High performance Wannier interpolation of Berry curvature and related quantities: WannierBerri code

Stepan S. Tsirkin
Department of Physics, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland
(Dated: August 19, 2020)

Wannier interpolation is a powerful tool for evaluation of Brillouin zone integrals over a dense grid of $k$ points, which is essential in e.g. anomalous Hall conductivity or Boltzmann transport coefficients. However new physical problems and new materials create new numerical challenges, and the computations with the existing codes become very computationally expensive, which often prevents reaching the desired accuracy. In this article I present a series of methods which allow to boost the speed of Wannier interpolation by several orders of magnitude compared to then the popular code Wannier90. The suggested methods include a combination of fast and slow Fourier transform, explicit use of symmetries, recursive adaptive grid refinement and some other techniques. The suggested methodology is implemented in the new python code WannierBerri, which also aims to serve as a convenient platform for development of further interpolation schemes for novel phenomena.

I. INTRODUCTION

Wannier functions [1, 2] (WFs) have proved extremely efficient in evaluation of multiple properties of solids, including the modern theory of polarization [3–5] and orbital magnetization [6–8] and topological properties [9, 10]. Moreover, WFs provide a way to describe a group of energy bands in a crystal by a compact Hamiltonian, thus allowing a fast evaluation of the bandstructure at any point of the Brillouin zone (BZ) without an extra call to the \textit{ab initio} code. This procedure called Wannier interpolation (WI) is similar in spirit to tight-binding calculations [11], but allows a systematic precise description of the bands without truncation error. WI is particularly useful in search of Weyl points [10] and evaluation of momentum-space integrals of rapidly varying functions, which appear, e.g. in calculations of anomalous Hall effect [12], orbital magnetization [13], Boltzmann transport coefficients [14], and optical properties [15]. WI schemes were also developed for, e.g., electron-phonon coupling [16, 17], gyrotropic effects [18], shift current [19] and spin Hall and effect [20, 21].

While the WFs may be constructed in multiple ways, the most popular technique is the maximal localization [2, 22], and the most established code for construction of maximally localized Wannier functions (MLWFs) is Wannier90 [23, 24], which is now driven by a broad community [25]. Generally, the construction of a good set of WFs requires a careful manual input, including a choice of an initial guess for the minimization process, as well as a choice of disentanglement energy windows [26]. However, recently a significant progress was achieved in automated construction of WFs, employing different techniques, such as optimized projection functions method [27], selected columns of the density matrix (SCDM) method [28–30], or automated choice of projections and energy windows [31, 32], and a database of Wannier Hamiltonians is being constructed [33]. These advances are a big step towards employing WI for high-throughput automated calculations of electronic properties of solids.

Most of the mentioned WI schemes are implemented within the popular codes – Wannier90 (namely its post-processing part \texttt{postw90.x}) and WannierTools [10]. These codes, being well-established and accepted by the community have quite a broad functionality. However, new materials and new physical effects create new numerical challenges, and calculation with those codes become quite heavy. In particular for a system with a large number of WFs and a complicated Fermi surface it maybe hard to achieve convergence with respect to momentum-space grid in a reasonable time. When speaking about high-throughput calculations the performance becomes even more important.

In this article I present a series of methodological improvements, which allow to perform Wannier interpolation much faster then \texttt{postw90.x}. The introduced methodology is implemented in the new python code WannierBerri [34] (WB), which is freely available to install [35] and is open for contributors [36]. The code has a broad functionality, and already a number of contributors. The present article covers only the core methodology of the code which procure its efficiency. More broad scope of the code can be found on its web page [34] and will be detailed in future publications with the co-authorship of co-developers. Interesting to note that WannierBerri may be equally used for WI and tight-binding calculations, and also offers a convenient platform for development of new features. The name of the code is derived from Wannier functions and the Basque word ”berri” which means ”new” and enters local toponyms (e.g. Lekunberri, Ekainberri) [37].

The high efficiency of calculations is achieved by implementing several numerical approaches. First, note that in a typical calculation using \texttt{postw90.x}, the bottleneck is the Fourier transform, which is implemented as a standard discrete Fourier transform. Therefore it looks appealing to use a fast Fourier transform (FFT) which is widely used in numerical calculations. However it is problematic to do it over a very dense grid of k-points which
may include up to $10^8$ points. This issue is overcome by a mixed scheme employing both fast and "slow" Fourier transform. Next the symmetries of the system may be used, to reduce the evaluation only to the symmetry-irreducible points. And obtain the result for the rest of points by applying symmetry operations. This also helps to obtain the result more symmetric (tensor components that should be equal or vanish will be exactly equal, or exactly vanish), even if the symmetries are slightly broken due to numerical inaccuracies in wannierization. Then I introduce an adaptive refinement algorithm, that chooses points that give the largest contribution to the integral, and makes the grid more dense in the vicinity of such points. This helps to have a more accurate description near special points, where the integrand is rapidly changing - e.g. near Weyl nodes or lines. This is close in spirit to adaptive refinement used in [12, 38], but is more automatic and requires less input from the user. Finally, I introduce methods which reduce the computational cost of the minimal-distance replica selection method [25] and scan of multiple Fermi levels to a negligible time.

The rest of the article is organized as follows. Sec. II describes the methods which help to boost the speed of the calculation. Sec. III demonstrates the usage of the code on the "textbook" example of the anomalous Hall conductivity of bcc iron, and in sec. IV based on that example the comparison the evaluation time with postw90 is made. Appendix A describes the list of implemented functionality.

II. METHODS

A. General equations for Wannier interpolation

The goal of this section is to introduce notation necessary for further discussion. For more details please refer to review [2] and original articles cited therein. The problem of Wannier interpolation is stated in the following way. First we evaluate the energies $E_{mn}$ and wavefunctions $\psi_{mnq}$ from first principles on a rather coarse grid of wavevectors $\mathbf{q}$. Next we want to find the energies and wavefunctions at points on a denser grid of wavevectors $\mathbf{k}$. Further we will consistently use $\mathbf{q}$ and $\mathbf{k}$ to denote the ab initio and interpolation grids respectively.

