# Database of Wannier tight-binding Hamiltonians using high-throughput density functional theory

Kevin F. Garrity[1], Kamal Choudhary[1,2]

1 Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA.

2 Theiss Research, La Jolla, CA, 92037, U.S.A.

**Abstract:**

Wannier tight-binding Hamiltonians (WTBH) provide a computationally efficient way to predict electronic properties of materials. In this work, we develop a computational workflow for high-throughput Wannierization of density functional theory (DFT) based electronic band structure calculations. We apply this workflow to 1771 materials (1406 3D and 365 2D), and we create a database with the resulting WTBHs. We evaluate the accuracy of the WTBHs by comparing the Wannier band structures to directly calculated spin-orbit coupling DFT band structures. Our testing includes k-points outside the grid used in the Wannierization, providing an out-of-sample test of accuracy. We illustrate the use of WTBHs with a few example applications. We also develop a web-app that can be used to predict electronic properties on-the-fly using WTBH from our database. The tools to generate the Hamiltonian and the database of the WTB parameters are made publicly available through the websites https://github.com/usnistgov/jarvis and https://jarvis.nist.gov/jarviswtb .

## Background & summary

Wannier functions (WF) were first introduced[1] in 1937, and have proven to be a powerful tool in the investigation of solid-state phenomenon such as polarization, topology, and magnetization[2]. Mathematically, WFs are a complete orthonormalized basis set that act as a bridge between a delocalized plane wave representation commonly used in electronic structure calculations and a localized atomic orbital basis that more naturally describes chemical bonds[2-8]. One of the most common ways of obtaining Wannier tight-binding Hamiltonians (WTBH)[9-11] is by using the Wannier90 software package[12] to generate maximally localized Wannier functions, based on underlying density functional theory (DFT) calculations. However, obtaining high-quality Wannier functions requires several choices by code users, including which bands and energy ranges to Wannierize, as well as a choice of starting orbitals. Therefore, in order to unlock the many materials properties that can be calculated with WTBH for use in high-throughput computations, we provide tools to automate the Wannierization of DFT band structures, and we generate a database of verified WTBH for use in future applications.

The computational advantage of Wannier functions comes from their localization, which allows the WTBH to be determined once on a relatively coarse real-space grid, and then Fourier transformed to obtain the Hamiltonian and its derivatives at arbitrary k-points in the Brillouin zone, allowing many expressions to be evaluated efficiently[13]. Many computationally expensive quantities such as the $Z_2$ index, Chern number, Fermi-surface, Weyl-chirality, Hall conductivity, spin-texture, photo-galvanic effect, thermoelectric coefficients, thermal properties, Landau level applications, gyrotropic effects, and shift-photocurrent[12,14-19] can be efficiently computed with WTBHs. In addition, many materials properties are based on localized phenomena[20-22] such as impurities[23], defects[24,25], excitons[26], polarons[27], screened electron-electron interaction[28], and

electron-phonon interactions[29], all of which can be modeled in a Wannier basis[30]. In addition, an examination of the Wannier Hamiltonian can provide intuition to help understand bonding that is difficult to get from examining the delocalized Kohn-Sham eigenvectors directly. They are also useful in second quantization based beyond-DFT calculations such as Dynamical Mean Field Theory (DMFT)[31,32].

Since its launch in 2011, the Materials Genome Initiative (MGI)[33] has spurred the generation of several high-throughput databases and tools such as from AFLOW[34], Materials Project[35], Open Quantum Materials Database (OQMD)[36], Materials Cloud[37], AiiDA[38], NOMAD[39], and NIST-JARVIS[40]. They have played key roles in the generation of electronic-property related databases to reduce the time between materials discovery and application. However, the development of WTBH databases and tools are still in the developing phases[41-46].

Since the work of Souza, Marzari, and Vanderbilt (SMV) [2,5-7], which requires an initial guess of the Wannier subspace and a minimization procedure to achieve maximum localization, there have been several methods proposed for determining localized Wannier functions with less human intervention. One method that has been applied in a high-throughput manner is the AFLOWπ projection method[41,42], which uses a projection of the Bloch states on localized atomic orbitals without minimization to construct a localized basis. A second is the Selected Columns of the Density Matrix (SCDM) method, which constructs a localized subspace without an initial guess, based on properties of the density matrix[44,45] . In this work, we instead use the original method of SMV, but we develop a workflow that can automatically construct the initial guess and set various parameters needed for Wannierization, as well as test the resulting WTBH.

The goal of this paper is to: a) develop a high throughput workflow for Wannierization of DFT calculations, b) develop a database of verified Wannier-based tight-binding Hamiltonians along

with all related input/output files, c) develop web-apps for convenient WTBH predictions. We use our Wannierization workflow on the JARVIS-DFT (https://jarvis.nist.gov/jarvisdft ) database which is a part of the MGI at NIST. The NIST-JARVIS[40] (https://jarvis.nist.gov) has several components such as JARVIS-FF[47,48], JARVIS-DFT[48-57], JARVIS-ML[49,51,57-59], JARVIS-STM[51], JARVIS-Heterostructure[53] and hosts material-properties such as lattice parameters[50], formation energies[60], 2D exfoliation energies[55], bandgaps, elastic constants[50], dielectric constants[59], infrared intensities[59], piezoelectric constants[59], thermoelectric properties[57], optoelectronic properties, solar-cell efficiencies[47,49], topological materials[17,21] , electric field gradient[61], and computational STM images[51]. The JARVIS-DFT database consists of ≈ 40000 3D and ≈1000 2D materials. As an initial step, we deploy our computational workflow on the materials that were recently predicted to be topologically non-trivial based on the spin-orbit spillage technique, including three dimensional (3D), two dimensional (2D), magnetic, non-magnetic, insulating, and metallic systems[52,60] including spin-orbit interactions. After obtaining the WTBH from DFT, we perform several checks to ensure the quality of the Hamiltonians. Although here we present results mainly for high-spillage materials, we will be extending this workflow to the entire JARVIS-DFT database. Currently, we have calculated Wannier Hamiltonians including spin-orbit coupling for 1406 3D and 365 2D materials, which can be used to efficiently calculate materials properties using either our software tools or other external software such as Wannier-tools[14], Z2Pack[62], WOPTIC[63], EPW[64]. We believe that releasing this database and toolset for use by the materials community should enable accelerated materials prediction and analysis.

## Methods

The methodology supporting the current project consists of several steps that are given in Fig. 1. The beginning of the procedure selects materials for Wannierization that we have prescreened to have strong spin-orbit coupling effects in our previous work and are therefore likely to be topological insulators or semimetals[52,60]. The main subject of the current work is the automation of the Wannierization, which proceeds by first selecting parameters for the Wannierization, including the initial guess for the Wannier functions and a "frozen window." We then perform the Wannierization and test the resulting WTBH. These steps are discussed in detail below.

