ARTICLE   OPEN

Machine learning sparse tight-binding parameters for defects

Christoph Schattauer [1], Milica Todorović[2,3], Kunal Ghosh [2,4], Patrick Rinke [2] and Florian Libisch [1] ✉

We employ machine learning to derive tight-binding parametrizations for the electronic structure of defects. We test several machine learning methods that map the atomic and electronic structure of a defect onto a sparse tight-binding parameterization. Since Multi-layer perceptrons (i.e., feed-forward neural networks) perform best we adopt them for our further investigations. We demonstrate the accuracy of our parameterizations for a range of important electronic structure properties such as band structure, local density of states, transport and level spacing simulations for two common defects in single layer graphene. Our machine learning approach achieves results comparable to maximally localized Wannier functions (i.e., DFT accuracy) without prior knowledge about the electronic structure of the defects while also allowing for a reduced interaction range which substantially reduces calculation time. It is general and can be applied to a wide range of other materials, enabling accurate large-scale simulations of material properties in the presence of different defects.

INTRODUCTION

Ab-initio calculations have become the method of choice for the atomistic description of many materials[1]. However, relevant physical properties of a crystalline material often depend not only on its pristine structure but on various lattice defects[2–6]. While advances in sample preparation for low-dimensional materials have concurrently improved control over the occurrence of such defects, their influence is still significant for many investigations. Conversely, tailoring of material properties by defect engineering[7], chemical doping or functionalization inherently introduces point defects into the material[8]. An accurate description of how these defects modify the electronic structure of a system is thus key for exploring potential applications of novel materials.

While density-functional theory (DFT) typically yields a high-level description for moderately sized systems, simulating realistic devices used for measurements involves system sizes beyond the realm of these methods. Tight-binding (TB) models offer a quantum mechanical description for coherent electronic structure simulations with a scalability far better suited to experimental length scales[9]. For bulk systems, empirical tight-binding parameters[10,11] can be fit against ab-initio (e.g., DFT) or measured band structures (BS). More rigorous approaches aim for directly calculating effective tight-binding Hamiltonians by projecting DFT orbitals onto a suitably spatially localized basis[12–14], as implemented by the associated program package PAOFLOW[15,16]. These approaches work very well for automatically finding accurate bulk tight-binding descriptions of occupied orbitals that are well described by a suitable, known localized basis. By contrast, iterative methods such as maximally localized Wannier functions (e.g., wannier90[17–19]) try to determine an optimal localized basis, which often requires cumbersome convergence procedures. Once converged, they typically yield TB models of the highest quality[20–24]. A recent ML approach for high-throughput investigation also produced accurate bulk TB parametrizations[25] for pristine materials.

Computing accurate TB parameters for defect structures presents a challenge to established parametrization approaches. Simulating a defect requires large supercells to prevent artifacts from interacting with periodic images and to ensure accurate geometry relaxation at the edge of the supercell. In addition, breaking translational as well as (some) point group symmetries at the defect site vastly increases the number of independent TB parameters. Empirical bulk parametrizations lack the flexibility to describe defect systems with different local environments (e.g., different coordination numbers) than the pristine system[26], while Wannier projections become increasingly difficult to converge for larger cells. The resulting TB Hamiltonians also lack sparsity, and typically include finite long-distance interactions beyond even 5th nearest neighbors[22]. Since the efficiency of TB models partly stems from operating with sparse matrices, there is motivation to find sparse TB representations with minimal loss of representability. Simply truncating long-range interactions generally produces a significant loss of accuracy[22]. A quantitative description of defects, which is key to understanding their influence on the electronic structure, thus seems out of reach using established parameterization techniques.

In recent years, machine learning (ML) has facilitated new research lines in materials science and chemistry[27–38]. Here, we apply ML methods to generate TB parametrizations for defect structures in novel materials. We aim for an ML based scheme that achieves Wannier TB accuracy, while being automated and thus easy to use. Ideally, we want to be able to tune the sparseness of our machine-learned TB parameters at will to obtain a desired balance of accuracy and efficiency. To remain accurate despite fewer tuning parameters implied by improved sparsity, we will adjust the parametrization to specific energy regions of interest (i.e., close to the Fermi edge). We benchmark several test cases to demonstrate the accuracy of our approach, its efficiency and the effect of sparseness on accuracy and speed.

For simplicity and focus on the ML technique, the graphene benchmark system we consider features a comparatively simple orbital structure, with only the p_z orbitals contributing close to the Fermi edge. To account for coupling between orbitals of different

[1]Institute for Theoretical Physics, TU Wien, 1040 Vienna, Austria. [2]Department of Applied Physics, Aalto University, P.O. Box 11100, Aalto FI-00076, Finland. [3]Department of Mechanical and Materials Engineering, University of Turku, FI-20014 Turku, Finland. [4]Department of Information and Computer Science, Aalto University, P.O. Box 15400, Aalto FI-00076, Finland. ✉email: florian.libisch@tuwien.ac.at
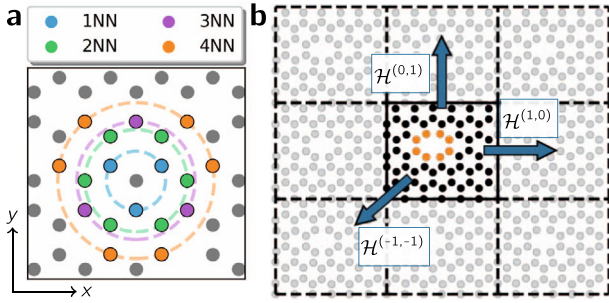
Fig. 1 Tight-binding model of graphene lattice. a Nearest-neighbor interactions in a hexagonal graphene lattice. The central gray atom has three first nearest neighbors (1NN, dark blue), 6 second-nearest neighbors (2NN, green), etc. **b** Interaction of a defect supercell with its periodic images (defect region highlighted in orange). The center cell itself is described by $\mathcal{H}^{(0,0)}$, the interactions to its neighboring cells by $H^{(\lambda_x, \lambda_y)}$ --- for large supercells only the nearest-neighbor interactions between cells, i.e., $\lambda_x, \lambda_y \in \{-1, 0, 1\}$ are non-zero.

angular momenta or different atomic species, it is straightforward to extend our scheme to distant-dependent, element-specific Slater–Koster parameters. All directional and orbital contributions are then accounted for by the well-known Slater–Koster formulas[10], while the distance-dependence still allows for the flexibility to accurately describe defect structures. We showcase this extension for a Se-divacancy in WSe_2 in Supplementary Note 6.

This paper is structured as follows: we introduce a model for mapping the desired TB Hamiltonian matrix to a vector of parameters that is compatible with ML algorithms. After establishing the necessary approximations and retrieval of DFT input data, we compare the efficiency and accuracy of different ML techniques for calculating TB parameters. We find multi-layer perceptrons (MLPs), i.e., neural networks, to be optimal for the task at hand. We present a detailed workflow of MLPs used for determining an optimal set of TB parameters for a given atomic structure. We explicitly generate parametrizations for two common defects (see insets in Table 2 and the methods section "Methods" for calculation details) in single layer graphene (SLG). The final section of this work focuses on validating and testing our machine-learned parametrizations. We consider the influence of defects on the local density of states, electronic transport as well as the level spectrum of a smoothly confined graphene quantum dot (GQD).

RESULTS

TB model

The TB approximation projects the Schrödinger equation for electrons — a partial differential equation — onto a basis of tightly bound (i.e., well localized) orbitals $|i\rangle$ at site i, yielding an algebraic equation. A system with n_o orbitals can then be described by a TB Hamiltonian

$$\mathcal{H} = \sum_i^{n_o} s_i \hat{c}_i^\dagger \hat{c}_i + \sum_{\langle i,j \rangle} \gamma_{ij} \hat{c}_i^\dagger \hat{c}_j. \tag{1}$$

$\hat{c}_i^\dagger (\hat{c}_i)$ are the creation (annihilation) operators of a quasiparticle at site i with position r_i, $s_i = \langle i | \mathcal{H} | i \rangle$ the onsite (diagonal) matrix elements and $\gamma_{ij} = \langle i | \mathcal{H} | j \rangle$ the hopping amplitudes between sites i and j. For sufficiently localized orbitals, the magnitude of γ_{ij} quickly decays for increasing distance $|r_i - r_j|$ between orbitals. Omitting such elements below a certain threshold (e.g., 1 meV) makes $\mathcal{H}$ sparse.