For a group of entangled bands one can define a set of $J$ Wannier functions defined as

$$|\mathbf{R}n\rangle = \frac{V}{(2\pi)^3} \int_{BZ} d\mathbf{q} e^{-i\mathbf{q} \cdot \mathbf{R}} \sum_{m=1}^{J_\mathbf{q}} |\psi_{mn\mathbf{q}}\rangle V_{mn}(\mathbf{q})$$

where $J_\mathbf{q} \geq J$. The matrices $V_{mn'}(\mathbf{q})$ are constrained by $\sum_{m=1}^{J_\mathbf{q}} V_{mn}(\mathbf{q})V_{mn'}(\mathbf{q}) = \delta_{mn'}$ and are chosen in such a way that the Wannier functions are localized, which yields that the Bloch wavefunctions in the Wannier gauge

$$|\psi_{nk}^W\rangle \equiv e^{i\mathbf{k} \cdot \mathbf{r}} |u_{nk}\rangle \equiv \sum_{\mathbf{R}} e^{i\mathbf{R} \cdot \mathbf{r}} |\mathbf{R}n\rangle$$

vary slowly with the $\mathbf{k}$ vector, unlike the true wavefunctions. Now let us see how Wannier functions may be used to interpolate the band energies. First, one evaluates the matrix elements of the Hamiltonian:

$$H_{mn}(\mathbf{R}) = \frac{1}{N} \sum_{\mathbf{q}} e^{-i\mathbf{q} \cdot \mathbf{R}} \langle \psi_{mn\mathbf{q}}^W |H|\psi_{mn\mathbf{q}}^W \rangle = \frac{1}{N} \sum_{\mathbf{q}} e^{-i\mathbf{q} \cdot \mathbf{R}} \sum_{l} V_{lm}(\mathbf{q}) E_{l\mathbf{q}} V_{ln}(\mathbf{q})$$

Next, to obtain energies at an arbitrary point $\mathbf{k}$ one needs to construct the Wannier Hamiltonian

$$H_{mn}^W(\mathbf{k}) = \sum_{\mathbf{R}} H_{mn}(\mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}}$$

which further may be diagonalized as

$$H_{mn}^W(\mathbf{k}) = U_{ml}(\mathbf{k}) E_{l\mathbf{q}} U_{nl}^*(\mathbf{k})$$

In a similar way, for any operator $\hat{X}$, for which the matrix elements are evaluated on the ab initio grid, one may obtain the real-space matrix elements

$$X_{mn}(\mathbf{k}) = \sum_{\mathbf{R}} X_{mn}(\mathbf{R}) e^{i\mathbf{k} \cdot \mathbf{R}}$$

and then may be interpolated to any $\mathbf{k}$ point in the Wannier gauge by

$$X_{mn}^W(\mathbf{k}) = \left(U^\dagger \cdot X^W \cdot U\right)_{mn}$$

For example, it was shown [12] that the total Berry curvature of the occupied manifold is interpolated via expression

$$\Omega_\gamma(\mathbf{k}) = \text{Re} \sum_n \Omega_{n\gamma}^{(H)} = 2\epsilon_{\alpha\beta\gamma} \text{Re} \sum_n \sum_l D_{nl,\alpha} A_{ln,\beta}^{(H)} + \epsilon_{\alpha\beta\gamma} \text{Im} \sum_n \sum_l D_{nl,\alpha} A_{ln,\beta}$$

where the ingredients of the equation are obtained using eqs. (7), (8) starting from $D_{nl,\alpha} \equiv \frac{\partial H_{mn}^W}{\partial E_{mn}}$, $H_{mn}^W \equiv \partial_{\mathbf{q}} H_{mn}^W$, $A_{mn,\alpha}(\mathbf{R}) \equiv (\langle 0|\hat{\mathbf{r}}_\alpha|\mathbf{R}n\rangle, \Omega_{n\gamma}^{(W)} \equiv \epsilon_{\alpha\beta\gamma} \partial_{\alpha} A_{\beta\gamma}^W$. The anomalous Hall conductivity is evaluated as an integral:

$$\sigma_{\alpha\beta}^{AHE} = -\frac{e^2}{h} \epsilon_{\alpha\beta\gamma} \frac{d}{(2\pi)^3} \Omega_\gamma(\mathbf{k})$$

Note, that while the direct Fourier transform (6) is performed only once for the calculation, and is not repeated for the multiple $\mathbf{k}$ point upon interpolation, the inverse Fourier transform (7) is repeated for every interpolation $\mathbf{k}$ point. And in fact it presents the most time-consuming part of the calculation involving Wannier interpolation as implemented in the Wannier90 code.
B. Mixed Fourier Transform

In this section we will see how the evaluation of Eq. (7) may be accelerated. It is easy to see that the computation time of a straightforward evaluation of discrete Fourier transform scales as \( t \propto N_k N_F T \), where we are typically interested in a case \( N_k \gg N_R \) (\( N_R \approx N_q \)).

When the Fourier transform is done on a regular grid of \( k \) points, it is usually appealing to use the fast Fourier transform (FFT). For that one needs to place the \( R \) vectors on a regular grid of size \( N_k \), fill the missing spots with zeros and perform the standard FFT, which will scale as \( t \propto N_k \log N_k \). However the drawbacks are manifold. First, FFT would give some advantage for moderate \( k \) grids, but not for really large ones (\( N_k \gg N_R \)), which are the main goal of doing Wannier interpolation at all. Moreover doing a FFT on a large grid means storing the data for all \( k \)-points in memory at the same time, which becomes a severe computational limitation. Also FFT does not allow reduce computation to only the symmetry-irreducible \( k \)-points, and is more difficult to do in parallel.

However there is a way to combine the advantages of both the FFT and the usual discrete Fourier transform, leading to the concept of mixed Fourier transform.

We want to evaluate Eq. (7) for a set of \( k \)-points,

\[
k_{n_1,n_2,n_3} = \frac{n_1}{N_k^1} b_1 + \frac{n_2}{N_k^2} b_2 + \frac{n_3}{N_k^3} b_3 \tag{11}
\]

where \( 0 \leq n_i < N_k^i \) – integers (\( i = 1, 2, 3 \)), \( N_k^i \) – size of interpolation grid, \( b_i \) — reciprocal lattice vectors.

