GEARS H: Accurate machine-learned Hamiltonians for next-generation device-scale modeling

Anubhab Haldar*, Ali K. Hamze*, Nikhil Sivadas, and Yongwoo Shin†

Advanced Materials Lab
Samsung Advanced Institute of Technology-America
Samsung Semiconductor Inc.
Cambridge, Massachusetts 02138, USA

June 13, 2025

Abstract

We introduce GEARS H, a state-of-the-art machine-learning Hamiltonian framework for large-scale electronic structure simulations. Using GEARS H, we present a statistical analysis of the hole concentration induced in defective WSe_2 interfaced with Ni-doped amorphous HfO_2 as a function of the Ni doping rate, system density, and Se vacancy rate in 72 systems ranging from 3326 to 4160 atoms—a quantity and scale of interface electronic structure calculation beyond the reach of conventional density functional theory codes and other machine-learning-based methods. We further demonstrate the versatility of our architecture by training models for a molecular system, 2D materials with and without defects, solid solution crystals, and bulk amorphous systems with covalent and ionic bonds. The mean absolute error of the inferred Hamiltonian matrix elements from the validation set is below $2.4\,\mathrm{meV}$ for all of these models. GEARS H outperforms other proposed machine-learning Hamiltonian frameworks, and our results indicate that machine-learning Hamiltonian methods, starting with GEARS H, are now production-ready techniques for DFT-accuracy device-scale simulation.

1 Introduction

Density functional theory (DFT) has proven to be the most widely applied computational technique in condensed matter physics. Indeed, two foundational DFT papers rank among the top 10 most-cited papers of all time [1]. The applications of DFT, however, have been limited to relatively small systems due to the high computational cost of calculations. Systems with $\mathcal{O}(10^0 - 10^1)$ atoms are readily accessible, while systems with $\mathcal{O}(10^2)$ atoms require researchers to consider whether they are necessary. Only in recent years, with the advent of GPUs and of more powerful CPUs have system with low-$\mathcal{O}(10^3)$ atoms become possible, but such calculations are rarely done due to their exorbitant cost.

Meanwhile, progress in semiconductor device manufacturing is becoming increasingly difficult due to material and process constraints. While ever increasing transistor densities were once taken for granted, now, other solutions must be sought. These include new materials like 2D transition metal dichalcogenides (TMDs) as channel materials and new device geometries like monolithic 3D integrated circuits [2, 3, 4]. Exploration of new materials and fabrication of devices with novel transistor geometries, however, requires large upfront investment. This presents an opportunity for new, low-cost computational methods to lead industry forward. Such new methods, ideally, will not sacrifice the accuracy of DFT in pursuit of device-scale simulation.

*These authors contributed equally
†email: yongwoo.s@samsung.com

In this work, we present GEARS Hamiltonian (GEARS H), our framework for machine learning Hamiltonians (MLH) in a linear combination of atomic orbitals (LCAO) basis. GEARS H is the first MLH framework to enable models of realistic, device-scale systems that are beyond the reach of traditional DFT, demonstrating the true strength of MLH methods. We show this by training a model on a combined system of Ni-doped amorphous $\mathrm{HfO_2}$ interfaced to $\mathrm{WSe_2}$. This system was recently proposed [5] for modulation doping of $\mathrm{WSe_2}$. We use the model to perform a statistical study of the hole concentration induced in the $\mathrm{WSe_2}$ layer in device-scale systems (3326 to 4160 atoms) as a function of Ni doping rate, Se vacancy rate, and system density.

We further demonstrate the broad applicability of GEARS H by applying it to 1) Lithium Bis(trifluoromethanesulfonyl)imide (LiTFSI), a molecular system with 6 elements, 2) 2D $\mathrm{WSe}_{2-x}\,(0.0 \leq x < 0.07)$, 3) a dataset of 9 different 2D materials featuring 8 distinct atomic species, 4) $\mathrm{Ag}_x\,\mathrm{Au}_{1-x}\,(0.34 < x < 0.72)$, a metal alloy, 5) amorphous SiO_2 (a-SiO_2), a covalent solid, and 5) amorphous HfO_2 (a-HfO_2), a mixed ionic-covalent solid.

