

Automated construction of symmetrized Wannier-like tight-binding models from *ab initio* calculations

Dominik Gresch,¹ QuanSheng Wu,^{2,*} Georg W. Winkler,^{3,*} Rico Häuselmann,⁴
Matthias Troyer,^{1,5} and Alexey A. Soluyanov^{1,6,7}

¹Theoretical Physics, ETH Zurich, CH-8093 Zurich, Switzerland

²Institute of Physics, École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

³Station Q, Microsoft Research, Santa Barbara, California 93106-6105, USA

⁴Theory and Simulation of Materials (THEOS) and National Centre for Computational Design and Discovery of Novel Materials (MARVEL), École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland

⁵Microsoft Research, One Microsoft Way, Redmond, Washington 98052, USA

⁶St. Petersburg State University, St. Petersburg, 199034, Russia

⁷Physik-Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland



(Received 4 June 2018; revised manuscript received 27 August 2018; published 30 October 2018)

Wannier tight-binding models are effective models constructed from first-principles calculations. As such, they bridge a gap between the accuracy of first-principles calculations and the computational simplicity of effective models. In this work, we extend the existing methodology of creating Wannier tight-binding models from first-principles calculations by introducing the symmetrization post-processing step, which enables the production of Wannier-like models that respect the symmetries of the considered crystal. Furthermore, we implement automatic workflows, which allow for producing a large number of tight-binding models for large classes of chemically and structurally similar compounds or materials subject to external influence such as strain. As a particular illustration, these workflows are applied to strained III-V semiconductor materials. These results can be used for further study of topological phase transitions in III-V quantum wells.

DOI: [10.1103/PhysRevMaterials.2.103805](https://doi.org/10.1103/PhysRevMaterials.2.103805)

I. INTRODUCTION

A significant part of materials science is devoted to the problem of finding the electronic structure of a given material. As a result, numerous computational techniques have been developed to study this problem. These techniques can roughly be classified into two kinds: *First-principles* methods solve the problem using the fundamental physical principles and properties of atoms comprising the material. For weakly-interacting systems, density functional theory (DFT) [1] is the dominant (mean field) technique for solving the electronic structure problem from first principles.

In contrast, *empirical* methods aim to capture the relevant physical properties using a simplified model. Such models are usually matched to known properties of the material, which can be obtained from either experiments or first-principles calculations. An example of such an empirical method is given by the tight-binding approximation, which describes a material as a set of localized orbitals and predefined electron hopping terms between them. While the first-principles methods typically have superior accuracy, empirical methods are often used due to their lower computational cost. In particular, calculations of complex device geometries are often inaccessible to a direct first-principles study. As such, the construction of reliable empirical models is of significant importance, and

the technique of creating Wannier tight-binding models [2,3] from first-principles calculations is arguably one of the most popular tools in modern computational materials science. The use of Wannier tight-binding models allows one to combine the simplicity of empirical methods with the correct wave function properties obtained from first principles.

In recent years, *high-throughput* techniques made a profound impact in various fields of materials science [4–7]. While the domain eludes a strict definition, a common feature of such techniques is that computational tools are applied to a wide range of candidate materials, or variations of a given material, in search of some beneficial property. Existing codes and techniques are combined and applied on a scale that was not previously possible. A range of automated frameworks [8,9] support this by facilitating the combination of separate calculations into logical workflows. The challenge in designing such a high-throughput workflow is to make it resilient to varying input parameters. Since the number of calculations performed is too large to be human controlled, many decisions—for example which calculation to perform based on the output of a previous calculation—need to be encoded into the automated workflow.

In this paper, we introduce steps for addressing two standardly known problems of using Wannier90 [10,11] in combination with any *ab initio* software to construct tight-binding models: the absence of symmetries present in the original compound in the obtained tight-binding model and the necessity to search for optimal inner and outer energy windows for projection of the first-principles energy bands. We do not,

*The work was carried out at Theoretical Physics, ETH Zurich, 8093 Zurich, Switzerland.

however, treat the issue of selecting the initial projections used by Wannier90. As such, we create automated workflows which are applicable to large classes of materials with similar orbital character of the bands of interest. However, these workflows are not yet applicable to high-throughput scenarios in the sense that they can trivially be applied to arbitrary compounds. Nevertheless, the presented workflows are written in a way that they could be combined with efforts to address the problem of selecting initial projections [12].

In Sec. II, we review the general process of calculating the Wannier tight-binding models by means of Wannier90 and explain the proposed and implemented symmetrization and automatic energy window choice procedures. Section III describes how these procedures are used for the development of an automated workflow using the AiiDA [8] framework. While this workflow automates the tight-binding calculation itself, there are still some tunable parameters which might be eliminated by a more sophisticated system. By using a modular design approach, we provide an extensible framework for implementing such improvements. In the final section, we illustrate the application of this workflow to calculate tight-binding models for strained III-V semiconductor materials. These are useful in the pursuit of Majorana devices [13–15], enabling the study of transport properties for different topological devices with III-V quantum wells, where strains play an important role in the topological transition.

II. CONSTRUCTION OF WANNIER-LIKE TIGHT-BINDING MODELS

In this section, we describe the process of generating *symmetrized* Wannier-like tight-binding (SWTB) models. First, we give a short description of the method for creating Wannier tight-binding models (WTB) as introduced in the works of Refs. [2,3] and implemented in the Wannier90 [10,11] software package. Next, we describe a method for symmetrizing these WTBs in a post-processing step. Finally, we describe a scheme to enhance the band-structure accuracy by *optimizing the energy windows* used by Wannier90.

A. Wannier tight-binding construction

Tight-binding models represent a common way to describe crystalline systems in a computationally cheap way. The material is described as a system of localized orbitals with positions \mathbf{t}_i in the unit cell and hopping terms $H^{ij}[\mathbf{R}]$ between the j th orbital in the unit cell at location \mathbf{R} and the i th orbital in the home unit cell $\mathbf{R} = \mathbf{0}$. From these parameters, the matrix Hamiltonian can be written as [16]

$$H^{ij}(\mathbf{k}) = \sum_{\mathbf{R}} H^{ij}[\mathbf{R}] e^{i\mathbf{k} \cdot (\mathbf{R} + \mathbf{t}_j - \mathbf{t}_i)}. \quad (1)$$

For the case of spinful systems, we choose the indices i, j to include the spin index for simplicity.

The Wannier tight-binding (WTB) method utilizes localized Wannier functions as basis orbitals to capture the compound's physics. These basis Wannier functions are obtained from first-principles simulations. This procedure is based on the work of Refs. [2,3] and implemented in the Wannier90 [10,11] code. After obtaining the necessary Wannier90

input files from a first-principles calculation, two steps are performed to construct these Wannier functions.

In a first step, the Bloch wave functions $|\psi_{n,\mathbf{k}}\rangle$ calculated by the first-principles code are *disentangled* to obtain M wave functions, where M is the target number of basis Wannier functions in WTB. For selecting the Bloch wave functions which are involved in this procedure, one needs to choose an *outer* energy window. Optionally, an *inner* energy window can be chosen. States inside this inner window will be preserved by the disentanglement. An optimization routine is performed to select the M states such that the “change of character” Ω_1 (defined in Ref. [3]) is minimized. As an initial guess for this optimization procedure, M localized trial orbitals $|g_m\rangle$ are used. Because the disentanglement procedure needs to discard some states, it usually changes both the symmetry and the energy bands of the model in comparison with first-principles results. Consequently, choosing good values for both the energy windows and the trial orbitals has a strong effect on the quality of the resulting model.

As a second (optional) step, another optimization is performed to find a unitary transformation such that the resulting Wannier functions are maximally localized [2]. Again, the trial orbitals $|g_m\rangle$ are used to create an initial guess for this optimization. Typically, these orbitals are chosen to be those chemical atomic orbitals that contribute most to the bands of interest. A method for constructing Wannier orbitals without the need for such a guess is described in Ref. [12].

B. Symmetrization

An important feature of tight-binding models, especially for studying topological effects, is that they preserve certain crystal symmetries. For a given symmetry group G , the symmetry constraint on the Hamiltonian matrix is given by [17]

$$\forall g \in G : \mathcal{H}(\mathbf{k}) = D^{\mathbf{k}}(g) \mathcal{H}(g^{-1}\mathbf{k}) D^{\mathbf{k}}(g^{-1}), \quad (2)$$

where $D^{\mathbf{k}}(g)$ is the \mathbf{k} -dependent representation of the symmetry g from the group G . We define the \mathbf{k} -independent part $D(g)$ of the representation as

$$D^{\mathbf{k}}(g) = e^{i\boldsymbol{\alpha}_g \cdot \mathbf{k}} D(g), \quad (3)$$

where $\boldsymbol{\alpha}_g$ is the translation vector of the symmetry.

For a Hamiltonian which does not fulfill these symmetry constraints, we define the *symmetrized* Hamiltonian as the group average

$$\tilde{\mathcal{H}}(\mathbf{k}) = \frac{1}{|G|} \sum_{g \in G} D^{\mathbf{k}}(g) \mathcal{H}(g^{-1}\mathbf{k}) D^{\mathbf{k}}(g^{-1}). \quad (4)$$

This procedure projects the Hamiltonian onto the symmetric subspace, meaning that the modified Hamiltonian respects Eq. (2), as shown in Appendix A. Furthermore, if the original Hamiltonian is already symmetric, the original and symmetrized Hamiltonians are identical. Since this construction does not explicitly construct the corresponding Wannier functions, we term these models *symmetrized Wannier-like* tight-binding models (SWTB).