Now suppose we can factorize \( N_k^i \) (\( i = 1, 2, 3 \)) as \( N_k^i = N_{FFT}^i N_k^i \). Then the set of points (11) is equivalent to a set of points \( k = K + \kappa \), where

\[
K_{l_1,l_2,l_3} = \frac{l_1}{N_k^1} b_1 + \frac{l_2}{N_k^2} b_2 + \frac{l_3}{N_k^3} b_3 \tag{12a}
\]

\[
k_{m_1,m_2,m_3} = \frac{m_1}{N_{FFT}^1} b_1 + \frac{m_2}{N_{FFT}^2} b_2 + \frac{m_3}{N_{FFT}^3} b_3 \tag{12b}
\]

where \( 0 \leq l_i < N_{FFT}^i \), \( 0 \leq m_i < N_{FFT}^i \). This separation is illustrated in Fig 1(a), which shows a \( 2 \times 2 \) K-grid, each corresponding to \( 4 \times 4 \) FFT grid (dots of a certain color). Now for each \( K \)-point we can define

\[
X_{mn}(K,R) \equiv X_{mn}(R)e^{iK \cdot R} \tag{13}
\]

and then (7) reads as

\[
X^W(K,\kappa) = \sum_R X_{mn}(K,R)e^{i\kappa \cdot R} \tag{14}
\]

The principle idea of mixed Fourier transform consists in performing the fourier transform (14) as FFT, while (13) is performed as a usual discrete Fourier transform. The advantages of this approach are the following. First, the computational time scales as \( t_1 \propto N_k N_{FFT} \) for (13) and \( t_2 \propto N_k N_{FFT} \log N_{FFT} \) for (14). Because it is required that \( N_{FFT} \geq N_R \) (to fill all \( R \)-vectors in the FFT box), we have \( t_1 \ll t_2 \propto N_k \log N_{FFT} \), which scales better then both the Fast and 'slow' Fourier transform. Next, we can perform Eqs. (13) and (14) independently for different \( K \)-points. This saves us memory, and also offers a simple parallelization scheme. Also we can further restrict evaluation only to symmetry irreducible \( K \)-points (Sec. II C) and also perform adaptive refinement over \( K \)-points (Sec. II D).

Moreover, the evaluation time of a mixed Fourier transform only logarithmically depends on the size of the \( ab \) \( initio \) grid (recall that \( N_{FFT} \sim N_R \sim N_q \)), while for the slow Fourier transform, the dependence is linear. However, in practice we will see (sec. IV) that the Fourier transform in the present implementation consumes only a small portion of computational time, and therefore the overall computational time is practically independent of the size of the \( ab \) \( initio \) grid.

C. Symmetries

When we integrate some quantity over the BZ, at every \( K \)-point (after summing over \( \kappa \) points) we obtain the result as a rank-\( m \) tensor \( X_{I_1,\ldots,I_m}(K) \). Then the BZ integral is expressed as a sum

\[
\bar{X} = \sum_K X(K)w_K \tag{15}
\]

and initially set \( \{K\} \) as a regular grid of \( N_k = N_k^1 \times N_k^2 \times N_k^3 \) points, and \( w_k = 1/N_k \). Suppose \( G \) is the point group of the system. We define the set of symmetry-irreducible \( K \)-points as a set of points that \( \forall (K,K') \in \text{irr} \), \( gK \neq K' \), unless \( g = E \) (identity). Then we can split the sum (15) as

\[
\bar{X} = \sum_{\text{irr}} X(K)w_K + \sum_{\text{all} \setminus \text{irr}} X(K)w_K = \sum_{\text{irr}} X(K)w_K + \sum_{\text{all} \setminus \text{irr}} g_K X(g_K^{-1} K)w_K \tag{16}
\]

Where in the second line we choose \( g_K \) such that \( g_K^{-1} K \in \text{irr} \). (this choice may be not unique). Thus, only the irreducible \( K \)-points need to be evaluated. Next, to make sure that the result respects the symmetries, despite possible numerical inaccuracies, we symmetrize the result as:

\[
\bar{X} = \frac{1}{|G|} \sum_f f X \tag{17}
\]

Note, that \( \bar{X} = X \) if the model respects the symmetry precisely (e.g. when symmetry-adapted Wannier functions [40] are used). Combining (16) and (17) and using \( \sum_f f \circ g_K = \sum_f g_K \) we get

\[
X = \frac{1}{|G|} \sum_f \left[ \sum_{\text{irr}} X(K) \left( \sum_{K'} w_{K'} \right) \right] \tag{18}
\]
where \( G \cdot \mathbf{K} \) denotes the orbit of \( \mathbf{K} \) under action of group \( G \). The latter equation reflects the implementation in the \( \text{WB} \) code. Starting from a regular grid of \( \mathbf{K} \) points we search for pairs symmetry-equivalent points. Whenever such a pair is found, one of the points is excluded and its weight is transferred to the other point. Compare fig 1(a) and (b): the red points are removed and their weight is moved to green points. Thus we end with a set of irreducible \( \mathbf{K} \)-point with weights \( \sum_{G \cdot \mathbf{K} \rightarrow \mathbf{K}'} w_{\mathbf{K}'} \). Next we evaluate \( X(\mathbf{K}) \) (employing the corresponding interpolation scheme) only at symmetry-irreducible \( \mathbf{K} \)-points. Note, that although some \( \mathbf{k} \)-points corresponding to the same \( \mathbf{K} \)-point (same color in fig. 1) are equivalent, we have to evaluate them all to be able to use the FFT. Finally, after summation, we symmetrize the result. The described procedure achieves two goals: (i) reduce the computational costs and (ii) make the result precisely symmetric, even if the Wannier functions are not perfectly symmetric.