Our results suggest that, with the advancements presented in our framework, MLH models are now production-ready tools for next-generation device modeling. There has been a dramatic acceleration in the search for new crystalline structures through the successful development and deployment of machine-learning-based interatomic potentials (MLIPs) [6, 7, 8]. We hope that GEARS H leads to a similar phenomenon in the field of electronic structure.

GEARS H builds on previous work towards MLHs. The earliest attempts include those by Hegde and Bowen [9] and Schutt *et al.* [10]. Advances were made by Li and colleagues [11] and the related work by Gong *et al* [12]. Unke and colleagues [13] have demonstrated highly accurate learning of molecular Hamiltonians and provide mathematical details for the construction of such models. Nigam and colleagues [14] demonstrate linear models of molecular Hamiltonians with rigorous mathematical analysis. Several `e3nn`-based [15] models have also been developed including DeepH-E3, [12], DeePTB [16], QHNet [17], and HamGNN [18]. To the best of our knowledge, GEARS H has the fewest number of parameters of any MLH model reported in the literature, and is the only model that has been successfully applied to amorphous systems.

The Hamiltonian architecture of GEARS H is inspired by architectural decisions in PhiSNet, ACEhamiltonians, and DeepH-E3. We present the ideas underlying the GEARS Hamiltonian, provide a user- and performance-focused implementation of our model using E3x [19], and provide an interface to GPAW [20](chosen because it is open-source, written in Python, easy to install, and allows for low-cost training data generation using strictly confined numerical atomic orbitals and projector-augmented waves to describe core electrons [21]). Interfaces to other LCAO codes are possible and we welcome community contributions to implement data conversion to the format required for GEARS H. Our work is a part of a greater ongoing effort which we call GEARS (**G**iant-scale **E**lectronic structure and **A**tomic configuration **R**esearch **S**olution) that will be further detailed in subsequent publications.

2 Results and Discussion

2.1 Model architecture

An overview of the GEARS H architecture is shown in Fig. 1(a). Here, we briefly describe the model inputs and outputs and present the details of three pieces of the model architecture: the Atom Centered descriptor, the Bond Centered descriptor, and the Scale Shift layers. Detailed layer architecture diagrams, descriptions of the other layers [22], as well as additional discussion of the layers described here, can be found in the Supplement.

The input data consists of an array of atomic numbers Z_i, sparse neighbor lists, and the corresponding pairwise Cartesian vectors. The training and output data consists of two arrays of Hamiltonian irreducible representations (irreps) in direct-sum form corresponding to atom-centered and atom-pair interaction Hamiltonian blocks similar to the approach used in [13].

2.1.1 The Atom-centered descriptor

Atom-centered descriptors have been extensively studied in the context of MLIPs; we refer the reader to the review by Musil *et al.* [23] on their design choices. The inputs to the atom-centered descriptor in GEARS H are expansions of the local neighborhoods of atoms using a 2-body (2B) basis consisting of radial and angular functions.

