Machine Learning: Science and Technology

ACCEPTED MANUSCRIPT • OPEN ACCESS

Machine learned environment-dependent corrections for a *spds*[*] empirical tight-binding basis

To cite this article before publication: Daniele Soccodato *et al* 2024 *Mach. Learn.: Sci. Technol.* in press https://doi.org/10.1088/2632-2153/ad4510

Machine learned environment-dependent corrections for a $spds^*$ empirical tight-binding basis

Daniele Soccodato,[1, a)] Gabriele Penazzi,[2] Alessandro Pecchia,[3] Anh-Luan Phan,[1] and Matthias Auf der Maur[1]

[1)] *Department of Electronic Engineering, University of Rome 'Tor Vergata', Via del Politecnico 1, 00133 Rome*

[2)] *Synopsys Denmark, Fruebjergvej 3, Postbox 4, DK-2100 Copenhagen, Denmark*

[3)] *CNR-ISMN, Via Salaria 29,300, 00017 Monterotondo, Roma*

Empirical tight-binding (ETB) methods have become a common choice to simulate electronic and transport properties for systems composed of thousands of atoms. However, their performance is profoundly dependent on the way the empirical parameters were fitted, and the found parametrizations often exhibit poor transferability. In order to mitigate some of the the criticalities of this method, we introduce a novel Δ-learning scheme, called MLΔTB. After being trained on a custom data set composed of *ab-initio* band structures, the framework is able to correlate the local atomistic environment to a correction on the on-site ETB parameters, for each atom in the system. The converged algorithm is applied to simulate the electronic properties of random GaAsSb alloys, and displays remarkable agreement both with experimental and *ab-initio* test data. Some noteworthy characteristics of MLΔTB include the ability to be trained on few instances, to be applied on 3D supercells of arbitrary size, to be rotationally invariant, and to predict physical properties that are not exhibited by the training set.

Keywords: Δ-Learning, Empirical Tight-Binding, Atomistic simulations, Electronic band structure, Antimonides, III-V materials

I. INTRODUCTION

Over the past decades, the empirical tight binding (ETB) method has become a valuable alternative whenever first-principle calculations require too much computational resources to be practically feasible. In fact, since the original article published by *Slater* and *Koster*[1] describing the method for the first time, its application grew substantially over the years, arriving to be applied to a wide range of materials and nano-structures. In particular, it has gained interest for the simulation of disordered materials like III-V and III-nitride alloys[2,3]. In the orthogonal ETB scheme, the Hamiltonian of the system is described using an atomic-like localized basis, in combination with a one-electron mean-field approximation[4]:

$$H = \sum_{m,i} |\phi_m(\mathbf{r}_i)\rangle \, E_m^{(i)} \, \langle \phi_m(\mathbf{r}_i)| + \sum_{\substack{m,n \\ i\neq j}} |\phi_m(\mathbf{r}_i)\rangle \, V_{m,n}^{(i)}(\mathbf{r}_i - \mathbf{r}_j) \, \langle \phi_n(\mathbf{r}_j)| \tag{1}$$

In 1, the indices m, n run over the orbitals that define the TB basis $\{|\phi_m\rangle\}$, whereas i, j denote the atomic sites within the chosen unit cell. The empirical character comes from the choice to not compute directly the integrals $E_m^{(i)}, V_{m,n}^{(i)}$ but rather to fit them to *ab-initio* or experimental results. Depending on the size of the chosen basis and on the parametrization scheme, the number of these parameters can be considerably high. The energies $E_m^{(i)}$ and $V_{m,n}^{(i)}$ are usually called *on-site* (or ionization potentials) and *off-site* parameters, respectively. Many different basis sets were proposed in the past, ranging from sets containing

[a)] Electronic mail: daniele.soccodato@uniroma2.it

sp^3s^* orbitals[5–7] to more complete ones that include also d^5 states[8,9]. However, regardless of the complexity, the ETB method has some shortcomings. For one thing, empirical tight-binding models are typically parametrized for specific elements or classes of materials, and their application outside of the scope intended during fitting may not be accurate. Secondly, the parameters are often not transferable, so that a set found for a specific element could be inapplicable if said element is simulated in an alloy or a compound. Moreover, constructing accurate tight-binding models can be challenging, as the fitting process is quite time consuming and requires a considerable amount of experimental or *ab-initio* data. All these limitations suggest the need to find a way of improving the tight-binding method in terms of transferability and reliability, without increasing the number of parameters to fit.

The term "machine learning" (ML) is used to delineate the set of statistics-based procedures aimed at finding patterns or predicting data without the need of providing a specific rule. In one of the most used approaches, a ML algorithm is given a task and a training set, and it has to learn how to perform the task by using the features of the training set data points. This is usually referred to as *supervised training*. A particular subset of supervised learning techniques is the so-called Δ-*learning*, in which a model is trained to adjust on a small degree a method that is considered to be fast but inaccurate. Δ-learning approaches have been recently used especially in the field of quantum chemistry, where they have been employed to compute electronic and molecular properties of several chemical species[10,11]. In this context, we developed a new Δ-machine learning framework (henceforth called MLΔTB) that corrects an established empirical tight-binding parametrization, using local information from the neighbourhood of each atom in the unit cell. More specifically, we implemented a technique designed to overcome one of the most critical points of the ETB method, namely the transferability of a set of parameters to a material that was not included in the original fitting scope. To do this, our framework is trained to correlate the local environment of each atom to a correction on the on-site parameters. The framework is applied to transfer a GaAs-GaSb parametrization to the case where the two materials are combined to simulate an alloy, $GaAs_{1-x}Sb_x$. We highlight that the application of the framework is extendable to any arbitrary parameter set and material.

The paper is structured as follows. In Section II the procedure to generate the data set is explained, with a focus on the choice of the target data and on the relevance of the simulated materials. Section III introduces and defines the baseline ETB parametrization that the framework is tasked to improve. In Section IV the components of the machine learning framework are described in detail, as well as all the relevant training settings. Section V shows the results and the performance of the trained ML scheme, while Section VI is reserved for the final considerations.

II. DATA SET GENERATION

Like any ML supervised model, our framework works by automatically finding a correlation between a $(\boldsymbol{x}, y)$ couple for each data set point, $\boldsymbol{x}$ and y being respectively the vector of features and the target of the training procedure. This section has the purpose of explaining the details for the choice of the target, which was picked to be the band structure (BS) of randomly generated alloy supercells. The particulars regarding the input features $\boldsymbol{x}$, and how they are related to the target, are given in Section IV.

A. Materials

At the present time, III-V materials are among the most studied compound semiconductors in scientific and industrial research. Among these, both GaAs and GaSb have attracted considerable attention during the years. GaAs is employed for manufacturing various types

of transistors, detectors, and high efficiency - high cost solar cells. Its lattice parameters and band structure characteristics have been very well known for decades. The same can be said for GaSb, which even if less widespread in its large-scale application has lately seen a surge regarding its study and use. Indeed, optical and electronic properties of short period InAs/GaSb superlattices were investigated in[12,13], while antimony-based high electron mobility transistors, resonant tunneling diodes, and heterojunction bipolar transistors are all examples of Sb devices researched in the past years[14]. Finally, alloys of $GaAs_{1-x}Sb_x$ at different Sb concentrations have been the subject of various articles regarding its potential employment in the fabrication of nanowires, quantum wells and photodetectors[15,16].

In this regard, it is quite relevant to try and develop an accurate Ga-, As-, Sb- empirical tight-binding parametrization; which ideally is computationally efficient and at the same time meets the requirements in terms of transferability from bulks to alloys, clusters and heterostructures. As anticipated, this work focuses on developing a ML method suited particularly for bulk GaAs, GaSb and random $GaAs_{1-x}Sb_x$ alloys.

B. Target and ab-initio simulation details

The band structure is not a novel choice as the target for the task of automatic learning related to ETB. Indeed, the k-dependent Hamiltonian eigenvalues often determine the loss function on which this kind of algorithms are trained[17–20]. Even in the classical approach of fitting the tight-binding parameters, a combination of BS data and other quantities (such as eigenfunctions) is often employed to solve the problem.
To our knowledge, there is no data set publicly available regarding the band structure of antimony-based alloys. For this reason, we generated a custom data set using the QuantumATK software[21]. In particular, the software was employed in its *U-2022.12-SP1* version to create $GaAs_{1-x}Sb_x$ supercells of $N = 54$ atoms ($3 \times 3 \times 3$ repetitions of a zincblende primitive unit cell), with $x = 0.1,\ 0.3,\ 0.4,\ 0.6,\ 0.8$. At every Sb concentration, 10 random realizations were drawn, in order to generate a total ensemble of 50 configurations. Also, two additional supercells were added to include the bulk GaAs and GaSb structures ($x = 0, 1$). All supercells were internally relaxed using a Abell-Tersoff empirical potential[22], with a limited memory BFGS optimizer (LBFGS)[23] and a tolerance on the atomic forces of 0.05 eV/Å.

Subsequently, a DFT-LCAO calculation[24] was performed for each configuration, using the *Medium* basis as described in[21]. For all computations, the Brillouin zone was sampled on a $3 \times 3 \times 3$ Monkhorst-Pack grid, and the exchange-correlation term was modeled using a custom HSE functional (from now on addressed as c-HSE)[25]. The α parameter (mixing between the exact Hartree-Fock and PBE0 exchange energies) was first tuned on the two GaAs and GaSb bulks in order to match the experimental band gaps at 300 K reported by *Madelung*[26], resulting in values of respectively $\alpha = 0.33,\ 0.39$.
Starting from these values on the extremes, α was then linearly interpolated according to the Sb concentration of the alloy:

$$\alpha(x) = (1 - x) \cdot \alpha^{\mathrm{GaAs}} + x \cdot \alpha^{\mathrm{GaSb}} \tag{2}$$

The same linear interpolation was performed for the lattice constant (with extreme values taken from[26], $a^{\mathrm{GaAs}} = 5.6536$ Å, $a^{\mathrm{GaSb}} = 6.0960$ Å at 300 K).
A discussion should be made about the choice of linearly interpolating these two parameters. On one hand, the lattice constant choice is motivated by experimental results, which show a linear trend as a function of the concentration[27] (see also the Supplementary Information for the corresponding plot).
On the other hand, several works show that the mixing parameter can be linearly interpolated if a chosen common reference energy level does not vary as a function of α[28–30]. We checked this condition on a test structure of the dataset, and we provided additional details in the Supplementary Information.