D. Recursive adaptive refinement

It is well known that in calculations of quantities involving Berry curvature, orbital moments, etc, one performs integration over \( \mathbf{k} \)-space of a function that strongly depends on \( \mathbf{k} \). As a result, small areas of \( \mathbf{k} \)-space give the major contribution to the integral. Such areas often appear in the vicinity of Weyl points, nodal lines, as well as avoided crossings. To accelerate convergence with respect to the number of \( \mathbf{k} \) points, we utilize adaptive mesh refinement similar to Refs. [12, 38]. The authors of [12, 38] assumed a pre-defined threshold, and the \( \mathbf{k} \)-points yielding Berry curvature above the threshold were refined. This is inconvenient because one needs a good intuition to guess an optimal value for this threshold, because it depends both on the quantity one wants to calculate, and the material considered.

In \( \text{WB} \) it is implemented in a way that does not require initial guess from the user. This procedure, in combination with symmetrization described above, is illustrated in Fig. 1. After excluding symmetry-equivalent \( \mathbf{K} \)-points (Fig. 1b) the results are evaluated for every \( \mathbf{K} \) point and stored. Then we pick a few "most important \( \mathbf{K} \)-points". The criteria of importance may be different - either the Maximal value for any \( E_F \), or maximal value summed over all \( E_F \), or yielding most variation over the \( E_F \). Suppose it is the magenta point. Then those points are refined — replaced with 8 points around it (in Fig. 1c 4 new blue \( \mathbf{K} \)-points in the 2D case), with weight of the initial point distributed over the new points. Then the symmetrization is applied again (the four blue points are connected by 4-fold rotation) to exclude the
equivalent points. After the new \( \mathbf{K} \)-points are evaluated, we go to the next iteration of refinement. On each iteration any point may be refined, including both those from the initial regular grid, and those created during previous refinement iterations. The procedure stops after the pre-defined number of iterations was performed. Fig 1(g) shows how undesired artificial peaks of the the AHC curve are removed iteration by iteration, yielding a smooth curve. (See sec. III for details)

E. Minimal-distance replica selection method

Minimal-distance replica selection (MDRS) method was shown to be very efficient in the Wannier interpolation [25], in particular when moderate \( \mathbf{q} \)-grids are used in the abinitio calculations. With MDRS method is the Fourier transform (7) is modified in the following way:

\[
X_{mn}^W(k) = \sum_{\mathbf{R}} \frac{1}{N_{mnR}} \sum_{j=1}^{N_{mnR}} e^{i[k+\mathbf{T}^{(j)}_{mnR}]}X_{mnR}
\]

(19)

Where \( \mathbf{T}^{(j)}_{mnR} \) are \( N_{mnR} \) lattice vectors that minimise the distance \( |\mathbf{r}_m-(\mathbf{r}_n+\mathbf{R}+\mathbf{T})| \) for a given set \( m,n,\mathbf{R} \). However, the evaluation of Eq. (19) is quite slower then (7), because every \( \mathbf{k},m,n,\mathbf{R} \) an extra loop over \( j \) is needed. Therefore calculations employing MDRS in postw90.x (which is enabled by default) takes more time. Instead it is convenient to re-define the modified real-space matrix elements as

\[
\tilde{X}_{mnR} = \sum_{\mathbf{R}'} \frac{1}{N_{mnR'}} \sum_{j=1}^{N_{mnR'}} X_{mnR'}\delta_{\mathbf{R},\mathbf{R}'+\mathbf{T}^{(j)}_{mnR'}}
\]

(20)

only once for the calculation, and then the transformation to \( \mathbf{k} \)-space is performed via

\[
X_{mn}^W(k) = \sum_{\mathbf{R}} e^{ik\mathbf{R}}\tilde{X}_{mn}(\mathbf{R})
\]

(21)

The latter expression, having essentially same form as (7), can be evaluated via mixed Fourier transform, as described in Sec. II B.

Thus the MDRS method implemented in \( \text{WB} \) via Eqs. (20)-(21), and has no extra computational cost, while giving notable accuracy improvement.

F. Evaluation of Fermi-sea integrals

It is often needed to study anomalous Hall conductivity (AHC) not only for the pristine Fermi level \( E_F \), but considering it as a free parameter \( \epsilon \). On the one hand it gives an estimate of the accuracy of the calculation, e.g. sharp spikes may indicate that the result is not converge. On the other hand \( \epsilon \)-dependence gives access to the question of the influence of doping and temperature, and also allows calculation of anomalous Nernst effect (see eq. (A1)). As implemented in postw90.x, evaluation of multiple Fermi levels has a large computational cost. However there is a way to perform the computation of AHC for multiple Fermi levels without extra computational costs. To show this let’s rewrite (9), (10) as

\[
\sigma_{\alpha\beta}(\epsilon) = -\frac{2}{h}\sum_{\mathbf{K}} w_{\mathbf{K}} \Omega_{\gamma}(\mathbf{K},\epsilon)
\]

with

\[
\Omega(\mathbf{K},\epsilon) = \sum_{\kappa} \left( \sum_{n} P_{n}(\epsilon) + \sum_{l} \sum_{n} Q_{ln}(\epsilon) \right)
\]

(22)

where \( \mathbf{k} = \mathbf{K} + \kappa \), and the definitions of \( P_{n} \) and \( Q_{ln} \) straightly follow from (9). Now suppose we want to evaluate \( \mathcal{O}(\epsilon_i) \) for a series of Fermi levels \( \epsilon_i \). For different \( \mathbf{k} \)-points and Fermi levels \( \epsilon \) the sets of occupied \( \mathcal{O}(\mathbf{k},\epsilon) \) and unoccupied states \( U(\mathbf{k},\epsilon) \) change and repeating this summations many times may be computationally heavy. Instead we note that when going from one Fermi level \( \epsilon_i \) to another \( \epsilon_{i+1} \) only a few states at a few \( \mathbf{k} \)-points change from unoccupied to occupied. Let’s denote the set of such \( \mathbf{k} \)-points as \( \delta\kappa_i \) then, the change of the total Berry curvature is

\[
\delta\Omega_i = \Omega(\epsilon_{i+1}) - \Omega(\epsilon_i) = \sum_{\mathbf{k}} \left( \sum_{n} P_{n}(\epsilon_{i+1}) U_{\mathbf{k},\epsilon_{i+1}}(\mathbf{k}) - \sum_{n} P_{n}(\epsilon_i) U_{\mathbf{k},\epsilon_i}(\mathbf{k}) \right) = \sum_{\mathbf{k}} \left( \sum_{n} P_{n} + \sum_{l} \sum_{n} Q_{ln}(\epsilon_{i+1}) \delta\Omega_{\gamma}(\mathbf{K},\epsilon_{i+1}) \right) = \sum_{\mathbf{k}} \left( \sum_{n} P_{n} + \sum_{l} \sum_{n} Q_{ln}(\epsilon_{i+1}) \delta\Omega_{\gamma}(\mathbf{K},\epsilon_{i+1}) \right)
\]