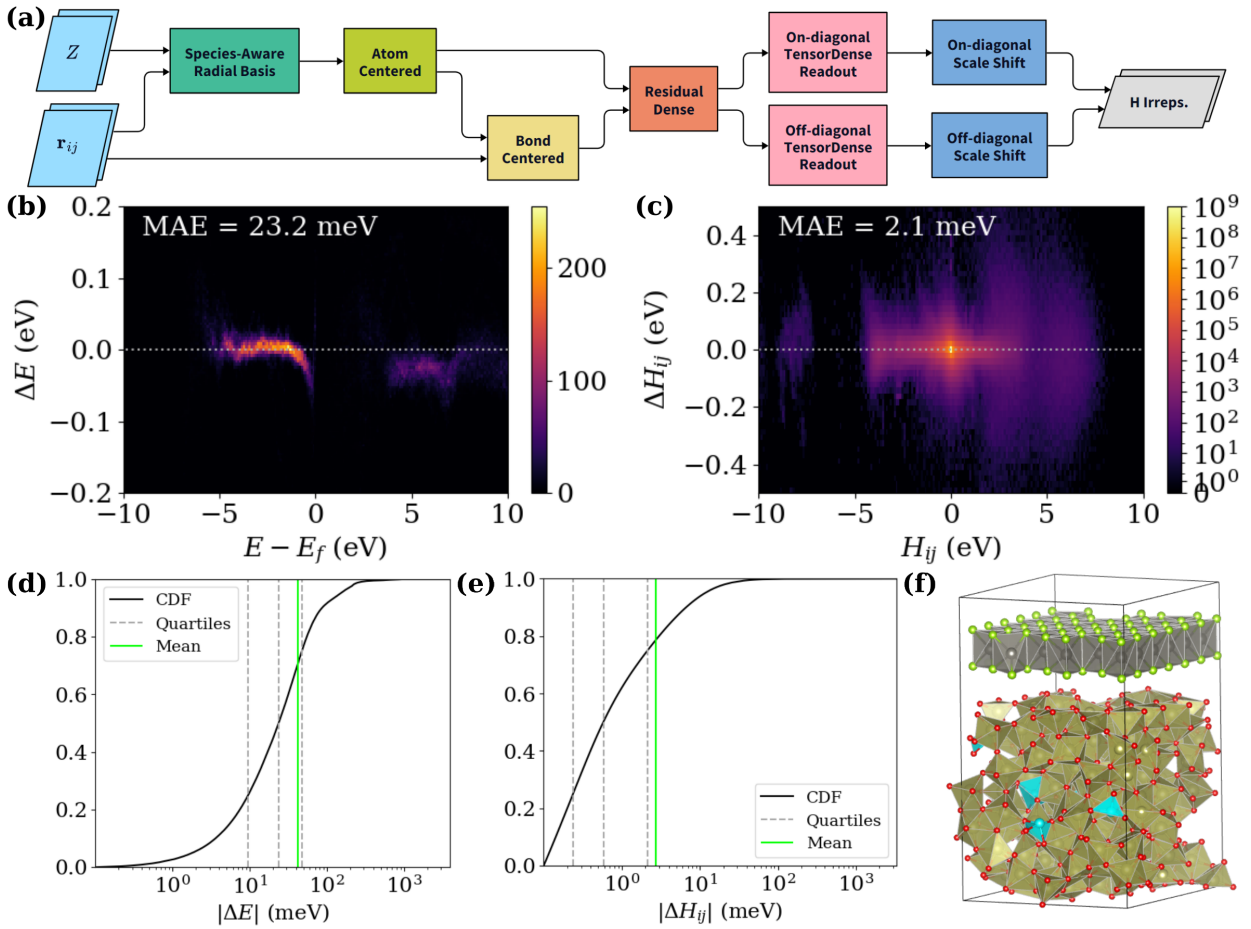

Figure 1: (a) GEARS H architecture overview. (b) Validation set eigenvalue errors relative to the reference eigenvalues. (c) Validation set Hamiltonian matrix element errors relative to the reference matrix elements with an `asinh` scale to help resolve bins that have smaller counts. (d) Cumulative distribution function of the eigenvalue errors larger than $0.1\,\mathrm{meV}$. (e) Cumulative distribution function of the Hamiltonian matrix element errors larger than $0.1\,\mathrm{meV}$. (f) Sample structure from the training dataset. Hf are gold, Ni are turquoise, O are red, W are gray, and Se are green. The MAEs shown in (b-c) are averaged across training set structures, whereas the mean values in (d-e) are taken across all errors. GEARS H performs very well on this highly complex system.

To make our descriptor many-body and increase its sensitivity to the local atomic environment, we use the density trick[24, 25, 23]: the outer product of *pooled* 2B features leads to 3-body features, and subsequent outer products lead to higher body-order features. The descriptors created from these outer products are known as atom-centered density correlations (ACDC)[26]. We focus on learning a dense subspace since higher-order descriptors in the Atomic Cluster Expansion (ACE) model are known to be relatively sparse[27]. We do this using a `TensorDense` layer as implemented in `E3x` [19] to learn a feature-wise tensor product of two linear projections of the input features. 2B descriptors are passed through `TensorDense` layers (optionally, although we recommend at least one–otherwise, the descriptor does not have many-body information) to get 3B descriptors, and so on. The pooling operation is a sum over the atoms j in the neighborhood of a given atom i, which we implement using the `indexed_sum` operation as implemented in `E3x`. Empirically, a body order of 3-5 is sufficient for acceptable accuracy of learned quantities like energies and forces[28]—we find the same is true for GEARS H.