TABLE I. Summary of the data set generated using the QuantumATK software. The mean band gap E_g is computed by averaging the values over the 10 ensemble configurations, for each Sb concentration. Each average value is paired with its computed standard deviation, σ_{E_g}.

Sb fraction	Nr. of structures	Average E_g [eV]	σ_{E_g} [eV]
0.0	1	1.422	0.000
0.1	10	1.177	0.019
0.3	10	0.784	0.062
0.4	10	0.663	0.097
0.6	10	0.479	0.062
0.8	10	0.487	0.070
1.0	1	0.762	0.000

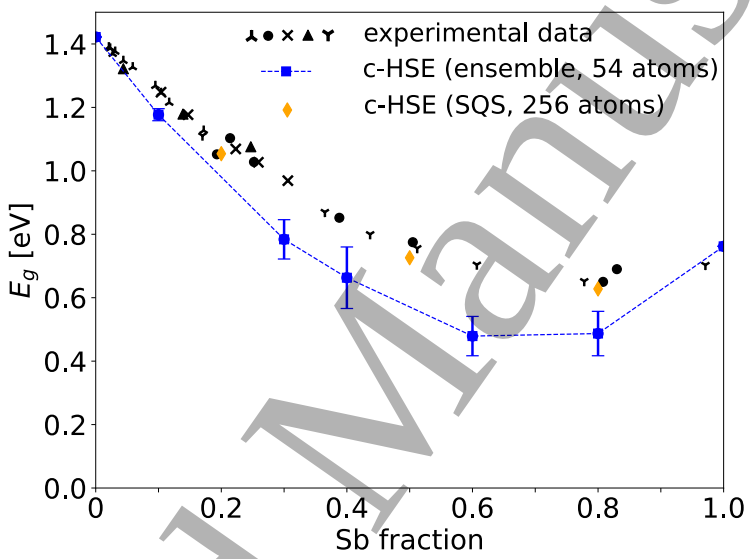

FIG. 1. Compositional bowing of the direct band gap for the $GaAs_{1-x}Sb_x$ 54-atom supercells, simulated using the custom c-HSE functional. The squared marks correspond to ensemble averages, while the vertical bars represent the standard deviation reported in table I. Results are shown together with experimental measurements from[32–36] (different shapes correspond to different data sets), and with DFT computations of three 256-atom SQS supercells at $x = 0.2,\ 0.5,\ 0.8$.

The other HSE characteristic parameter, the screening length μ, was set to 0.2078 Å^{-1} for all calculations. Regarding the pseudopotential, the choice fell on the norm-conserving PseudoDojo potentials[31].
Finally, unless otherwise specified, the target band structures were computed along the $L - \Gamma - X$ path, with a total of 21 uniformly distanced k-points. A summary of the generated data set is reported in table I

Figure 1 shows the computed band gaps as a function of the antimony concentration. The DFT calculations are reported together with experimental values of the band gap collected from 5 different data sets[32–36]. As is evident, there is a discrepancy between the experimental and computed points. Indeed, all the band gap values seem to be consistently underestimated, and this reflects also on the *compositional bowing* parameter, which is reported should lie in the range of $1.0 - 1.44$[27], in contrast with the value of our calculations ($B = 2.18$).

An explanation for this seemingly incorrect result can be found in the limited size of the supercell. Indeed, because of computational limitations and in order to have a dis-

tinguishable target for the model (i.e. a BS that is not exceedingly folded), the supercell dimensions had to be restricted. This is in turn far from ideal for simulating random alloys: the periodic images of the cells decrease the randomness, and this results in underestimated ensemble average values and bigger variances. Equivalently, a decrease on randomness due to clustering has been shown to have a similar effect in III-Nitride alloys[2].
To provide further evidence that the small cell size is the cause behind the overestimation of the bowing, figure 1 also shows the band gap values at $x = 0.2,\ 0.5,\ 0.8$ resulting from a 256-atom HSE calculation. Again, the details for this simulation are the same previously discussed, with the distinction that in this case, the alloys were generated using the *Special Quasi-Random Structure* (SQS) technique. This method aims to find the most random configuration possible, by minimizing the distance between the correlation functions of the candidate structure and those of a perfectly random alloy[37]. With this method, there is no need to sample an ensemble of realizations in order to get the average band gap value. SQS techniques were previously applied with success to other random alloy studies[21,38]. Finally, to give one ulterior proof to the argument, we repeated the ensemble calculation at $x = 0.6$ with an increased supercell size (10 supercells of 128 atoms). The values of the mean band gap and standard deviation resulted in $E_g = 0.608$ eV, $\sigma_{E_g} = 0.063$; which are indeed higher than those at the same concentration in Table I

Despite the apparent limitation in the description of the compositional bowing, we argue that this data set still constitutes a suitable candidate for our machine learning task, especially considered the local nature of the target corrections, which is not affected by this global effect. Even more importantly, we will demonstrate how our MLΔTB method can overcome this problem, effectively predicting a characteristic of the alloy that the very same data on which it was trained failed to show.

III. INITIAL PARAMETRIZATION

As stated in Section I, the proposed algorithm is an example of Δ-machine learning. In the context of ETB, this means designing a method that improves an existing Slater-Koster (S-K) parametrization, such that the new set can accurately reproduce ab-initio results. In this section we focus on the S-K parameter set that determines the starting point of our approach.
We begin by considering a nearest-neighbour $sp^3d^5s^*$ orthogonal basis for GaAs and GaSb, as proposed by *Jancu et al.*[8]. This basis, that includes also spin-orbit coupling, has been proved to accurately describe a wide range of bulk semiconductors and III-V materials, and in our case consists of 29 independent parameters for each binary compound. Since the interest is in modeling an alloy of GaAs-GaSb, a problem arises in the choice of the 4 on-site parameters for the Ga atom. Of course, these parameters have different (although very similar) values depending on whether the Ga atom belongs to GaAs or to GaSb. This distinction very clearly loses meaning when dealing with structures like random alloys, interfaces or clusters, for which it makes no sense to distinguish between Ga atoms belonging to an individual type of material. Efforts to overcome such limitation have been tried in the form of on-site averaging, or with more complex parametrizations that take into account the local atomistic environment of each atom[39], although this results in quite an increase of the numbers of parameters that require to be fitted.

While all the previous options are in principle suitable for defining the initial set, we opted to re-fit the Jancu parameters by enforcing shared Ga on-site integrals. This means that the two bulk GaAs and GaSb materials were simultaneously fitted on DFT band structures, constraining the Ga on-sites to have the same value for both compounds. The initial guess for the parameters was taken from[8], the Ga on-sites being arbitrarily selected from the GaAs column. We included also a simple Harrison scaling[40] for the hopping integrals to take internal strain into account. The details of the DFT simulation for the two

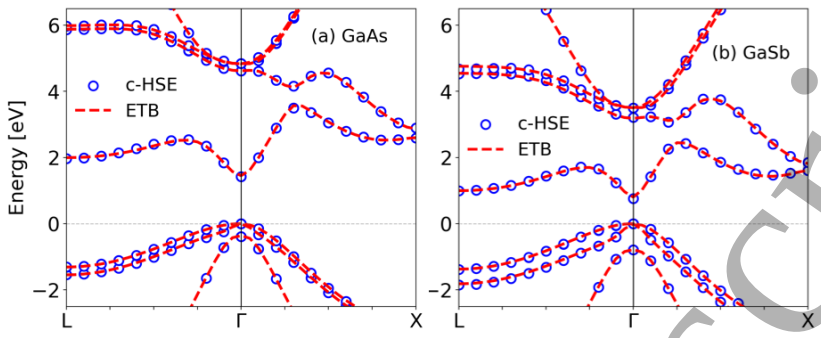

FIG. 2. The result of re-fitting the parameters from[8] after enforcing a shared Ga- parameter set. The computed ETB band structure is compared with the DFT reference.

TABLE II. The ionization potentials of the Ga atom for the *spds** basis (in eV). In the first two columns, the parameters as reported in[8]. In the third column, the parameters resulting from the simultaneous re-fitting of the GaAs and GaSb band structures.

Ga on-site parameter	GaAs	GaSb	GaAs-GaSb
E_s	-0.4028	-0.4003	-2.7219
E_p	6.3853	6.3801	3.8666
E_d	13.1023	11.5944	19.2383
E_{s^*}	19.4220	16.6388	15.7063

bulk references follow Section II B, with the only difference being that only a 2-atom unit cell was created, in contrast of the 54-atoms supercells that make up the data set. As a consequence, the Brillouin zone was in this case sampled on a $8\times8\times8$ Monkhorst-Pack grid.

Fitting was performed using a least square optimizer from the *scipy* package[41]. The results of the fitting procedure can be seen in figure 2, while table II reports the original ionization potentials for Ga compared to the re-fitted ones. We highlight the fact that the band structures were defined up to an arbitrary reference, so that there is no real connection between the re-fitted values and typically used reference energies (such as the work function).

Although the newly found parametrization agrees very well with the DFT references, its application to the $GaAs_{1-x}Sb_x$ alloy is severely underperforming. Indeed, figure 3 shows the comparison between the $GaAs_{1-x}Sb_x$ DFT and ETB band structures, calculated for some selected x on 4 instances of our generated data set.

It is evident how the shared-Ga parametrization fails to be transferred for any of the alloys. This is most probably due to the fact that the parameters do not contain any information about the local atomistic environment, nor they capture the valence band offset (VBO) between the two bulk materials. Moreover, the simple d^{-2} Harrison law is too crude of a correction to account for the internal strain consequent to the relaxation of the supercells. Clearly, the starting parametrization could in principle include these informations as well. For example, one could consider material and/or alloy-concentration specific corrections, like adding a VBO parameter, fitting the exponents of Harrison's law, or using the scheme defined in[39]. Nevertheless, we chose to refrain from doing so, as one of our aims was to define a ML model which requires a description of the local atomic environment only, and as few free parameters as possible.

