straightforward. The fact that the corresponding crystal structure has a lower symmetry of course cannot disappear. It is reflected in the shape of the WFs (see Fig. 9) and the fact that WFs centered at the two Fe sites in the original unit cell are not connected by a translation, but a translation combined with a 90 deg rotation along the z-axis.
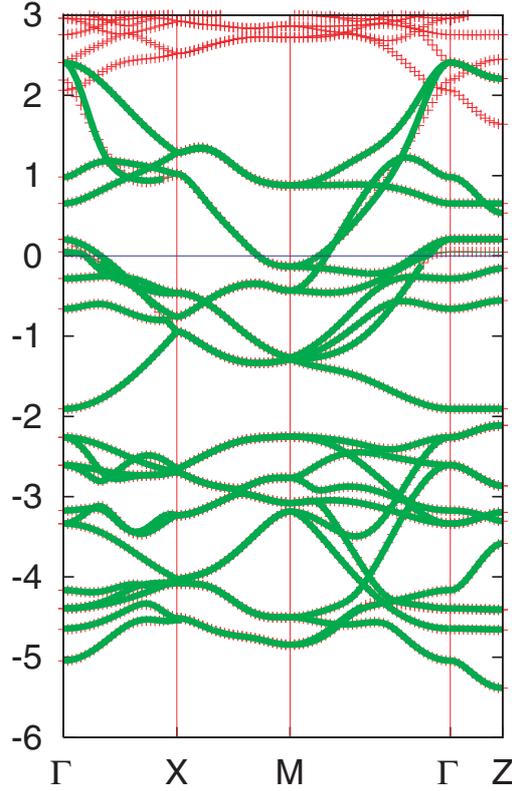
Fig. 10: *LDA band structure of LaOFeAs (red) compared with tight-binding bands obtained from WFs (green) for the $d - p$ model. Note the difference above 2 eV.*

5.3   Entangled bands

Staying with LaOFeAs, we will demonstrate a typical application of the band disentanglement technique. As before, we are interested in constructing an effective Hamiltonian of LaOFeAs in terms of Fe-d and As-p WFs. However, we see from Fig. 10 that the Fe-d bands overlap with the higher lying bands (above $\sim$2 eV). The hybridization between the overlapping bands leads to admixing of Fe-d character to these higher lying bands and to band anti-crossing, which modifies the band topology, e.g., around the Γ point. In order to reproduce the bands around 2 eV at Γ precisely, we would have to either enlarge our model to include other orbitals or work with more extended Fe-d WFs. However, the details of band topology at 2 eV above the Fermi level are of little physical significance. It is preferable to work with an approximate bandstructure, which can be obtained with localized Fe-d WFs in a $d - p$ model. What we want to achieve can be viewed roughly as follows. Imagine we build a WF representation for a much larger energy window than we are interested in. Subsequently we switch off all the hoppings between our target Fe-d and As-p WFs and anything else. The resulting band structure would closely resemble the original one around the chemical potential where the bands are dominated by Fe-d and As-p character. Deviations will appear at energies and k-space regions where the neglected hoppings play a role, e.g. 2 eV at Γ-point. The disentanglement procedure performs essentially the same task, however, without the need to construct the large tight-binding model

first. In fact, it does even better. Using the above picture, the disentanglement procedure does not put all the hoppings between Fe-d and As-p strictly to zero, but minimizes them in such a way that the band dispersion in a specified energy window is reproduced precisely, which in practice means it produces Fe-d and As-p WFs that are a bit more extended than they would have been in the large tight-binding model.

From the discussion it is clear that the disentanglement procedure may find its use even in some cases with isolated (composite) bands when one decides to trade the accuracy away from the region of interest (Fermi level) for having more localized WFs.

5.4 Spin-orbit coupling: Sr_2IrO_4

The electronic states in condensed matter physics are commonly described in the quasi-relativistic approximation, i.e., using two component spinors. In many applications, in particular using lattice models, the spin projection is conserved. Therefore we can work in a basis in which the Bloch eigenstates have only one non-zero spinor component and forget about the spinor structure. The most common situation where non-trivial spinor structure survives are systems with spin-orbit coupling (another less common example are broken-symmetry systems with a non-collinear magnetic order). However, since the construction of WFs amounts to a unitary transformation from the space indexed by $(n, \mathbf{k})$, the band indices and $\mathbf{k}$-vectors, to the space, indexed by $(m, \mathbf{R})$, the orbital indices and the lattice vectors, the spinor character of the electronic functions does not change the discussion in the previous sections. It merely changes the definition of the dot product in (23) and (38) from

$$\langle\psi|\psi'\rangle = \int d\mathbf{r}\, \psi^*(\mathbf{r})\psi'(\mathbf{r}) \tag{45}$$

to

$$\langle\psi|\psi'\rangle = \sum_\sigma \int d\mathbf{r}\, \psi_\sigma^*(\mathbf{r})\psi_\sigma'(\mathbf{r}), \tag{46}$$

where σ indexes the spinor components

$$|\psi\rangle = \begin{pmatrix} \psi_\uparrow \\ \psi_\downarrow \end{pmatrix}. \tag{47}$$

The transformations (13) or (16) now connect the Bloch spinor functions to Wannier spinor functions. The spinors must be viewed as the basic objects. For example a MALOC procedure applied to the spinor components separately will not lead to a sensible result.