(23)

where \( \delta\Omega_i(\mathbf{k}) \equiv \mathcal{O}(\mathbf{k},\epsilon_{i+1}) - \mathcal{O}(\mathbf{k},\epsilon_i) \). Note that if the step \( \epsilon_{i+1} - \epsilon_i \) is small, then \( \delta\kappa_i \) and \( \delta\Omega_i(\mathbf{k}) \) include only
few elements, if not empty. Hence the evaluation of 23 will be very fast. Thus, the full summation Eq. (22) is needed only for the first Fermi level.

In a similar way this approach may be applied to orbital magnetization and other Fermi-sea properties. E.g. the orbital magnetization may be written as

$$
M_r(k) = \sum_{n}^{\text{occ}} \text{Re} \left[ C_{n,\gamma}^{(H)} + E_n A_{n,\gamma}^{(H)} \right] - 2\epsilon_{\alpha\beta} \sum_{l}^{\text{unocc}} \sum_{n}^{\text{occ}} D_{nl,n}(A_{l,\beta}^{(H)} + A_{l,\beta}^{(H)}) E_n + \epsilon_{\alpha\beta} \sum_{l}^{\text{unocc}} \sum_{n}^{\text{occ}} D_{nl,n}(E_l + E_n) D_{l,\beta}^{(H)}
$$

where $C_{n,\gamma}(R) \equiv \langle 0| r_\alpha \cdot \hat{H} \cdot (r_\beta - R_\beta)|Rn\rangle$ and the other ingredients were explained under eq. (9). Eq. (24) is written following the approach of Ref. 13, but the result has a different form, which can be straightforwardly processed by analogy with eqs. (22), (23).

### III. EXAMPLE: AHC OF BCC FE

In this section the usage of the *WannierBerri* code is demonstrated on a simple example – anomalous Hall conductivity of bcc iron. First we performed the ab-initio calculations using the QuantumEspresso (QE) code[41] on a grid $8 \times 8 \times 8 \cdot q$-points, fixing the magnetization along [001] next. We construct 18 WFs. These two steps follow exactly the Tutorial#18 of Wannier90 (W90) [42], ad the reader is addressed to the documentation of Wannier90 and QE for details.

After that, the calculation is performed by the following short python script. First, we import the needed packages:

```python
import wannierberri as WB
import numpy as np
```

Then we read the information about the system and Wannier functions:

```python
system=WB.System_w90('Fe', berry=True)
```

from files `Fe.chk`, `Fe.eig`, `Fe.mmn` [43], or we can read all information from a file `Fe_tb.dat`, which is also written by Wannier90, or maybe composed by user from any tight-binding model:

```python
system=WB.System_tb('Fe_tb.dat', getAA=True)
```

Next, we define the symmetries of the system that we wish to take into account. In the *ab initio* calculation we have specified the magnetization along the $z$ axis, hence the symmetries that are preserved are inversion $I$, 4-fold rotation around the $z$ axis $C_{4z}$, and a combination of time-reversal $\mathcal{T}$ and 2-fold rotation around the $x$ axis $C_{2x}$. Here we need only the generators of the symmetry group.

```python
system.set_symmetries(['Inversion','C4z','TimeReversal*C2x'])
```

The other symmetries will be automatically obtained by taking products of these generators. E.g. the mirror $M_x = (C_{4z})^2 \cdot \mathcal{T}$.

Next we need to set the grids of $k$, $K$ and $\kappa$ points. Most conveniently it can be done by setting the 'length' parameter:

```python
grid=WB.Grid(system, length=100)
```

This will guarantee the grid to be consistent with the symmetries, and the spacing of $k$-points will be $\Delta k \approx 2\pi/\text{length}$. In this particular case a $k$-grid of $52 \times 52 \times 52$ points ($13 \times 13 \times 13$ $K$-grid, $4 \times 4 \times 4$ $\kappa$-grid) will be generated, but that depends on the size of the unit cell.

Next, we want to integrate the Berry curvature to get the AHC. This is done by the `WB.integrate` method.

```python
WB.integrate(system, grid,
Efermi=np.linspace(12.,13.,1001),
smearEf=10, # 10K
quantities=['ahc','cumdos'],
umproc=16,
adpt_num_iter=20,
fout_name="Fe")
```

and in addition to AHC we evaluate the cumulative density of states (cDOS) (A6). We consider Fermi level as a free parameter, scanning over a set of Fermi levels from 12 to 13 eV with a step of 1 meV, and a small smearing over the Fermi level corresponding to temperature 10K (~1 meV) is used. It is known, that in the BZ integration, some $k$ points may give a large contribution to the integral. This is especially strong for Berry curvature, which blows up near band degeneracies and avoided crossings, that fall close to the Fermi level. This is reflected as huge spikes in the $E_F$-resolved curves – see blue curve in Fig 1(g). To make the calculation more precise around such points, an adaptive recursive refinement algorithm is used, and we set the number of iterations to 20. The integration is done in parallel by means of the `multiprocessing` module and the parameter `numproc` specifies that a Pool of 16 worker processes is used.

From the cDOS we can find the precise position of the Fermi level $E_F = 12.610$ eV — the energy at which the cumulative DOS reaches 8 electrons per unit cell. This is more accurate, then the result evaluated from a coarse *ab initio* grid. Next, it is instructive to plot the AHC after each iteration. In Fig. 1g one can see that already after 20 iterations the chaotic peaks are removed, and we can get a reasonably smooth curve, which we have started from a rather coarse grid of only $52 \times 52 \times 52$ k-points.