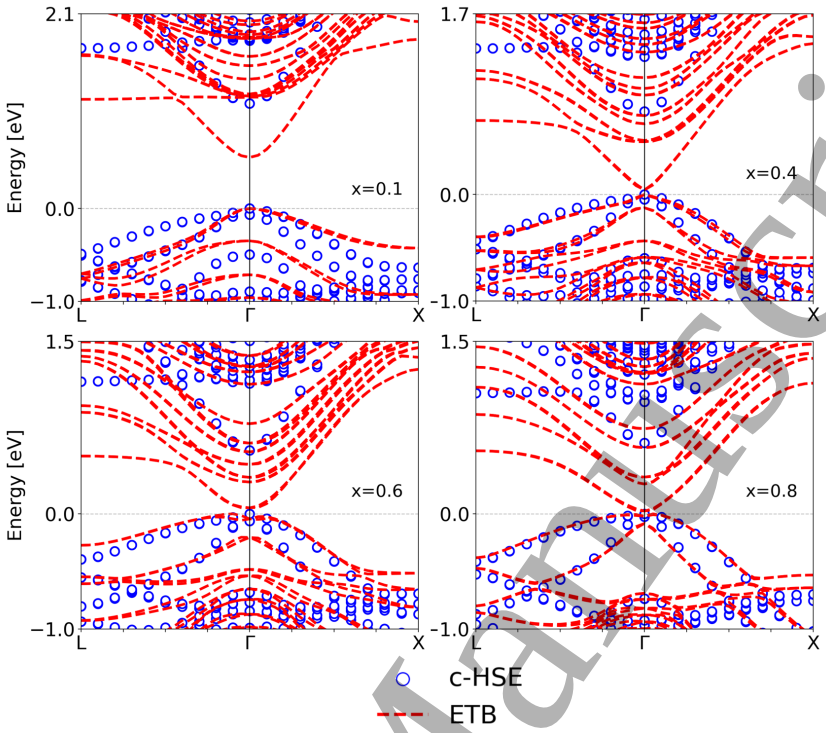

FIG. 3. The ETB-computed band structure of four instances of $GaAs_{1-x}Sb_x$, for $x = 0.1, 0.4, 0.6, 0.8$, compared with the *ab-initio* reference.

IV. ML∆TB MODEL

The overall ML∆TB framework is shown in figure 4. The input to the framework is the atomistic structure generated with the QuantumATK software. For each atom in the configuration, a representation of its local environment is defined (figure 4-a). This representation, also called *descriptor* in the literature[42], constitutes a fingerprint of the atom, and usually depends on the chemical species of its neighbours (and itself), as well as on the distances between the neighbour atoms and the central one. Being a common concept especially in the field of computational chemistry, the descriptor can have many definitions, but it is required to satisfy some properties: it must be invariant under translation, rotations, reflections and permutations of two equivalent atoms, and it must be complete (i.e. it must uniquely determine the atomic environment). We highlight that our scheme is built in such a way that any kind of input descriptors can be given. For this reason, in Section IV A we provide a brief overview of some possible definitions of such input, as well as introduce our choice for the task of $GaAs_{1-x}Sb_x$ alloy fitting.

After having created all the descriptors in each structure of the training set (i.e. the set of configurations used for training, details are given in Section IV C), the next step in the pipeline of our model is to pass (sequentially) these vectors to a feed-forward neural network (NN), which is assigned to output 4 real values (figure 4-b). These quantities constitute the corrections, in eV, to the 4 on-site terms of the initial S-K parametrization, introduced in Section III. Indeed, they represent the Δ- values to be added to each ionization potential reported in the third column of table II, as well as to the As, Sb onsite parameters (not shown in the table). Since the network acts on the atomic level, the following stage consists in rearranging the sequence of 4-dimensional vectors to be able to correct the tight-binding Hamiltonian. This process is depicted in figure 4-c. The outputs of the network are lined

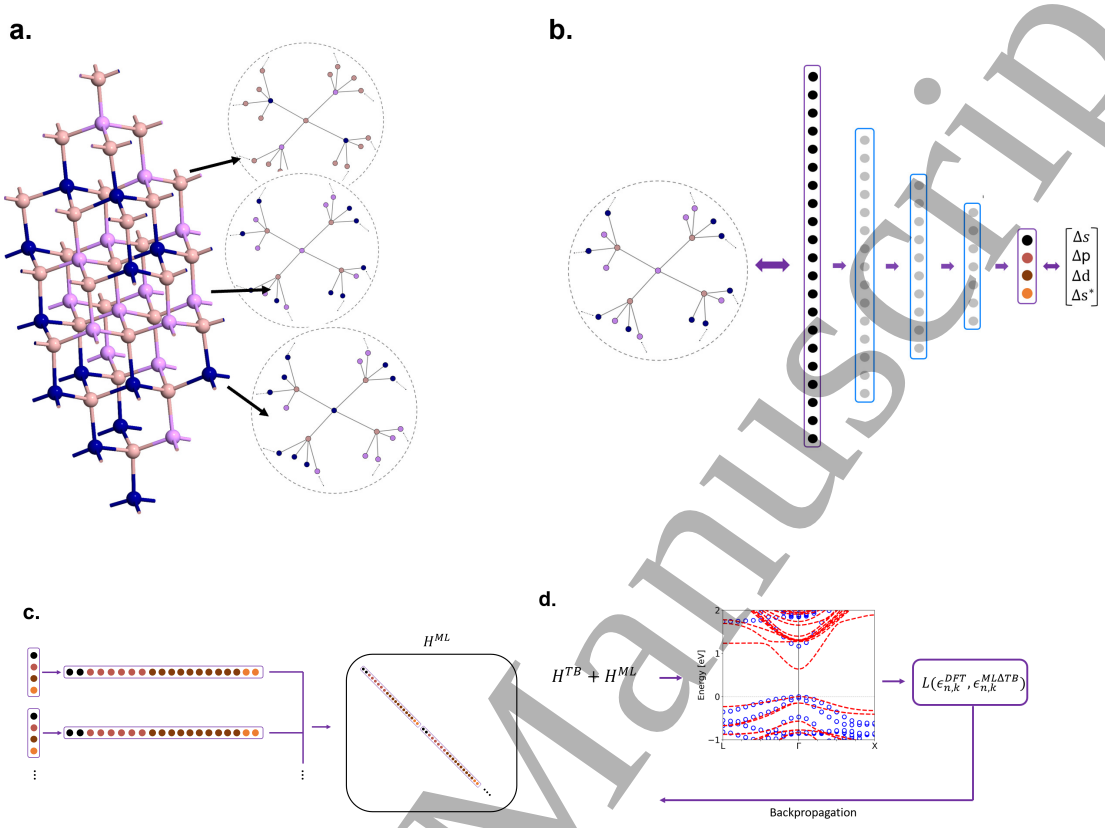

FIG. 4. The presented MLΔTB framework. In (a), for every atom in the considered structure a local descriptor is computed. In (b), all the descriptors calculated in the previous step are given as an input to a feed-forward NN. For every descriptor, the network outputs a 4-dimensional vector containing the corrections to the s-, p-, d-, s^*- orbitals. In (c), the 4-valued output is mapped into a 20 dimensional vector. All $N = 54$ vectors are lined up and placed on the diagonal of a zero-filled matrix. In (d), the matrix computed at the previous step is summed to the ETB Hamiltonian, resulting in a correction of the on-site elements. The resulting matrix is used to compute the band structure. Finally, in case of training, the computed BS is used in combination with the reference to update the model weights through the process of *backpropagation*.

up and each on-site value is repeated as many times as the degeneracy of the corresponding orbital. Spin-orbit is accounted for by repeating all the values of an additional factor 2. The result is a sequence of N 20-dimensional vectors (recall that $N = 54$ for every structure in the data set), which are placed on the diagonal of a 0-filled matrix, H^{ML}. Some technical details about the neural network and the nature of the correction are given in Section IV B.

Finally, the network-generated diagonal Hamiltonian is added to the ETB one, the resulting matrix is diagonalized and the produced band structure is compared with the target reference (figure 4-d), through the evaluation of a loss function. The process of *backpropagation*, by means of which the weights of the network are updated for the next training cycle, is performed through automatic differentiation starting from these loss function values. This last phase differs from the more traditional definition of target quantity in data science, and deserves an in-depth examination in Section IV C.

A. Atomistic descriptors

There exist several methods to create a local representation of the atomistic environment. Yet, all possible definitions start from the choice of what constitutes a neighbourhood. Given an atom i in a bulk configuration or molecule $C = 1, 2, ..., i, ...N$, we can define its neighbourhood K_i as the subset of C containing all atoms within a certain range, R_{cut}:

$$K_i = \{j\} \in C, \quad |R_i - R_j| < R_{cut} \tag{3}$$

The choice of R_{cut} plays an important role in learning the proper correction for the on-site S-K parameters. The correction is supposed to be highly dependent from the Sb concentration in the alloy, therefore we need a descriptor informative enough to correlate with this dependency. This argument motivated us in choosing a range that goes up to the third neighbour (for our data set, a proper choice for all structures is $R_{cut} = 4.6$ Å).
Another important property that the descriptor should have is the dependence from relative distances between the central atom and its neighbours, which become of considerable importance when internal strain is introduced. Many, modern methods are able to take distances into account. *Bartok et al.* defined the so-called smooth overlap of atomic positions (SOAP), a similarity measure between atomic neighbour densities expanded in an harmonic basis[43]. Another alternative is given in[44], where a local version of the Coulomb matrix is used to predict atomic potentials. It is even possible to refrain from defining a functional form altogether, and rely instead on an additional machine learning model to learn the best representation for the task at hand[19,45].