An essential difference between the well-explored previous energy predictions and our approach is that our model does *not* average over all rotations of these ACDCs.

These 2B, 3B, ..., $(2N-1)$B descriptors are then separately (and optionally) message-passed between atoms using self-attention (SA) to carry out the learnable coupling across all incoming messages. In this work, none of the models presented make use of SA, so we leave discussion of it to the supplement. Crucially, the omission of message-passing does not reduce our accuracy and simultaneously greatly reduces our parameter count, contributing to the status of GEARS H as the smallest reported MLH model in the literature.

The separate (optionally) message-passed descriptors are then sent through a nonlinear block, for which we use a two-layer perceptron with a residual connection. The dense layers are interleaved with a `LayerNorm` and `mish` activation function to refine the atom-centered features and add functional expressivity.

Finally, the resulting descriptors are reduced to a user-controlled maximum angular momentum and then concatenated along the feature dimension (F in `E3x` convention). By keeping the descriptors of distinct body-order separate until the very end, the intervening message-passing and nonlinear blocks remain small (block diagonal in body order), which further helps reduce parameter counts and speed up training.

2.1.2 The bond-centered descriptor

Off-diagonal terms in Hamiltonian or overlap matrix blocks can be predicted as a function of atom-centered features of the two atoms comprising a 'bond'. Here, a bond refers to any two atoms with significant basis function overlap (and corresponding interaction strength). To calculate atom-pairwise features for predicting off-diagonal matrix blocks, we sum pairs of atom-centered features, which is similar to the approach used in PhiSNet [13]. To add more functional expressivity, the pooled features are refined using a two-layer perceptron with a `LayerNorm` and `mish` activation between layers and a residual connection between layers, akin to the nonlinear block in the atom-centered descriptor. For bond orientation information, we expand the bond vector into radial and angular basis functions, which we pass through a `Dense` layer as a learnable linear projection to refine features. Finally, we take a feature-wise tensor product of linearly-projected bond vector expansion with the pooled atom-pair features.

2.1.3 The scale-shift layers

The scale-shift layers are non-learnable blocks that scale and shift the parity-symmetric scalars in the readout output. This allows (scalar) outputs from the readout to be approximately zero-centered and unit-variance by mapping the outputs to the physical values, which can vary greatly in magnitude. These parameters can be extracted from the training dataset, a functionality which we have built into GEARS H.

2.2 Case study: Modulation doping of WSe_2 with Ni-doped a-HfO_2

Transition metal dichalogenides (TMDs) like WSe_2 are attractive candidates for post-silicon channel materials because they are atomically thin and can have mobilities comparable to Si. Conventional substitution doping strategies of TMDs, however, lead to reduced mobilities due to the introduction of scattering centers, and do not contribute enough carriers to the TMD. Recently, Sivadas and Shin [5] proposed modulation doping of WSe_2 through doping an interfaced HfO_2 gate dielectric layer. However, their work was plane-wave DFT-based, which limited the accessible doping rates and dopant distributions, prevented the consideration of the effect of Se defects in WSe_2 (which are known to form during synthesis) , and restricted their study to crystalline HfO_2 for the gate dielectric, despite the ubiquity of amorphous gate oxides in real devices.

GEARS H does not suffer from these constraints. As a proof of concept of its utility in modeling multi-component systems of engineering interest, we train a model for amorphous, Ni-doped HfO_2 interfaced with WSe_2 containing Se vacancies. This system presents both geometric and chemical challenges for ML-based modeling and provides a testbed for the atomistic modeling of device-scale geometries. A large number of diverse chemical environments are present in this system, ranging from 2D crystalline WSe_2 to the amorphous HfO_2 bulk, which is further complicated by the presence of Ni dopants, Se vacancies, and the interface with WSe_2. To our knowledge, no other MLH framework has been applied to a system of this complexity.

2.2.1 Validation set

A sample training structure for this system is shown in Fig. 1(f). 200 structures total were generated, which were split into 160 training structures and 40 validation structures (see Methods for more details).