It is important to note that the eigenstates and eigenvalues of the symmetrized Hamiltonian may differ significantly from those of the nonsymmetrized Hamiltonian. In fact, for an

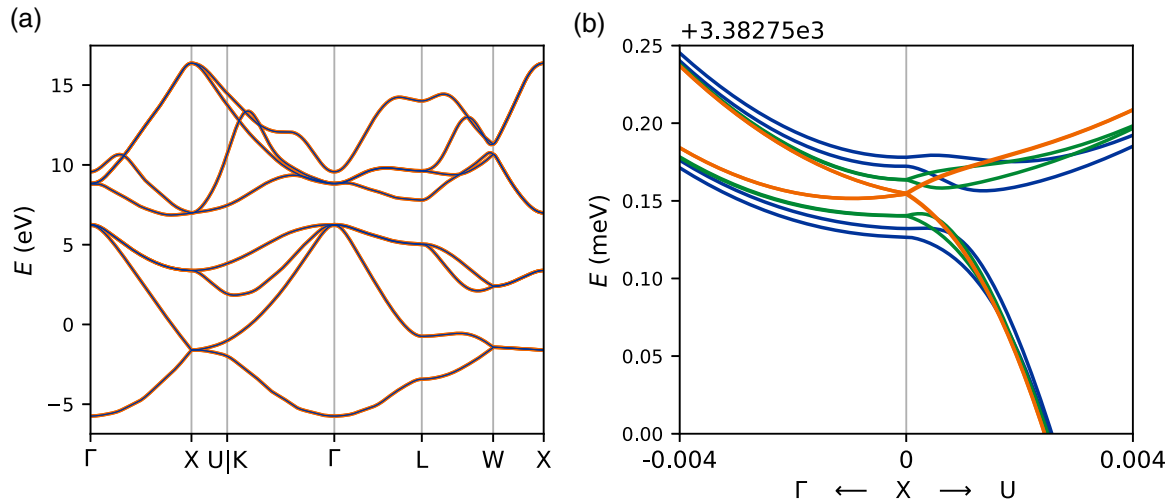


FIG. 1. Comparison of the initial (blue) and symmetrized (orange) band structure for a tight-binding model of silicon with atom-centered sp^3 orbitals. (a) In the eV scale, there are no visible differences between the two models. (b) A zoom in around the X point on the meV scale reveals a slight lifting of the band degeneracies in the initial model. This incorrectness is resolved in the symmetrized model. For comparison, a symmetrized band structure taking into account only symmorphic symmetries (green) is also shown.

antisymmetric initial Hamiltonian, meaning that

$$D^{\mathbf{k}}(g)\mathcal{H}(g^{-1}\mathbf{k})D^{\mathbf{k}}(g^{-1}) = -\mathcal{H}(\mathbf{k}) \quad (5)$$

for some symmetry g , the symmetrized result vanishes completely. However, given a Hamiltonian which *almost* respects the symmetry, this technique can effectively eliminate small symmetry-breaking terms.

In the context of tight-binding models, this symmetrization technique can only straightforwardly be applied when the underlying basis set is symmetric. If the tight-binding basis contains an orbital $|\alpha\rangle$ centered around the position \mathbf{r} , it must also contain $g|\alpha\rangle$ centered around $g\mathbf{r}$ for all symmetries $g \in G$. For example, if the model for a material which has C_4^x symmetry contains a p_x orbital at the origin, it must also contain a p_y orbital at the origin.

For Wannier tight-binding models, this means that the technique can generally only be applied when the step of maximally localizing the Wannier functions is omitted, and pre-defined atomic orbitals are used. When this condition is met however, the method can be applied for both unitary and antiunitary symmetries, as well as nonsymmorphic symmetry groups.

To apply the group average to tight-binding models, it is convenient to rewrite Eq. (4) directly in terms of the hopping matrices $H[\mathbf{R}]$ (see Appendix B for derivation):

$$\tilde{H}^{ij}[\mathbf{R}] = \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m}} D_{il}(g)H^{lm}[S_g^{-1}(\mathbf{R} - \mathbf{T}_{ij}^{ml})]D_{mj}(g^{-1}), \quad (6)$$

where S_g is the real-space rotation matrix of the symmetry g , $\mathbf{T}_{ij}^{ml} = S_g(\mathbf{t}_m - \mathbf{t}_l) - \mathbf{t}_j - \mathbf{t}_i$, and the indices m, l only go over values for which \mathbf{T}_{ij}^{ml} is a lattice vector. Note that we use the \mathbf{k} -independent part $D(g)$ of the representation here.

Figure 1 shows the results of this symmetrization procedure on a tight-binding model for bulk silicon in the diamond cubic crystal structure, with atom-centered sp^3 orbitals. The

initial model already approximately fulfills the symmetry condition, which is reflected in the fact that the band structure does not change in the electronvolt scale. However, at the submillielectronvolt scale the band degeneracies are lifted in the original model but restored after the symmetrization procedure. Since the symmetry group of the diamond cubic structure $Fd\bar{3}m$ (no. 227) is nonsymmorphic, this example demonstrates that the symmetrization technique is capable also of enforcing such symmetries. In panel (b) of Fig. 1, we compare the symmetrization using the full symmetry group to a partial symmetrization enforcing only the symmorphic subgroup. Adding nonsymmorphic symmetries enforces the fourfold degeneracy at the X point and twofold degeneracy on the X-U line, whereas symmorphic symmetries only enforce a twofold degeneracy on the Γ -X line.

To determine the matrix representations $D(g)$, we use the fact that Wannier90 allows one to manually choose the trial orbitals $|g_m\rangle$. As a result, the basis after the disentanglement procedure corresponds to the chosen orbitals, up to some numerical error. Since the behavior of the basis orbitals under symmetries is known, $D(g)$ can be determined in this way. For the treatment of spin, we use the rotation matrices as given in Ref. [18]. The action of time reversal on the spin basis $\{|\uparrow\rangle, |\downarrow\rangle\}$ is given by $\sigma_y \hat{K}$, where \hat{K} represents complex conjugation. An automated method for generating the representation matrices for given atomic orbitals is available in the symmetry-representation package. Importantly, we used Wannier90 *without* performing the maximal localization step. It is the case in the illustrated application of Sec. IV, where this allows us to preserve the orbital basis. Alternatively, one could use the basis transformation matrices $U^{(\mathbf{k})}$ provided by Wannier90 [10] to transform $D(g)$ into the maximally-localized basis. While this approach produces computationally cheaper localized models, the drawback is that the basis is different for each produced tight-binding model. As a result, comparing models is more difficult. Also, linear interpolation between models, as described in Sec. IV C, would require a change of basis.

Another approach to obtaining symmetric tight-binding models is to use the site-symmetry mode implemented in Wannier90 [19]. However, this method is limited to symmetries which leave a given real-space coordinate invariant (site symmetries) and does not include time reversal. The method presented here has no such limitation but is instead limited to models which have a symmetric set of basis functions as described above. The site-symmetry mode also relies on obtaining the symmetry information from the first-principles code, which is currently implemented only for Quantum Espresso [20,21]. The workflow described in Sec. III could be adapted to allow using this approach with only minimal changes.

C. Optimization for band-structure fit

As described above, an important parameter in running Wannier90 is the choice of the so-called energy windows [10]. There are two such windows: The *outer* window determines which states are taken into account for the disentanglement procedure. At every \mathbf{k} point, it must contain at least M bands, where M is the desired number of bands in the tight-binding model. The *inner* (or frozen) window on the other hand determines which states should not be modified during disentanglement. It can contain at most M bands at any given \mathbf{k} .

Since the quality of the resulting tight-binding model depends sensitively on the choice of energy windows, a strategy for reliably choosing good windows is required. A straightforward way of achieving this is by iteratively optimizing the window values. Having constructed and symmetrized a tight-binding model, its quality can be determined by comparing its band structure to a reference computed directly from first principles [22]. As a measure of their mismatch, we choose the average difference between the energy eigenvalues

$$\Delta = \frac{1}{M} \frac{1}{N_{\mathbf{k}}} \sum_{i=1}^M \sum_{\mathbf{k}} |\varepsilon_{i,\mathbf{k}}^{\text{DFT}} - \varepsilon_{i,\mathbf{k}}^{\text{TB}}|. \quad (7)$$

Some values of the energy windows cannot produce a tight-binding model, for example if the outer window contains less than M bands. As a result, finding appropriate energy windows is a constrained, four-dimensional optimization problem. The Nelder-Mead (downhill simplex) algorithm [23] can be used to solve this problem [24].

Figure 2 shows the result of such an optimization procedure for unstrained InSb, as described in Sec. IV. A clear improvement is visible between the tight-binding model obtained with the initial windows chosen by hand [panel (a)] and the optimized window values [panel (b)]. In particular, the conduction bands at the X and Z points are represented more accurately in the optimized model. Since the given bands for InSb are not entangled, it is also possible to skip the disentanglement step completely by using the `exclude_bands` parameter of Wannier90 to ignore all other energy bands. The resulting band structure is shown in Fig. 2(c). Nevertheless, we find that the band structure using optimized disentanglement is slightly better ($\Delta = 0.0327$) than the one without disentanglement ($\Delta = 0.0375$), especially for the four lowest conduction bands on the Z - Γ - X line. Hence, it can be useful to apply the disentanglement procedure and energy

window optimization even in cases where the bands are not inherently entangled, especially when the time required to run the tight-binding calculation is short compared to the initial first-principles calculation.

III. IMPLEMENTATION IN AiiDA WORKFLOWS

The AiiDA [8] platform is a Python framework for performing high-throughput calculations, focused on the field of materials physics. It enables reproducible research by keeping track of inputs, outputs, and settings for each calculation. On top of this provenance layer, it provides a tool set for automatically chaining calculations into user-defined workflows.

In this section, we describe the implementation of the Wannier tight-binding extraction scheme as an AiiDA workflow. This automation enables the application to the study of strain effects (described in Sec. IV). Special care has been taken to design the workflow in a modular way, which enables re-using parts of the workflow for purposes other than tight-binding extraction. We first discuss these design principles before showing how they are applied in the tight-binding workflows. The code for the AiiDA workflows is available in the open-source `aaida-tbextraction` package and provided as Supplementary Material.