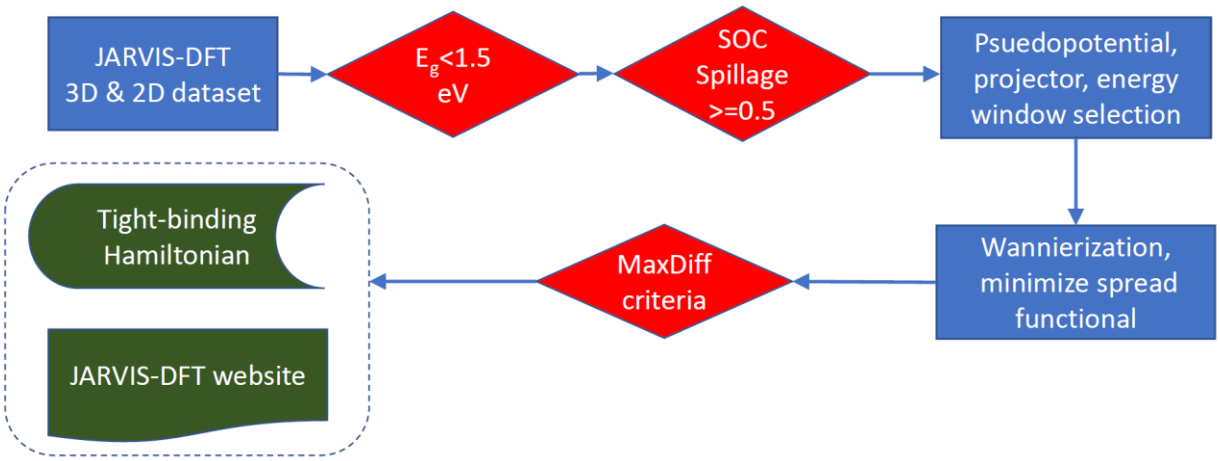



*Fig. 1 Workflow showing the Wannierization from using the DFT calculations.*

DFT calculations were carried out using the Vienna Ab-initio simulation package (VASP)[65] software using the workflow given on our JARVIS-Tools github page (https://github.com/usnistgov/jarvis)[66]. We use the OptB88vdW functional[67], which gives accurate lattice parameters for both vdW and non-vdW (3D-bulk) solids[50,55]. We optimize the crystal-structures of the bulk and monolayer phases using VASP with OptB88vdW. Because spin-orbit coupling (SOC) is not currently implemented for OptB88vdW in VASP, we carry out spin-orbit

PBE calculations. Such an approach has been validated by Refs. [60,68]. The crystal structure was optimized until the forces on the ions were less than 0.01 eV/Å and energy less than $10^{-6}$ eV. We use Wannier90[12] to construct Maximally-Localized Wannier Functions (MLWF) based TB-Hamiltonians.

The basic formalism of Wannierization is well-established. We briefly review some aspects here, interested readers can see longer discussions in [5,19]. For a set of Bloch eigenvectors $|\psi_{n,k}\rangle$, a general set of WFs $|\mathbf{R}n\rangle$ (n=1...N) can be written as:

$$|\mathbf{R}n\rangle = \frac{V}{(2\pi)^3}\int_{BZ}\sum_{m=1}^{N} U_{mn}^{(\mathbf{k})}\,|\psi_{mk}\rangle e^{-i\mathbf{k}.\mathbf{R}}d\mathbf{k} \qquad (1)$$

where **R** labels the unit cell of the WF, *V* is the volume of the unit cell, and $U_{mn}^{(\mathbf{k})}$ is an arbitrary unitary matrix. To construct maximally-localized WFs, $U_{mn}^{(\mathbf{k})}$ is chosen to minimize the following spread functional:

$$\Omega = \sum_n\left[\langle r^2\rangle_n - \bar{r}_n^2\right] \qquad (2)$$

where $\bar{\mathbf{r}}_n = \langle 0n|\mathbf{r}|0n\rangle$ and $\langle r^2\rangle_n = \langle 0n|r^2|0n\rangle$. The minimization proceeds iteratively, based on an initial guess of localized orbitals.

For the case of interest in this work, we wish to describe both the valence and conduction bands near the Fermi level. Therefore, it is necessary to first select a set of bands to Wannierize, and to separate these bands from the free-electron-like bands that overlap energetically with the conduction bands[62]. The procedure to determine this localized subspace of Bloch wavefunctions proceeds similarly to minimization described above, where after an initial guess, the subspace is iteratively updated in order to minimize the spread function in Eq. 2. After this initial disentanglement step, the Wannierization of the selected subspace proceeds as described above.

Due to the iterative non-linear minimization employed during both the disentanglement and Wannierization steps, the localization and utility of the final Wannier functions depend in practice on the initial choice of orbitals that are used to begin the disentanglement procedure, and which are then used as the initial guess for the Wannierization. Our initial guesses consist of a set of atomic orbitals we have chosen to describe all the chemically relevant orbitals for each element in typical elemental systems and compounds. We provide the list of the orbitals we select for each element in Table S1. For many specific materials, it may be possible to select a smaller set of orbitals while still maintaining high-quality WFs that describe the bands of interest; however, our fairly inclusive set of orbitals is able Wannierize nearly all compounds in a high-throughput manner without human intervention. Because most applications of WFs are computationally inexpensive compared to the DFT calculations used to construct the WFs, in practice, our larger Wannier basis has only minimal computational cost. However, it is necessary to have enough empty bands in the underlying DFT calculation such that any empty orbitals chosen are included in the Bloch basis. We do not include any semicore orbitals in our Wannier basis, as they are generally well-separated in energy from the valence orbitals and are not necessary to describe bands near the Fermi level.

During the disentanglement step, it is possible to choose an energy range that is included exactly ("the frozen window")[12], with the result that the Wannier band structure will exactly match the DFT band structure in this energy range and at the grid of k-points used in the Wannierization (see discussion in Sec. 2.I in Ref.[2] ). We use a default frozen window of $\pm$ 2 eV around the Fermi-energy. This window ensures that bands near the Fermi level are well described by the WTBH. Outside the frozen window, disagreement will tend to increase, as the procedure will select the most localized set of Wannier functions possible given the frozen window constraint, rather than

reproduce additional bands exactly. This disagreement outside the frozen window should not affect most properties computed using Wannier interpolation, which depend on bands near the Fermi level, but other choices may work better for some applications. For cases where the original WFs were unsatisfactory (see below), we found that lowering the lower bound of this window to include all the valence bands often improves that WTBH, and we use this as a second possible Wannierization setting.

In order to validate our WTBH, we calculate the maximum absolute difference ($\mu$) between the Wannier and DFT eigenvalues within an energy range of ± 2eV around the Fermi level:

$$\mu = \max_{n\mathbf{k}} \left( \left| E_{n\mathbf{k}}^{DFT} - E_{n\mathbf{k}}^{WTB} \right| \right) \quad (3)$$

As discussed above, at the grid of k-points used in the construction of the WFs and within the frozen window, the eigenvalues should match exactly by construction. Therefore, we require a different set of k-points to meaningfully test the WTBH. We choose to evaluate Eq. 3 on the dense lines of k-points along high symmetry directions that we already use to generate band structures. A weakness of this evaluation method is that highly dispersive energy bands (high $\frac{dE_{nk}}{dk}$) can result in high $\mu$ values even if the WTBH is of good quality because any slight shift in the *k*-direction of a dispersive band will result in a large energy error. We consider that systems with $\mu$ less than 0.1 eV to useful for most applications, and we provide data for the user to evaluate individual WTBH for their own applications.

Another failure mode for the Wannierization can be because the initial guess does not describe the DFT wavefunctions included in the Wannierization. This can happen either because important orbitals near the Fermi level were neglected, or if orbitals that were included have energies above

the energy range included in the DFT calculation. However, as demonstrated below, we find that our chosen set of initial orbitals works well in most cases.