In the following we use Sr_2IrO_4 as an example of a material where spin-orbit coupling substantially modifies the band structure and leads to Wannier orbitals in which both spin projections are mixed. For the sake of simplicity, we have performed the calculations using an idealized double-perovskite crystal structure. The electronic structure (see Fig. 11) can be understood by considering crystal-field splitting, spin-orbit coupling and inter-site hopping between Ir-d orbitals. The crystal-field splitting, being the largest of the three, opens a gap between the t_{2g} and e_g bands, rendering the latter empty, while the former accommodate one hole per Ir atom.
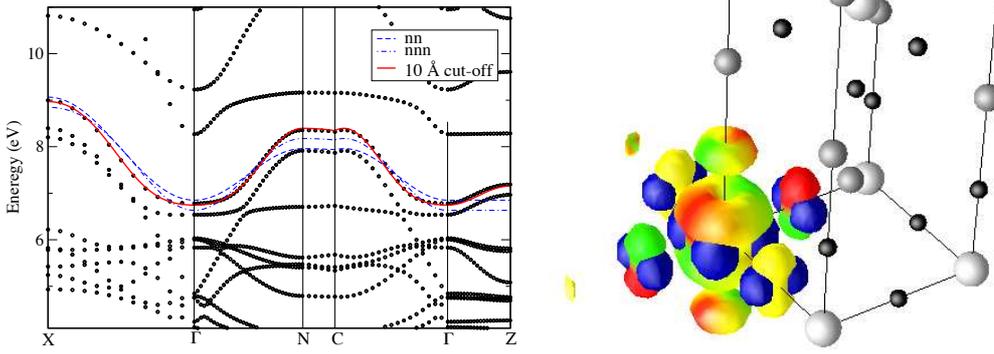
Fig. 11: *Ir $J = 1/2$ Wannier orbital $w_{+,[0,0,0]}(\mathbf{r})$ visualized as a $|w|^2$ isosurface. The almost real $\downarrow$-spin component is blue, the $\uparrow$-spin component is colored with cosine of its phase (red=real positive, green=real negative, yellow=0).*

The t_{2g} orbitals may be labeled with a pseudo-spin $I = 1$. The spin-orbit coupling splits the t_{2g} manifold into a quadruplet and doublet with pseudo-spins $J = 3/2$ and $J = 1/2$, respectively. Since the spin-orbit splitting is rather large, the inter-site hopping leads only to moderate mixing of the states with different J. Therefore we may expect the isolated band doublet at the top of the t_{2g} manifold to be predominantly of $J = 1/2$ character. We construct the MALOC WFs for this two-fold degenerate band. Expressed in the local coordinate system pointing towards the O atoms, the $J = 1/2$ spinor functions adopt the form

$$|\phi_+\rangle \sim \begin{pmatrix} -2Y_{21} \\ Y_{2-2} - Y_{22} \end{pmatrix}, \quad |\phi_-\rangle \sim \begin{pmatrix} Y_{2-2} - Y_{22} \\ 2Y_{2-1} \end{pmatrix}. \tag{48}$$

The MALOC WFs are indeed very close to this form as shown in Fig. 11. In particular, the WF, to a very good approximation, consists of a real xy orbital in one spin channel and a complex $(x \pm iy)z$ orbital in the other ($\pm$) spin channel. Note that the relative phase of the two components is not arbitrary and that the corresponding charge density has approximately cubic symmetry as expected for a $J = 1/2$ orbital. Also, similar to WFs in $SrVO_3$ the anti-bonding character with respect to oxygen is reflected in the sign-change (red-green) between the lobes on Ir and its O neighbors.