Starting from a full DFT Hamiltonian, optimal values for s_i, γ_{ij} can be directly and exactly calculated using maximally localized Wannier functions[17–19,39]. In practice, however, the final degree of localization — i.e., the distance beyond which overlaps between

orbitals are smaller than the defined threshold — may be several unit cells[22]. To obtain a more sparse description, one can directly fit a small set of TB parameters s_i, γ_{ij} to reproduce the DFT BS in an energy region of interest. The second sum in Eq. (1) then only runs over the n-th nearest-neighbor (NN) sites (Fig. 1a), where $|r_i - r_j| < r_{NN}$ with a cutoff radius r_{NN} controlling the sparseness.

Without loss of generality, we restrict our analysis to two-dimensional systems. We account for the Bloch phase of the periodic wave function by adding corresponding phase factors in the periodic images of the Hamiltonian. The periodic Hamiltonian matrices $\mathcal{H}^{(\lambda_x, \lambda_y)}$ determine the interaction of sites in the original cell $(0, 0)$ with sites in the periodic image of the cell (λ_x, λ_y) translated along a linear combination of lattice vectors $\{\lambda_x \cdot \mathbf{R}_x, \lambda_y \cdot \mathbf{R}_y\}$. The entire Hamiltonian then reads

$$\mathcal{H}(\mathbf{k}) = \sum_{\lambda_x, \lambda_y} e^{i\mathbf{k} \cdot (\lambda_x \mathbf{R}_x + \lambda_y \mathbf{R}_y)} \mathcal{H}^{(\lambda_x, \lambda_y)}. \tag{2}$$

Note that the set of s_i, γ_{ij} entirely determines the matrix elements of $\mathcal{H}^{(\lambda_x, \lambda_y)}$ while the grouping into periodic cells just accounts for the periodicity of the lattice. A system of interest is thus fully described by a set of lattice vectors and parameters s_i, γ_{ij} yielding the Hamiltonian matrices $\{\mathcal{H}^{(\lambda_x, \lambda_y)}\}$. The indices $\lambda_x, \lambda_y \in [-m, m]$ with $m \in \mathbb{N}_0$ determine the range of non-zero interactions between periodically shifted unit cells. In practice, we truncate at $|m| = 1$ given the large defect super cells in this work (see Fig. 1b).

Our objective is to use our TB Hamiltonian for transport calculations of SLG in realistic device settings, i.e., SLG including defects. We can therefore restrict the TB Hamiltonian to the carbon p_z orbitals, which determine the electronic structure of SLG close to the Fermi energy (see Supplementary Information for details).

Having reduced the TB Hamiltonian to only the p_z orbitals of carbon, we now consider a further reduction of the number of free parameters for the TB Hamiltonian. If we were to only enforce hermiticity, our TB Hamiltonian of Eq. (2) would feature $\frac{n_o(n_o+1)}{2} + 4n_o^2$ independent parameters s_i, γ_{ij}, which quickly gets out of hand. Considering a medium-sized defect supercell with 70 orbitals this would require ~25,000 independent parameters. We can however employ the residual symmetries of a defect structure to further reduce the number of parameters our ML model needs to optimize. To obtain a robust framework, we aim for a simple mapping between the hopping matrix elements γ_{ij} and local geometry information.

Finding such a simple mapping seems daunting as coordination numbers of atoms around the defect site will in general differ substantially from those in the bulk. A general mapping therefore seems to require detailed information about the local chemical environment. We avoid additional, complex geometrical parameters by exploiting that for the pristine bulk lattice, there are only a few distances (the nearest-neighbor spacings, Fig. 4a) while a relaxed defect geometry features many different distances. We generate the γ_{ij} purely as a mapping of distance $\gamma_{ij} = \gamma(|r_i - r_j|)$ to obtain an efficient and compact representation of the final TB Hamiltonian. A sufficiently fine, discontinuous mapping between atomic distance and hopping parameters essentially implies assigning an individual hopping parameter to each unique distance — except for degeneracies implied by symmetries, which should, indeed, have the same hopping interaction. A parametrization on distance alone thus yields a hermitian Hamiltonian correctly accounting for symmetries by construction. We can also simply choose a cutoff length r_{NN} above which no orbitals share a finite hopping value, to obtain a more sparse description. We discretize the interval $[0, r_{NN}]$ into n_c equidistant bins l with $l \in [1, n_c]$ using

$$\gamma_{ij} = \gamma(|r_i - r_j|) = \gamma_l, \quad l = \text{ceil}\frac{|r_i - r_j|}{\Delta r}, \tag{3}$$
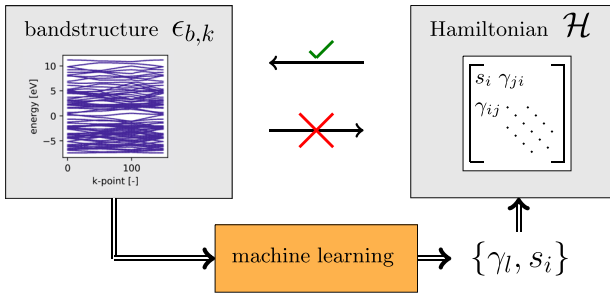
Fig. 2 Schematic of the inverse BS problem. For a given Hamiltonian, calculating a BS is trivial. By contrast, there is no constructive algorithm to obtain a Hamiltonian from a BS. Using ML, we aim to find such an inverse mapping from BS data (scalar energy values $\epsilon_{b,k}$ for each band b and k-point k) to a minimal list of TB parameters $\{\gamma_l, s_i\}$ (for each distance and onsite class l) which describes a full TB Hamiltonian $\mathcal{H}$.

with $\Delta r = r_{NN}/n_c$ the discretization step, and ceil(x) the ceiling function picking the smallest integer l with $l \geq x$.

We append a minimal set of onsite terms $\{s_i\}$ (accounting for symmetries) to the set of hopping values $\{\gamma_l\}$ with $l \in [1, n_c]$ to obtain a full TB parameterization, denoted for brevity as $\{\gamma_l\}$. We can then establish a bijective mapping from this list of interactions to full Hamiltonian matrices and vice versa. r_{NN} provides a tunable parameter for the desired sparseness of our TB model (up to how distant a neighboring orbital interacts with another one).

The number of bins n_c controls the coarseness of the discretization and can be adapted depending on the distribution of inter-orbital distances in a given structure. As long as the discretization Δr is fine enough, we only establish a convenient way of simultaneously addressing all symmetry-related interactions. For the two SLG defects we choose as benchmark systems, we decrease Δr until the number of different γ_l no longer increases (i.e., each value γ_l only addresses the hopping terms connected by symmetry, $\Delta r \approx 10^{-4}$ Å). At first glance, this prescription for grouping and setting the relevant interaction elements in a TB Hamiltonian seems quite similar to introducing an exponential dependence on distance in Slater–Koster parametrizations[10,11,40]. However, the discrete distance-hopping map only decouples symmetries and hermiticity from the parameter search and introduces little to no unnecessary simplification — in particular, it does not enforce a specific functional dependence on the distance. We do not need to consider the local geometric configuration (screening) of interacting orbital pairs as long as the discretization is fine enough to distinguish all different hoppings not related by symmetry. Indeed, we do not aim for a smooth mapping $\gamma(r_{ij})$, but rather for a distinct hopping parameter for all different couplings. Consequently, two neighboring values γ_l and γ_{l+1} can in principle take entirely different values.

From TB parameters $\{\gamma_l\}$ one can easily calculate a TB BS by diagonalizing the **k**-space Hamiltonian of Eq. (2) to obtain band energies $\epsilon_{b,\mathbf{k}}^{TB}$ and eigenfunctions $|\psi_{b,\mathbf{k}}\rangle$ via the eigenvalue problem:

$$\mathcal{H}(\mathbf{k})|\psi_{b,\mathbf{k}}\rangle = \epsilon_{b,\mathbf{k}}|\psi_{b,\mathbf{k}}\rangle \qquad (4)$$

The full set of TB parameters thus straightforwardly yields a BS with minimal numerical cost, $(\{\gamma_l\} \to \mathcal{H} \to \epsilon_{b,\mathbf{k}}^{TB}[\gamma_l])$.