This short example demonstrates that the calculations with `WB` may be run with a few lines of Python script. Appendix A describes more of the implemented functionality, and more options are under development or under testing. For more detailed and updated information please refer to the online documentation at [34].
IV. COMPUTATION TIME

In this section we will compare the calculation time for the calculations of anomalous Hall conductivity using postw90.x and WannierBerri. We will take the example of bcc Fe and vary different parameters. Calculations were performed on identical 32-core virtual nodes of the ScienceCloud cluster at University of Zürich. The nodes are based on AMD EPYC 7702 64-Core Processors with frequency 2GHz and 128 GB RAM per node, and one node was used per task.

The computation consists of two phases. First, some preliminary operations are done. Those include reading the input files and performing Fourier transform from \textit{ab initio} grid \textbf{q} to real-space vectors \textbf{R} : eqs. (3) and (6). This operation takes in \textit{WB} (postw90.x) between 2 (3) seconds for the small \textbf{q} grid $4 \times 4 \times 4$ and 2 (3) minutes for a large grid of $16 \times 16 \times 16$. This time is mostly taken by reading the large formatted text file \texttt{Fe.mmn}, and it is done only once and does not scale with the density of the interpolation grid. In \textit{WB} this is done in the constructor of the \texttt{System\_w90} class, and the object can be saved on disk using a pickle module, so that this operation does not repeat for further calculations.

Next comes the interpolation part itself, for which the evaluation time scales linearly with the number of \textbf{k}-points used. Further the time for an interpolation grid $200 \times 200 \times 200$ is given, which is a rather good grid to make an accurate calculation for this material.

We start with comparing time with the MDRS switched off and without use of symmetries in \textit{WB}. As can be seen in Fig. 2, for a small \textbf{q}-grid $4 \times 4 \times 4$ \textit{WB} is just slightly faster then postw90.x. However, for dense \textbf{q}-grids the computational time of postw90.x grows linearly with the number of \textbf{q} points, while in \textit{WB} it stays almost the same. This happens because in postw90.x the Fourier transform is major time-consuming routine. On the other hand, in \textit{WB}, although cost of the mixed Fourier transform is expected to grow logarithmically with the \textit{ab initio} grid, we do not see it because Fourier transform amounts only to $\sim 10\%$ of the computational time.

Next, we switch on the MDRS method in postw90.x, and the computational time grows by a factor of 5. On the other hand the computational time does not change (not shown).

Finally let’s switch on the use of symmetries in \textit{WB}. Thus the computational time decreases by a factor of 8. In the ultra-dense grid limit one would expect the speedup to be approximately equal to the number of elements in the group – 16 in the present example, due to exclusion of symmetry-equivalent \textbf{K}-points. But this does not happen, because we use an FFT grid of $25 \times 25 \times 25$ \textbf{k}-points, hence the \textbf{K}-grid is only $8 \times 8 \times 8$, and a considerable part of \textbf{K}-points are at high-symmetry positions. Therefore they do not have symmetric partners to be excluded from the calculation.

Thus we can see that the difference in computational time with postw90.x and \textit{WB} reaches 3 orders of magnitude for this example. Note that the examples above were performed only for the pristine Fermi level. Now let’s see what happens upon scanning the Fermi levels (Fig. 3). In \textit{WB} the computational time remains practically unchanged when we use upto $N_{\epsilon} \approx 1000$ Fermi levels, and only start to grow considerably at $N_{\epsilon} \sim 10^4$. On the other hand in postw90.x the computational time significantly grows with $N_{\epsilon}$, which is especially remarkable for small \textbf{q}-grids.

In this section we did not use the adaptive refinement procedure. However when on starts from a rather dense grid of \textbf{K}-points, the new \textbf{K}-points coming from the re-
finement procedure constitute only a small portion of the initial grid, and hence do not contribute much into computation time.

In this section we have shown that the methods suggested in this article help to significantly reduce the computation time from days to minutes. However, bcc iron is a simple example with only 1 atom per unit cell, and only 18 WFs are needed. More complicated systems will require more time. E.g. to obtain a converged value of ANE in PrAlGe [45] using wannier19 (early version of WB), the calculation took approximately 30 hours on the same computation node. Estimates predict that the same calculation with postw90.x could take several months. Thus it is the case where the numerical advance not only saves time, but also brings the calculation from the unreasonably time-consuming area, where most people would avoid working, to a reasonably feasible computation time.

V. SUMMARY

In this article I have presented a number of methods that allow to boost the performance of Wannier interpolation to a new level. The methods are implemented in the new Python code WannierBerri. It is important to note that the mixed Fourier transform and the optimization of MDRS method and Fermi-level iteration, while giving a large computational advantage, do not affect the result within machine precision. Hence the new code can be easily benchmarked with the established postw90.x code. The code not only allows to perform high-speed and high-precision calculations of AHC and a palette of other properties, it also serves as a platform for implementation new features involving Wannier interpolation. Thus it has a potential to become a new community code. Interestingly, the code uses the same routines to perform calculations both from Wannier Functions and for tight-binding models models. Finally, in combination with recent advances in automated construction of Wannier functions, it paves a way to high-throughput calculations of properties of solids that require Wannier interpolation.

ACKNOWLEDGEMENTS

I thank Ivo Souza for useful discussions, and also Xiaoxiong Liu, Miguel Angel Jimenez Herrera, Patrick Lannggenhager, Jae-Mo Lihm, Minsu Ghim for fixing bugs and further development of the code, which will be described elsewhere. I acknowledge support from the Swiss National Science Foundation (grant number: PF00P2_176877) and from the European Unions Horizon 2020 research and innovation program (ERC-StG-Neupert-757867-PARATOP) and the NCCR MARVEL.