**Data records**

After the calculations, the TB Hamiltonians, Wannier90 input and outputs files are stored as tar files and distributed through the Figshare repository[69]. Each 'zip' file consists of wannier90.win, wannier90.wout, wannier90_hr.dat files. The wannier90.win and wannier90.wout are the input and output files for Wannier90 code respectively. The wannier90_hr.dat file can be loaded as WanHam class with scripts in the JARVIS-Tools (https://github.com/usnistgov/jarvis) and similar packages to apply post-processing analysis such as calculating band-structures. There are also a JavaScript Object Notation (JSON) and Portable Network Graphic (PNG) file for comparing DFT bandstructure to WTBH.

**Technical validation**

To validate the WTBHs generated in this work, we compare the Wannier electronic bands with directly calculated DFT bands and measure the differences using Eq. 3 on two different k-point grids. As an example, in Fig. 2, we show an evaluation of the WTBH for $Bi_2Se_3$. In this figure, the top two panels show the WTBH evaluated on the same k-point grid used to generate the WFs, while the lower two panels show the evaluation on a typical set of high-symmetry k-points and lines, which includes k-points not used in the construction of the WFs. Figs. 2a and 2c show the eigenvalue comparison at separated k-points, with the WTBH bands in red and the DFT bands in blue, while Figure 2b and 2d show the eigenvalue differences as a function of energy.

As expected, the agreement within the frozen window and on the dense k-point grid is almost exact, but quickly increases up to 0.25 eV when leaving the window. We find a larger but still

small energy difference on the high symmetry grid Fig.2c-2d, with a maximal error in the frozen window of 9 meV. This test shows that this WTBH can be used to interpolate the band structure accurately.

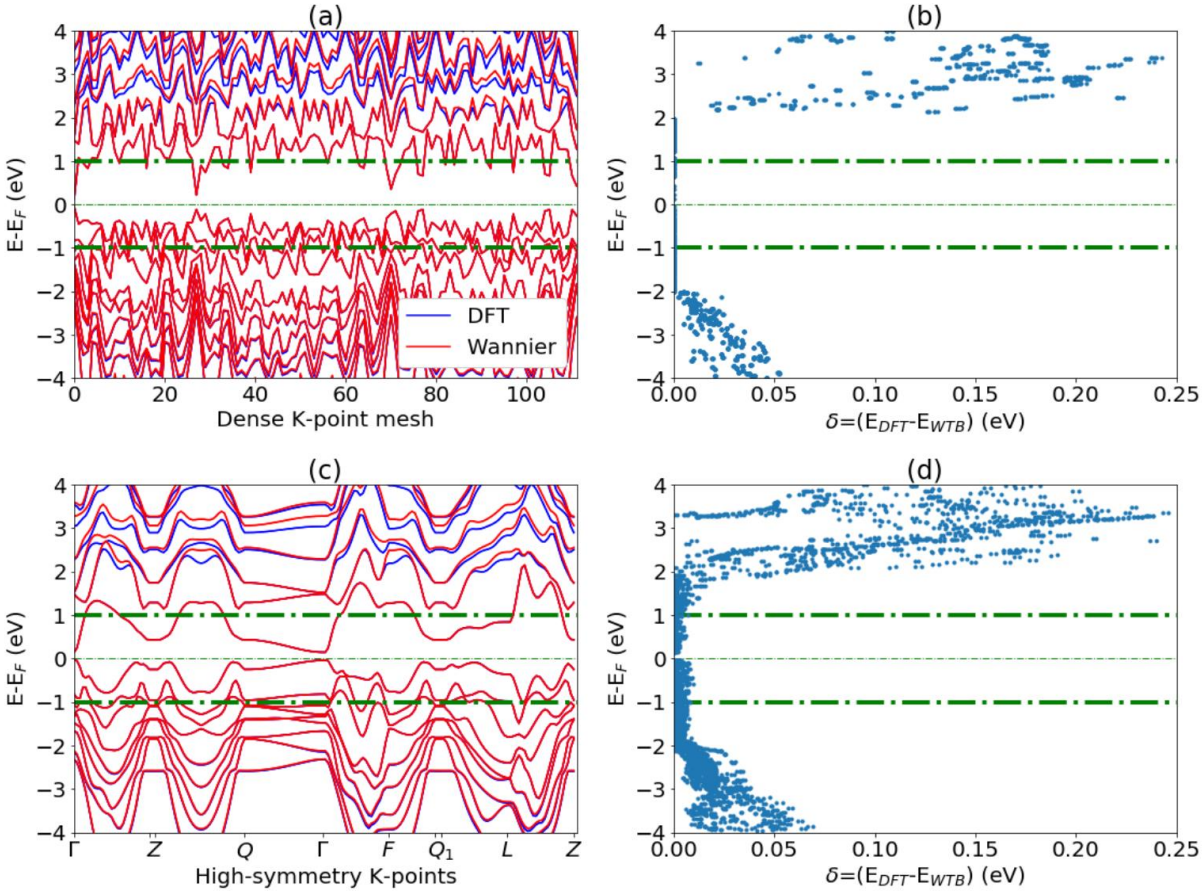



*Fig. 2 Comparison of DFT and WTB bandstructures for $Bi_2Se_3$. a-b) on dense k-grid, c-d) high-symmetry Brillouin zone points.*