We found that a rather effective, out-of-the-box and cheap method to create local atomic embeddings is a class of functions called *moment tensors* (MTs). First introduced by *Shapeev et al.*[46], moment tensors were defined in the context of a linear regression procedure designed to predict interatomic potentials. This class of functions is systematically improvable, meaning that their accuracy can be increased by increasing the number of the defined basis functions. Although originally used for a different (linear) problem, this basis set can be computed in a more general way and used as an input for the MLΔTB neural network. Indeed, following the practical implementation from[47], we can define these descriptors by first introducing the concept of *moments*:

$$M_{\mu,\nu}(K_i) = \sum_{j \in K_i} f_{\mu,\nu}(|r_{ij}|, z_i, z_j) \underbrace{\boldsymbol{r}_{ij} \otimes ... \otimes \boldsymbol{r}_{ij}}_{\nu \text{ times}} \tag{4}$$

These functions are comprised of a radial and an angular component, and depend on the two indices μ, ν. In our case, the radial function $f_{\mu,\nu}(|r_{ij}|, z_i, z_j)$ is chosen to be the product between a Chebyshev polynomial and a cutoff function defined on a compact support, multiplied by an arbitrary constant:

$$f_{\mu,\nu}(|r_{ij}|, z_i, z_j) := \kappa_{z_i,z_j,\mu} C_\mu(|r_{ij}|) \hat{f}_{\mu,\nu}(|r_{ij}|) \tag{5}$$

$$\hat{f}_{\mu,\nu}(r) := \begin{cases} r^{-\nu-2+\mu}(R_{cut} - r)^2 & r < R_{cut} \\ 0 & r \geq R_{cut} \end{cases} \tag{6}$$

In 5, the term $C_\mu(r)$ is the μ-th Chebyschev polynomial, while the coefficient $\kappa_{z_i,z_j,\mu}$ is specified as the product of the atomic masses of atoms i and j:

$$\kappa_{z_i,z_j,\mu} = z_i z_j \quad \forall \mu \tag{7}$$

The radial function above guarantees that the moment tensors decay to 0 outside of the neighborhood, as is desirable, and that they do so in a smooth fashion.
Moving to the angular component, equation 4 shows the outer product $\underbrace{\boldsymbol{r}_{ij} \otimes ... \otimes \boldsymbol{r}_{ij}}_{\nu \text{ times}}$ of the distance vector between atoms i and j to be dependent from the ν parameter. This

function encodes the angular information between atom i and its neighbours, and determines the rank of the moment tensor through the value of ν.

As is evident, the choice of the maximum values ν_{max} and μ_{max} univocally fixes the complexity of the moments' calculations. It is possible to constrain the choice of the two parameters by introducing the *level* of the moments: $\text{lev}M_{\mu,\nu} := 8 + 2\mu + \nu$, so that by choosing lev_{max} both μ and ν have a fixed limit.

The last step in order to obtain the MT descriptors is to define a contraction of the moments, in the form of tensorial operations, to reduce all the M_{μ_i,ν_i} into scalar values. The level of an operation $OP(M_{\mu_1,\nu_1}, M_{\mu_2,\nu_2}, ..., M_{\mu_k,\nu_k})$ on the moments is defined as the sum of the moments' levels:

$$\begin{aligned} &\text{lev}OP(M_{\mu_1,\nu_1}, M_{\mu_2,\nu_2}, ..., M_{\mu_k,\nu_k}) = \\ &= \text{lev}M_{\mu_1,\nu_1} + \text{lev}M_{\mu_2,\nu_2} + ... + \text{lev}M_{\mu_k,\nu_k} = \\ &= 2(\mu_1 + \mu_2 + ... + \mu_k) + \nu_1 + \nu_2 + ... + \nu_k + 8k \end{aligned} \tag{8}$$

$OP(\cdot)$ being any kind of operation that reduces the rank ν of the tensors to 0 (scalar product for two moments that have $\nu = 0$, dot product "$\cdot$" for $\nu = 1$, Frobenius product of two matrices ":" for $\nu = 2$ ecc...). Of course, the larger the number k of tensors involved in $OP(\cdot)$, the more the possible sequences of operations to reduce the total rank to 0.
Finally, we define the basis set B_α as all the possible operations on the k MTs such that $\text{lev}B_\alpha \leq \text{lev}_{max}$. Let, for example, be $\text{lev}_{max} = 16$. Then, the resulting MT descriptor would be:

$$D(K_i) = [M_{0,0}, M_{1,0}, M_{2,0}, M_{3,0}, M_{0,0}^2, M_{4,0}] \tag{9}$$

and the components of the descriptor would have levels $[8, 10, 12, 14, 16, 16]$.

In the context of MLΔTB, a maximum level of $\text{lev}_{max} = 24$ and a restriction to the first 20 basis functions have been found ideal in terms of performances for training. We make one final remark about the set B_α. Defined as it is, the method does not differentiate for symmetric atomic compositions of a neighbourhood. Indeed, given the choice of the coefficients in 7, the descriptors of two neighbourhoods such as {Ga, As, As, As, As} and {As, Ga, Ga, Ga, Ga} would have the same values in a perfect crystal. To account for this, the definitive form of the descriptor used in MLΔTB includes the atomic number of the central atom in its first entry:

$$D(K_i) = [z_i, B_1(K_i), B_2(K_i), ..., B_{20}(K_i)] \tag{10}$$

B. Network and on-site correction

The feed-forward NN introduced in the start of this section constitutes the regressor of the MLΔTB framework, responsible for outputting the 4-orbital corrections to the ETB model. Recall that a feed-forward neural network can be seen as L subsequent applications of a non-linear function $f(\cdot)$ to the dot-product of a feature vector $\boldsymbol{x}$ and a weight matrix W (plus a bias $\boldsymbol{w}$). Using the MT descriptors as input, this would translate to:

$$\begin{aligned} &\boldsymbol{x}^{(0)} = D(K_i); \\ &\boldsymbol{x}^{(l)} = f(W^{(l)}\boldsymbol{x}^{(l-1)} + \boldsymbol{w}^{(l)}), \qquad l = 1, ..., L; \end{aligned} \tag{11}$$

Equation 11 implies that $\boldsymbol{x}$ is a column-vector. The index l is usually referred to as the *hidden layer* number, while the $n \times m$ dimensions of $W^{(l)}$ determine the number of *neurons* (or units) of layer l and $l-1$ respectively. As in most common applications, the function f is chosen to be the *Rectified Linear Unit* ($\text{ReLU}(x) = \max(0, x)$)[48]. The network structure is composed of $L = 3$ hidden layers, and an output layer with no activation function, to ensure a regression on the whole $\mathbb{R}$ set:

$$[\Delta s, \Delta p, \Delta d, \Delta s^*] = \boldsymbol{x}^{(4)} = W^{(4)}\boldsymbol{x}^{(3)} + \boldsymbol{w}^{(4)} \tag{12}$$

TABLE III. The structure of the MLΔTB network as also shown in figure 4-b. The NN has a total of 599 trainable weights.

Layer (l)	Units	Activation function	W+$\boldsymbol{w}$ shape
1	15	ReLU	(15,21)+(15,1)
2	10	ReLU	(10,15)+(10,1)
3	7	ReLU	(7,10)+(7,1)
4	4	Identity	(4,7) + (4,1)

The details of the network structure are summarized in table III. The feed-forward NN is demonstrated to be a universal approximator[49]. For this reason, and for the relatively few parameters that need to be learned with the architecture defined above, it is the ideal choice for the task of predicting the 4-dimensional vector of corrections.

An important step to perform before training the network is *weight initialization*. Indeed, the convergence of the loss function to its minimum is critically dependent on the initial state of $\{W^{(l)}, \boldsymbol{w}^{(l)},\ l = 1, ..., L+1\}$[50]. In this case, following the spirit of any Δ-learning algorithm, we assume that the output of the network needs to be small with respect to the corrected values. This means, that an appropriate initialization of the weights is one which initially outputs a vector of 4 zeroes. To obtain such an initial state, we performed a round of *pre-training*, in which dummy 4-dimensional, zero-valued targets were used in combination with an ADAM optimizer[51] (learning rate: 0.001) and a *mean squared error* (MSE) loss to force the network output to be $[0, 0, 0, 0]$, for all input descriptors in the training set. Given the cheap cost of this strategy (total runtime took $\simeq 2$ minutes), the pre-training stage consisted of 3000 epochs. This procedure can be interpreted as an instance of *physics-informed* learning[52]. Indeed, we use physical information from the system (the initial ETB parametrization and the knowledge that the corrections have to be small) to guide the learning procedure in a way that saves time and computational resources.

A few remarks should also be made regarding the correction scheme. The shell-resolved shift on the Hamiltonian diagonal is the simplest among all possible models designed to adjust an existing parametrization. As such, one may think it has a limitation on the performance, especially when compared to more complex methods that contemplate for example a orbital-resolved diagonal correction or an addition on the hopping integrals. Nonetheless, it can be argued that such a model can still incorporate the effects that we are trying to take into account, namely the strain effects due to internal relaxation and the local effects of alloying. Consider, in fact, that an orthonormal basis such as the one introduced in Section III is defined through the Löwdin procedure[53]. In this process, the overlap matrix S of a non-orthogonal basis $\{|\eta_j\rangle\,;\ j = 1, ..., n_{\text{basis}}\}$ is used to construct a set $\{|\phi_j\rangle\,;\ j = 1, ..., n_{\text{basis}}\}$ of orthonormal functions:

$$|\phi_j\rangle = \sum_{m=1}^{n_{\text{basis}}} \left[S^{-1/2}\right]_{m,j} |\eta_j\rangle\,; \qquad [S]_{m,n} = \langle\eta_m|\eta_n\rangle \tag{13}$$

When strain is introduced in the crystal, the overlap matrix elements are changed as a result of the displacement of the atoms. This change reflects of course in the hopping integrals, but less intuitively also on the on-site elements $\langle\phi_j|H|\phi_j\rangle$, having used S to construct the set $\{|\phi_j\rangle\}$[54]. Therefore, it is theoretically possible to account by some extent for the strain, through the addition of a machine-learned correction on the ionisation potentials alone. A more sophisticated treatment of strain, although not the main focus of this work, would require energy splitting between orbitals in the same subshell, or the inclusion of intra-atomic interactions, as well as defining additional fittable parameters[55,56].