A. Modular workflow design

The basic principle of modular workflow design is to split up a single monolithic workflow into minimal subworkflows or calculations that perform exactly one task. For example, the tight-binding model created by Wannier90 is post-processed by parsing it to an HDF5 format, followed by optionally changing the order of the basis and symmetrizing the model. While this could easily be implemented in a single script, splitting these three steps up into separate calculations allows separately reusing each of the steps.

More complex workflows are created by combining multiple subworkflows into a logical unit at a higher abstraction level. Inputs to the subworkflow are either forwarded directly from the input to the parent workflow or created within the parent workflow. Similarly, outputs from the subworkflow can either be forwarded to be an output of the parent workflow or consumed directly to guide the further execution of the parent workflow.

Since a complex workflow can consist of multiple layers of wrapped subworkflows, this modular approach is maintainable only if the overhead of forwarding input and output is minimal. Following the single responsibility principle, a parent workflow should not have to change if an input or output parameter of a subworkflow changes, unless it directly interacts with this parameter. To achieve this, a syntax is needed to specify that a parent workflow will *inherit* inputs or outputs of a subworkflow, without explicitly listing each parameter. In AiiDA, such a feature is available in the newly-introduced `expose` functionality, as described in Appendix C.

The modular architecture improves not only the reusability but also the flexibility of workflows. Often, a given part of a workflow could be performed in different ways. For example, many different codes can perform the first-principles calculations in the tight-binding extraction workflows. Additionally,

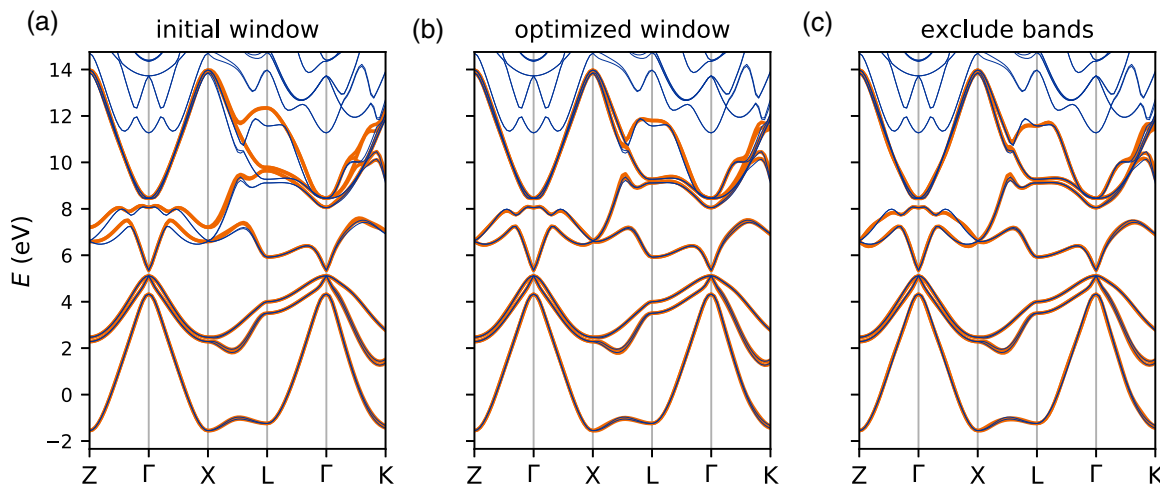


FIG. 2. Comparison between the reference first-principles band structure (blue) and band structures calculated from tight-binding models (orange) for InSb. The tight-binding model in (a) was calculated with the initial energy window, whereas (b) shows the model using the optimized energy window as detailed in Table II. The model in (c) was calculated without the disentanglement procedure, using the `exclude_bands` parameter.

one might want to add steps such as relaxation or cutoff energy convergence.

To allow for this, the parent workflow can allow for dynamically selecting a workflow for performing a given task by passing it as an input [25]. An abstract workflow class defines the interface that a workflow must fulfill so that it can be used to perform the task. If needed, the parent workflow can allow for dynamic inputs, which are just forwarded to the specific workflow implementing the interface. In this way, the parent workflow can act as a template that defines an abstract series of steps, without knowledge of the detailed input flags available on each step.

B. Tight-binding extraction workflow

Having discussed the design principles for modular workflows, we now show how these are applied to create a workflow for the construction of tight-binding models. This workflow is implemented in the `OptimizeFirstPrinciplesTightBinding` class as sketched in Fig. 3. At the uppermost level, the workflow has two parts: `FirstPrinciplesRunBase`, which executes the first-principles calculations, and `WindowSearch` which calculates the tight-binding model with energy window optimization.

Since different first-principles codes can produce the input files required by Wannier90, `FirstPrinciplesRunBase` defines only the minimum interface needed to perform this task. As described in the previous section, a workflow that implements this interface for a specific first-principles code can then be chosen dynamically. As a result, the subsequent parts of the workflow are independent of which first-principles code is used.

The `WindowSearch` workflow performs the Nelder-Mead algorithm for finding the optimal energy window. Because optimization schemes are useful outside of this specific application, we implemented the Nelder-Mead method in a general way. The `OptimizationWorkChain`, defined in the

`aiida-optimize` module, can be used to solve generic optimization problems in the context of AiiDA workflows. It requires two inputs: a workflow which defines the function to be optimized and an engine that implements the optimization method. Consequently, changing the whole workflow to use a different optimization method would be a simple matter of using a different engine.

Because AiiDA workflows need to be able to stop and restart after any given step, the engine is written in an object-oriented instead of a procedural way. While this complicates

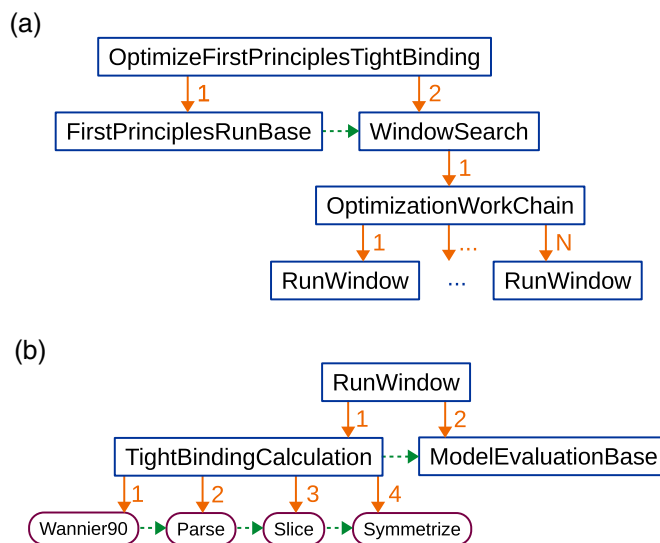


FIG. 3. Schematic of the AiiDA workflow for creating tight-binding models with energy window optimization. Workflows are shown in blue and calculations in purple. Orange arrows show calls from parent to child workflows (or calculations). Dashed green arrows show the implicit data dependency between workflows of the same level. In calculation names, the suffix `Calculation` is omitted for brevity.

implementing the Nelder-Mead method, it allows for serializing and storing the state of the engine.

The function which is optimized by the `OptimizationWorkChain` is implemented in the `RunWindow` workflow. It again consists of two parts: `TightBindingCalculation` creates the tight-binding model itself, and `ModelEvaluationBase` evaluates the quality of the model. The first step in the `TightBindingCalculation` workflow is to run Wannier90 on the given input parameters. In a second step, the Wannier90 output is parsed and converted into the TBmodels [26] HDF5 format. A third, optional, “slicing” step is used to either permute the basis orbitals or discard some orbitals. Finally, the (also optional) symmetrization procedure is performed. Both the `Slice` and the `Symmetrize` calculation have a TBmodels HDF5 file as both input and output, meaning that they could be chained arbitrarily with other such post-processing steps.

For the evaluation of the tight-binding model, we again use an abstract interface class, `ModelEvaluationBase`. While for the purposes of this paper we used the average difference of band energies [Eq. (7)] as a measure of model quality, other quantities might be more appropriate for different applications.

IV. STRAIN-DEPENDENT TIGHT-BINDING MODELS FOR MAJORANA DEVICES

The quest for Majorana zero modes (MZMs) in condensed matter systems has recently attracted a lot of interest [13–15,27–31]. The non-Abelian exchange statistics of Majorana Fermions makes these zero modes promising candidates for the realization of topological quantum computation devices [13,32]. Experimental investigations of possible MZMs focus on the proposal by Lutchyn *et al.* and Oreg *et al.* [14,15] in which MZMs appear on the boundaries of proximitized spin-orbit coupled quantum wires. Current experimental setups include semiconducting InAs nanowires with epitaxial superconducting Al [33] and InAs/GaSb heterostructures in which the quantum spin Hall effect [34,35] can be realized providing the possibility to proximity couple the helical edge state [28,30]. While there is a good deal of evidence suggesting that MZMs exist in the wire-based setups [36,37], a conclusive proof requires directly showing the braiding statistics of MZMs. An important step in realizing braiding with the systems based on the helical edge state is the search for optimized device and material properties. For optimizing the topological gap, a better theoretical understanding of the electronic structure in such devices is required. In this section, we show how the workflows can be used to generate tight-binding models which form the basis for accurate device simulations. While these device simulations themselves are outside the scope of this work, this shows the potential use of the method for a topic of active research in current condensed matter physics.

Highly accurate first-principles methods, using hybrid functionals [38], or the *GW* approximation [39], are computationally too demanding for the simulation of realistic device geometries and heterostructures. State of the art simulations of such structures use the $\mathbf{k}\cdot\mathbf{p}$ method [40] or empirical tight-binding (ETB) methods [41]. In both of these methods the

Hamiltonian is parametrized by a small number of parameters which are obtained empirically, for example via fitting to the first-principles band structure. For both of these methods the choice of parameters is ambiguous and one can obtain a good fit of the band structure while at the same time the electronic wave function might be wrongly represented. This might lead to unphysical solutions in confined geometries [42,43] and low transferability of the bulk models to the heterostructure in general. Recently, it was shown that better matching the ETB with the first-principles calculations can improve their transferability [43,44].