Inverse band structure problem
Obtaining a BS from Eqs. (2) and (4) for a given Hamiltonian $\mathcal{H}^{(\lambda_x,\lambda_y)}$ is straightforward. However, to find the optimal Hamiltonian that best reproduces a given DFT BS $\{\epsilon_{b,\mathbf{k}}^{DFT}\}$ we need to solve the inverse problem $(\{\epsilon_{b,\mathbf{k}}\} \to \mathcal{H}$, Fig. 2). There is no straightforward (or unique) solution to this problem as highlighted by the plethora of TB parametrizations for any given material. Since

$\{\epsilon_{b,\mathbf{k}}^{TB}\}$ can be quickly evaluated, generating pairs of (arbitrary) sets $\{\gamma_l, s_i\}$ and the resulting BS $\{\epsilon_{b,\mathbf{k}}^{TB}\}$ on the TB level is easy. We can then use ML algorithms to identify the set of TB parameters which produces a TB BS in closest agreement with DFT.

To select a ML algorithm suitable for the inverse problem, we need to quantitatively compare different approaches. We grade several ML approaches both in terms of computational efficiency (how quickly do we arrive at an answer) as well as quality. To obtain a quantitative criterion for the quality of a parametrization we evaluate the difference of the final converged result to the DFT BS $\{\epsilon_{b,\mathbf{k}}^{DFT}\}$,

$$\delta_\epsilon[\{\epsilon_{b,\mathbf{k}}\}] = \sum_j^{n_k} \sum_b^{n_b} \left(\epsilon_{b,\mathbf{k}_j} - \epsilon_{b,\mathbf{k}_j}^{DFT}\right)^2. \qquad (5)$$

To tackle such a relatively high-dimensional, non-uniquely solvable inversion problem, we test variations of gradientless descent methods[41,42] (GLD), both multilayer perceptrons (MLPs) and convolutional neural networks (CNN) and Bayesian optimization via Gaussian process regression (GPR)[43] as possible alternative methods. We include the conceptually most simple gradientless descent as reference method to assess the benefit of more intricate approaches. All our ML methods produce reasonable parameter sets as exemplified by the small errors (δ_ϵ) in Table 1. Comparing also the time required to obtain a parametrization, we observe considerable differences between the approaches and therefore selected only the MLP for our final benchmarks. Below we briefly introduce each approach and discuss its pros and cons.

a. Bayesian Optimization trains a Gaussian process that maps input TB parameters $\{\gamma_l\}$ to the BS mismatch δ_ϵ. An acquisition function (see Supplementary Information for details) tailored to minimize δ_ϵ then decides which new $(\{\gamma_l\}, \{\epsilon_{b,\mathbf{k}}^{TB}\})$-pair is added to the data set. Such an active learning strategy results in compact datasets. However, given the low computational cost of generating $(\{\gamma_l\}, \{\epsilon_{b,\mathbf{k}}^{TB}\})$-pairs, the Bayesian optimization is dominated by the high cost of GPR training (Table 1). In the high-dimensional search space, the time saved by avoiding unnecessary evaluations of the forward problem (i.e., TB $\to \delta_\epsilon$ mapping) is smaller than the additional time needed to fit the Gaussian process.

b. Gradientless Descent is a zeroth-order, model-free optimization technique[41,42] that does not rely on an underlying gradient estimate (such an estimate can get expensive to come by in high dimensional spaces). It solves the inverse problem by repeated application of the forward problem. Despite reasonable δ_ϵ, the extracted parametrizations seem to perform less convincing for derived quantities (see Supplementary Information).

c. Multilayer Perceptrons are shallow feed-forward neural networks. In previous work, some of us have shown that neural networks can accurately predict spectra from the atomic positions alone[34]. Here we demonstrate that multilayer perceptrons (MLPs) can also solve the inverse problem directly by mapping band structures onto TB parameters. We add regularization via dropout layers and train them on $(\{\epsilon_{b,\mathbf{k}}^{TB}\}, \{\gamma_l\})$-pairs. We then make a final TB parameter prediction for $\{\epsilon_{b,\mathbf{k}}^{DFT}\}$. In our investigations, MLPs outperform all alternative approaches in accuracy at approximately equal or even lower computational cost (Table 1). We attribute this to the strong interdependence between the different TB parameters: an almost identical BS can be described by several different parameter sets, while changing only a single parameter (with the others fixed) will substantially change the BS. Such a structure is better represented by the fully connected network as opposed to model-free optimization schemes optimizing the different parameters individually.

d. Convolutional Neural Networks are reasonably deep, sparsely connected neural networks that are designed for automatic feature extraction from the input BS. CNNs excel at exploiting

correlations in their input data (e.g., the continuous lines forming a BS). Despite a reduction of trainable parameters compared to MLPs, the convolutional setups we benchmarked resulted in significantly longer training times but slightly worse BS losses.

Table 1. ML comparison: comparing performance [in terms of BS error δ_ε, see Eq. (5)] and time efficiency of several ML approaches to the inverse BS problem of the double vacancy in SLG.

Method	$\delta_\varepsilon[-]$	Runtime [hh:mm]
Wannier	1.23	$N_{trial} \times 00:50^a$
MLP	3.09	02:20
Convolutional Neural Network	3.48	60:00
Gradientless Descent	3.73	01:30
Bayesian optimization via GPR	6.2	> 24:00
Slater–Koster	54.72	–
Linear interpolation[b]	13.92	–

While the underlying data sets are not necessarily equal for the different optimization algorithms we still find this to be a legitimate comparison.
[a]While wannier90 is impressively efficient for a single run, it will typically require of the order of $N_{trial} \approx 20$ trial runs to identify correct parameters to achieve convergence in our experience.
[b]Introduced in Section "Data set generation" and Fig. 4.

We provide more details on all candidate methods in Supplementary Note 4, and focus on our final, most efficient algorithm below.

In the following we provide a step-by-step guide to our ML approach shown graphically in Fig. 3.

Data set generation

Before we can query our MLP to predict hopping parameters for the DFT BS of a defect system we need to procure appropriate training data in the form of BSs and their corresponding parameter lists. We do so entirely on the TB level, i.e., without requiring any DFT input by randomly sampling the vicinity of a reasonable initial guess in TB parameter space. We first determine an initial distance-hopping map $\gamma(|\mathbf{r}_i - \mathbf{r}_j|)$ (used to create $\{\gamma_l, s_i\}$ following Eq. (3)) based on the TB parameters of the pristine material, to obtain an initial TB Hamiltonian $H_{TB}^{(0)}$. We assume some reasonable parametrization of the pristine material exists - it is far simpler to extract a 10th-NN TB description for the bulk material than it is for a defect structure. For materials where even the bulk cell proves challenging to wannierize, one could resort to empirical or recent machine-learning approaches[25] for the initial parameter set. We initialize the distance-hopping map $\gamma^{(0)}(\delta r)$ as a piece-wise linear interpolation between the ten distance-hopping pairs extracted for the bulk material (see blue line and red markers in Fig. 4). We have validated this interpolated initialization for several defects in graphene and found that already such a (physically unmotivated) prescription for a
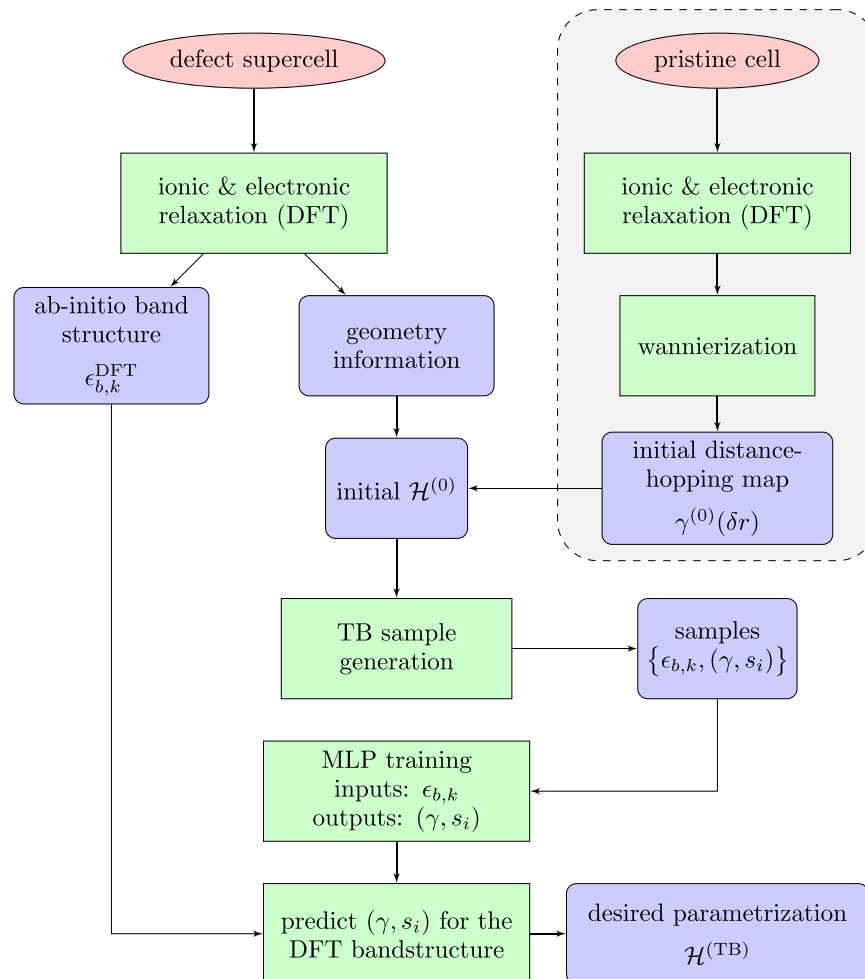