As for the effects of alloying, it is more straightforward to argue that a correction that directly depends on the local atomic environment, can incorporate the effects that random fluctuations of Sb have on measurable quantities such as the band structure or the density of states (DOS). Indeed, effects like clustering of Sb atoms or simply different local Sb concentrations, are not included in the original ETB parametrizations, which treat the

parameters as related to the material rather than to the atom and its neighbourhood. In turn, these local effects reflect in the global quantities that are computed from the Hamiltonian.

A further advantage of our model is that it still is rotationally invariant. Indeed, since we chose to correct at the level of the Slater-Koster formalism, the shift of the diagonal elements of H^{TB} does not depend on the orientation of the supercell. This would not be true anymore when dealing with the inter-orbital hopping matrix elements or with the orbital-resolved diagonal ionization potentials, that by definition contain information about the angles between all couples of atoms. This apparently simple feature is actually very important, because it allows the trained network to be used without effort on new structures, without the need of readjusting the orientation. Moreover, the training set can in this way be extended in a quite simple fashion by just adding more structures in the data set.

Nonetheless, one might argue that a large part of the environmental effect captured by our method relates to charge transfer effects and could be described effectively in terms of *Self-Consistent field* (SCF) techniques. These techniques are based upon the definition of a loop, in which an extra term to the Hamiltonian is added, that depends on the solution of the ETB problem at the previous step[57–59]. The loop then continues until the new addition becomes smaller than a pre-defined threshold. Many of such techniques were indeed devised to tackle the investigation of electronic properties of alloys, as done for example in the work from *Goyhenex and Tréglia*[60]. Even though such techniques are indeed more accurate, and are much more physically justified than our approach, they still suffer from a critical problem, that is the need of fully diagonalizing the Hamiltonian at each step. Of course, the diagonalization process soon becomes unfeasible for large systems, so that a non-SCF method acquires value, even though it loses physical transparency.

C. Training and loss function details

The complete MLΔTB code can be found at: `https://github.com/DanSoccodato/mlDeltatb.git`. It is written using a combination of QuantumATK and *tensorflow* (v.2.12.0)[61], with customized classes and a custom training loop. The model was trained for 1500 epochs on a total of 12 structures, meaning the 2 bulk structures at $x = 0, 1$ and 2/10 configurations for all Sb concentrations reported in table I, randomly selected. The training set was then shuffled and divided into two batches of 6 configurations each. The remaining 40 structures were left for the evaluation of the model. This is unorthodox compared to classical ML algorithms, where the train/test split ratio usually favours the training set. The fact that the model can be trained on a small set is actually a notable feature of our framework, making it an example of a *few-shot* learning procedure[62]. To ensure a better convergence, the training set was normalized using the computed mean and variance across all descriptors populating it. These two values were then saved in order to normalize future test instances. The optimization on the model's weights was again carried out using ADAM, with a learning rate of 0.001.

In ordinary machine learning tasks, during the training loop the update on the model's weights is performed by directly comparing the output of the network with the target quantity. As has become clear by now, this is not the case for MLΔTB. In our framework, the individual (atomistic) outputs of the NN do not have a ground truth, because there are no known labels for what the atoms' corrections should be. Instead, the only possible ground truth is an aggregate quantity, the band structure, which depends on the N consecutive outputs of the network. This peculiarity leads to the need of defining a custom training loop within tensorflow, including the coding of the BS computation using the framework's built-in functions. This is needed in order to exploit the backpropagation algorithm and the built-in optimizers, since to compute the gradients tensorflow needs to keep track of all operations that occur between the evaluation of the loss function and the input of the network.

As already mentioned, the loss function L is computed on the eigenvalues of the k-dependent Hamiltonian:

$$H_k^{(ML\Delta TB)} \left|\psi_{n,k}\right\rangle^{(s)} = \varepsilon_{n,k}^{(s)} \left|\psi_{n,k}\right\rangle^{(s)} \tag{14}$$

where n is the band index and s refers to a generic structure in the training set. For the functional form of L, we opted for a structure-weighted MSE:

$$L[\{\varepsilon_{n,k}\}, \{\varepsilon_{n,k}^{(DFT)}\}] = \frac{1}{F} \sum_{s=1}^{N_{train}} \omega_s \sum_{n=1}^{n_b} \sum_{k=1}^{21} \left(\varepsilon_{n,k}^{(s)} - \varepsilon_{n,k}^{(DFT,s)}\right)^2 \tag{15}$$

where N_{train} is the number of structures in the training set, n_b is the number of considered bands, ω_s is the weight associated to structure s and factor $F = N_{train} \cdot n_b \cdot 21$ is the total number of eigenvalues. The value of ω_s was set to 1 for all alloys of $GaAs_{1-x}Sb_x$, and 2 for the bulks. This is to compensate for the presence of two structures for each Sb concentration, as opposed to the inclusion of just one structure for $x = 0, 1$.
To select n_b, more relevance was given to the bands closest to the Fermi level. Therefore, we defined:

$$\begin{aligned} E_{min} &= \text{VB} - 1.0 \text{ eV}, \\ E_{max} &= \text{CB} + 1.0 \text{ eV} \end{aligned} \tag{16}$$

and all eigenvalues below or above the range $[E_{min}, E_{max}]$, were clipped to have values E_{min}, E_{max} respectively. In 16, VB and CB are the valence and conduction band edges.

V. RESULTS AND DISCUSSION

The classical way of evaluating the performance of a machine learning model is to define a *metric function*, which can arbitrarily coincide with the loss function used to optimize the weights during training. We chose it to be the same of 15, but with $\omega_s = 1$ for all structures in the training and test set:

$$MSE[\{\varepsilon_{n,k}\}, \{\varepsilon_{n,k}^{(DFT)}\}] = \frac{1}{F} \sum_{s=1}^{N_{set}} \sum_{n=1}^{n_b} \sum_{k=1}^{21} \left(\varepsilon_{n,k}^{(s)} - \varepsilon_{n,k}^{(DFT,s)}\right)^2 \tag{17}$$

where $N_{set} = N_{train} = 12$ for the evaluation on the training set, while $N_{set} = N_{test} = 40$ for the evaluation on the test set. Figure 5 shows the mean squared error 17 as a function of the training epochs. The model clearly converged to a minimum, as is exemplified by the trend on the training set curve. At the same time, the comparable error on the test set (also taking into account that $N_{test} > N_{train}$) and the monotonically decreasing trend show that the model did not overfit the data. Table IV reports the final value of the MSE, as well as an additional goodness-of-fit performance indicator (the R^2-score[63]), for the individual Sb concentrations and for the aggregated data sets. The value of R^2 very close to 1 quantitatively confirms that the network correctly managed to generalize the atomistic corrections to unseen structures. It is interesting to notice how the model seems to work better on the Sb fractions closest to the two bulk GaAs and GaSb extremes. Indeed, the performance of the framework on the test set has the highest error at $x = 0.4,\ 0.6$. This is also visible in figure 6, where the predicted eigenvalues are plotted against the *ab-initio* ones; and where indeed the two central concentrations exhibit the largest spread from the linear trend. Nonetheless, the results on the test band structures mark a great improvement over the original tight-binding parametrization. As a proof of this, in figure 7 we reported the band structures of the same configurations plotted in figure 3. The four structures all belong to the test set, so they were not seen by the network during training. Using the original ETB method and parameters, both the valence and conduction bands in the range $[E_{min}, E_{max}]$ were severely off, with three out of the four alloys even turning out to be metallic. The enhancement of these band structures using the MLΔTB correction is

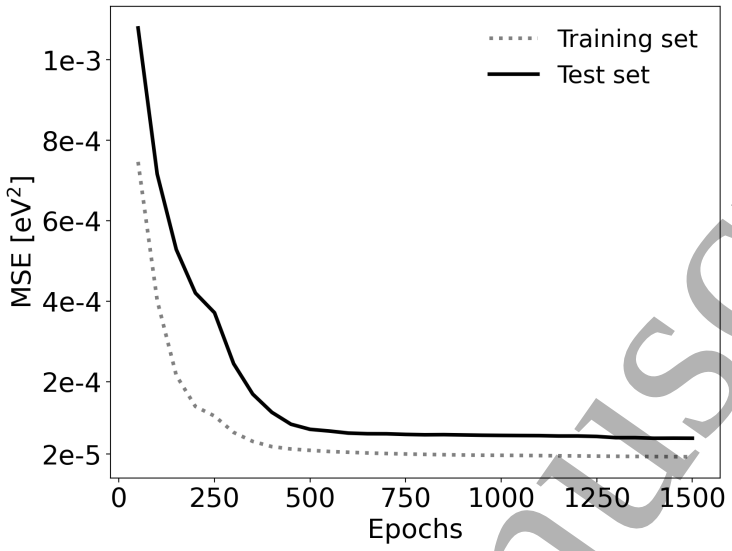

FIG. 5. The MSE metric as a function of the number of training epochs. The evaluation was performed on respectively 12 training and 40 test configurations.

TABLE IV. The evaluation of the MLΔTB framework on the training and test set. The metrics used to measure the performances are the MSE and the R^2 goodness-of-fit score.

	Sb fraction	MSE [eV^2]	R^2	N_{set}
Training set	0.0	1.575e-5	0.999983	1
	0.1	1.460e-5	0.999982	2
	0.3	2.840e-5	0.999956	2
	0.4	1.405e-5	0.999973	2
	0.6	3.541e-5	0.999929	2
	0.8	3.180e-5	0.999935	2
	1.0	1.438e-5	0.999973	1
Total		2.322e-5	0.999962	12
Test set	0.1	2.608e-5	0.999967	8
	0.3	8.023e-5	0.999870	8
	0.4	1.233e-4	0.999788	8
	0.6	1.207e-4	0.999757	8
	0.8	6.230e-5	0.999872	8
Total		8.255e-5	0.999862	40

evident, and it is clearly visible also for $x = 0.4,\ 0.6$, where the metrics show the lowest performance. Given the large number of test instances, the results on the whole set are reported in the Supplementary Information.