5.5 Wannier functions in dynamical mean-field theory

WFs and the techniques described in these notes are often used to construct Hubbard type Hamiltonians for real materials

$$H = \sum_{\mathbf{k}} \left([h^{dd}(\mathbf{k})]_{\alpha\beta} d^{\dagger}_{\mathbf{k}\alpha} d_{\mathbf{k}\beta} + [h^{pp}(\mathbf{k})]_{\gamma\delta} p^{\dagger}_{\mathbf{k}\gamma} p_{\mathbf{k}\delta} + [h^{dp}(\mathbf{k})]_{\alpha\gamma} d^{\dagger}_{\mathbf{k}\alpha} p_{\mathbf{k}\gamma} + [h^{pd}(\mathbf{k})]_{\gamma\alpha} p^{\dagger}_{\mathbf{k}\gamma} d_{\mathbf{k}\alpha} \right)$$
$$+ \sum_i U_{\alpha\beta\gamma\delta} d^{\dagger}_{i\alpha} d^{\dagger}_{i\beta} d_{i\gamma} d_{i\delta} - H_{\mathrm{dc}}, \tag{49}$$

where we have denoted electrons subject to local interactions as d-type and non-interacting electrons as p-type. The $\mathbf{k}$-indexed operators are Bloch sums of the corresponding direct space

operators and summation over the Greek indices is assumed. The double-counting term H_{dc} is a one-particle operator discussed below. Obviously, the results of any theory should depend only on physical parameters and not on the coordinate system or basis used to formulate the problem. How can this requirement be reconciled with the non-uniqueness in the definition of the Wannier orbitals?

First, let us emphasize that the first term in (49) does not depend on the choice of the WFs. Its form is the same in any WF basis. While the numerical values of the matrices $h^{dd}(\mathbf{k})$, $h^{pp}(\mathbf{k})$, $h^{pd}(\mathbf{k})$ and $h^{pp}(\mathbf{k})$ are basis dependent, the spectrum of the operator is not. The same is not true for the second term. Here a change of the basis leads to terms, which are not present in (49) such as inter-site interactions. Therefore the model (49) is always basis dependent and we can only require that the dependence is weak and that the choice of a particular basis is physically motivated. We can heuristically reason as follows. The electron-electron scattering exists between any four states (subject to conservation laws). However, to describe the physics on the eV scale, the majority of the interaction terms can be treated at mean-field level. These terms are implicitly included in the one-particle part of (49). Only those terms leading to size-able dynamical correlation effects have to be treated explicitly as two-particle operators. The choice of the WF basis thus amounts to deciding which part of the electron-electron interaction can be approximated by a mean-field decoupling and which part should be kept explicitly. The double-counting term H_{dc} is used to avoid counting the same interaction twice, once on a mean-field level in the first term and once explicitly. Following this reasoning, it appears physically well justified to work with as localized WF orbitals as possible since this amounts to the best possible separation of intra- and inter-atomic electron-electron interaction. We emphasize that we have just presented a physical picture rather than a mathematical framework. In practice the starting band structures are not obtained from Hartree-Fock theory (i.e., a self-consistent mean-field solution of the electron Hamiltonian), but from density functional theory, which has a similar mathematical structure, but uses essentially an empirical one-particle potential, which is known to yield better results than Hartree-Fock theory. Using the empirical potential makes it impossible to rigorously define what the double-counting term H_{dc} should be. Therefore, we are left with heuristic arguments to define H_{dc} and various forms are being used with varied success for different groups of materials.

In Fig. 12 a flow chart of a typical LDA+DMFT [26] calculation is shown. It consists of two more or less independent steps. First, density-functional theory (usually in LDA or similar approximation) is used to obtain the band structure of a given material. This involves iteration of the charge distribution. Once the LDA calculation is converged, results are postprocessed to obtain the model Hamiltonian (49). This involves the physically motivated selection of the relevant part of the Hilbert space (specified through the energy window or band indices) and the construction of the corresponding Wannier functions. The one-particle part of (49) is obtained by a straightforward unitary transformation between the Bloch and Wannier functions. Determination of the interaction parameters is less straightforward since it has to account for various screening processes, in particular those arising from electronic states excluded from the model. Several approaches are possible: (i) $U_{\alpha\beta\gamma\delta}$ is treated as a free parameter or adjusted to some ex-
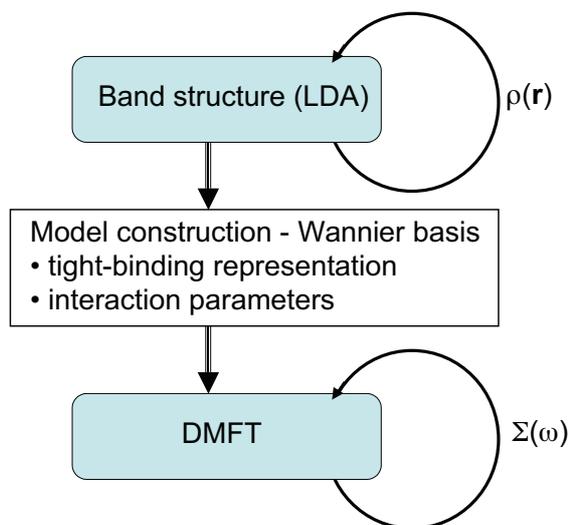
Fig. 12: *A flow chart of typical LDA+DMFT calculation. Two separate self-consistency loops over the charge density $\rho(\mathbf{r})$ and self-energy $\Sigma(\omega)$ are involved.*