Realistic simulations of heterostructures require a correct treatment of strains at interfaces. In the $\mathbf{k}\cdot\mathbf{p}$ and the ETB method this is usually done by strain-dependent parameter sets. However, often the symmetries are not broken correctly. In this context, the Wannier or Wannier-like tight-binding models can offer a significant improvement by accurately representing the first-principles wave function and correctly capturing the effect of strain. As a demonstration of the AiiDA workflows, we construct SWTB models for the III-V semiconductors InSb, InAs and GaSb.

Including spin-orbit coupling (SOC), we require only 14 basis functions, namely s and p orbitals centered on the In/Ga atom, and p orbitals centered on the As/Sb atom. The popular $sp^3d^5s^*$ ETB models on the other hand require 40 [45] basis functions. The reason for this is that WTb models generally include longer-range neighbor interactions, whereas ETB is typically limited to nearest-neighbor (or next-nearest-neighbor in some cases [46]) interactions to keep the number of parameters manageable. As illustrated in Fig. 4, the produced tight-binding models include long-range hopping parameters, with amplitudes quickly decaying with distance.

To account for strain, we construct tight-binding models with biaxial (001), (110), and (111) strains, and the uniaxial [110] strain, as described in Appendix D. For each material and strain direction, we calculated 16 models in the range of $\pm 4\%$ strain. Including the unstrained models, we constructed a total of 195 tight-binding models, showing the applicability of the AiiDA workflow to a large number of chemically and structurally similar compounds.

A. Strained tight-binding workflow

To automatically extract tight-binding models for different strain directions and strengths, we define an additional workflow, `OptimizeStrainedFirstPrinciplesTightBinding`, as shown in Fig. 5. The first step in this workflow, `ApplyStrainsWithSymmetry`, creates the strained structures from the initial structure and strain parameters. Since strain can break crystal symmetries, the symmetries of the unstrained system are tested against the strained structure. With the strained structures and the remaining symmetries, we then use the `OptimizeFirstPrinciplesTightBinding` workflow to create a tight-binding model for each strain value.

B. First-principles calculations

In the first step of generating the SWTB we need to carry out a first-principles calculation of the bulk semiconductor structure. We performed all first-principles calculations using

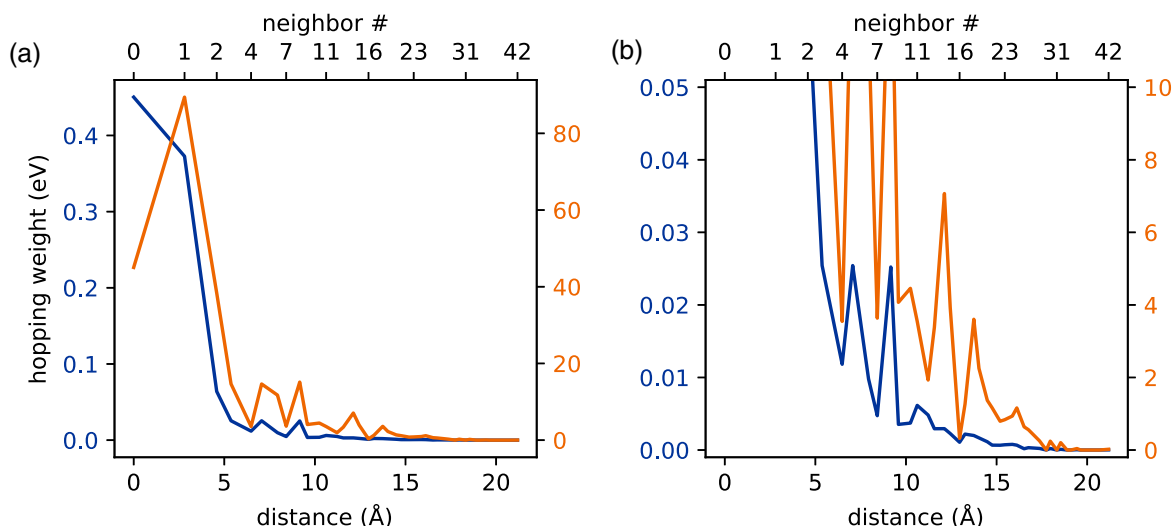


FIG. 4. Average (blue, left axis) and total (orange, right axis) weights of the hopping parameters for the unstrained InSb tight-binding model, as a function of distance.

the Vienna *ab initio* simulation package (VASP) utilizing projector augmented-wave (PAW) basis sets [47]. To obtain an accurate prediction of the band gap we employed hybrid functionals [48]. The HSE03/HSE06 hybrid functionals proved to be successful in computing band structures of III-V semiconductors [49]. These hybrid functionals are constructed by replacing a quarter of the density functional short-range exchange (which is the Perdew-Burke-Ernzerhof functional in our case [50]) with its Hartree-Fock counterpart. The screening parameter μ defines the separation into long- and short-range parts. In the popular HSE06 scheme, it is set to $\mu = 0.2 \text{ \AA}^{-1}$. We treated μ as an empirical parameter such that the calculated band gap is fitted to the experimental value. In this work, we used $\mu_{\text{InAs}} = 0.20 \text{ \AA}^{-1}$, $\mu_{\text{GaSb}} = 0.15 \text{ \AA}^{-1}$, and $\mu_{\text{InSb}} = 0.23 \text{ \AA}^{-1}$, following the prescriptions of Ref. [51]. Since the SOC of III-V semiconductors is significant, we accounted for it by using scalar-relativistic PAW potentials.

InAs, GaSb, and InSb crystallize in the zincblende structure with space group T_d^2 (no. 216). For the unstrained structures we perform the first-principles calculation with the experimental lattice constant a at 300 K, that is $a_{\text{InAs}} = 6.058 \text{ \AA}$, $a_{\text{GaSb}} = 6.096 \text{ \AA}$, $a_{\text{InSb}} = 6.479 \text{ \AA}$, from Ref. [52]. A plane-wave energy cutoff of 380 eV was used for all calculations. The Brillouin-zone integrations were sampled by a $6 \times 6 \times 6$ Γ -centered k -points mesh.

To get optimal results from the Wannier90 code in conjunction with VASP [47] we found that it is necessary to turn

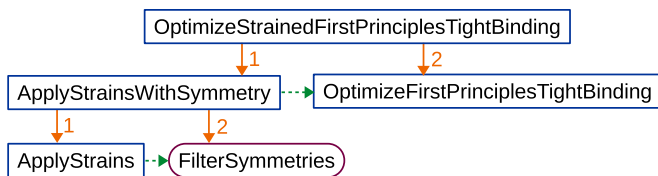


FIG. 5. Sketch of the workflow for constructing strained tight-binding models. The color scheme is the same as in Fig. 3.

symmetries off in VASP, that is setting the ISYM-tag to 0. Since the states are obtained by a numerical diagonalization routine, they obtain a random phase at each \mathbf{k} point. When symmetries are enabled however, the phases are the same for all vectors forming the star of \mathbf{k} . Since the convergence of Wannier90 is better if the numerical phases are random, turning symmetries off generally results in more localized Wannier functions after the projection step.

The interface for running first-principles calculations in the tight-binding extraction workflow is defined in the `FirstPrinciplesRunBase` class (see Sec. III B). Here, we describe the specific subclass used to implement these calculations with VASP [47], `VaspFirstPrinciplesRun` (see Fig. 6). In a first step, this workflow performs a self-consistent calculation. The resulting wave function is then passed to calculations for the reference band structure and the input files for Wannier90. Two workflows `VaspReferenceBands` and `VaspWannierInput` are used to perform these calculations. The workflows are thin wrappers around the corresponding calculations from the `aiida-vasp` plugin [53], providing additional input and output validation. For the band-structure calculation, the workflow also adds the \mathbf{k} -point grid needed for hybrid functional calculations.

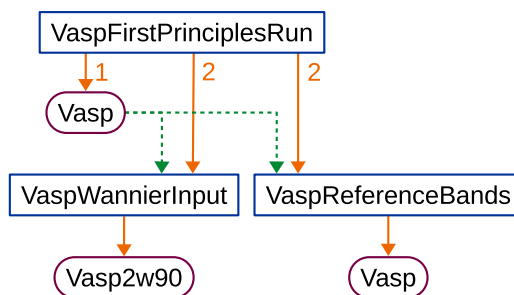


FIG. 6. Sketch of the `FirstPrinciplesRunBase` subclass used for calculating the Wannier90 input and reference bands with VASP and hybrid functionals.

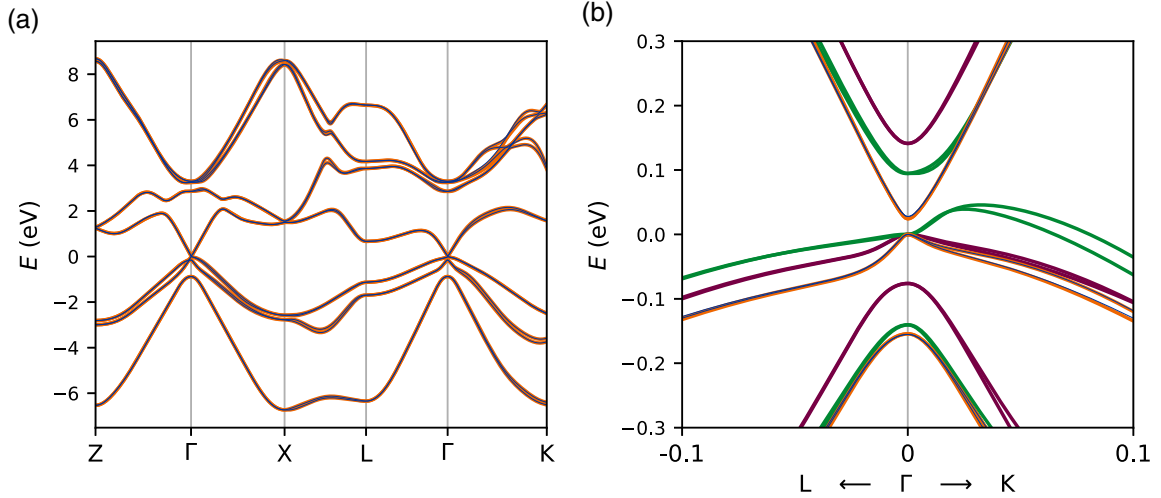


FIG. 7. Comparison between the InSb band structure obtained directly from the tight-binding model with 2% biaxial (001) strain (blue) and from the linear interpolation (orange) between models with 1% and 3% strain. The energy scale is fixed by setting the top of the valence bands at Γ to zero. (a) At the electron-volt scale, the only visible difference is in the upper bands along the Γ -K line. (b) Closeup of the bands around Γ . The bands for 1% (purple) and 3% strain (green) are also shown.