In Fig. 1(b), we show the validation set errors of eigenvalues from inferred Hamiltonians. Within $\pm 5\,\mathrm{eV}$ of the Fermi level E_f, the eigenvalue mean absolute error (MAE) averaged across validation set systems is $23.2\,\mathrm{meV}$. While a small increase in the error is visible at the valence band maximum, this is in fact a numerical artifact arising from uncorrelated sorting of the eigenvalues between the eigenvalues of the reference Hamiltonian and the eigenvalues of the inferred Hamiltonian. To provide another view of the eigenvalue errors, in Fig. 1(d), we show the cumulative distribution function (CDF) of the absolute eigenvalue errors larger than $0.1\,\mathrm{meV}$, and plot the quartiles and MAE of all validation set eigenvalues taken together. 50% of the eigenvalue errors are smaller than $23.6\,\mathrm{meV}$, which is below thermal fluctuations at room temperature ($k_B T|_{T=298\,\mathrm{K}} = 25.7\,\mathrm{meV}$). The MAE in Fig. 1(d) is higher than that shown in Fig. 1(b) because, in the former, the MAE is calculated across all eigenvalues, while in the latter, we only include eigenvalues within $E_f \pm 5\,\mathrm{eV}$. In other words, even when considering states more than $5\,\mathrm{eV}$ from the Fermi level, which will have correspondingly smaller impact on observables, our errors for this complex system will not impact the application of our model.

We show the Hamiltonian matrix element errors in Fig. 1(c). Note that the color map was created using an `asinh` normalization. The MAE of matrix elements averaged across validation set systems is $2.1\,\mathrm{meV}$, and errors are within $\pm 1\,\mathrm{eV}$ across the full range of matrix element values ($-15\,\mathrm{eV}$ to $55\,\mathrm{eV}$). In Fig. 1(e), we show the CDF of the absolute Hamiltonian matrix element errors, after filtering out errors smaller than $0.1\,\mathrm{meV}$. Over 90% of errors are smaller than $6\,\mathrm{meV}$, and 50% are smaller than $0.58\,\mathrm{meV}$. Had all the errors been considered, the error at these quartiles would be even lower.

Altogether, these figures indicate that while the model performs quite well, there is room for improvement towards minimizing outliers. Conversely, outliers have an outsized effect on the MAE. The full CDFs reveal that the model performs very well across the validation set, and can therefore be trusted for studying modulation doping of WSe_2 interfaced with a Ni-doped a-HfO_2 gate dielectric. A systematic study of the effect of random outliers and random errors in general on the eigenvalues of matrices will be critical for enhancing trust in MLH model predictions as their reliability and usage grows.

2.2.2 Application to device-scale structures

Given the good performance of our model across the validation set, we now use it to perform a statistical study of hole concentrations in the WSe_2 layer. We considered systems sizes ranging from 3326 to 4160 atoms with systems with side lengths ranging from $3.4\,\mathrm{nm}$ to $4.5\,\mathrm{nm}$. These systems are comparable in size to candidate next-generation 2D field effect transistors under active research [29, 30]. 72 structures were generated for the statistical study with Ni doping rates ranging from $\mathrm{Ni:Hf} = 3.23 \times 10^{-3}$ to 16.86×10^{-3}, Se vacancy rates ranging from 0%-1.04% (0-21.9 vacancies/cm^2), and system densities ranging from $6.6\,\mathrm{g/cm^3}$ to $8.4\,\mathrm{g/cm^3}$. The distribution of Ni doping rates, Se vacancy rates, and system densities considered are shown in the histograms in the diagonal subplots of Fig. 2(a), along pair plots colored with the corresponding hole concentrations in the off-diagonal subplots.

We emphasize that the full process of generating this data (the generation of all the structures, the inference of all their electronic structures, and the diagonalization of the inferred Hamiltonians) took less than 12 hours on a single GPU workstation with 8 Nvidia L40S GPUs. The inference of the Hamiltonians itself was the fastest part of the process took approximately $13\,\mathrm{s}$ per structure (see additional discussion on inference in the Supplement). Investigations of realistic systems of this complexity and length scale would not be feasible without GEARS H.

With the WSe_2 layer hole concentration data from large-scale systems as our target variable, we now perform a Bayesian study to find the effect of several experimentally-controllable parameters. We report the parameters as A^C_B, where A is the mean value of the parameter, B is 3% high-density interval, and C is 97% high-density interval.