perimental input, or it is calculated using (ii) constrained LDA method [27] or (iii) constrained RPA [28]. The computational approaches (ii) and (iii) are based on the response of the effective LDA band structure to external perturbations. The second step in the LDA+DMFT approach consists in DMFT treatment of model (49), which involves iteration over the one-particle self-energy. Since the dynamical effects in DMFT often lead to some charge redistribution, it may be desirable to have the charge and self-energy self-consistency in the same loop [19].

6   Summary

Wannier functions provide an extremely useful set of orthogonal functions, which can represent isolated single or composite bands exactly or entangled bands with adjustable accuracy or resolution. They are, in a sense, an optimal compromise between localization in direct space and in the energy domain. The long-range behavior of WFs reflects the smoothness of the Bloch functions as functions of quasi-momentum $\mathbf{k}$. Existence of exponentially localized WFs is a topological property of a given (composite) band. While in commonly studied cases, e.g., time reversal symmetry and no spin-orbit coupling, exponentially localized WFs can be constructed, the bands for which this is not possible are of great interest as they exhibit anomalies such as topologically protected surface currents.

Acknowledgment

References

[1] for review see R.M. Dreizler and E.K.U. Gross, *Density functional theory* (Springer-Verlag, Berlin, 1990)

[2] for example in Ch. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996)

[3] R.D. King-Smith and D. Vanderbilt, Phys. Rev. B **47**, 1651 (1993)

[4] G.H. Wannier, Phys. Rev. **52**, 191 (1937)

[5] for details see e.g. G. P. Tolstov, *Fourier Series* (Dover Publications, 1962)

[6] www-math.mit.edu/ stevenj/bump-saddle.pdf

[7] W. Kohn, Phys. Rev. **115**, 809 (1959)

[8] J. des Cloizeaux, Phys. Rev. **135**, A685 (1964)

[9] J. des Cloizeaux, Phys. Rev. **135**, A698 (1964)

[10] M. Kohmoto, Ann. Phys. **160**, 343 (1985)

[11] G. Nenciu, Commun. Math. Phys. **91**, 81 (1983)

[12] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, Phys. Rev. Lett. **98**, 046402 (2007)

[13] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997)

[14] G. Panati, Ann. Henri Poincarè **8**, 995 (2007)

[15] M.Z. Hasan and C.L. Kane, Rev. Mod. Phys. **82**, 3045 (2010)

[16] J. des Cloizeaux, Phys. Rev. **129**, 554 (1963)

[17] Wei Ku, H. Rosner, W.E. Pickett, and R.T. Scalettar, Phys. Rev. Lett. **89**, 167204 (2002)

[18] V.I. Anisimov *et al.*, Phys. Rev. B **71**, 125119 (2005)

[19] M. Aichhorn *et al.*, Phys. Rev. B **80**, 085101 (2009)

[20] I. Souza, N. Marzari, and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001)

[21] W. Kohn and J.R. Onffroy, Phys. Rev. B **8**, 2485 (1973)

[22] W. Kohn, Phys. Rev. B **10**, 382 (1974)

[23] A.A. Mostofi, J.R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt and N. Marzari, Comput. Phys. Commun. **178**, 685 (2008)

[24] P. Blaha, K. Schwarz, G.K.H. Madsen, D. Kvasnicka, J. Luitz: *WIEN2K, An Augmented Plane Wave and Local Orbitals Program for Calculating Crystal Properties* (Vienna University of Technology, Austria, 2001), K. Schwarz, P. Sorantin, S.B. Trickey, Comput. Phys. Commun. **59**, 399 (1990)

[25] J. Kuneš, R. Arita, P. Wissgott, A. Toschi, H. Ikeda, and K. Held, Comput. Phys. Commun. **181**, 1888 (2010)

[26] a recent review G. Kotliar *et al.*, Rev. Mod. Phys. **78**, 865 (2006)

[27] O. Gunnarsson, O.K. Andersen, O. Jepsen, and J. Zaanen, Phys. Rev. B **39**, 1708 (1989) V.I. Anisimov and O. Gunnarsson, Phys. Rev. B **43**, 7570 (1991)

[28] F. Aryasetiawan *et al.*, Phys. Rev. B **70**, 195104 (2004)