Some other works in the literature applied a machine learning technique in the context of empirical tight-binding. For example, *Schattauer et al.*[17] performed an inverse mapping of the BS to the tight-binding Hamiltonian for structures with defects, starting from Hamiltonians of the same defect-less structures. In[19], a Graph Convolutional Network (GCN) was used to learn the best local descriptor for graphene nano-ribbons, and used in combination with two feed-forward NNs for predicting on-site- and off-site matrix elements. Yet another work from *Nakhaee et al.*[18] showed the feasibility of performing linear regression of TB parameters after generating a training set of uniaxially strained systems. All these papers have some characteristics in common: they are all restricted to a quite small basis set, that goes up to the p-orbital, and the methods are applied to small (mainly two-dimensional)

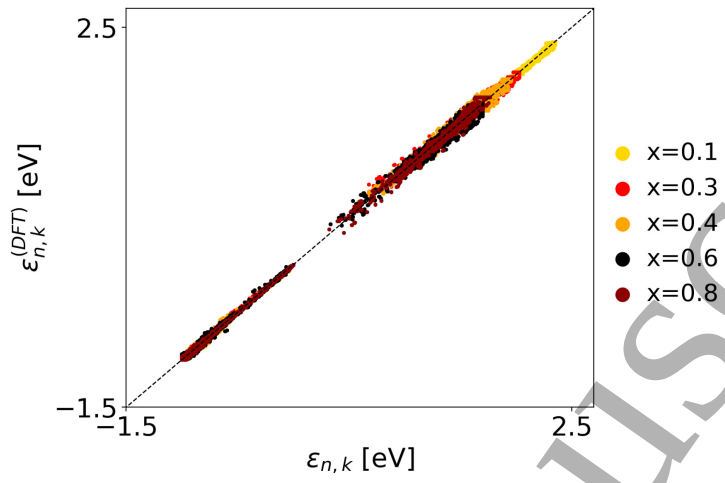

FIG. 6. The predicted eigenvalues plotted against the *ab-initio* references, for all $GaAs_{1-x}Sb_x$ systems in the test set.

systems. To our knowledge, the one presented in this manuscript is the first ML framework that is able to exploit a $sp^3d^5s^*$ tight-binding basis for simulating fully 3D bulk supercells. This comes of course at the price of relying on an already existing parametrization, for which our model represents a very useful correction.

It may also come natural to make a connection between the performances of this method with the SCF techniques cited in section IV B. Indeed, one could compare the output corrections of the MLΔTB network with the local electrostatic potential computed after a SCF loop, and look for similar trends that could give insight on the nature of the corrections, and on their connection with a local charge transfer between atoms. We performed such an analysis, whose details are reported in the Supplementary Information. To give a summary, we found that there is a correlation between the MLΔTB output and the electrostatic potential computed on the inequivalent sites, especially for low concentrations of antimony. This correlation decreases for the structures with higher values of x, which could be a symptom of the model learning to give a unphysical correction to some of the input atomistic environments. We did not proceed to further investigate the issue, as the comparison with SCF techniques was beyond the scope of our work. However, we highlight again the value of our method lying in the possibility to investigate the properties of bigger structures, than what is possible to simulate with SCF approaches.

In the remainder of this section we present some more evidence of the good generalization capabilities of the MLΔTB framework. So far, the model's performance has simply been evaluated on the same configurations and BSs populating the data set, only unseen during training. But there are other ways to ensure that the environment-dependent corrections are physically meaningful and do not represent a case of overfitting. Figure 8 shows the BS computation of a $GaAs_{0.9}Sb_{0.1}$ alloy, with a domain on energy and k-points that goes beyond the one used for training. More specifically, whereas the model was trained on the $L-\Gamma-X$ route and within the $[E_{min}, E_{max}]$ energy range, the figure shows the prediction of the band structure when such window in the k-E space is extended. The result shows good agreement with the *ab-initio* simulation, especially for the valence bands, where the error appears to be very small. This result is evidence of a good generalization performance not only at the level of unseen structures, but also in terms of prediction on unseen k-E values.

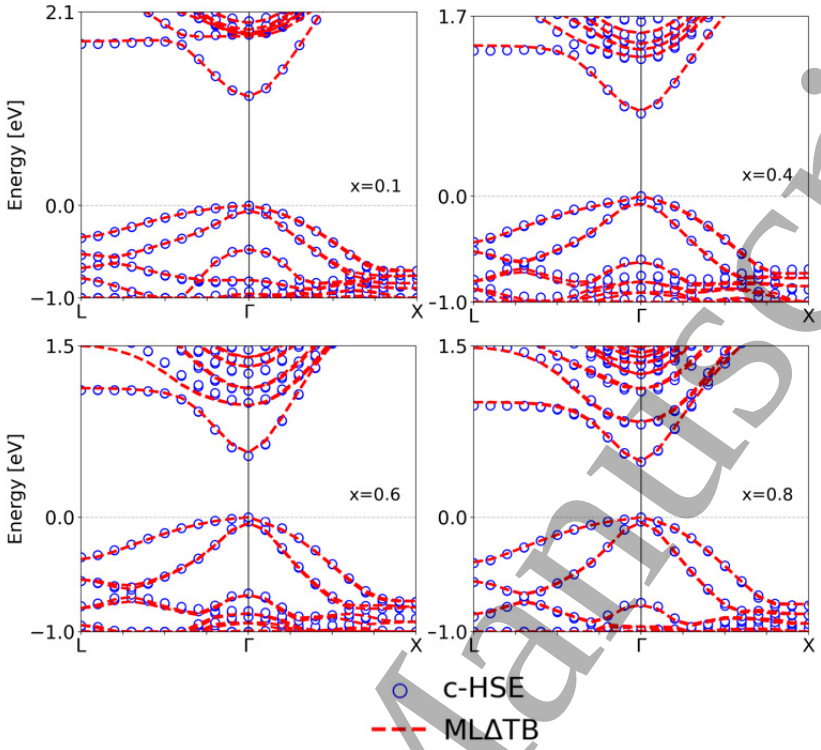

FIG. 7. The MLΔTB-computed band structure of four instances of $GaAs_{1-x}Sb_x$, for $x = 0.1, 0.4, 0.6, 0.8$, compared with the *ab-initio* reference. See figure 3 for the corresponding ETB simulations.

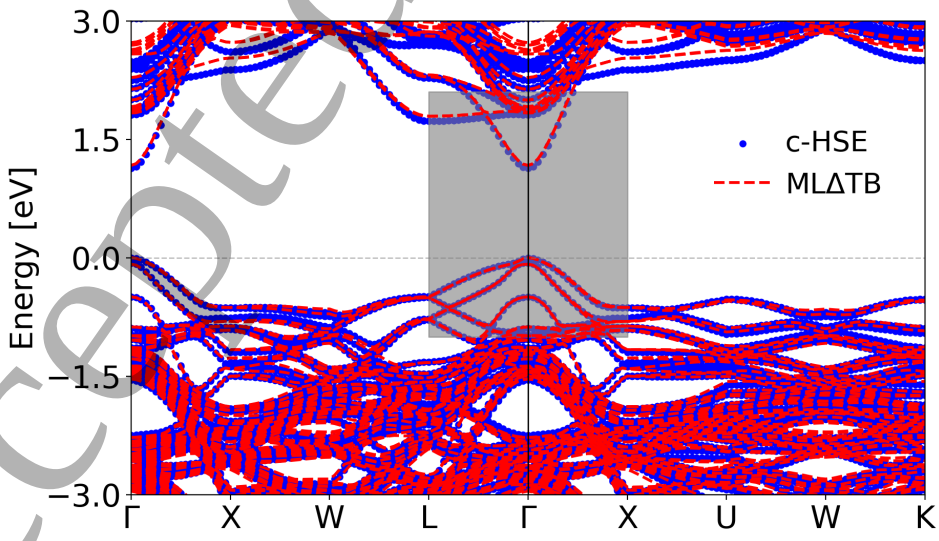

FIG. 8. The application of the MLΔTB correction for the BS computation of a $GaAs_{0.9}Sb_{0.1}$ alloy. The grey shaded area represents the k-E domain on which the network was trained.

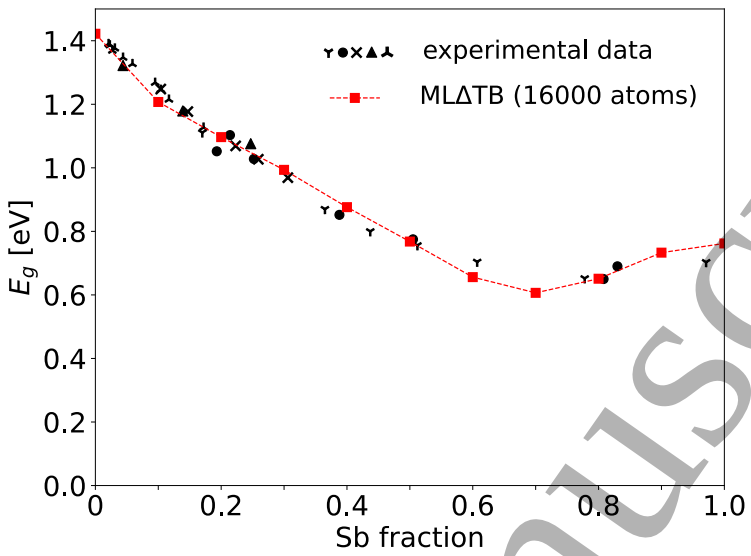

FIG. 9. Compositional bowing of the direct band gap for the $GaAs_{1-x}Sb_x$ 16000-atom supercells, simulated using the MLΔTB framework, for $x = 0.0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9,\ 1.0$. Results are shown together with experimental measurements from[32–36] (different shapes correspond to different data sets). To be compared with figure 1.