Appendix A: Capabilities of WannierBerri

This appendix outlines the functionality that is currently implemented in WannierBerri. For more detailed and updated information please refer to the online documentation [34]

1. Integration

The code may be used to evaluate the following quantities, represented as Brillouin zone integrals:
- ‘ahc’ : intrinsic anomalous Hall conductivity via eq. (10).
- Anomalous Nernst conductivity [47] ααβ may be obtained from \( \sigma_{\alpha\beta}(\varepsilon) \) evaluated over a dense grid of Fermi levels \( \varepsilon \).
\[
\alpha_{\alpha\beta} = -\frac{1}{e} \int d\varepsilon \frac{\partial f}{\partial \varepsilon} \sigma_{\alpha\beta}(\varepsilon) \frac{\varepsilon - \mu}{T} \quad (A1)
\]
where \( f(\varepsilon) = 1/\left(1 + e^{\varepsilon/\mu} \right) \)
- ‘Morb’ : orbital magnetization
\[
M^\alpha_n(k) = \frac{e}{2\hbar} \text{Im} \epsilon_{\alpha\beta\gamma} \int [dk] \sum \limits_n \langle \partial_a u_{nk} | H_k + E_{nk} - 2E_F | \partial_b u_{nk} \rangle \quad (A2)
\]
- ‘berry_dipole’ : berry curvature dipole
\[
D_{\alpha\beta}(\mu) = \int [dk] \sum \limits_n \partial_\alpha \Omega^\beta_n \quad (A3)
\]
which describes, nonlinear Hall effect [48]
- ‘gyrotropic.Korb’ and ‘gyrotropic.Kspin’ : gyrotropic magnetoelectric effect (GME) [49] tensor (orbital and smin contributions):
\[
K_{\alpha\beta}(\mu) = \int [dk] \sum \limits_n \partial_\alpha m^\beta_n \quad (A4)
\]
- ‘conductivity.Ohmic’ : ohmic conductivity within the Boltzmann transport theory in constant relaxation time \( \tau \) approximation.
\[
\sigma^\text{Ohm}_{\alpha\beta}(\mu) = \tau \int [dk] \sum \limits_n \partial_\alpha E_{nk} \partial_\beta E_{nk} \delta(E_{nk} - \mu) = \tau \int [dk] \sum \limits_{E_{nk} < \mu} \partial^2_{\alpha\beta} E_{nk} \quad (A5)
\]
- ‘dos’ : density of states \( n(E) \)
- ‘cumdos’ : cumulative density of states
\[
N(E) = \int \limits_{-\infty}^E n(\varepsilon) d\varepsilon \quad (A6)
\]
2. Tabulating

WB can also tabulate certain band-resolved quantities over the Brillouin zone. This feature is called by, e.g.

```
WB.tabulate(system, grid, 
  quantities=['berry'],
  fout_name='Fe',
  numproc=16,
  ibands=ns.p.arange(4,10),
  Ef0=12.610)
```

which will produce files `Fe_berry-?.frmsf`, containing the Energies and Berry curvature of bands 4-9 (band counting starts from zero). The format of the files allows to be directly passed to the FermiSurfer visualization tool [50] which can produce a plot like Fig 4. Transformation of files to other visualization software is straightforward.

Currently the following quantities are available to tabulate:

- `'berry'`: Berry curvature
  \[ \Omega^2(k) = -\epsilon_{\alpha\beta\gamma} \text{Im} \langle \partial_\alpha u_{n\k} | \partial_\beta u_{n\k} \rangle \]  
  \[ (A7) \]

- `'morb'`: orbital moment of Bloch states
  \[ m^\gamma_{\alpha\beta}(k) = \frac{e}{2\hbar} \epsilon_{\alpha\beta\gamma} \text{Im} \langle \partial_\alpha u_{n\k} | H_k - E_{n\k} | \partial_\beta u_{n\k} \rangle \]  
  \[ (A8) \]

- `'spin'`: the expectation value of the Pauli operator:
  \[ s_\gamma(k) = \langle u_{n\k} | \sigma^\gamma | u_{n\k} \rangle \]  
  \[ (A9) \]

- `'V'`: the band gradients \[ \nabla_k E_{n\k} \]

3. Evaluation of additional matrix elements

Wannier interpolation starts from certain matrix elements defined on the \textit{ab initio} (\( \mathbf{q} \)) grid. Those matrix elements should be evaluated within the \textit{ab initio} code, namely within its interface to Wannier90. However, only QuantumEspresso [41] has the most complete interface \texttt{pw2wannier90.x}. The other codes provide only the basic interface, which includes the eigenenergies \( E_{n\k} \) and overlaps

\[ M^b_{in\k} = \langle u_{n\k} | u_{n\k+b} \rangle \]  
\[ (A10) \]

(file \texttt{.mmm}), where \( \mathbf{b} \) vector connects neighbouring \( \mathbf{q} \)-points. This information allows to interpolate the band energies (and their derivatives of any order) as well as Berry connections [15] and Berry curvature [12]. However, to evaluate the orbital moment of a Bloch state, one needs matrix elements of the Hamiltonian [13] (\texttt{.uHu} file)

\[ C^{b_1,b_2}_{in\k} = \langle u_{n\k+b_1} | H | u_{n\k+b_2} \rangle \]  
\[ (A11) \]

The evaluation of eq. (A11) is very specific to the details of the \textit{ab initio} code, and implemented only in \texttt{pw2wannier90.x}. To make it possible study of orbital magnetization with other \textit{ab initio} codes, the following workaround may be employed. By inserting a complete set of Bloch states at a particular \( \mathbf{q} \) point \( \mathbf{I} = \sum_{\mathbf{q}} |u_{n\k}(\mathbf{q})\rangle \langle u_{n\k}(\mathbf{q})| \) we can rewrite (A11) as

\[ C^{b_1,b_2}_{in\k} \approx \sum_{\mathbf{q}} \left(M^{b_1}_{in\k}\right)^* E_{n\k} M^{b_2}_{in\k} \]  
\[ (A12) \]

This equation is implemented within the wannierberri.mmn2uHu submodule, which allows to generate the \texttt{.uHu} file out of \texttt{.mmm} and \texttt{.eig} files. The equality in (A12) is exact only in the limit \( l_{\text{max}} \to \infty \) and infinitely large basis set for the wavefunctions representation. So in practice one has to check convergence for a particular system. However, in practice the term depending on \( C^{b_1,b_2}_{in\k} \) usually gives a minor contribution (although it is not guaranteed and depends on the choice of WFs), and the convergence of the total result may be easily achieved.