C. Strain interpolation

Using the AiiDA workflow, we obtained tight-binding models for strains in the range of $\pm 4\%$, in steps of 0.5%. However, it is sometimes useful to have a finer control over the strain value without having to run additional first-principles calculations. A common way of obtaining this is by linear interpolation of the hopping parameters. Given two strain values s_1 and s_2 , for which the hopping parameter $H^{s_i}[\mathbf{R}]$ are known, the hopping parameters for an unknown s^* can be calculated as

$$H^{s^*}[\mathbf{R}] = \alpha H^{s_1}[\mathbf{R}] + (1 - \alpha) H^{s_2}[\mathbf{R}], \quad (8)$$

where

$$\alpha = \frac{s^* - s_2}{s_1 - s_2}. \quad (9)$$

Since this method assumes that the hopping parameters are a linear function of strain value, it becomes unreliable when s^* is too far away from s_1 and s_2 . For this reason, we compared

TABLE I. Effective masses of light hole (LH), heavy hole (HH), split-off hole, and electron at the Γ point along the [100] direction in the unstrained case. Values for symmetrized Wannier-like tight-binding models (SWTB) are compared to first-principles (HSE_{bgfit}) [51] and experimental results [51,54].

Material	Method	$ m_{\text{SO}}^* $	$ m_{\text{LH}}^* $	$ m_{\text{HH}}^* $	$ m_{\text{c}}^* $	g factor
InSb	HSE _{bgfit}	0.129	0.018	0.245	0.017	
	SWTB	0.118	0.016	0.219	0.015	-49.8
	Expt.	0.110	0.015	0.263	0.014	-50.6
InAs	HSE _{bgfit}	0.112	0.033	0.343	0.027	
	SWTB	0.118	0.036	0.340	0.029	-15.3
	Expt.	0.140	0.027	0.333	0.026	-15
GaSb	HSE _{bgfit}	0.143	0.047	0.235	0.042	
	SWTB	0.124	0.039	0.20	0.036	-15.1
	Expt.	0.120	0.044	0.250	0.039	-7.8

a tight-binding model for InSb with 2% biaxial (001) strain obtained from linear interpolation of 1% and 3% strain models with one calculated directly from first principles. Figure 7 shows a comparison of the two band structures, which we find to be almost identical.

Important to note is that while linear interpolation works well for strains of the same kind, this is not necessarily the case when combining two models with different strain directions. The reason for this is that the symmetries of a particular structure depend on the direction of the applied strain but (unless it is zero) not on its strength. As a result, a tight-binding model resulting from linear interpolation between two models of a different strain direction would not have the correct symmetries.

D. Results

To validate the tight-binding models obtained using the aiiida-tbextraction workflows, several material parameters were calculated. Table I shows effective masses and g factors for the unstrained models, in comparison to first-principles [51] and experimental [51,54] values. Effective masses for the tight-binding models were calculated using

TABLE II. Initial and optimized energy windows used for calculating unstrained tight-binding models and the corresponding band-structure mismatch as defined in Eq. (7).

Material	Energy Windows (eV)	Δ	
InSb	initial	(-4.5, [-4, 6.5], 16)	0.107
	optimized	(-4.44, [-3.24, 8.67], 14.01)	0.033
InAs	initial	(-4.5, [-4, 6.5], 16)	0.113
	optimized	(-4.44, [-3.59, 7.34], 15.04)	0.046
GaSb	initial	(-4.5, [-4.5, 7], 16)	0.082
	optimized	(-5.35, [-3.34, 7.90], 14.27)	0.043

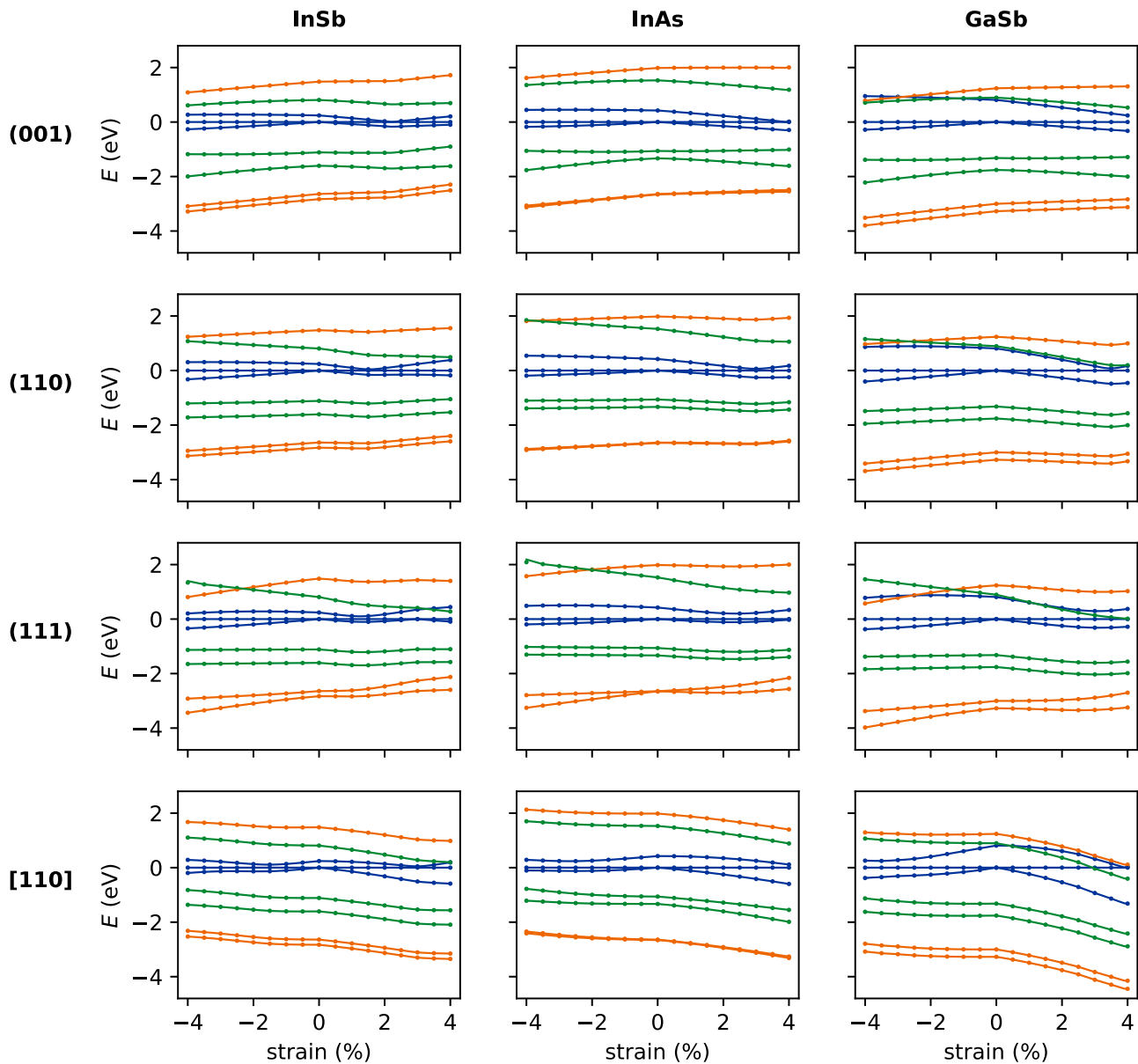


FIG. 8. Strain dependence of band energies. The two highest valence bands and the lowest conduction band are shown at the Γ (blue), X (orange), and L (green) points, where each band is doubly degenerate. Energy values are shifted such that the valence band maximum at Γ is zero. The line represents values calculated from the tight-binding models with linear interpolation [Eq. (8)] in steps of 0.1%. For comparison, the points show values calculated from first principles. We find a good agreement between the tight-binding and first-principles values, except for the conduction band value at the L point at -4% biaxial (111) strain.

a second-order polynomial fit with range 0.001 \AA^{-1} . The g -factor calculations were performed using both perturbation theory and a Landau level calculation [55], with good agreement ($<0.5\%$ difference) between the two methods.

The effect of the energy window optimization is shown in Table II, which lists the initial and optimized windows, as well as the corresponding band-structure mismatch. As previously shown in Fig. 2, it can be seen that the mismatch is substantially reduced after optimization.

Finally, the effect of strain on the energy levels at high-symmetry points is shown in Fig. 8. The numerical data is listed in the Supplemental Material [56].

In the Supplemental Material of this paper, an export of the AiiDA database is given. This database contains the full

provenance of each calculation performed to create the tight-binding models. For ease of accessibility, a separate data set containing only the 195 strained tight-binding models is also given.

V. CONCLUSION AND OUTLOOK

We have implemented a workflow for an *automatic* construction of Wannier tight-binding models from first-principles calculations. Building on the known procedure for calculating these models, we introduced a post-processing step to symmetrize the models and an optimization of the energy windows used for disentanglement. These workflows are implemented in the `aidda-tbextraction` package, which

is a free and open-source plugin for the AiiDA framework. As a test case, tight-binding models for strained III-V semiconductor materials were calculated. These results should enable device simulations for Majorana designs and other quantum devices.