Another test that can be performed concerns the compositional bowing of the band gap. Recall from figure 1 that the custom HSE functional, which was used to generate the data set, has an issue in reproducing the bowing parameter. Indeed the band gaps for all Sb fractions are underestimated, and we postulated that this is the result of the small size of the supercell, which contains too few atoms to correctly model random alloys. Truly, in order to approach the statistical limit for which the physical properties of the alloys are correctly reproduced, it is necessary to either generate many configurations (to the order of 10^6) or large supercells (containing thousands of atoms)[64]. Both these conditions are quite difficult to be satisfied using a first-principle method. Instead, a non-self consistent ETB calculation can reach a cell size of millions of atoms. For this reason, it comes natural to test the MLΔTB framework on large $GaAs_{1-x}Sb_x$ supercells and check the resulting band gaps. Figure 9 shows the band gaps computed on 11 supercells composed of 16000 atoms ($20 \times 20 \times 20$ repetitions of a zincblende primitive unit cell), one every 10% of antimony fraction. The MLΔTB-computed band gaps definitely show a better agreement than the ones simulated using the smaller supercells with DFT. The bowing parameter for the curve shown in the figure is $B = 1.35$. Recall that the value computed with the *ab-initio* computations is $B = 2.18$, and that the parameter is expected to be in the range $1.0 - 1.44$. This is yet another point for the validation of the method, because this physical property was in no way present in the training set, but was instead predicted by our model.
This result recovers the SQS values showed in figure 1, but with considerably less computational burden. Indeed, simulating a 256-atom supercell using a HSE functional took more than 17 hours on a HPC node with 40 processes, as opposed to ~ 45 minutes using 30 processes for one of the red points in figure 9.

Finally, one last quantity that can be analyzed is the valence band offset between the bulk GaAs and GaSb materials. The initial re-fitted ETB parametrization is completely agnostic about the VBO between the two bulks, and this is one of the reasons for which it performs quite well on the two individual band structures, but does not fit the alloys BS. To account for this, the offset is routinely included as a parameter to fit. Instead of adding one more parameter, and given the good results on the band structure fitting of MLΔTB, one can wonder if the method implicitly learned the VBO during training. To

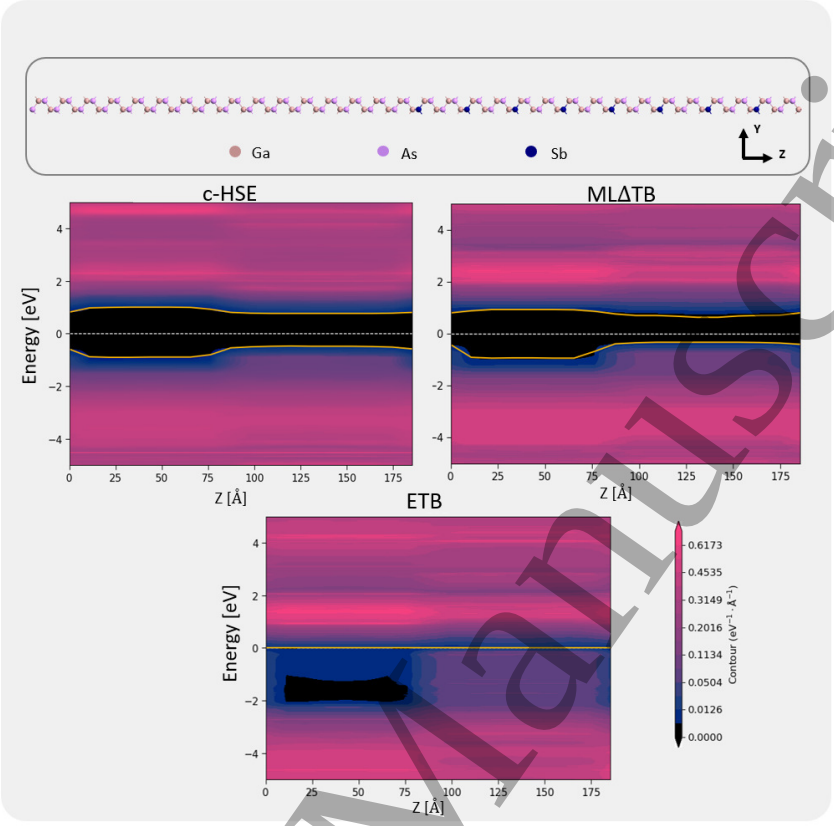

FIG. 10. The local density of states of a GaAs/$GaAs_{0.75}Sb_{0.25}$ interface (128 atoms) along the Z-direction. The yellow lines mark the conduction and valence band edges.

check this, we performed a computation of the local density of states (LDOS) on an ideal 128-atom GaAs/$GaAs_{0.75}Sb_{0.25}$ interface. The system was relaxed using again a Abell-Tersoff empirical potential, with a constraint on the X and Y directions. The comparison of the DFT, MLΔTB and ETB methods is reported in figure 10. As is evident, the ETB method fails in reproducing the *ab-initio* profile of the LDOS, whereas the c-HSE and MLΔTB contour plots exhibit quite similar values and band edges. For comparison, the DFT calculation took $\sim$ 17 hours on a HPC node with 20 processes, while the ML-corrected TB simulation could be run on a laptop with 4 processes in about 24 minutes.

VI. CONCLUSION

To summarize, a novel Δ-machine learning framework was introduced, for the task of correcting the empirical tight-binding on-site parameters of a $sp^3d^5s^*$ basis, applied to $GaAs_{1-x}Sb_x$. After training, the model is able to generalize the mapping from a local atomistic environment to a orbital-resolved correction and exhibits good transferability properties. All tests show good agreement either with the *ab-initio* references, or with experimental data. Notable features of our model, that also differentiate it from other similar works, include the possibility of training on few instances (few-shot learning), the application on a larger basis than previously ever done, the application on bulk 3D structures, and its rotational invariance, which allows for easy employment of the trained network on new instances, as well as an effortless possible expansion of the training set to new alloys. We also highlight how the nature of the scheme, which is designed to act at the atomic level,

allows to cheaply apply the trained network to configurations of arbitrary size. Finally, we stress how the application of the framework to large structures allowed for the computation of band gaps that agree substantially better with the experimental data, as opposed to the very same *ab-initio* training set, which was affected by the limited supercell size.

ACKNOWLEDGMENTS

This work has been supported by the European Union under the PON Ricerca e Innovazione - FSE REACT-EU9 and H2020 Marie Skłodowska-Curie grant n° 956548.

We also wish to thank Søren Smidstrup, Troels Markussen and Julian Schneider (Synopsys-QuantumATK) for useful insight and for clarifications about Moment Tensors.

REFERENCES

[1] J. C. Slater and G. F. Koster. Simplified lcao method for the periodic potential problem. *Phys. Rev.*, 94:1498–1524, Jun 1954.

[2] A. Di Vito, A. Pecchia, A. Di Carlo, and M. Auf der Maur. Impact of compositional nonuniformity in (In, Ga)N-based light-emitting diodes. *Phys. Rev. Appl.*, 12:014055, Jul 2019.

[3] Robert Finn and Stefan Schulz. Impact of random alloy fluctuations on the electronic and optical properties of (Al,Ga)N quantum wells: Insights from tight-binding calculations. *The Journal of Chemical Physics*, 157(24):244705, 12 2022.

[4] Aldo Di Carlo. Microscopic theory of nanostructured semiconductor devices: beyond the envelope-function approximation. *Semiconductor Science and Technology*, 18(1):R1, dec 2002.

[5] P. Vogl, Harold P. Hjalmarson, and John D. Dow. A semi-empirical tight-binding theory of the electronic structure of semiconductors†. *Journal of Physics and Chemistry of Solids*, 44(5):365–378, 1983.

[6] Gerhard Klimeck, R.Chris Bowen, Timothy B Boykin, Carlos Salazar-Lazaro, Thomas A Cwik, and Adrian Stoica. Si tight-binding parameters from genetic algorithm fitting. *Superlattices and Microstructures*, 27(2):77–88, 2000.

[7] J. N. Schulman and Yia-Chung Chang. Band mixing in semiconductor superlattices. *Phys. Rev. B*, 31:2056–2068, Feb 1985.

[8] Jean-Marc Jancu, Reinhard Scholz, Fabio Beltram, and Franco Bassani. Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters. *Phys. Rev. B*, 57:6493–6507, Mar 1998.

[9] J.-M. Jancu, F. Bassani, F. Della Sala, and R. Scholz. Transferable tight-binding parametrization for the group-III nitrides. *Applied Physics Letters*, 81(25):4838–4840, 12 2002.

[10] Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld. Big data meets quantum chemistry approximations: The δ-machine learning approach. *Journal of Chemical Theory and Computation*, 11(5):2087–2096, May 2015.

[11] Kenneth Atz, Clemens Isert, Markus N. A. Böcker, José Jiménez-Luna, and Gisbert Schneider. δ-quantum machine-learning for medicinal chemistry. *Phys. Chem. Chem. Phys.*, 24:10775–10783, 2022.

[12] Jun Otsuka, Takashi Kato, Hirofumi Sakakibara, and Takao Kotani. Band structures for short-period (inas)n(gasb)n superlattices calculated by the quasiparticle self-consistent gw method. *Japanese Journal of Applied Physics*, 56(2):021201, jan 2017.

[13] Messaoud Caid, D. Rached, Oualid Cheref, Righi Haroun, Habib RACHED, S. Benalia, Mostefa MERABET, and Lakhdar Djoudi. Full potential study of the structural, electronic and optical properties of (inas)m/(gasb)n superlattices. *Computational Condensed Matter*, 21:e00394, 04 2019.

[14] Brian R. Bennett, Richard Magno, J. Brad Boos, Walter Kruppa, and Mario G. Ancona. Antimonide-based compound semiconductors for electronic devices: A review. *Solid-State Electronics*, 49(12):1875–1895, 2005.

[15] Hossein Anabestani, Rassel Shazzad, Md Fahim Al Fattah, Joel Therrien, and Dayan Ban. Review on gaassb nanowire potentials for future 1d heterostructures: Properties and applications. *Materials Today Communications*, 28:102542, 2021.

[16] Liang Ma, Xuehong Zhang, Honglai Li, Huang Tan, Yankun Yang, Yadan Xu, Wei Hu, Xiaoli Zhu, Xiujuan Zhuang, and Anlian Pan. Bandgap-engineered gaassb alloy nanowires for near-infrared photodetection at 1.31 μm. *Semiconductor Science and Technology*, 30(10):105033, sep 2015.