To interpolate the spin operator expectation value, the matrix \( S_{n\k} = \langle u_{n\k} | \sigma^\gamma | u_{n\k} \rangle \) is needed. To facilitate study of spin-dependent properties within VASP [51] code, a submodule wannierberri.vaspqp is included, which computes \( S_{n\k} \) based on the normalized pseudo-wavefunction read from the WAVECAR file. Note that the use of pseudo-wavefunction instead of the full PAW[52] wavefunction is an approximation, which however in practice gives a rather accurate interpolation of spin.

The \texttt{mmn2uHu} and \texttt{vasqp} modules were initially developed and used in [18] as separate scripts, but were not published so far. Now they are included in the WannierBerri package with a hope of being useful for the community.
Appendix B: Software availability

All software used and developed in this article is open-source and available for free. WannierBerri [34] is available via pip [35] and GitHub [36]. External libraries used in WannierBerri include NumPy [53], SciPy [54], pyFFTW[55] wrapper of FFTW3 library [56] and lazy-property [57]. Wannier90 [23–25] and QuantumEspresso [41] are available at [42] and [58]. Figures were produced using matplotlib [50] (Figs. 1-3) and FermiSurfer[50] (Fig. 4), and also with help of Inkscape [60].
nierisation, npj Computational Materials 6, 66 (2020).

31. Z. Zhang, R.-W. Zhang, X. Li, K. Koepernik, Y. Yao, and H. Zhang, High-throughput screening and automated processing toward novel topological insulators, The Journal of Physical Chemistry Letters 9, 6224 (2018), https://doi.org/10.1021/acs.jpclett.8b02800.

32. K. F. Garrity and K. Choudhary, Database of wannier tight-binding hamiltonians using high-throughput density functional theory (2020), arXiv:2007.01205 [cond-mat.matrl-sci].

33. JARVIS-WannierTB database, https://www.ctcms.nist.gov/jarviswtb/.

34. WannierBerri code, https://wannier-berri.org/.

35. PyPI repository, https://pypi.org/project/wannierberri.

36. WannierBerri git repository, https://github.com/stepan-tsirkin/wannier-berri.

37. During the work on the code and on the manuscript the author did not receive any funding from Basque institutions of foundations. Usage of Basque language is motivated by a coincidental similarity to the Berry phase, and good memories of the years that the author spent in the Basque region.

38. Y. Yao, L. Kleinman, A. H. MacDonald, J. Sinova, T. Jungwirth, D.-s. Wang, E. Wang, and Q. Niu. First principles calculation of anomalous hall conductivity in ferromagnetic bcc Fe, Phys. Rev. Lett. 92, 037204 (2004).

39. This is always possible unless $N_k$ is a prime number. But for really dense grids, we can adjust $N_k$ a bit, to be factorizable in any way we want.

40. R. Sakuma, Symmetry-adapted wannier functions in the maximal localization procedure, Phys. Rev. B 87, 235109 (2013).

41. P. Giannozzi, O. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. Chiarotti, A. Coc ca, M. Dederichs, S. de Gironcoli, S. Dal Corso, O. Dell’Agnello, A. Dorigo, A. Marzari, T. J. Schneider, P. G. Putz, A.分别. foundation, The Journal of Chemical Physics 128, 154105 (2008).

42. Wannier90 code, http://www.wannier.org/.

43. The first is written by Wannier90, the other two by the interface of the ab initio code (e.g. pw2wannier90.x).

44. multiprocessing python module, https://pypi.org/project/multiprocessing.

45. D. Destraz, L. Das, S. S. Tsirkin, Y. Xu, T. Neupert, J. Chang, A. Schilling, A. G. Grushin, J. Kohlbrecher, L. Keller, P. Puphal, E. Ponomkushina, and J. S. White, Magnetism and anomalous transport in the weyl semimetal pralge: possible route to axial gauge fields, npj Quantum Materials 5, 5 (2020).

46. N. Nagaosa, J. Sinova, S. Onoda, A. H. MacDonald, and N. P. Ong. Anomalous hall effect, Rev. Mod. Phys. 82, 1539 (2010).

47. D. Xiao, Y. Yao, Z. Fang, and Q. Niu. Berry-phase effect in anomalous thermoelectric transport, Phys. Rev. Lett. 97, 026603 (2006).

48. I. Sodemann and L. Fu, Quantum nonlinear hall effect induced by berry curvature dipole in time-reversal invariant materials, Phys. Rev. Lett. 115, 216806 (2015).

49. S. Zhong, J. E. Moore, and I. Souza, Gyrotropic magnetic effect and the magnetic moment on the fermi surface, Phys. Rev. Lett. 116, 077201 (2016).

50. M. Kawamura, Fermisurfer: Fermi-surface viewer providing multiple representation schemes, Computer Physics Communications 239, 197 (2019).

51. VASP code, https://www.vasp.at/.

52. P. E. Blöchl, Projector augmented-wave method, Phys. Rev. B 50, 17953 (1994).

53. T. E. Oliphant, A guide to NumPy, Vol. 1 (Trelgen Publishing USA, 2006).

54. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovskiy, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. Jarrod Millman, N. Mayorov, A. R. J. Nelder, E. Jones, R. Kern, E. Larson, C. Carey, I. Polat, Y. Feng, E. W. Moore, J. Vand erPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and S. . . Contributors, SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods 17, 261 (2020).

55. pyFFTW python module, https://github.com/pyFFTW.

56. M. Frigo and S. G. Johnson, The design and implementation of FFTW3, Proceedings of the IEEE 93, 216 (2005), special issue on “Program Generation, Optimization, and Platform Adaptation”.

57. lazy-property python module, https://pypi.org/project/lazy-property.

58. QuantumEspresso code, https://www.quantum-espresso.org/.

59. J. D. Hunter, Matplotlib: A 2d graphics environment, Computing in Science & Engineering 9, 90 (2007).

60. Inkscape vector graphics editor, https://inkscape.org.