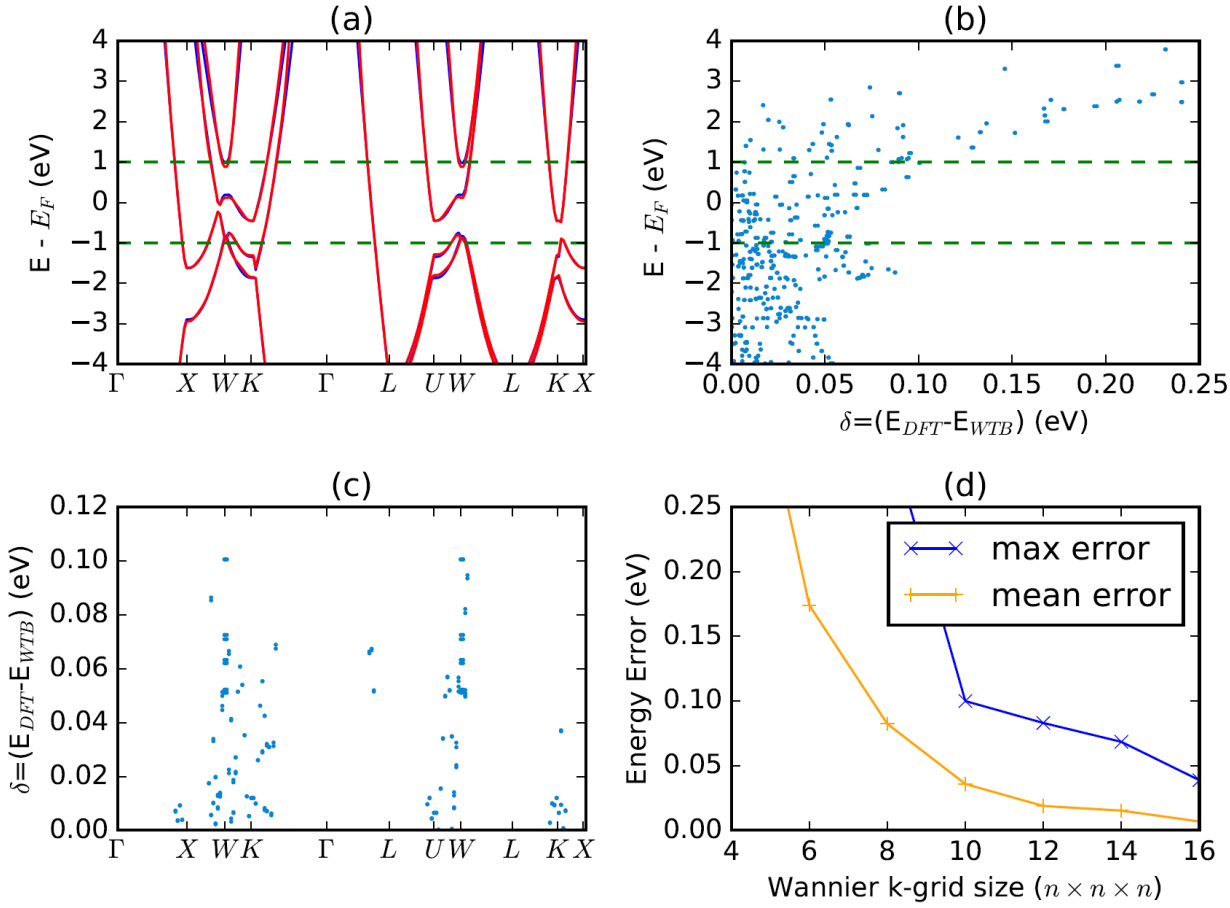



*Fig. 3 Comparison of DFT and WTBH bandstructures for Al (JVASP-816) on high-symmetry Brillouin zone. WTBH in a)-c) calculated using a 10×10×10 k-point grid. a) Band structure b) Energy error vs. DFT energy (eV) c) Energy error (eV) at each k-point for eigenvalues with ± 1 eV of the Fermi level. d) Maximum (blue) and mean (orange) energy error (eV) for WTBH made with different k-grids.*

Next, we consider the Wannierization of *fcc*-Al, a free-electron-like metal that is more difficult to Wannierize. In Fig. 3a-c, we show a comparison between the DFT and WTBH bands constructed using a 10×10×10 k-point grid during the Wannierization, which is the value used in our workflow (Fig. 1). We find very good agreement for the general band shape. However, in Figs. 3b-c, we show a detailed look at the error near the Fermi level, finding maximum errors of nearly 0.1 eV, with large errors occurring where dispersive bands cross the Fermi level, such as between Γ and

L. This higher error is due to the longer-range behavior of wavefunctions near the Fermi level in metals, as compared to the exponential decay in insulators[70,71], causing the Wannierization to require a higher density of k-points to converge. In Fig. 3d, we show the maximum and average errors for WTBH constructed with different k-point grids. While the average error decays reasonably quickly, the maximum error requires a very dense mesh to converge. During a high-throughput study, it is necessary to make reasonable tradeoffs between convergence and computational time. Therefore, we pick reasonable convergence parameters for our Wannierization and report an error assessment for each WTBH, allowing users to assess the suitability of each WTBH for their applications.

We show a few more examples of 3D WTBH in Fig. 4 for Si, PbTe, $Sb_2Te_3$ and $Na_3Bi$, this time focusing only on the difference for the high-symmetry k-point grids. Similar to the $Bi_2Se_3$ case discussed above, they show the minimal difference, and the WTBH are able to reproduce features such as the Dirac point band crossing of $Na_3Bi$ between Γ and A.

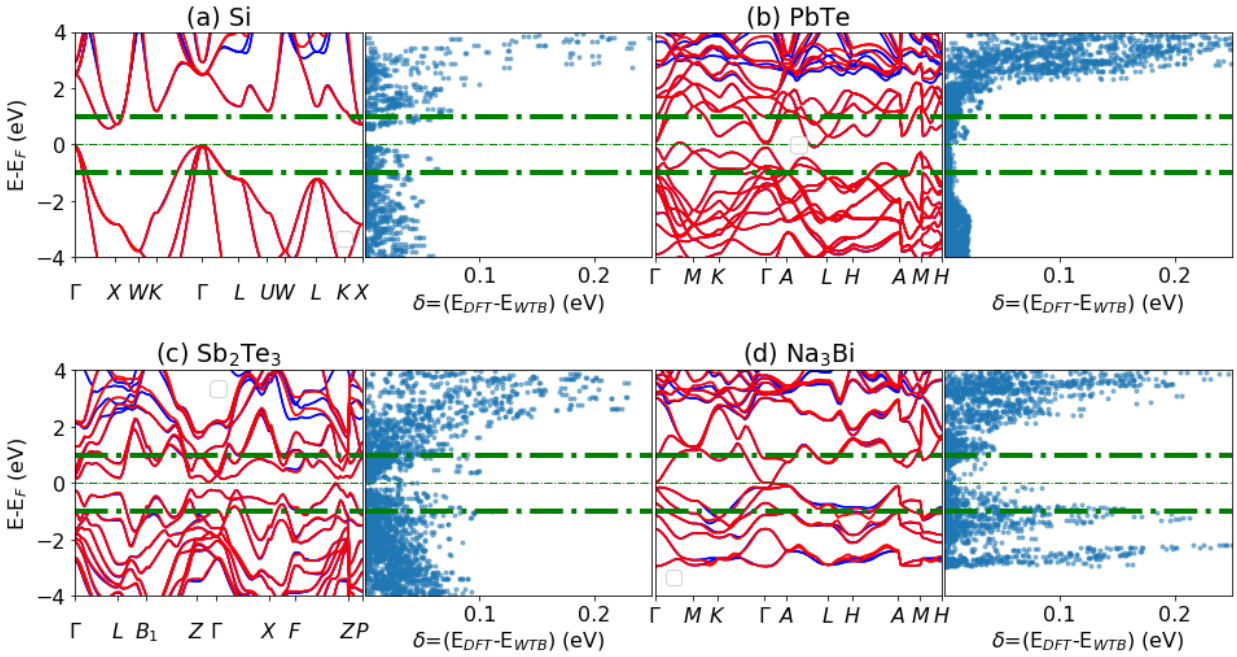


*Fig. 4 Examples of Wannier and DFT bandstructure and their energy difference plot for example 3D materials. a) Si, b) PbTe, c) $Sb_2Te_3$, and d) $Na_3Bi$.*

$Bi_2Se_3$, shown in Fig. 2, is a classic example of a 3D topological insulator. We show similar examples of 2D topological materials for graphene, $ZrFeCl_6$, $Ti_2Te_2P$, and $VAg(PSe_3)_2$ in Fig. 5. A detailed topological analysis of these materials can be found in our previous works[60]. Similar to the $Bi_2Se_3$ case, we observe that the DFT and WTBH bands overlap within the $\pm$ 2eV window and start to separate for outside these ranges. We again find excellent agreement between the DFT and the Wannier bands. Similar figures will be available for all the WTBH produced in this work on our website, so that the user can evaluate the WTBH for their own applications.