Fig. 3 Schematic flow chart of ML algorithm. Produces a machine-learned TB parametrization of a defect system. The hued section on the right can be replaced with an initial $\gamma^{(SK)}$ from Slater–Koster theory for materials with challenging bulk cells. Cornered green nodes represent calculation processes and rounded blue nodes represent data.
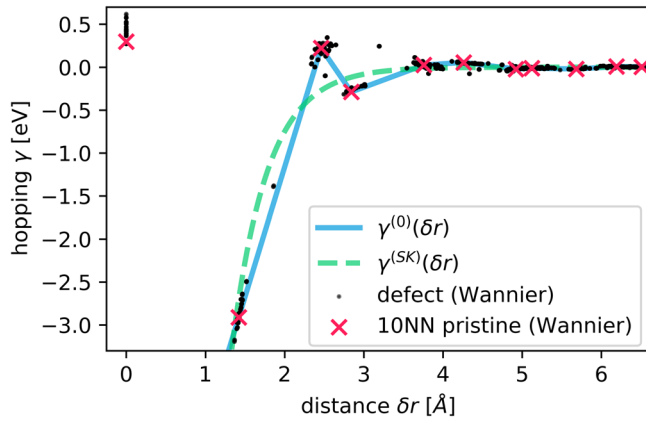
Fig. 4 Distance-hopping map of interaction elements. We plot all entries of a Wannier parametrization for the double vacancy defect in SLG (black dots) with both a Slater–Koster based initialization (taken from[40]) $\gamma^{(SK)}(\delta r)$ (dashed green line) as well as another initialization $\gamma^{(0)}(\delta r)$ defined as piece-wise linear interpolation (solid blue line) between the 10th-NN distance-hopping values (red crosses) of a bulk singler layer graphene cell. The hoppings at distance zero represent the onsite energies s_i.

TB parametrization outperforms a common Slater–Koster parametrization of graphene (see Table 1 and dashed green line in Fig. 4).

Having obtained a starting guess for $H_{TB}^{(0)}$, we calculate the training dataset by solving the forward problem ($\mathcal{H} \rightarrow \epsilon_{b,\mathbf{k}}$) many times with random fluctuations added to $H_{TB}^{(0)}$ (see Supplementary Information for details). We generate samples until further increase of the dataset size no longer reduces the BS error. We add relative and absolute noise to randomly selected parameters, carefully choosing noise amplitudes to sufficiently explore the relevant search space (for details see Supplementary Note 1). We then train the MLP to correlate changes in the shape of bands to corresponding modifications of values for specific TB parameters.
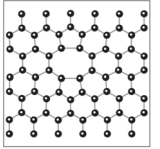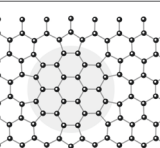
Multilayer perceptron model

As alluded to in Section "Inverse band structure problem", we adopt a multilayer perceptron to map BSs to TB parameters. The MLP takes all BS data $\{\epsilon_{b,\mathbf{k}}\}$ as 1D vector ($\overrightarrow{\epsilon}_{k_0}, \ldots, \overrightarrow{\epsilon}_{k_n}$) and outputs TB parameters as another 1D vector $\{\gamma_l\}$ holding the different hopping values for every distance as well as the minimal set of onsite energies necessary for building the entire TB Hamiltonian. We find optimal performance using three hidden network layers and choose their sizes via linear interpolation of the sizes for input- and output layers (see Supplementary Information and Methods section for details).

The range of BS inputs and TB parameter outputs covers several orders of magnitude. This wide spread necessitates Gaussian scaling of both inputs and outputs across all samples. Drop-out regularization (20% at the input layer) effectively avoids overfitting. By applying the distance-hopping map procedure to TB Hamiltonians of the two defect structures, we obtain a number of output parameters that strongly varies with the desired sparseness of the model (see Table 2). The sampling density of our BS in k space determines the number of input neurons in our network. We find sampling the Brillouin zone path with 30 points (i.e., $30 \times n_o$ input values for the network) to be a sufficient compromise between resolving BS features while keeping the input layer size manageable.

We emphasize that we aim to train a single-use network that is specifically tailored to one specific defect in a given material, as opposed to training a general MLP for predicting parameters for different defects. Such an approach would fail to capture the peculiarities and details of the individual defects. Our training approach is very robust and straightforward, enabling a much faster workflow than manually converging a well-behaved

Table 2. Sparseness of TB parametrization: number of independent TB parameters for a given sparseness in both defect structures under consideration.

structure	n_o	sparseness (neighbours)	n_p (# of parameters)
double vacancy	70	10NN	475
		5NN	168
		3NN	120
flower defect	128	10NN	842
		5NN	344
		3NN	173

Wannier parametrization. Indeed, for large systems converging a Wannier parametrization can even prove quite elusive, while our MLP-based approach should still work.

Training

We train the MLP on $N_s = 150{,}000$ data points, since performance converges and does not improve further by providing more samples (see Supplementary Information). We use a custom loss function that accounts for both parameter loss and BS mismatch of the predictions:

$$\mathcal{L} = \mathcal{L}_\gamma + a_\epsilon \mathcal{L}_\epsilon \tag{6}$$

$$\mathcal{L}_\epsilon = \sum_{b=1}^{n_b} \sum_{j=1}^{n_k} \left(\epsilon_{b,k_j}^{(p)} - \epsilon_{b,k_j}^{(t)} \right)^2 \tag{7}$$

$$\mathcal{L}_\gamma = \sum_l^{n_p} \left(\gamma_l^{(p)} - \gamma_l^{(t)} \right)^2 \tag{8}$$

With a_ϵ as a weighting factor, $\epsilon_b^{(t)}(k_j)(\epsilon_b^{(p)}(k_j))$ the true (predicted) value of band b at k-point j and $\gamma_l^{(t)}$, $(\gamma_l^{(p)})$ the true (predicted) value for the hopping (or s_i) of distance l, which we know for each pair of random Hamiltonian and associated BS in the training set. While an exact solution of the inverse band structure problem implies zero parameter loss, $\mathcal{L}_\gamma = 0$, we find that adding a physical observable, i.e., the actual BS mismatch $\mathcal{L}_\epsilon$ to the loss function improves convergence. We achieve optimal performance for $a_\epsilon \approx 5 \times 10^{-4}$ (see Supplementary Information).

Models for sparse parametrizations

The numerical effort in using a given TB parametrization strongly depends on the sparsity of the TB Hamiltonian, i.e., the number of non-zero hopping elements γ_{ij}. To improve performance, one can introduce a smaller cutoff length r_{NN} requiring that all interactions beyond the NN-th nearest neighbor are set to zero. We denote this as xNN for the models generated in this work. Generating sparser TB models barely requires changes to our ML workflow yet enables vast performance gains for subsequent application of the TB models (Eq. (11)). The initial parameters $\gamma^{(0)}(\delta r)$ can again be

taken from the piece-wise linearly interpolated bulk parameters (but cut off at r_{NN}). We will end up with fewer individual parameters (see Table 2) in a sparser TB description, generally allowing for a less accurate fit. However, in many applications the interesting physics is confined to a specific energy region, most commonly around the Fermi edge. Depending on the desired sparseness it proved beneficial to introduce additional weighting $w(\overline{\epsilon}_b^{(t)})$ into the BS loss function:

$$\mathcal{L}_\epsilon = \sum_{b=1}^{n_b} \sum_{j=1}^{n_k} \left(\epsilon_{b,k_j}^{(p)} - \epsilon_{b,k_j}^{(t)} \right)^2 w\left(\overline{\epsilon}_b^{(t)}\right) \qquad (9)$$

$$\overline{\epsilon}_b^{(t)} = \frac{1}{n_k} \sum_{j=1}^{n_k} \epsilon_{b,k_j}^{(t)} \qquad (10)$$

Restricting long-range interactions increasingly compromises the accurate reconstruction of the entire band structure. We achieved best results by focusing on the energy bands close to the charge neutrality point ($E = 0$) by reducing the number of input bands for the MLP (i.e., this mimics a step function for $w(\overline{\epsilon}_b^{(t)})$) all together and thus reduce both network size and computational cost for training. Employing a zero-centered Gaussian distribution with appropriate width for $w(\overline{\epsilon}_b^{(t)})$ achieves similar results at higher computational costs.