[17] Christoph Schattauer, Milica Todorović, Kunal Ghosh, Patrick Rinke, and Florian Libisch. Machine learning sparse tight-binding parameters for defects. *npj Computational Materials*, 8(1):116, May 2022.

[18] M. Nakhaee, S. A. Ketabi, and F. M. Peeters. Machine learning approach to constructing tight binding models for solids with application to BiTeCl. *Journal of Applied Physics*, 128(21):215107, 12 2020.

[19]Zifeng Wang, Shizhuo Ye, Hao Wang, Qijun Huang, Jin He, and Sheng Chang. Graph representation-based machine learning framework for predicting electronic band structures of quantum-confined nanostructures. *Science China Materials*, 65(11):3157–3170, Nov 2022.

[20]Zifeng Wang, Shizhuo Ye, Hao Wang, Jin He, Qijun Huang, and Sheng Chang. Machine learning method for tight-binding hamiltonian parameterization from ab-initio band structure. *npj Computational Materials*, 7(1):11, Jan 2021.

[21]Søren Smidstrup, Troels Markussen, Pieter Vancraeyveld, Jess Wellendorff, Julian Schneider, Tue Gunst, Brecht Verstichel, Daniele Stradi, Petr A Khomyakov, Ulrik G Vej-Hansen, Maeng-Eun Lee, Samuel T Chill, Filip Rasmussen, Gabriele Penazzi, Fabiano Corsetti, Ari Ojanperä, Kristian Jensen, Mattias L N Palsgaard, Umberto Martinez, Anders Blom, Mads Brandbyge, and Kurt Stokbro. Quantumatk: an integrated platform of electronic and atomic-scale modelling tools. *Journal of Physics: Condensed Matter*, 32(1):015901, oct 2019.

[22]D. Powell, M. A. Migliorato, and A. G. Cullis. Optimized tersoff potential parameters for tetrahedrally bonded iii-v semiconductors. *Phys. Rev. B*, 75:115202, Mar 2007.

[23]Dong C. Liu and Jorge Nocedal. On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1):503–528, Aug 1989.

[24]Søren Smidstrup, Daniele Stradi, Jess Wellendorff, Petr A Khomyakov, Ulrik G Vej-Hansen, Maeng-Eun Lee, Tushar Ghosh, Elvar Jónsson, Hannes Jónsson, and Kurt Stokbro. First-principles green's-function method for surface calculations: A pseudopotential localized basis set approach. *Physical Review B*, 96(19):195309, 2017.

[25]Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics*, 118(18):8207–8215, 04 2003.

[26]Otfried Madelung. *Semiconductors: Data Handbook*. Springer Berlin, Heidelberg, second edition, 2001.

[27]I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III–V compound semiconductors and their alloys. *Journal of Applied Physics*, 89(11):5815–5875, 06 2001.

[28]Alexandros Kyrtsos, Masahiko Matsubara, and Enrico Bellotti. First-principles study of the impact of the atomic configuration on the electronic properties of $al_xga_{1-x}N$ alloys. *Phys. Rev. B*, 99:035201, Jan 2019.

[29]Alexandros Kyrtsos, Masahiko Matsubara, and Enrico Bellotti. Band offsets of alxga1xn alloys using first-principles calculations. *Journal of Physics: Condensed Matter*, 32(36):365504, jun 2020.

[30]Alexandros Kyrtsos, Masahiko Matsubara, and Enrico Bellotti. Investigation of the band gaps and bowing parameter of $inas_{1-x}sb_x$ alloys using the modified becke-johnson potential. *Phys. Rev. Mater.*, 4:014603, Jan 2020.

[31]M.J. van Setten, M. Giantomassi, E. Bousquet, M.J. Verstraete, D.R. Hamann, X. Gonze, and G.-M. Rignanese. The pseudodojo: Training and grading a 85 element optimized norm-conserving pseudopotential table. *Computer Physics Communications*, 226:39–54, 2018.

[32]G. A. Antypas and L. W. James. Liquid Epitaxial Growth of GaAsSb and Its Use as a High-Efficiency, Long-Wavelength Threshold Photoemitter. *Journal of Applied Physics*, 41(5):2165–2171, 11 2003.

[33]R. E. Nahory, M. A. Pollack, J. C. DeWinter, and K. M. Williams. Growth and properties of liquid-phase epitaxial GaAs1-xSbx. *Journal of Applied Physics*, 48(4):1607–1614, 08 2008.

[34]H. Sakaki, L. L. Chang, R. Ludeke, Chin-An Chang, G. A. Sai-Halasz, and L. Esaki. In1-xGaxAs-GaSb1-yAsy heterojunctions by molecular beam epitaxy. *Applied Physics Letters*, 31(3):211–213, 08 2008.

[35]T.S. Wang, J.T. Tsai, K.I. Lin, J.S. Hwang, H.H. Lin, and L.C. Chou. Characterization of band gap in gaassb/gaas heterojunction and band alignment in gaassb/gaas multiple quantum wells. *Materials Science and Engineering: B*, 147(2):131–135, 2008. E-MRS 2007, Symposium B: Semiconductor Nanostructures towards Electronic and Optoelectronic Device Applications.

[36]Mitsuaki Yano, Yukio Suzuki, Tetsuo Ishii, Yuichi Matsushima, and Morihiko Kimata. Molecular beam epitaxy of gasb and gasbxas1-x. *Japanese Journal of Applied Physics*, 17(12):2091, dec 1978.

[37]A. van de Walle, P. Tiwary, M. de Jong, D.L. Olmsted, M. Asta, A. Dick, D. Shin, Y. Wang, L.-Q. Chen, and Z.-K. Liu. Efficient stochastic generation of special quasirandom structures. *Calphad*, 42:13–18, 2013.

[38]Petr A. Khomyakov, Mathieu Luisier, and Andreas Schenk. Compositional bowing of band energies and their deformation potentials in strained InGaAs ternary alloys: A first-principles study. *Applied Physics Letters*, 107(6):062104, 08 2015.

[39]Yaohua Tan, Michael Povolotskyi, Tillmann Kubis, Timothy B. Boykin, and Gerhard Klimeck. Transferable tight-binding model for strained group iv and iii-v materials and heterostructures. *Phys. Rev. B*, 94:045311, Jul 2016.

[40]W. A. Harrison. *Elementary Electronic Structure*. World Scientific, 1999.

[41]Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.

[42]Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Phys. Rev. B*, 87:184115, May 2013.

[43]Sandip De, Albert P. Bartók, Gábor Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Phys. Chem. Chem. Phys.*, 18:13754–13769, 2016.

[44]James Barker, Johannes Bulin, Jan Hamaekers, and Sonja Mathias. *LC-GAP: Localized Coulomb Descriptors for the Gaussian Approximation Potential*, pages 25–42. Springer International Publishing, Cham, 2017.

[45]Patrick Reiser, Marlen Neubert, André Eberhard, Luca Torresi, Chen Zhou, Chen Shao, Houssam Metni, Clint van Hoesel, Henrik Schopmans, Timo Sommer, and Pascal Friederich. Graph neural networks for materials science and chemistry. *Communications Materials*, 3(1):93, Nov 2022.

[46]Alexander V. Shapeev. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation*, 14(3):1153–1173, 2016.

[47]Ivan S Novikov, Konstantin Gubaev, Evgeny V Podryabinkin, and Alexander V Shapeev. The mlip package: moment tensor potentials with mpi and active learning. *Machine Learning: Science and Technology*, 2(2):025002, dec 2020.

[48]Vinod Nair and Geoffrey E. Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th International Conference on International Conference on Machine Learning*, ICML'10, page 807–814, Madison, WI, USA, 2010. Omnipress.

[49]Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006.

[50]Meenal V. Narkhede, Prashant P. Bartakke, and Mukul S. Sutaone. A review on weight initialization strategies for neural networks. *Artificial Intelligence Review*, 55(1):291–322, Jan 2022.

[51]Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[52]George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, Jun 2021.

[53]Per-Olov Lowdin. On the Non-Orthogonality Problem Connected with the Use of Atomic Wave Functions in the Theory of Molecules and Crystals. *The Journal of Chemical Physics*, 18(3):365–375, 12 2004.

[54]Gerhard Klimeck and Timothy Boykin. *Tight-Binding Models, Their Applications to Device Modeling, and Deployment to a Global Community*, pages 1601–1640. Springer International Publishing, Cham, 2023.

[55]Timothy B. Boykin, Gerhard Klimeck, R. Chris Bowen, and Fabiano Oyafuso. Diagonal parameter shifts due to nearest-neighbor displacements in empirical tight-binding theory. *Phys. Rev. B*, 66:125207, Sep 2002.

[56]Y. M. Niquet, D. Rideau, C. Tavernier, H. Jaouen, and X. Blase. Onsite matrix elements of the tight-binding hamiltonian of a strained crystal: Application to silicon, germanium, and their alloys. *Phys. Rev. B*, 79:245201, Jun 2009.

[57]M-C Desjonqueres and Daniel Spanjaard. *Concepts in surface physics*. Springer Science & Business Media, 2012.

[58]M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, Th. Frauenheim, S. Suhai, and G. Seifert. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys. Rev. B*, 58:7260–7268, Sep 1998.

[59]Kurt Stokbro, Dan Erik Petersen, Søren Smidstrup, Anders Blom, Mads Ipsen, and Kristen Kaasbjerg. Semiempirical model for nanoscale device simulations. *Phys. Rev. B*, 82:075420, Aug 2010.

[60]C. Goyhenex and G. Tréglia. Unified picture of d-band and core-level shifts in transition metal alloys. *Phys. Rev. B*, 83:075101, Feb 2011.

[61]Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[62]Archit Parnami and Minwoo Lee. Learning from few examples: A summary of approaches to few-shot learning, 2022.

[63]Davide Chicco, Matthijs J Warrens, and Giuseppe Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7:e623, 2021.

[64]S.-H. Wei, L. G. Ferreira, James E. Bernard, and Alex Zunger. Electronic properties of random alloys: Special quasirandom structures. *Phys. Rev. B*, 42:9622–9649, Nov 1990.