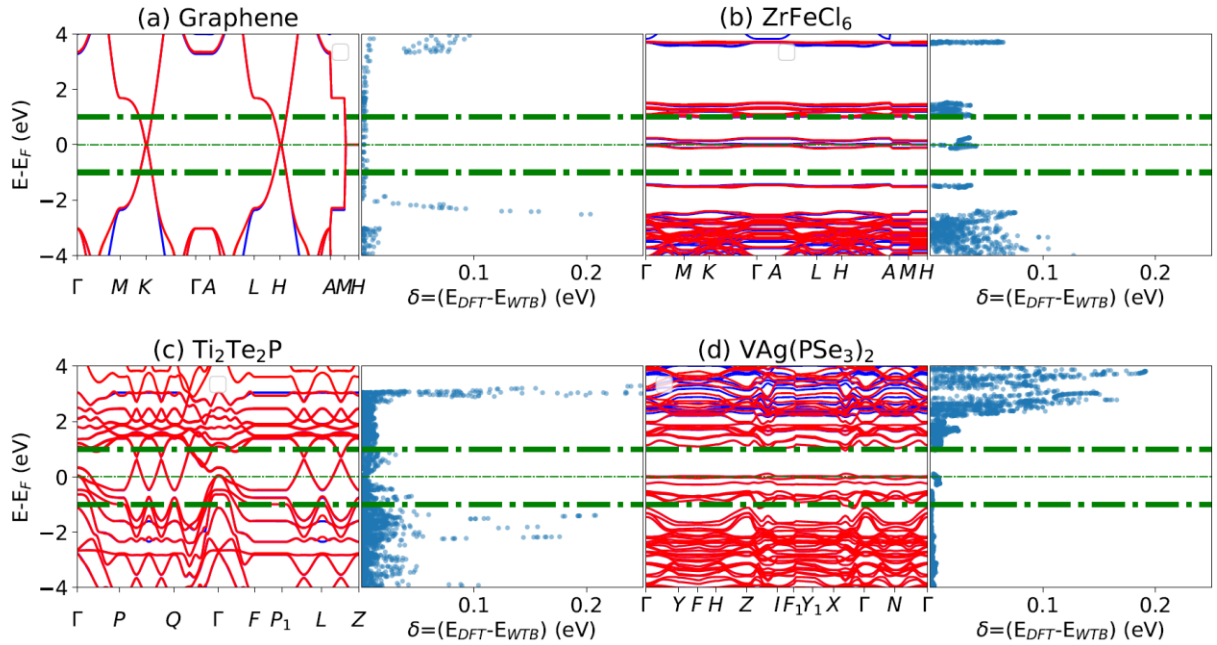



*Fig. 5 Examples of Wannier and DFT bandstructure and their energy difference plot for example 2D materials. a) for graphene, b) for $ZrFeCl_6$, c) for $Ti_2Te_2P$, and d) for $VAg(PSe_3)_2$.*

As is clear from the above examples, it is important to evaluate the energy difference between the DFT and WTBH bands to ensure a high-quality Wannierization. We use the maximum value of these differences (MaxDiff) for each k-point and in the disentanglement window range ($\pm$2 eV) as the measure of the quality of WTBHs (see Eq. 3). We calculate these differences for both the k-point grid and high-symmetry BZ points. Choosing a tolerance of 0.1 eV as the maximum energy difference, we find that 93.0 % of materials have a dense k-mesh MaxDiff less than the tolerance, while only 64 % of materials have high-symmetry BZ MaxDiff less than the tolerance as shown in Fig. 6a and 6b respectively. These larger discrepancies mainly occur for metallic systems such as Al, which have very dispersive electronic bands that naturally result in larger errors as discussed earlier (see Fig. 3). In the supplementary section (Table S2), we include the MaxDiffs of all materials we tried to Wannierize to help demonstrate the utility and limitations of this high-throughput approach.

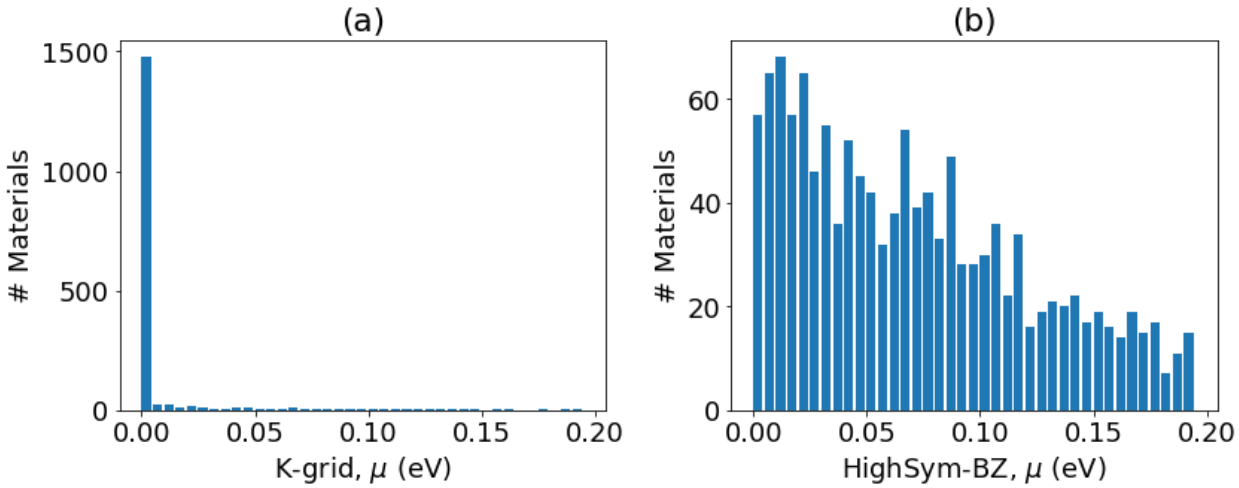



*Fig. 6 DFT-TB maximum difference ($\mu$) distribution for all the Wannier Tight-binding Hamiltonians (WTBHs). A) on a regular k-point grid, b) on high-symmetry k-points.*

In Fig. 7 we analyze the Wannier spread following Eqn. 2 for all the materials in the database. We find that most of the Wannier orbitals are well-localized with average spread of less than 3 $Å^2$. We do find a long tail of high Wannier spread in Fig. 7a. However, in Fig. 7b, we find little relationship between the Wannier spread and the accuracy of Wannier tight binding bands versus DFT. While high spread orbitals can indicate a failure of Wannierization, they can also be a result of including high energy states in the Wannierization, and these high spread Wannier functions may not affect the bands near the Fermi level.

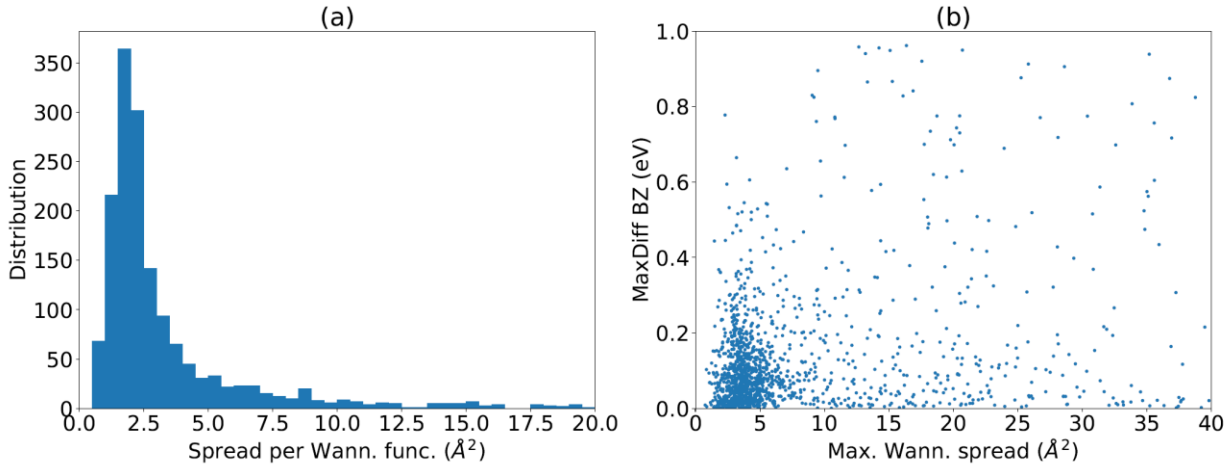



*Fig. 7 Analysis of Wannier function spread. a) distribution of average Wannier function spread for every material, b) comparison of maximum Wannier spread and the maximum Wannier and DFT band energy difference in a material.*

Next, we show a few example applications to demonstrate the usefulness of the WTB Hamiltonians. In Fig. 8a, we show the total and Bi (p) projected density of states in the $Bi_2Se_3$ system. The DOS can be evaluated with a very dense k-point grid at low computational cost using WFs, allowing detailed features to be converged. As mentioned in the introduction section, the WTB Hamiltonians can also be used to study defect phenomenon, especially if the defect only

removes weak vdW bonds. For example, in Fig. 8b, we show the (001) surface bandstructure of $Bi_2Se_3$. As expected for a $Z_2$ topological insulator, there is a bulk gap and a surface Dirac cone feature at Γ. Similarly, we show the edge band structure of a 2D monolayer of $VAg(PSe_3)_2$ with ferromagnetic spin ordering. $VAg(PSe_3)_2$ is a 2D Chern insulator[60], and the resulting spin-polarized conducing edge channel can be visualized in Fig. 8c.