Our machine-learned TB parameters cannot be directly verified as they are no physical observables. Their exact values are not necessarily unique so long as they are capable of accurately reproducing derived quantities. We thus test the quality and validity of our extracted parametrizations with respect to BS, local density of states (LDOS), quantum transport and GQD-spectra which we found to be highly sensitive to the local electronic configuration of defects in recent work[44].

Benchmarks

For each defect, we calculate the LDOS on both the TB and DFT level thus enabling direct comparison to DFT results (as compared to the additional benchmarks discussed below in which the Wannier TB parametrization is the only reference). LDOS and BS are shown for the double vacancy and flower defect in Figs. 5 and 6, respectively.

Our 10th-NN ML TB model displays excellent agreement with the DFT BS (Fig. 5a) over a large energy window. While exact symmetries are captured via the distance-hopping map, noticeable disagreement regarding the exact width of some avoided crossings prove as the most challenging aspects for the MLP. In terms of the total density of states (DOS) the 10th-NN ML-TB-model is on par with the Wannier-TB-model. While neither can capture all the features of the ab-initio DOS both reproduce it much better than general Slater–Koster models (see Figs. 5b and
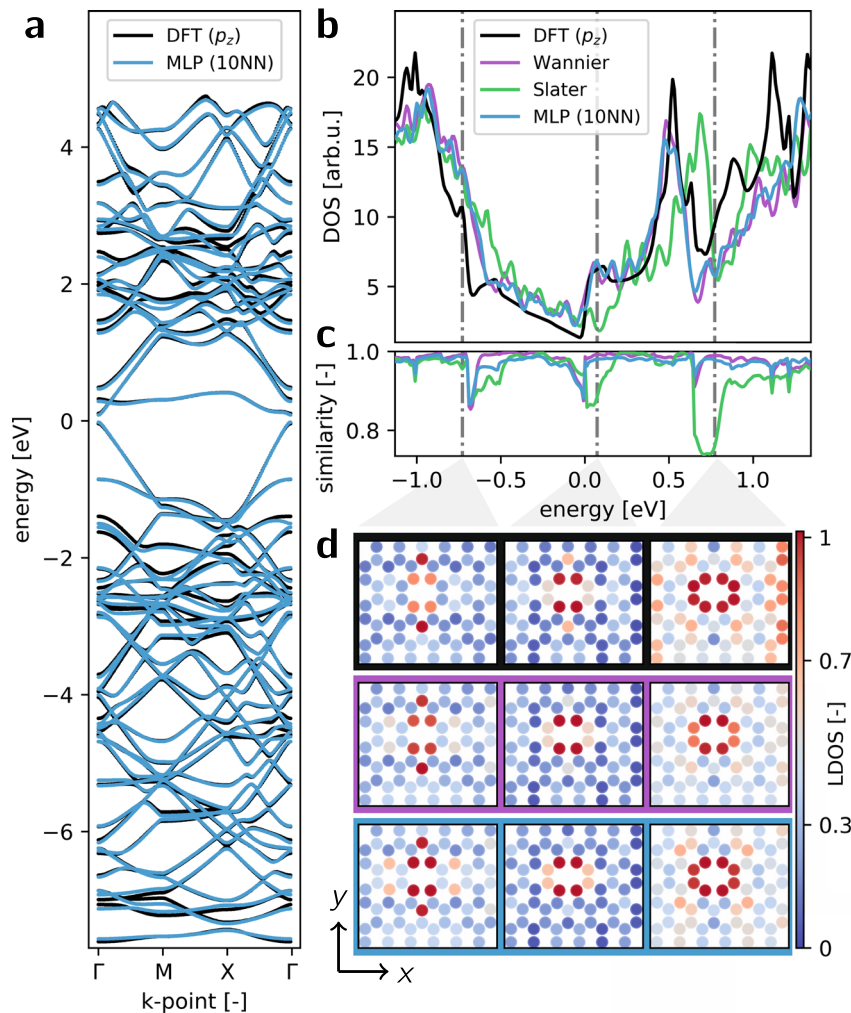

Fig. 5 Electronic structure analysis of vacancy defect. a BS of the SLG double vacancy supercell along $\Gamma M X \Gamma$ of both DFT calculation and MLP TB-model. **b** p_z-projected density of states of the supercell. **c** Cosine similarity of the local density of states between different TB models and the DFT result. **d** LDOS at the three energies (left to right) indicated by vertical dash-dotted gray lines in **b** and **c** for DFT, Wannier, MLP respectively (top to bottom, colored boxes match line colors in **b** and **c**).
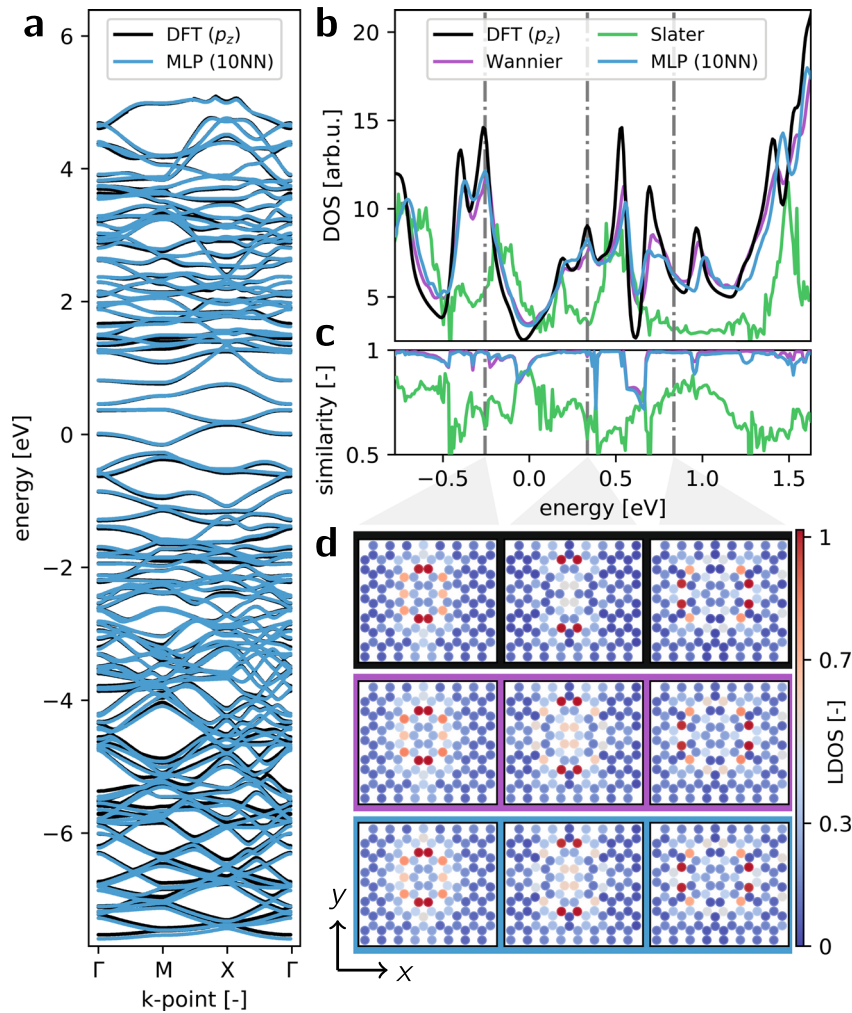
Fig. 6 Electronic structure analysis of flower defect. a BS of the SLG flower defect supercell along $\Gamma M X \Gamma$ of both DFT calculation and MLP TB-model. **b** p_z-projected density of states of the supercell. **c** Cosine similarity of the local density of states between different TB models and the DFT result. **d** LDOS at the three energies (left to right) indicated by veritcal dash-dotted gray lines in **b** and **c** for DFT, Wannier, MLP respectively (top to bottom).