The workflows have been implemented in a modular and extensible way. As a result, they can be used as building blocks for further improvements in automating the process of generating Wannier tight-binding models. Possible directions include extending the number of first-principles codes which are compatible with the plugin, adding different fitness criteria for the energy window optimization, and further minimizing the number of tunable parameters. For example, the need for choosing initial trial orbitals could be eliminated either by using another optimization step or by utilizing the method of Ref. [12].

ACKNOWLEDGMENTS

We would like to thank S. Huber, M. Uhrin, G. Pizzi, A. Marrazzo, M. Könz, and D. Rodic for helpful discussions. The authors were supported by Microsoft Research and the Swiss National Science Foundation through the National Competence Centers in Research MARVEL and QSIT. A.A.S. also acknowledges the support of the Swiss National Science Foundation Professorship. Calculations were performed on the Mönch cluster of ETH Zurich.

APPENDIX A: PROPERTIES OF THE SYMMETRIZED HAMILTONIAN

The symmetrized Hamiltonian is defined as [Eq. (4)]

$$\tilde{\mathcal{H}}(\mathbf{k}) = \frac{1}{|G|} \sum_{g \in G} D^{\mathbf{k}}(g) \mathcal{H}(g^{-1}\mathbf{k}) D^{\mathbf{k}}(g^{-1}). \quad (\text{A1})$$

We first show that this Hamiltonian respects the symmetries in G . Let $g' \in G$:

$$\begin{aligned} & D^{\mathbf{k}}(g') \tilde{\mathcal{H}}([g']^{-1}\mathbf{k}) D^{\mathbf{k}}([g']^{-1}) \\ &= \frac{1}{|G|} \sum_{g \in G} D^{\mathbf{k}}(g') D^{\mathbf{k}}(g) \mathcal{H}(g^{-1}[g']^{-1}\mathbf{k}) D^{\mathbf{k}}(g^{-1}) D^{\mathbf{k}}([g']^{-1}) \\ &= \frac{1}{|G|} \sum_{g \in G} D^{\mathbf{k}}(g'g) \mathcal{H}([g'g]^{-1}\mathbf{k}) D^{\mathbf{k}}([g'g]^{-1}) \\ &\stackrel{g''=g'g}{=} \frac{1}{|G|} \sum_{g'' \in G} D^{\mathbf{k}}(g'') \mathcal{H}([g'']^{-1}\mathbf{k}) D^{\mathbf{k}}([g'']^{-1}) = \tilde{\mathcal{H}}(\mathbf{k}). \end{aligned} \quad (\text{A2})$$

Also, it is easily shown that symmetrizing a Hamiltonian $\mathcal{H}_{\mathbf{k}}^{\text{symm.}}$ which already respects the symmetry does not change it:

$$\begin{aligned} \tilde{\mathcal{H}}^{\text{symm.}}(\mathbf{k}) &= \frac{1}{|G|} \sum_{g \in G} D^{\mathbf{k}}(g) \mathcal{H}^{\text{symm.}}(g^{-1}\mathbf{k}) D^{\mathbf{k}}(g^{-1}) \\ &\stackrel{\text{Eq. (2)}}{=} \frac{1}{|G|} \sum_{g \in G} \mathcal{H}^{\text{symm.}}(\mathbf{k}) = \mathcal{H}^{\text{symm.}}(\mathbf{k}). \end{aligned} \quad (\text{A3})$$

APPENDIX B: SYMMETRIZED HAMILTONIAN IN TERMS OF REAL-SPACE MATRICES

In this Appendix, we show how the symmetrized Hamiltonian $\tilde{\mathcal{H}}(\mathbf{k})$ can be expressed in terms of real-space hopping matrices $H[\mathbf{R}]$. In the following, we will assume that the representation matrix $D^{\mathbf{k}}(g)$ of unitary operations is given in the form

$$D^{\mathbf{k}}(g) = e^{i\alpha_g \cdot \mathbf{k}} D(g) = e^{i\alpha_g \cdot \mathbf{k}} U_g, \quad (\text{B1})$$

where α_g is the translation vector and U_g is a unitary matrix. For antiunitary operations, we assume they are of the form

$$D^{\mathbf{k}}(g) = e^{i\alpha_g \cdot \mathbf{k}} D(g) = e^{i\alpha_g \cdot \mathbf{k}} U_g \hat{K}, \quad (\text{B2})$$

where \hat{K} represents complex conjugation.

First, we notice that $D_{ij}^{\mathbf{k}}(g) \neq 0$ only if $g\mathbf{t}_i - \mathbf{t}_j \in \mathbb{Z}^d$, meaning that orbitals centered at \mathbf{t}_i are mapped onto \mathbf{t}_j up to a possible lattice translation. Using Eqs. (1) and (4), we can write the symmetrized Hamiltonian as

$$\tilde{\mathcal{H}}^{ij}(\mathbf{k}) = \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}^{\mathbf{k}}(g) H^{lm}[\mathbf{R}] e^{i(g^{-1}\mathbf{k}) \cdot (\mathbf{R} + \mathbf{t}_m - \mathbf{t}_i)} D_{mj}^{\mathbf{k}}(g^{-1}), \quad (\text{B3})$$

where the indices l, m only go over nonzero $D_{il}^{\mathbf{k}}(g)$ and $D_{mj}^{\mathbf{k}}(g^{-1})$. Writing the real-space operator for g in Seitz notation [58]

$$g_{\mathbf{r}} = \{S_g | \alpha_g\}, \quad g_{\mathbf{r}}^{-1} = \{S_g^{-1} | -S_g^{-1}\alpha_g\}, \quad (\text{B4})$$

where S_g is the rotational part, and α_g is the translation vector of the symmetry, this means that

$$g\mathbf{t}_i - \mathbf{t}_i = S_g\mathbf{t}_i + \alpha_g - \mathbf{t}_i \in \mathbb{Z}^d \quad (\text{B5})$$

$$\begin{aligned} g^{-1}\mathbf{t}_j - \mathbf{t}_m &= S_g^{-1}\mathbf{t}_j - S_g^{-1}\alpha_g - \mathbf{t}_m \in \mathbb{Z}^d \\ &\Rightarrow \mathbf{t}_j - \alpha_g - S_g\mathbf{t}_m \in \mathbb{Z}^d \end{aligned} \quad (\text{B6})$$

$$\stackrel{-(\text{B5})-(\text{B6})}{\Rightarrow} \mathbf{T}_{ij}^{ml} = S_g(\mathbf{t}_m - \mathbf{t}_i) - (\mathbf{t}_j - \mathbf{t}_i) \in \mathbb{Z}^d. \quad (\text{B7})$$

Next, we must consider how g acts on the reciprocal-space vector \mathbf{k} . For an (anti)unitary operator, we know that

$$\langle \mathbf{k}, \mathbf{r} \rangle = \pm \langle g\mathbf{k}, g\mathbf{r} \rangle, \quad (\text{B8})$$

where the positive (negative) sign corresponds to the unitary (antiunitary) case. Since g acts on \mathbf{r} with S_g , it follows that

$$\begin{aligned} \langle \mathbf{k}, \mathbf{r} \rangle &= \pm \langle A\mathbf{k}, S_g\mathbf{r} \rangle \\ \mathbf{k}^T \mathbf{r} &= \pm \mathbf{k}^T A^T S_g \mathbf{r}, \end{aligned} \quad (\text{B9})$$

where A is the operation which acts upon \mathbf{k} when g is applied. Since this is true for all \mathbf{k} and \mathbf{r} ,

$$A = \pm (S_g^T)^{-1}, \quad (\text{B10})$$

and thus

$$g^{-1}\mathbf{k} = \pm S_g^T \mathbf{k}. \quad (\text{B11})$$

For the next step, we treat the unitary and antiunitary cases separately for clarity.

(1) *Unitary case*

By applying Eq. (B11) to Eq. (B3), we get

$$\begin{aligned} \tilde{\mathcal{H}}^{ij}(\mathbf{k}) &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}^{\mathbf{k}}(g) H^{lm}[\mathbf{R}] e^{i(S_g^T \mathbf{k}) \cdot (\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)} D_{mj}^{\mathbf{k}}(g^{-1}) \\ &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}^{\mathbf{k}}(g) H^{lm}[\mathbf{R}] e^{i\mathbf{k} \cdot [S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]} D_{mj}^{\mathbf{k}}(g^{-1}). \end{aligned} \quad (\text{B12})$$

Applying Eq. (B1), we obtain

$$\begin{aligned} \tilde{\mathcal{H}}^{ij}(\mathbf{k}) &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} e^{i\alpha_g \cdot \mathbf{k}} (U_g)_{il} H^{lm}[\mathbf{R}] \\ &\quad \times e^{i\mathbf{k} \cdot [S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]} (U_g^\dagger)_{mj} e^{-i\alpha_g \cdot \mathbf{k}} \\ &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} (U_g)_{il} H^{lm}[\mathbf{R}] (U_g^\dagger)_{mj} e^{i\mathbf{k} \cdot [S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]} \\ &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}(g) H^{lm}[\mathbf{R}] D_{mj}(g^{-1}) e^{i\mathbf{k} \cdot [S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]}. \end{aligned} \quad (\text{B13})$$

(2) *Antiunitary case*

In the antiunitary case, we get

$$\begin{aligned} \tilde{\mathcal{H}}^{ij}(\mathbf{k}) &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}^{\mathbf{k}}(g) H^{lm}[\mathbf{R}] e^{i(-S_g^T \mathbf{k}) \cdot (\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)} D_{mj}^{\mathbf{k}}(g^{-1}) \\ &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}^{\mathbf{k}}(g) H^{lm}[\mathbf{R}] e^{-i\mathbf{k} \cdot [S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]} D_{mj}^{\mathbf{k}}(g^{-1}). \end{aligned} \quad (\text{B14})$$