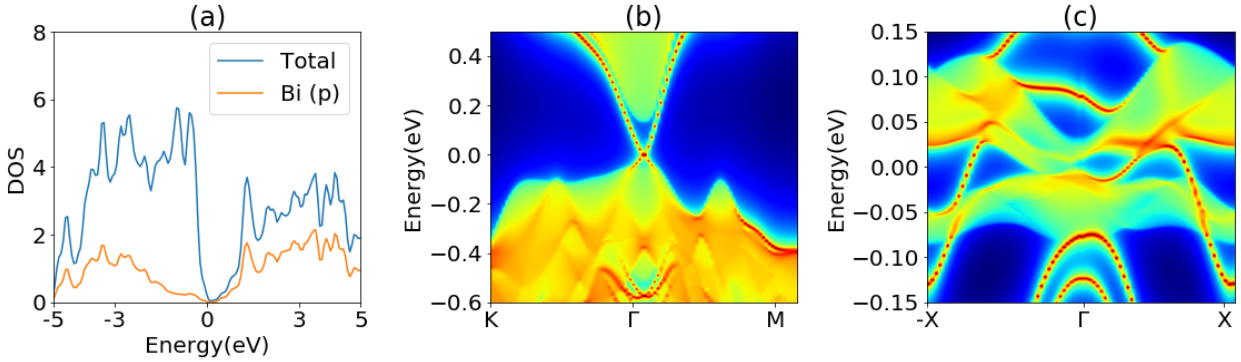
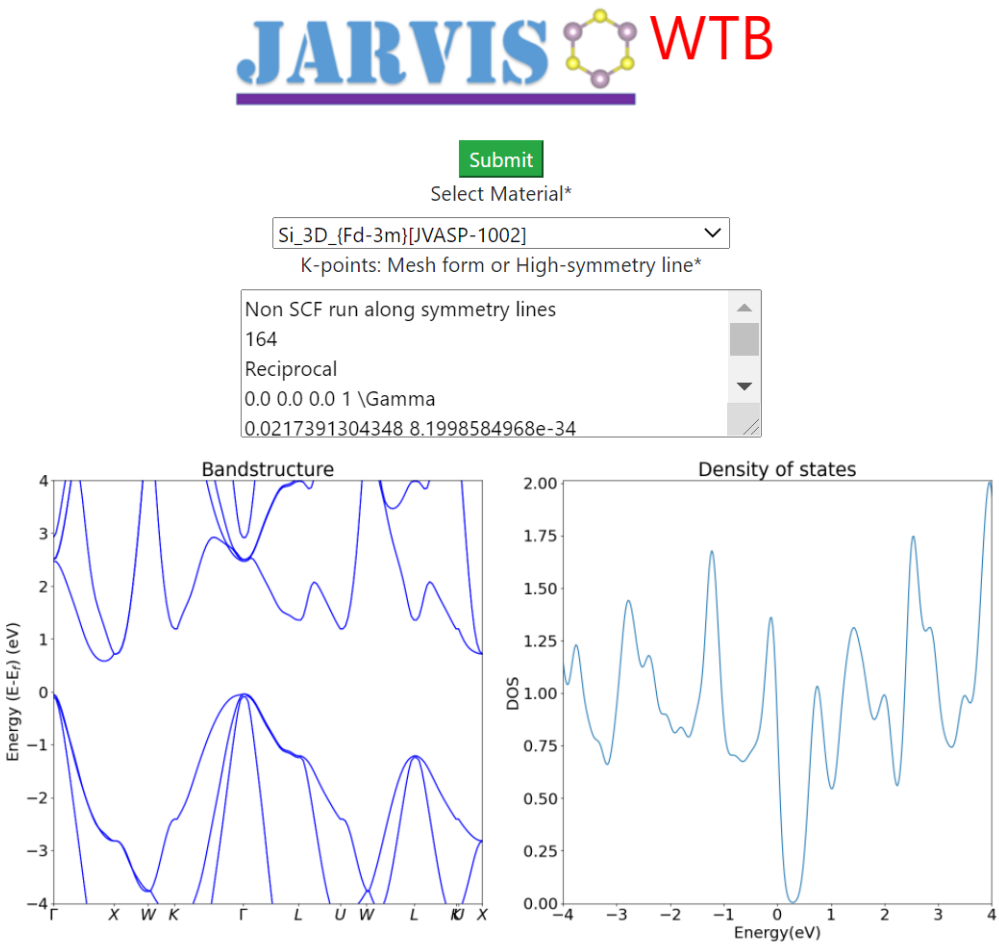



*Fig. 8 A few example applications of the WTB Hamiltonians. A) total and projected density of states, b) (001) surface band-structure of $Bi_2Se_3$, c) edge bandstructure of $VAg(PSe_3)_2$.*

Finally, in Fig. 9 we show a screenshot of a web-app we are developing to allow users to calculate materials properties using WTBH directly from our database, without downloading the Hamiltonians themselves. We curate the list of materials on the app to only include materials with MaxDiff <0.1 eV, but all of the WTBH are available to download. Currently, we support the calculation of Wannier-projected band structures for arbitrary k-points, as well as projected DOS. In addition, we provide plots to evaluate the accuracy of the WTBH. We plan to add other WTBH related functionalities in the app soon.




| Your selected material: | Si ,Fd-3m, JVASP-1002 |
|---|---|
| Maxdiff (K-Mesh, eV) | 0.0 |
| Maxdiff (High-symmetry, eV) | 0.066 |
| Calculation | Bandstructure and DOS |
| Nwann | 16 |
| Fermi-energy set to zero | 5.5427 |
| Computation-time (s) | 4.0919 |

*Fig. 9 Snapshot of the web-app available at https://jarvis.nist.gov/jarviswtb/ .*

**Usage notes**

The database presented here represents the largest collection of consistently calculated Wannier tight binding Hamiltonians of materials using density functional theory assembled to date. We anticipate that this dataset, and the methods provided for access will provide a useful tool in fundamental and application-related studies of materials. Our actual DFT verification provides

insight into understanding the applicability and limitation of our the WTBH data. The WTBH can be used to obtain important electronic properties such as band-structures, density of states, and topological invariants in a computationally efficient way. Data-analytics tools can also be applied on the generated dataset.

## Author Contributions

Both KFG and KC contributed in developing the workflow, analyzing data and writing the manuscript.

## Competing interests

The authors declare no competing interests.