6b). Since the deviations to the DFT DOS are present for both the machine learned and the Wannier parametrization we ascribe them to approximations of the TB formalism rather than a deficiency of our MLP algorithm.

The spatial information of the LDOS provides an even more detailed comparison, which we analyze both visually (Figs. 5d–f and 6d–f) at relevant energies (indicated as dash-dotted vertical gray lines in (Figs. 5b, c and 6b, c) and numerically via the cosine-similarity of individually normalized LDOS distributions with respect to the DFT results over the entire energy range (Figs. 5c and 6c). The results show that the MLP parametrizations not only very well capture the total DOS but also its spatial distribution (on par with Wannier) over a wide energy range (see Supplementary Information).

State-of-the-art modular recursive Green's function methods (MRGM)[45] (see methods section "Methods") profit immensely from sparse Hamiltonian matrices. Applying our sparse ML-TB-parametrizations to electronic transport calculations is therefore especially interesting. We study the different TB-parametrizations by embedding the defect supercells at five random but reproducible positions within a 15nm wide zig-zag SLG ribbon of length ≈130 nm (Fig. 7b). Employing our MGRM code we obtain the energy-dependent transmission $T(E)$ which uniquely portrays the multiple scattering events occuring in systems with several defects and compare $T(E)$ for the different parametrizations.

The 10th-NN ML-TB parametrizations accurately reproduce the transmission signature $T(E)$ for both defects (Fig. 7a, c). Our results also highlight the limited transferability[26] of Slater–Koster para-metrizations to different defect geometries: While the SK-TB-parameters for the double vacancy (Fig. 7a) produce a somewhat useful transmission curve its performance degrades drastically when applied to the flower defect (Fig. 7c).

Our sparser ML-TB parametrizations with interactions only up to the 3rd- or 5th-nearest neighbor still outperform the SK-parametrization. The loss in accuracy when enforcing very sparse Hamiltonians (3rd-NN) is a priori hard to quantify. While the TB description of the double vacancy seems more robust with respect to restraining long-range interaction than that of the flower vacancy (compare Fig. 7a and Fig. 7c) the 5th-nearest neighbor parametrization seems to strike an appropriate balance between computational performance gain

$$
\begin{aligned}
t_{\text{Transport}}^{\text{10NN}} &: 10\text{m}42\text{s} \\
t_{\text{Transport}}^{\text{5NN}} &: 1\text{m}26\text{s} \\
t_{\text{Transport}}^{\text{3NN}} &: 0\text{m}49\text{s}
\end{aligned}
\tag{11}
$$

and accuracy.

Another highly sensitive probe of our parametrizations comes in the form of smoothly-confined SLG quantum dots[46,47]. We
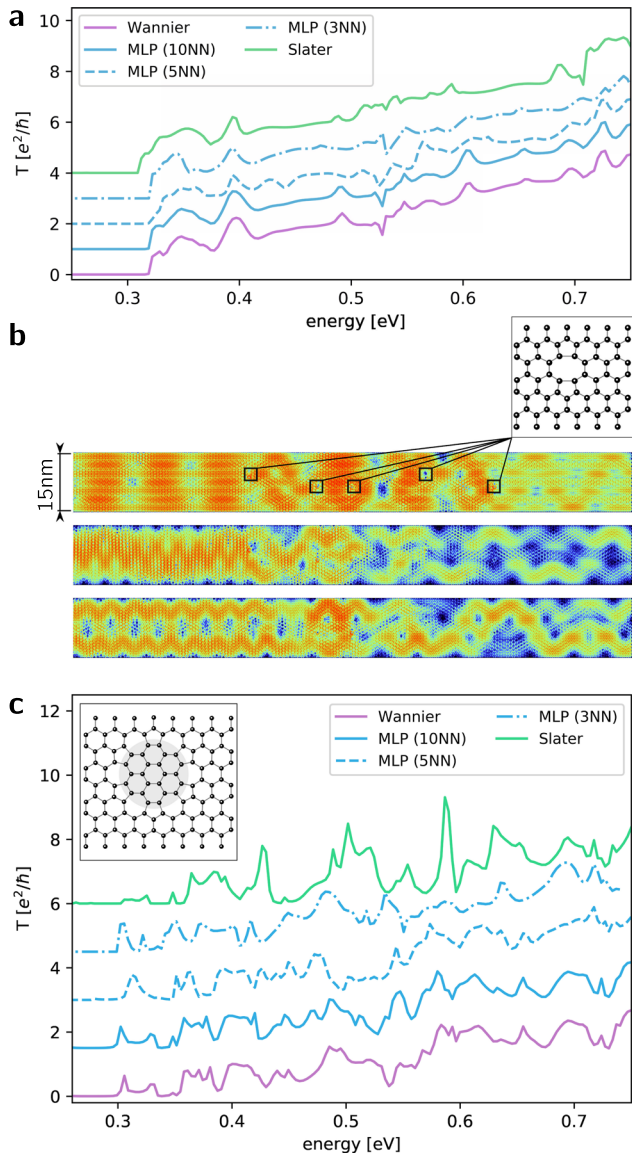
Fig. 7 Transport benchmark. Energy-dependent transmission $T(E)$ for different TB parametrizations of the **a** double vacancy and **c** flower defect in SLG (vertically offset for clarity). **b** Scattering density plots for the three lowest modes at $E = 0.7$ eV in the double vacancy setup with ribbon-width and embedded defect positions indicated.

Fig. 8 Quantum dot benchmark of vacancy defect. Level spectrum landscapes calculated with different TB parametrizations of the double vacancy in SLG compared against the Wannier parametrization [**a** MLP(10NN), **b** Slater–Koster, **c** MLP(5NN), **d** MLP(3NN)]. Inset shows schematic sketch of the underlying system: we calculate the level spectrum (orbital and valley quantum number, spin is omitted) as a function of the position of an STM-tip (brown) induced (smoothly confined) GQD relative to an embedded defect in a large graphene flake (gray rectangle). Dotted gray lines represent the level structure of a pristine GQD with doubly degenerate orbitals.

consider the influence of nearby lattice defects on the level spectrum of GQD's[44] as a benchmark for how well different TB-parametrizations model the local electronic configuration. Smoothly confining electrons in SLG retains the valley degeneracy which, omitting spin, yields doubly degenerate states. In the vicinity of a lattice defect this degeneracy is lifted as a function of defect-GQD distance[44] (see Fig. 8). The resulting level spectra as a function of GQD displacement X_T work as a unique fingerprint of the electronic structure of a defect.

We again find excellent agreement between the Wannier and the 10th-NN ML-TB parametrization. Conventional approaches such as Slater–Koster heavily underestimate the induced valley splittings Δ^τ and fail to capture the characteristic asymmetry of the lowest splitting for the double vacancy (Fig. 8c, d). The sparse ML parametrizations (3rd-NN or 5th-NN) still work quite well. Both slightly underestimate the induced splittings but manage to reproduce some of the asymmetry of the splittings for the double
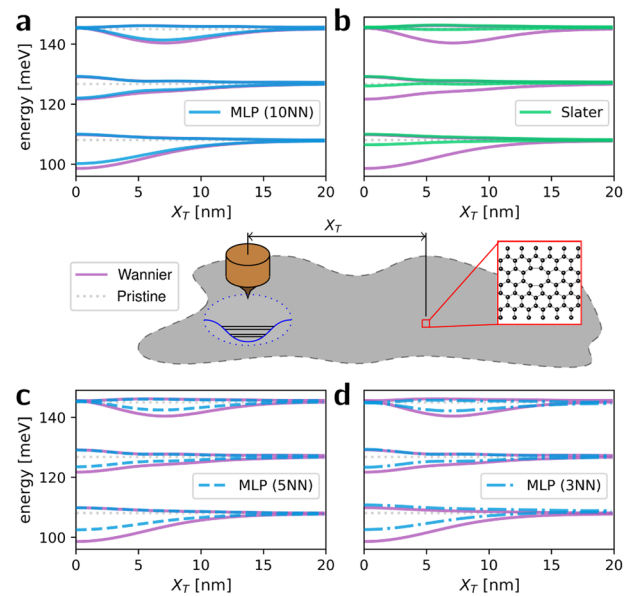
vacancy. The sparse ML-TB descriptions work especially well for the flower defect in this benchmark: qualitative agreement remains excellent and the quantitative changes to the induced valley splittings with increasing sparseness remain minor. The Slater–Koster model highly overestimates splittings and fails to reproduce several of the sharp avoided crossings.