When applying Eq. (B2), it is important to note that the representation of the inverse is given by

$$\begin{aligned} D^{\mathbf{k}}(g^{-1}) &= (D^{\mathbf{k}}(g))^{-1} = (e^{i\alpha_g \cdot \mathbf{k}} U_g \hat{K})^{-1} = \hat{K} U_g^\dagger e^{-i\alpha_g \cdot \mathbf{k}} \\ &= e^{i\alpha_g \cdot \mathbf{k}} \hat{K} U_g^\dagger = e^{i\alpha_g \cdot \mathbf{k}} D(g^{-1}). \end{aligned} \quad (\text{B15})$$

Applying Eqs. (B2) and (B15) to Eq. (B14), we get

$$\begin{aligned} \tilde{\mathcal{H}}^{ij}(\mathbf{k}) &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} e^{i\alpha_g \cdot \mathbf{k}} (U_g)_{il} \hat{K} H^{lm}[\mathbf{R}] \\ &\quad \times e^{i\mathbf{k} \cdot [-S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]} e^{i\alpha_g \cdot \mathbf{k}} \hat{K} (U_g^\dagger)_{mj} \\ &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} (U_g)_{il} \hat{K} H^{lm}[\mathbf{R}] \hat{K} (U_g^\dagger)_{mj} e^{i\mathbf{k} \cdot [S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]} \\ &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}(g) H^{lm}[\mathbf{R}] D_{mj}(g^{-1}) e^{i\mathbf{k} \cdot [S_g(\mathbf{R} + \mathbf{t}_m - \mathbf{t}_l)]}. \end{aligned} \quad (\text{B16})$$

We observe that the result is the same for the unitary and antiunitary cases and treat them together in the following. Note that the \mathbf{k} -dependent part of the representation cancels with its inverse in both cases [59].

Next, we substitute $\mathbf{t}_m - \mathbf{t}_l$ using \mathbf{T}_{ij}^{ml} defined above and define $\mathbf{R}' = S_g \mathbf{R} + \mathbf{T}_{ij}^{ml}$. Since \mathbf{R}' is again a lattice vector, we can change the summation from \mathbf{R} to \mathbf{R}' :

$$\tilde{\mathcal{H}}^{ij}(\mathbf{k}) = \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}}} D_{il}(g) H^{lm}[\mathbf{R}] D_{mj}(g^{-1}) e^{i\mathbf{k} \cdot [S_g \mathbf{R} + \mathbf{T}_{ij}^{ml} + \mathbf{t}_j - \mathbf{t}_l]} \quad (\text{B17})$$

$$\begin{aligned} &= \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m \\ \mathbf{R}'}} D_{il}(g) H^{lm}[S_g^{-1}(\mathbf{R}' - \mathbf{T}_{ij}^{ml})] D_{mj}(g^{-1}) \\ &\quad \times e^{i\mathbf{k} \cdot (\mathbf{R}' + \mathbf{t}_j - \mathbf{t}_l)}. \end{aligned} \quad (\text{B18})$$

Finally, we again use Eq. (1) to obtain the symmetrized real-space hopping matrices

$$\tilde{H}^{ij}[\mathbf{R}'] = \frac{1}{|G|} \sum_{\substack{g \in G \\ l, m}} D_{il}(g) H^{lm}[S_g^{-1}(\mathbf{R}' - \mathbf{T}_{ij}^{ml})] D_{mj}(g^{-1}). \quad (\text{B19})$$

APPENDIX C: AiiDA *expose* FUNCTIONALITY

In this Appendix, we illustrate how the AiiDA *expose* functionality simplifies writing modular workflows. It allows implicitly forwarding input and output values of a subworkflow instead of having to explicitly specify each value. Listing 1 shows a simple workflow with two inputs a and b and one output c. A parent workflow that only wraps this workflow is shown in listings 2 and 3 with and without using the *expose* functionality, respectively. Import statements are omitted in all listings for brevity.

Besides reducing the boilerplate code in the parent workflow, this enables adhering to the single responsibility principle: The parent workflow does not need to

change if the inputs or outputs of the wrapped workflow change, unless it directly impacts the parent workflow logic.

Listing 1. A simple workflow with inputs **a** and **b**, and output **c**. The steps executing the workflow are omitted.

```
class SubWF(WorkChain):
    @classmethod
    def define(cls, spec):
        spec.input('a', valid_type=Int)
        spec.input('b', valid_type=Int)
        spec.output('c', valid_type=Int)
    ...
```

Listing 2. A workflow that wraps SubWF by using the *expose* functionality.

```
class ParentWF(WorkChain):
    @classmethod
    def define(cls, spec):
        spec.expose_inputs(SubWF)
        spec.expose_outputs(SubWF)

        spec.outline(
            cls.invoke_subwf,
            cls.write_outputs
        )

    def invoke_subwf(self):
        return ToContext(
            sub_wf=self.submit(SubWF, **self.exposed_inputs(SubWF))
        )

    def write_outputs(self):
        self.out_many(self.exposed_outputs(self.ctx.sub_wf))
```

Listing 3. A workflow that wraps SubWF without using the *expose* functionality.

```
class ParentWF(WorkChain):
    @classmethod
    def define(cls, spec):
        spec.input('a', valid_type=Int)
        spec.input('b', valid_type=Int)
        spec.output('c', valid_type=Int)

        spec.outline(
            cls.invoke_subwf,
            cls.write_outputs
        )

    def invoke_subwf(self):
        return ToContext(
            sub_wf=self.submit(SubWF, a=self.inputs.a, b=self.inputs.b)
        )

    def write_outputs(self):
        self.out('c', self.ctx.sub_wf.out.c)
```

APPENDIX D: STRAIN TENSOR AND STRAINED ATOM POSITION

In this section, we list all the strain tensors that we used in the above simulations. Under a small homogeneous and elastic strain, the lattice vectors \mathbf{R} transform (in Cartesian coordinates) into [60,61]

$$\mathbf{R}' = (1 + \boldsymbol{\epsilon})\mathbf{R}, \quad (\text{D1})$$

where

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xz} \\ \epsilon_{yx} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{zx} & \epsilon_{zy} & \epsilon_{zz} \end{pmatrix} \quad (\text{D2})$$

is the strain tensor. Due to the stress-strain relation, the strain tensor under different kinds of strain can be obtained as listed below [61–63]:

(1) (001) plane biaxial strain

$$\boldsymbol{\epsilon}_{001}^{bi} = \begin{pmatrix} \epsilon_{xx} & 0 & 0 \\ 0 & \epsilon_{yy} & 0 \\ 0 & 0 & \epsilon_{zz} \end{pmatrix}, \quad (\text{D3})$$

where $\epsilon_{xx} = \epsilon_{yy} = \epsilon'$, $\epsilon_{zz} = -2\frac{C_{12}}{C_{11}}\epsilon'$.

(2) (110) plane biaxial strain

$$\boldsymbol{\epsilon}_{110}^{bi} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & 0 \\ \epsilon_{xy} & \epsilon_{xx} & 0 \\ 0 & 0 & \epsilon_{zz} \end{pmatrix}, \quad (\text{D4})$$

where

$$\begin{aligned} \epsilon_{zz} &= \epsilon' \\ \epsilon_{xx} &= \frac{2C_{44} - C_{12}}{2C_{44} + C_{11} + C_{12}}\epsilon' \\ \epsilon_{xy} &= \frac{-C_{11} - 2C_{12}}{2C_{44} + C_{11} + C_{12}}\epsilon'. \end{aligned} \quad (\text{D5})$$

(3) (111) plane biaxial strain

$$\boldsymbol{\epsilon}_{111}^{bi} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{xy} \\ \epsilon_{xy} & \epsilon_{xx} & \epsilon_{xy} \\ \epsilon_{xy} & \epsilon_{xy} & \epsilon_{xx} \end{pmatrix}, \quad (\text{D6})$$

where

$$\begin{aligned} \epsilon_{xx} &= \frac{4C_{44}}{4C_{44} + C_{11} + 2C_{12}}\epsilon' \\ \epsilon_{xy} &= \frac{-C_{11} - 2C_{12}}{4C_{44} + C_{11} + 2C_{12}}\epsilon'. \end{aligned} \quad (\text{D7})$$

TABLE III. Strain parameters used in this work.

Quantity	Symbol	Unit	InAs	GaSb	InSb
Elastic constant ^a	C_{11}	$10^{11}\text{dyn cm}^{-2}$	8.329	8.834	6.918
	C_{12}	$10^{11}\text{dyn cm}^{-2}$	4.526	4.023	3.788
	C_{44}	$10^{11}\text{dyn cm}^{-2}$	3.959	4.322	3.132
Internal dis. ^b	ζ		0.58	0.99	0.9

^aFrom Ref. [64].

^bFrom Ref. [65].

^cFrom Ref. [66].

(4) [110] uniaxial strain

$$\boldsymbol{\epsilon}_{110}^{\text{uni}} = \begin{pmatrix} \epsilon_{xx} & \epsilon_{xy} & 0 \\ \epsilon_{xy} & \epsilon_{xx} & 0 \\ 0 & 0 & \epsilon_{zz} \end{pmatrix}, \quad (\text{D8})$$

where

$$\begin{aligned} \epsilon_{zz} &= \epsilon' \\ \epsilon_{xx} &= -\frac{C_{11}}{2C_{12}}\epsilon' \\ \epsilon_{xy} &= -\frac{(C_{11} - C_{12})(C_{11} + 2C_{12})}{4C_{44}C_{12}}\epsilon'. \end{aligned} \quad (\text{D9})$$

In the distorted system, the position of the atoms also changes with the strain tensor. In the unstrained InAs, GaSb, and InSb system, the cation is located in the (0, 0, 0) site and the anion is located at $\hat{\tau} = (1/4, 1/4, 1/4)$ in primitive lattice vectors. In Cartesian coordinates, τ changes by the following [61–63]:

(1) (001) biaxial strain

$$\hat{\tau}' = (1 + \boldsymbol{\epsilon})\hat{\tau} \quad (\text{D10})$$

(2) (111) biaxial strain

$$\hat{\tau}' = (1 + \boldsymbol{\epsilon})\hat{\tau} - \frac{a_0}{2}\epsilon_{xy}\zeta \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = (1 + \boldsymbol{\epsilon} - 2\epsilon_{xy}\zeta)\hat{\tau}, \quad (\text{D11})$$

where a_0 is the lattice constant without strain.