## Code availability

Python-language based scripts for obtaining and analyzing the dataset are available at https://github.com/usnistgov/jarvis

## References


1 Wannier, G. H. The structure of electronic excitation levels in insulating crystals. *Phys. Rev*. **52**, 191 (1937).

2 Marzari, N., Mostofi, A. A., Yates, J. R., Souza, I. & Vanderbilt, D. Maximally localized Wannier functions: Theory and applications. *Rev. Md. Phys.* **84**, 1419 (2012).

3 Kohn, W. in *Condensed Matter Theories* 13 (Springer, 1993).

4 Kohn, W. J. Density functional/Wannier function theory for systems of very many atoms. *Chem. Phys. Lett.* **208**, 167-172 (1993).

5 Marzari, N. & Vanderbilt, D. Maximally localized generalized Wannier functions for composite energy bands. *Phys. Rev. B* **56**, 12847 (1997).

6 Mustafa, J. I., Coh, S., Cohen, M. L. & Louie, S. G. Automated construction of maximally localized Wannier functions: Optimized projection functions method. *Phys. Rev. B* **92**, 165134 (2015).

7 Cancès, É., Levitt, A., Panati, G. & Stoltz, G. Robust determination of maximally localized Wannier functions. *Phys. Rev. B* **95**, 075114 (2017).

8 Souza, I., Marzari, N. & Vanderbilt, D. Maximally localized Wannier functions for entangled energy bands. *Phys. Rev. B* **65**, 035109 (2001).

9 Papaconstantopoulos, D. & Mehl, M. J. The Slater–Koster tight-binding method: a computationally efficient and accurate approach. *J. Phys. Cond. Matt.* **15**, R413 (2003).

10 Goringe, C., Bowler, D. & Hernandez, E. Tight-binding modelling of materials. *Report Prog. Phys.* **60**, 1447 (1997).

11 Foulkes, W. M. C. & Haydock, R. Tight-binding models and density-functional theory. *Phys. Rev. B* **39**, 12520 (1989).

12 Mostofi, A. A. *et al.* wannier90: A tool for obtaining maximally-localised Wannier functions. *Comp. Phys. Comm.* **178**, 685-699 (2008).

13 Yates, J. R., Wang, X., Vanderbilt, D. & Souza, I. Spectral and Fermi surface properties from Wannier interpolation. *Phys. Rev. B* **75**, 195121 (2007).

14 Wu, Q., Zhang, S., Song, H.-F., Troyer, M. & Soluyanov, A. A. WannierTools: An open-source software package for novel topological materials. *Comp. Phys. Comm.* **224**, 405-416 (2018).

15 Vanderbilt, D. *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators*. (Cambridge University Press, 2018).

16 Pizzi, G., Volja, D., Kozinsky, B., Fornari, M. & Marzari, N. BoltzWann: A code for the evaluation of thermoelectric and electronic transport properties with a maximally-localized Wannier functions basis. *Comp.Phys. Comm.* **185**, 422-429 (2014).

17 Paul, I. & Kotliar, G. Thermal transport for many-body tight-binding models. *Phys. Rev. B* **67**, 115131 (2003).

18 Ibañez-Azpiroz, J., Tsirkin, S. S. & Souza, I. Ab initio calculation of the shift photocurrent by Wannier interpolation. *Phys. Rev. B* **97**, 245143 (2018).

19 Pizzi, G. *et al.* Wannier90 as a community code: new features and applications. *J. Phys. Cond. Matt.* **32**, 165902 (2020).

20 Silvestrelli, P. L., Marzari, N., Vanderbilt, D. & Parrinello, M. Maximally-localized Wannier functions for disordered systems: Application to amorphous silicon. *Solid State Comm*., 108, 7 (1998).

21 Busch, K., Mingaleev, S. F., Garcia-Martin, A., Schillinger, M. & Hermann, D. J. The Wannier function approach to photonic crystal circuits. *J. Phys. Cond. Matt.* **15**, R1233 (2003).

22 Pacchioni, G. J. T. Modeling doped and defective oxides in catalysis with density functional theory methods: Room for improvements. *J. Chem. Phys.* **128**, 182505 (2008).

23 Fitzhenry, P., Bilek, M., Marks, N., Cooper, N. & McKenzie, D. J. Wannier function analysis of silicon–carbon alloys. *J. Phys. Cond. Matt.* **15**, 165 (2002).

24 Leung, K.-M. Defect modes in photonic band structures: a Green's function approach using vector Wannier functions. *JOSA B* **10**, 303-306 (1993).

25 Gay, J. & Smith, J. Generalized-Wannier-function solutions to model surface potentials. *Phys. Rev. B* **11**, 4906 (1975).

26 Dvorak, M., Wei, S.-H. & Wu, Z. Origin of the variation of exciton binding energy in semiconductors. *Phys. Rev. Lett.* **110**, 016402 (2013).

27 Motta, C. & Sanvito, S. Electron–phonon coupling and polaron mobility in hybrid perovskites from first principles. *J. Phys. Chem. C* **122**, 1361-1366 (2018).

28 Şaşıoğlu, E., Schindlmayr, A., Friedrich, C., Freimuth, F. & Blügel, S. Wannier-function approach to spin excitations in solids. *Phys. Rev. B* **81**, 054434 (2010).

29 Agapito, L. A. & Bernardi, M. Ab initio electron-phonon interactions using atomic orbital wave functions. *Phys. Rev. B* **97**, 235146 (2018).

30 Lu, W.-C., Wang, C.-Z., Chan, T.-L., Ruedenberg, K. & Ho, K.-M. Representation of electronic structures in crystals in terms of highly localized quasiatomic minimal basis orbitals. *Phys. Rev. B* **70**, 041101 (2004).

31 Lechermann et al., Dynamical mean-field theory using Wannier functions: A flexible route to electronic structure calculations of strongly correlated materials, *Phys. Rev. B*, **74**, 125120 (2006).

32 Singh, V. *et al.* DMFTwDFT: An open-source code combining Dynamical Mean Field Theory with various Density Functional Theory packages. *arXiv 2002.00068* (2020).

33 de Pablo, Juan J., et al. The materials genome initiative, the interplay of experiment, theory and computation. *Curr. Opin. Solid State Mater. Sci.* **18**, 99 (2014).

34 Curtarolo, S. *et al.* AFLOWLIB. ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mat. Sci.* **58**, 227 (2012).

35 Jain, A. *et al.* Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *Apl Materials* **1**, 011002 (2013).

36 Kirklin, S. *et al.* The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comp. Mat.* **1**, 15010 (2015).

37 Talirz, L. et al., Materials Cloud, a platform for open computational science. *Sci. Data* **7**, 299 (2020).

38 Pizzi, G., Cepellotti, A., Sabatini, R., Marzari, N. & Kozinsky, B. AiiDA: automated interactive infrastructure and database for computational science. *Comp. Mat. Sci.* **111**, 218-230 (2016).

39 Draxl, C. & Scheffler, M. J. The NOMAD laboratory: from data sharing to artificial intelligence. *J. Phys. Mat.* **2**, 036001 (2019).

40 Choudhary, K. *et al.* JARVIS: An Integrated Infrastructure for Data-driven Materials Design. *npj Comput. Mater.* ***6****, 173 (2020)*.

41 Agapito, L. A., Ismail-Beigi, S., Curtarolo, S., Fornari, M. & Nardelli, M. B. Accurate tight-binding Hamiltonian matrices from ab initio calculations: Minimal basis sets. *Phys. Rev. B* **93**, 035104 (2016).