DISCUSSION

The ML TB parametrizations yield accuracy on par with a full Wannier description, yet at substantially reduced cost. Once the sparseness levels are set, no human input or convergence issues appear during the parametrization step, and the improved sparseness greatly reduces computational demands in applications. The learning phase proceeds in an automated way, allowing for high-throughput simulations of different defects. More complicated materials such as transition-metal dichalcogenides will require even more parameters, and thus grouping of interactions by atom and orbital type. The same general algorithm should again work to provide tailored defect models.

The comparatively poor performance (see Table 1) of effective Slater–Koster methods strongly highlights the need for more accurate defect descriptions tailored to the corresponding electronic structure, which simply cannot be captured without additional DFT calculations. The remaining minor discrepancies in the highly sensitive GQD benchmark underline how the long-range interactions dictated by the underlying physics ultimately determine the accuracy of effective short-range descriptions: since we cut off long-range hoppings in the TB Hamiltonian, the sparse parametrization underestimates the range of the change in electronic structure induced by the defect. As a consequence, energy splittings between the two valley states are underestimated for small point defects like a vacancy (Fig. 8): only a tiny fraction of the quantum dot wavefunction (those few orbitals close to the defect) can actually
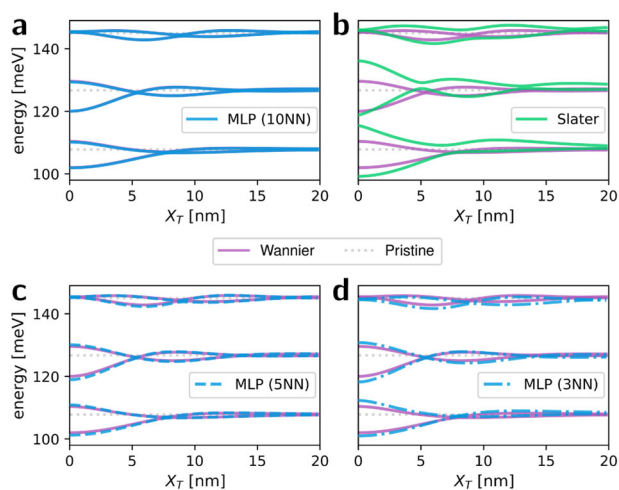
Fig. 9 Quantum dot benchmark of flower defect. Level spectrum landscapes calculated with different TB parametrizations of the flower defect in SLG compared against the Wannier parametrization [**a** MLP(10NN), **b** Slater–Koster, **c** MLP(5NN), **d** MLP(3NN)].

contribute to the defect-induced energy shift. By contrast, an extended defect like the flower (Fig. 9) is much better described.

Our comprehensive benchmarks (LDOS, transport, quantum states directly influenced by the defects) clearly outline the prowess of ML in obtaining DFT-quality results of defects in devices without substantial additional cost beyond the initial DFT calculation of the defect. Our sparse description of a defect system can be understood as a constrained optimization problem where ML offers elegant ways to find the sparse description with an optimal balance between accuracy and efficiency.

We have successfully implemented a ML algorithm to derive a TB Hamiltonian that accurately reproduces the BS details for general defect supercell structures in SLG. Given our universal treatment of symmetries and geometry information (distance-hopping map) this method can be applied to arbitrary material classes. This model requires a target BS and geometry information as inputs and allows for optimization towards a predefined sparseness of the desired TB description.

Our approach can be generalized to systems with relevant spin texture by either introducing additional distance-hopping maps ($\gamma_{\uparrow\uparrow}, \gamma_{\uparrow\downarrow}, \gamma_{\downarrow\downarrow}$) or employ a split off spin-orbit coupling term. For materials with a richer orbital structure (e.g., TMDs with dominant contributions from five d-orbitals on the metal site and three p orbitals on the chalcogen site) one may adopt a mixture of Slater–Koster[10] and discrete-distance-hopping-map approach by following the usual scheme for the angle-dependent assignment of interactions (i.e., direction cosines for the spherical harmonic nature of the respective orbitals) but promoting the typical Slater–Koster parameters (e.g., $V_{pp-\sigma}$, $V_{pp-\pi}$, $V_{pd-\sigma}$, $V_{pd-\pi}$, $V_{dd-\sigma}$, $V_{dd-\pi}$, $V_{dd-\delta}$, ...) to discretized distance-dependent maps (in principal identical to γ). An MLP can then learn these maps following the same algorithm as outlined above. Using such a scheme for Se divacancies in WSe_2 accurately reproduces all midgap defect states, including their different orbital characters (see Supplementary Note 6).

The conducted benchmarks included DOS analysis, multi-defect scattering in electronic transport calculations as well as simulations of the defect-induced splittings in a GQD. We found both qualitative and quantitative agreement of Wannier-TB-parameters (reference system) and the ML TB parameters of our MLP based approach. Given the considerably less complex input (energy values and atomic positions) than required by state-of-the-art iterative projection based methods (full DFT solution including Bloch states) our method should prove better suited for high-throughput material analysis.

METHODS

Machine learning

Our proposed neural network architectures (MLPs and CNNs) may be conveniently implemented via all common ML packages. We build our model via TensorFlow (v2.2.0) and the KERAS API (v2.3.0-tf). Furthermore, we use the Adam[48] optimizer with a learning rate $\beta \approx 10^{-5}$. We use a train/validation split of 75/25 of in total 200,000 samples. Learning rates $\beta \approx 10^{-5}$ with batch sizes of 2048 result in a fully trained model after roughly 1500 epochs. The Gaussian process regression employed in our Bayesian optimization scheme are implemented via the scikit-learn python package[49].

Density functional calculations

We perform DFT structural and electronic optimization with the VASP software package[50–53]. The double vacancy real space cell measures 6×6 pristine unit cells whereas the flower defect is modeled in an 8×8 cell. Both calculations encompass 25 Å vacuum in z direction and use a $3 \times 3 \times 1$ Monkhorst-Pack **k**-space grid. Our exchange-correlation functional of choice is Perdew–Burke–Ernzerhof (PBE) in a generalized gradient approximation. Both geometries are fully relaxed (using a conjugate gradient algorithm) to residual forces less than 10^{-2} eV Å^{-1}. Plane-wave energy cutoff is set to 500 eV and the systems are electronically converged to $\delta E \approx 10^{-9}$ eV.

Maximally localized Wannier transformation

The benchmark TB descriptions for the defects in this work have been generated with the Wannier90[17–19,39] software package. The double vacancy requires 175 Wannier functions initialized as atom-centered p_z and bond-centered s orbitals optimized with an outer energy window of $[-28.5$ eV,12.4 eV$]$ and an inner window of $[-28.5$ eV, -0.12 eV$]$. Disentangling the conduction bands from those virtual bands not included in the localized basis converges after 440 iterations while spread minimization converges after 187,089 iterations. The slightly larger flower defect requires 320 Wannier functions again initialized as atom-centered p_z and bond-centered s orbitals optimized with an outer energy window of $[-28.5$ eV, 12.4 eV$]$ and an inner window of $[-28.5$ eV, -0.12089 eV$]$. Disentangling converges after 599 iterations while spread minimization converges after 99,980 iterations. In both cases, Monkhorst **k**-space grids are taken over from the DFT calculations

Electronic transport

We evaluate transport in the Landau–Büttiker approximation using the energy-dependent Green's function $G(E)$ of the scattering structure[54]. By projecting $G|\chi_i\rangle$ onto the incoming wave in mode i we obtain a scattering state (see, e.g., Fig. 7b). By sandwiching G between incoming mode i and outgoing mode j we obtain the transmission $t_{ji} \propto \langle\chi_j|G|\chi_i\rangle$, where the proportionality factor is given by the square root of the relative group velocities $\sqrt{v_j/v_i}$. The total transmission is the sum of all squared transmission amplitudes $T = \sum |t_{ij}|^2$.

REFERENCES

1. Garrity, K. F. & Choudhary, K. Database of wannier tight-binding hamiltonians using high-throughput density functional theory. *Sci. Data* **8**, 106 (2021).
2. Zandiatashbar, A. et al. Effect of defects on the intrinsic strength and stiffness of graphene. *Nat. Commun.* **5**, 3186 (2014).
3. Linhart, L. et al. Localized intervalley defect excitons as single-photon emitters in wse₂. *Phys. Rev. Lett.* **123**, 146401 (2019).
4. Liu, Z. et al. Identification of active atomic defects in a monolayered tungsten disulphide nanoribbon. *Nat. Commun.* **2**, 213 (2011).
5. Liu, L. et al. Grain-boundary-rich polycrystalline monolayer WS2 film for attomolar-level hg2+ sensors. *Nat. Commun.* **12**, 3870 (2021).
6. Li, W. et al. The critical role of composition-dependent intragrain planar defects in the performance of MA1−xFAxPbI3 perovskite solar cells. *Nat. Energy* **6**, 624–632 (2021).