(3) (110) biaxial strain and [110] uniaxial strain

$$\begin{aligned} \hat{\tau}' &= (1 + \boldsymbol{\epsilon})\hat{\tau} - \frac{a_0}{2}\epsilon_{xy}\zeta \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \left(1 + \boldsymbol{\epsilon} - 2\epsilon_{xy}\zeta \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) \hat{\tau}. \end{aligned} \quad (\text{D12})$$

The internal displacement ζ and the stiffness constants C_{11} , C_{12} , C_{44} of InAs, GaSb, and InSb we used in the paper are listed in Table III.

[1] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).

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

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

- [4] A. Franceschetti and A. Zunger, *Nature (London)* **402**, 60 (1999).
- [5] G. H. Jóhannesson, T. Bligaard, A. V. Ruban, H. L. Skriver, K. W. Jacobsen, and J. K. Nørskov, *Phys. Rev. Lett.* **88**, 255506 (2002).
- [6] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, *Phys. Rev. Lett.* **91**, 135503 (2003).
- [7] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, and O. Levy, *Nat. Mater.* **12**, 191 (2013).
- [8] G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari, and B. Kozinsky, *Comput. Mater. Sci.* **111**, 218 (2016).
- [9] A. Jain, S. P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, and K. A. Persson, *Concurr. Comput. Pract. Exp.* **27**, 5037 (2015).
- [10] A. A. Mostofi, J. R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, *Comput. Phys. Commun.* **178**, 685 (2008).
- [11] A. A. Mostofi, J. R. Yates, G. Pizzi, Y.-S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, *Comput. Phys. Commun.* **185**, 2309 (2014).
- [12] J. I. Mustafa, S. Coh, M. L. Cohen, and S. G. Louie, *Phys. Rev. B* **92**, 165134 (2015).
- [13] A. Y. Kitaev, *Phys. Usp.* **44**, 131 (2001).
- [14] R. M. Lutchyn, J. D. Sau, and S. Das Sarma, *Phys. Rev. Lett.* **105**, 077001 (2010).
- [15] Y. Oreg, G. Refael, and F. von Oppen, *Phys. Rev. Lett.* **105**, 177002 (2010).
- [16] In this work, we use the tight-binding convention I of Ref. [57].
- [17] M. S. Dresselhaus, G. Dresselhaus, and A. Jorio, *Group Theory: Application to the Physics of Condensed Matter* (Springer Science & Business Media, Berlin, Heidelberg, 2007).
- [18] H. E. Haber, Three-dimensional proper and improper rotation matrices, http://scipp.ucsc.edu/~haber/ph116A/rotation_11.pdf (2011).
- [19] R. Sakuma, *Phys. Rev. B* **87**, 235109 (2013).
- [20] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, Davide Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougousis, Anton Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, Stefano Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, *J. Phys.: Condens. Matter* **21**, 395502 (2009).
- [21] P. Giannozzi, O. Andreussi, T. Brumme, O. Bunau, M. B. Nardelli, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, M. Cococcioni, N. Colonna, I. Carnimeo, A. D. Corso, S. de Gironcoli, P. Delugas, R. A. DiStasio Jr., A. Ferretti, A. Floris, G. Fratesi, G. Fugallo, R. Gebauer, U. Gerstmann, F. Giustino, T. Gorni, J. Jia, M. Kawamura, H.-Y. Ko, A. Kokalj, E. Küçükbenli, M. Lazzeri, M. Marsili, N. Marzari, F. Mauri, N. L. Nguyen, H.-V. Nguyen, A. Otero-de-la-Roza, L. Paulatto, S. Poncé, D. Rocca, R. Sabatini, B. Santra, M. Schlipf, A. P. Seitsonen, A. Smogunov, I. Timrov, T. Thonhauser, P. Umari, N. Vast, X. Wu, and S. Baroni, *J. Phys.: Condens. Matter* **29**, 465901 (2017).
- [22] Because the first-principles calculation usually contains more than M bands, we need to choose which bands should be represented by the tight-binding model.
- [23] J. A. Nelder and R. Mead, *Comput. J.* **7**, 308 (1965).
- [24] The constraint is implemented by assigning an infinite value of Δ to invalid energy windows.
- [25] For storing the workflow in the AiiDA database, it needs to be converted into an AiiDA data type. We chose to convert it into a string containing the fully qualified class name, from which we import the workflow when needed.
- [26] D. Gresch, TBmodels documentation, <http://z2pack.ethz.ch/tbmodels>
- [27] J. Alicea, *Phys. Rev. B* **81**, 125318 (2010).
- [28] D. I. Pikulin, J. P. Dahlhaus, M. Wimmer, H. Schomerus, and C. W. J. Beenakker, *New J. Phys.* **14**, 125011 (2012).
- [29] J. Alicea, *Rep. Prog. Phys.* **75**, 076501 (2012).
- [30] S. Mi, D. I. Pikulin, M. Wimmer, and C. W. J. Beenakker, *Phys. Rev. B* **87**, 241405 (2013).
- [31] C. Beenakker, *Annu. Rev. Condens. Matter Phys.* **4**, 113 (2013).
- [32] A. Y. Kitaev, *Ann. Phys.* **303**, 2 (2003).
- [33] M. T. Deng, S. Vaitiekėnas, E. B. Hansen, J. Danon, M. Leijnse, K. Flensberg, J. Nygård, P. Krogstrup, and C. M. Marcus, *Science* **354**, 1557 (2016).
- [34] C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
- [35] C. Liu, T. L. Hughes, X.-L. Qi, K. Wang, and S.-C. Zhang, *Phys. Rev. Lett.* **100**, 236601 (2008).
- [36] V. Mourik, K. Zuo, S. M. Frolov, S. R. Plissard, E. P. A. M. Bakkers, and L. P. Kouwenhoven, *Science* **336**, 1003 (2012).
- [37] H. O. H. Churchill, V. Fatemi, K. Grove-Rasmussen, M. T. Deng, P. Caroff, H. Q. Xu, and C. M. Marcus, *Phys. Rev. B* **87**, 241401 (2013).
- [38] A. D. Becke, *J. Chem. Phys.* **98**, 5648 (1993).
- [39] L. Hedin, *Phys. Rev.* **139**, A796 (1965).
- [40] E. O. Kane, in *Semiconductors and Semimetals*, Semiconductors and Semimetals Vol. 1, edited by R. K. Willardson and A. C. Beer (Elsevier, New York, 1966), pp. 75–100.
- [41] J. C. Slater and G. F. Koster, *Phys. Rev.* **94**, 1498 (1954).
- [42] Y. Tan, M. Povolotskyi, T. Kubis, Y. He, Z. Jiang, G. Klimeck, and T. B. Boykin, *J. Comput. Electron.* **12**, 56 (2013).
- [43] Y. P. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, *Phys. Rev. B* **92**, 085301 (2015).
- [44] Y. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, *Phys. Rev. B* **94**, 045311 (2016).
- [45] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, *Phys. Rev. B* **57**, 6493 (1998).
- [46] T. B. Boykin, *Phys. Rev. B* **56**, 9613 (1997).
- [47] G. Kresse and J. Furthmüller, *Phys. Rev. B* **54**, 11169 (1996).
- [48] J. Heyd, G. E. Scuseria, and M. Ernzerhof, *J. Chem. Phys.* **118**, 8207 (2003).
- [49] J. Heyd and G. E. Scuseria, *J. Chem. Phys.* **121**, 1187 (2004).
- [50] J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [51] Y.-S. Kim, M. Marsman, G. Kresse, F. Tran, and P. Blaha, *Phys. Rev. B* **82**, 205212 (2010).
- [52] *Group IV Elements, IV-IV and III-V Compounds. Part a - Lattice Properties*, Landolt-Börnstein - Group III Condensed Matter Vol. 41A1 α , edited by O. Madelung, U. Rössler, and M. Schulz (Springer, Berlin, Heidelberg, 2001).
- [53] R. Häuselmann, AiiDA-VASP documentation, <https://aiida-vasp.readthedocs.io/>
- [54] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, *J. Appl. Phys.* **89**, 5815 (2001).

- [55] M. Graf and P. Vogl, *Phys. Rev. B* **51**, 4940 (1995).
- [56] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevMaterials.2.103805> for tables containing the band energies at high-symmetry points for different values of strain, a full export of the AiiDA database containing the calculations of strained InSb, InAs, and GaSb tight-binding models and an archive containing the 195 strained tight-binding models of InSb, InAs, and GaSb.
- [57] T. Yusufaly, D. Vanderbilt, and S. Coh, Tight-binding formalism in the context of the PythTB package, http://www.physics.rutgers.edu/pythtb/_downloads/pythtb-formalism.pdf (2013).
- [58] D. B. Litvin and V. Kopský, *Acta Cryst.* **A67**, 415 (2011).
- [59] Following the same reasoning, we can see that an arbitrary phase of the matrix U_g does not change the result.
- [60] Y. Sun, T. Nishida, and S. E. Thompson, *Strain Effect in Semiconductors* (Springer, Boston, MA, 2010).
- [61] Q. M. Ma, K. L. Wang, and J. N. Schulman, *Phys. Rev. B* **47**, 1936 (1993).
- [62] C. G. Van de Walle and R. M. Martin, *Phys. Rev. B* **34**, 5621 (1986).
- [63] J. Sánchez-Dehesa, C. Tejedor, and J. A. Vergés, *Phys. Rev. B* **26**, 5960 (1982).
- [64] O. Madelung, *Semiconductors*, Data in Science and Technology (Springer, Berlin, Heidelberg, 1991).
- [65] L. Brey, N. E. Christensen, and M. Cardona, *Phys. Rev. B* **36**, 2638 (1987).
- [66] O. H. Nielsen and R. M. Martin, *Phys. Rev. B* **32**, 3792 (1985).