42 Supka, A. R. *et al.* AFLOWπ: A minimalist approach to high-throughput ab initio calculations including the generation of tight-binding hamiltonians. *Comp. Mat. Sci.* **136**, 76-84 (2017).

43 Agapito, L. A. *et al.* Accurate tight-binding Hamiltonians for two-dimensional and layered materials. *Phys. Rev. B* **93**, 125137 (2016).

44 Damle, A., Lin, L., Ying, L. J. Compressed representation of Kohn–Sham orbitals via selected columns of the density matrix. *J. Chem. Theory & Comp.* **11**, 1463-1469 (2015).

45 Damle, A., Lin, L. Disentanglement via entanglement: A unified method for Wannier localization. *Mult. Scale. Model. & Simul.* ***16***, 1392-1410 (2018).

46 Vitale, V. *et al.* Automated high-throughput wannierisation. *npj Comp. Mat*. **6**, 66 (2020).

47 Choudhary, K. *et al.* Evaluation and comparison of classical interatomic potentials through a user-friendly interactive web-interface. *Sci. Data* **4**, 1-12 (2017).

48 Choudhary, K. *et al.* High-throughput assessment of vacancy formation and surface energies of materials using classical force-fields. *J. Phys. Cond. Mat.* **30**, 395901 (2018).

49 Choudhary, K. *et al.* Accelerated Discovery of Efficient Solar-cell Materials using Quantum and Machine-learning Methods. *Chem. Mater.*, 31, 15, 5900 (2019).

50 Choudhary, K., Cheon, G., Reed, E. & Tavazza, F. Elastic properties of bulk and low-dimensional materials using van der Waals density functional. *Phys. Rev. B* **98**, 014107 (2018).

51 Choudhary, K. *et al.* Density Functional Theory and Deep-learning to Accelerate Data Analytics in Scanning Tunneling Microscopy. *arXiv* 1912.09027 (2019).

52 Choudhary, K., Garrity, K. F., Jiang, J., Pachter, R. & Tavazza, F. Computational search for magnetic and non-magnetic 2D topological materials using unified spin–orbit spillage screening. *npj Comp.Mat.* **6**, 1-8 (2020).

53 Choudhary, K., Garrity, K. F., Pilania, G. & Tavazza, F., Efficient Computational Design of 2D van der Waals Heterostructures: Band-Alignment, Lattice-Mismatch, Web-app Generation and Machine-learning. *arXiv*:2004.03025 (2020).

54 Choudhary, K., Garrity, K. F. & Tavazza, F. High-throughput Discovery of topologically Non-trivial Materials using spin-orbit spillage. *Sci. Rep.* **9**, 1-8 (2019).

55 Choudhary, K., Kalish, I., Beams, R. & Tavazza, F. High-throughput Identification and Characterization of Two-dimensional Materials using Density functional theory. *Sci. Rep.* **7**, 5179 (2017).

56 Choudhary, K. *et al.* Computational screening of high-performance optoelectronic materials using OptB88vdW and TB-mBJ formalisms. *Sci. Data* **5**, 180082 (2018).

57 Choudhary, K., Garrity, K. & Tavazza, F. Data-driven Discovery of 3D and 2D Thermoelectric Materials. *J. Phys. Cond. Mat.*, 32, 47, (2019).

58 Choudhary, K., DeCost, B. & Tavazza, F. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. *Phys. Rev. Mat.* **2**, 083801, doi:10.1103/PhysRevMaterials.2.083801 (2018).

59 Choudhary, K. *et al.* High-throughput Density Functional Perturbation Theory and Machine Learning Predictions of Infrared, Piezoelectric and Dielectric Responses. *npj Comp. Mat.*, 6, 64 (2019).

60 Choudhary, K., Garrity, K. F. & Tavazza, F. High-throughput Discovery of Topologically Non-trivial Materials using Spin-orbit Spillage. *Sci. Rep.* **9**, 8534 (2019).

61 Choudhary, K., Ansari, J. N., Mazin, I. I. & Sauer, K. L. Density functional theory-based electric field gradient database. *Sci. Data* **7**, 362, (2020).

62 Gresch, D. *et al.* Z2Pack: Numerical implementation of hybrid Wannier centers for identifying topological materials. *Phys. Rev. B* **95**, 075146 (2017).

63 Assmann, E. *et al.* woptic: Optical conductivity with Wannier functions and adaptive k-mesh refinement. *Comp. Phys. Comm.* **202**, 1-11 (2016).

64 Noffsinger, J. *et al.* EPW: A program for calculating the electron–phonon coupling using maximally localized Wannier functions. *Comp. Phys. Comm.* **181**, 2140-2148 (2010).

65 Kresse, G. & Furthmüller, J. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. C*omp. Mat. Sci.* **6**, 15-50 (1996).

66 Please note that commercial software is identified to specify procedures. Such identification does not imply recommendation by the National Institute of Standards and Technology. .

67 Klimeš, J., Bowler, D. R. & Michaelides, A. J. Chemical accuracy for the van der Waals density functional. *J. Phys. Cond. Mat.* **22**, 022201 (2009).

68 Cao, G. *et al.* Rhombohedral S b 2 S e 3 as an intrinsic topological insulator due to strong van der Waals interlayer coupling. *Phys. Rev. B* **97**, 075147 (2018).
69 Garrity, K, Choudhary, K, Database of Wannier tight-binding Hamiltonians using high-throughput density functional theory, *Figshare*,https://doi.org/10.6084/m9.figshare.c.5192276 .
70 Prodan, E. & Kohn, W. Nearsightedness of electronic matter. PNAS **102**, 11635-11638 (2005).
71 Brouder, C., Panati, G., Calandra, M., Mourougane, C. & Marzari, N. Exponential localization of Wannier functions in insulators. *Phys. Rev. Lett.* **98**, 046402 (2007).

**Figure legends:**

Fig. 1 Workflow showing the Wannierization from using the DFT calculations.

Fig. 2 Comparison of DFT and WTB bandstructures for $Bi_2Se_3$. a-b) on dense k-grid, c-d) high-symmetry Brillouin zone points.

Fig. 3 Comparison of DFT and WTBH bandstructures for Al on high-symmetry Brillouin zone. WTBH in a)-c) calculated using a 10×10×10 k-point grid. a) Band structure b) Energy error vs. DFT energy (eV) c) Energy error (eV) at each k-point for eigenvalues with ± 1 eV of the Fermi level. d) Maximum (blue) and mean (orange) energy error (eV) for WTBH made with different k-grids.

Fig. 4 Examples of Wannier and DFT bandstructure and their energy difference plot for example 3D materials. a) Si, b) PbTe, c) $Sb_2Te_3$, and d) $Na_3Bi$.

Fig. 5 Examples of Wannier and DFT bandstructure and their energy difference plot for example 2D materials. a) for graphene, b) for $ZrFeCl_6$, c) for $Ti_2Te_2P$, and d) for $VAg(PSe_3)_2$.

Fig. 6 DFT-TB maximum difference (μ) distribution for all the Wannier Tight-binding Hamiltonians (WTBHs). A) on a regular k-point grid, b) on high-symmetry k-points.

Fig. 7 Analysis of Wannier function spread. a) distribution of average Wannier function spread for every material, b) comparison of maximum Wannier spread and the maximum Wannier and DFT band energy difference in a material.

Fig. 8 A few example applications of the WTB Hamiltonians. A) total and projected density of states, b) (001) surface band-structure of $Bi_2Se_3$, c) edge bandstructure of $VAg(PSe_3)_2$.

Fig. 9 Snapshot of the web-app available at https://jarvis.nist.gov/jarviswtb/ .