7. Jiang, J., Xu, T., Lu, J., Sun, L. & Ni, Z. Defect engineering in 2d materials: precise manipulation and improved functionalities. *Research* **2019**, 1–14 (2019).

8. Feng, Y., Chen, Q., Cao, M., Ling, N. & Yao, J. Defect-tailoring and titanium substitution in metal–organic framework UiO-66-NH2 for the photocatalytic degradation of cr(VI) to cr(III). *ACS Appl. Nano Mater.* **2**, 5973–5980 (2019).

9. Goringe, C. M., Bowler, D. R. & Hernández, E. Tight-binding modelling of materials. *Rep. Prog. Phys.* **60**, 1447–1512 (1997).

10. Slater, J. C. & Koster, G. F. Simplified lcao method for the periodic potential problem. *Phys. Rev.* **94**, 1498–1524 (1954).

11. Papaconstantopoulos, D. A. & Mehl, M. J. The slater koster tight-binding method: a computationally efficient and accurate approach. *J. Phys. Condens. Matter* **15**, R413–R440 (2003).

12. 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).

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

14. D'Amico, P. et al. Accurate ab initio tight-binding hamiltonians: effective tools for electronic transport and optical spectroscopy from first principles. *Phys. Rev. B* **94**, 165166 (2016).

15. Cerasoli, F. T. et al. Advanced modeling of materials with paoflow 2.0: New features and software design. *Comput. Mater. Sci.* **200**, 110828 https://www.sciencedirect.com/science/article/pii/S0927025621005486 (2021).

16. Nardelli, M. B. et al. PAOFLOW: a utility to construct and operate on ab initio hamiltonians from the projections of electronic wavefunctions on atomic orbital bases, including characterization of topological materials. *Comput. Mater. Sci.* **143**, 462–472 (2018).

17. Marzari, N. & Vanderbilt, D. Maximally localized generalized wannier functions for composite energy bands. *Phys. Rev. B* **56**, 12847–12865 (1997).

18. Mostofi, A. A. et al. An updated version of wannier90: a tool for obtaining maximally-localised wannier functions. *Computer Phys. Commun.* **185**, 2309–2310 (2014).

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

20. Gao, F., Bylaska, E. J., El-Azab, A. & Weber, W. J. Wannier orbitals and bonding properties of interstitial and antisite defects in GaN. *Appl. Phys. Lett.* **85**, 5565–5567 (2004).

21. Lu, I.-T., Park, J., Zhou, J.-J. & Bernardi, M. Ab initio electron-defect interactions using wannier functions. *npj Comput. Mater.* **6**, 17 (2020).

22. Linhart, L., Burgdörfer, J. & Libisch, F. Accurate modeling of defects in graphene transport calculations. *Phys. Rev. B* **97**, 035430 (2018).

23. Damle, A. & Lin, L. Disentanglement via entanglement: a unified method for wannier localization. *Multiscale Model. Simul.* **16**, 1392–1410 (2018).

24. Gresch, D. et al. Automated construction of symmetrized wannier-like tight-binding models from ab initio calculations. *Phys. Rev. Mater.* **2**, 103805 (2018).

25. Wang, Z. et al. Machine learning method for tight-binding hamiltonian parameterization from ab-initio band structure. *npj Comput. Mater.* **7**, 11 (2021).

26. Lekka, C., Papanicolaou, N., Evangelakis, G. & Papaconstantopoulos, D. Transferability of slater-koster parameters. *J. Phys. Chem. Solids* **62**, 753–760 (2001).

27. Kranz, J. J., Kubillus, M., Ramakrishnan, R., von Lilienfeld, O. A. & Elstner, M. Generalized density-functional tight-binding repulsive potentials from unsupervised machine learning. *J. Chem. Theory Comput.* **14**, 2341–2352 (2018).

28. Himanen, L., Geurts, A., Foster, A. S. & Rinke, P. Data-driven materials science: status, challenges, and perspectives. *Adv. Sci.* **6**, 1900808 (2019).

29. Nakhaee, M., Ketabi, S. A. & Peeters, F. M. Tight-binding studio: a technical software package to find the parameters of tight-binding hamiltonian. *Computer Phys. Commun.* **254**, 107379 (2020).

30. Peano, V., Sapper, F. & Marquardt, F. Rapid exploration of topological band structures using deep learning. *Phys. Rev. X* **11**, 021052 (2021).

31. Panosetti, C., Anniés, S. B., Grosu, C., Seidlmayer, S. & Scheurer, C. DFTB modeling of lithium-intercalated graphite with machine-learned repulsive potential. *J. Phys. Chem. A* **125**, 691–699 (2021).

32. Drautz, R., Hammerschmidt, T., Čák, M. & Pettifor, D. G. Bond-order potentials: derivation and parameterization for refractory elements. *Model. Simul. Mater. Sci. Eng.* **23**, 074004 (2015).

33. Ladines, A., Hammerschmidt, T. & Drautz, R. BOPcat software package for the construction and testing of tight-binding models and bond-order potentials. *Comput. Mater. Sci.* **173**, 109455 (2020).

34. Ghosh, K. et al. Deep learning spectroscopy: neural networks for molecular excitation spectra. *Adv. Sci.* **6**, 1801367 (2019).

35. Hammerschmidt, T., Drautz, R. & Pettifor, D. G. Atomistic modelling of materials with bond-order potentials. *Int. J. Mater. Res.* **100**, 1479–1487 (2009).

36. Westermayr, J. & Maurer, R. J. Physically inspired deep learning of molecular excitations and photoemission spectra. *Chem. Sci.* **12**, 10755–10764 (2021).

37. Chmiela, S. et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).

38. Nakhaee, M., Ketabi, S. A. & Peeters, F. M. Machine learning approach to constructing tight binding models for solids with application to BiTeCl. *J. Appl. Phys.* **128**, 215107 (2020).

39. Marzari, N., Mostofi, A. A., Yates, J. R., Souza, I. & Vanderbilt, D. Maximally localized wannier functions: theory and applications. *Rev. Mod. Phys.* **84**, 1419–1475 (2012).

40. Koshino, M. Interlayer interaction in general incommensurate atomic layers. *N. J. Phys.* **17**, 015014 (2015).

41. Nesterov, Y. & Spokoiny, V. Random gradient-free minimization of convex functions. *Found. Comput. Math.* **17**, 527–566 (2015).

42. Golovin, D. et al. Gradientless descent: high-dimensional zeroth-order optimization. Preprint at https://arxiv.org/abs/1911.06317 (2020).

43. Williams, C. K. I. & Rasmussen, C. E. Gaussian processes for regression. In *Advances in neural information processing systems* **8**, 514–520 (MIT press, 1996).

44. Schattauer, C. et al. Graphene quantum dot states near defects. *Phys. Rev. B* **102**, 155430 (2020).

45. Rotter, S., Tang, J.-Z., Wirtz, L., Trost, J. & Burgdörfer, J. Modular recursive green's function method for ballistic quantum transport. *Phys. Rev. B* **62**, 1950–1960 (2000).

46. Subramaniam, D. et al. Wave-function mapping of graphene quantum dots with soft confinement. *Phys. Rev. Lett.* **108**, 046801 (2012).

47. Morgenstern, M., Freitag, N., Nent, A., Nemes-Incze, P. & Liebmann, M. Graphene quantum dots probed by scanning tunneling microscopy. *Ann. Phys.* **529**, 1700018 (2017).

48. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. Preprint at https://arxiv.org/abs/1412.6980 (2014).

49. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

50. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

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

52. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* **47**, 558–561 (1993).

53. Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys. Rev. B* **49**, 14251–14269 (1994).

54. Libisch, F., Rotter, S., Güttinger, J., Stampfer, C. & Burgdörfer, J. Transition to landau levels in graphene quantum dots. *Phys. Rev. B* **81**, 245411 https://journals.aps.org/prb/abstract/10.1103/PhysRevB.81.245411 (2010).

AUTHOR CONTRIBUTIONS

C.S. applied the ML algorithms of this work. F.L. conceived and supervised the project. All authors designed the ML workflow, discussed the results and edited the manuscript before submission.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-022-00791-x.

Correspondence and requests for materials should be addressed to Florian Libisch.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.