We *ansatz* a simple linear model dependent on relevant, controllable design variables: the Ni:Hf doping rate (ρ_{Ni}), the total system density (ρ), and the Se vacancy rate ($\rho_{\mathrm{V_{Se}}}$). Concretely,

$$\rho_h = c_{\mathrm{Ni}}\rho_{\mathrm{Ni}} + c_\rho \rho + c_{\mathrm{V_{Se}}}\rho_{\mathrm{V_{Se}}}, \tag{1}$$

where we have shifted the densities such that they are centered at approximately $0\,\mathrm{g/cm^3}$.

The results of the Bayesian analysis are shown in Fig. 2(b). The diagonal subplots are distributions of each parameter in the model and the residual, and the off-diagonal subplots are correlations between the parameters and residual themselves.

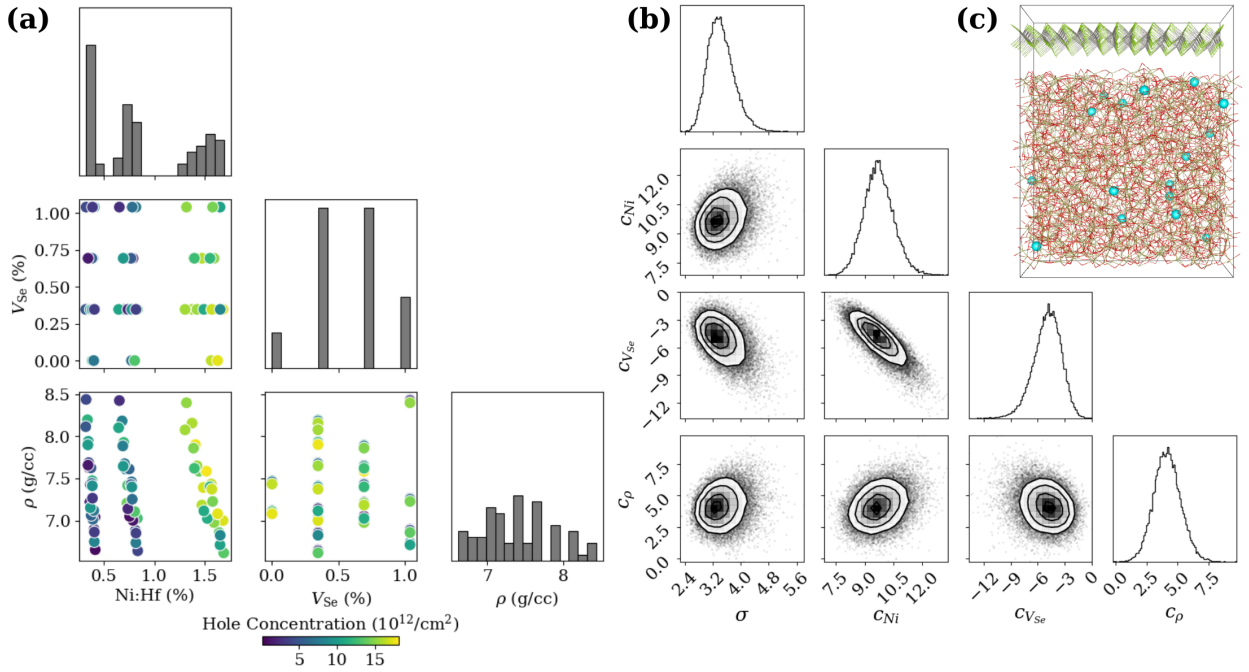

Figure 2: (a) Ni doping rates, Se vacancy rates, and total system density of the large scale systems used in the statistical analysis. Diagonal figures are histograms with 20 bins of the distribution of each individual quantity, while the off-diagonal figures show pair plots with the corresponding hole concentrations. (b) Results of the Bayesian analysis showing interactions between the posterior likelihood of the parameters of the model. Diagonal histograms show the distribution of each parameter and the residual σ. Off-diagonal subplots show marginal joint distributions of the parameters with iso-likelihood contours. Most interactions are weak with the exception of c_{Ni} and $c_{V_{\mathrm{Se}}}$, indicating that both the p-doping due to Ni and the n-doping due to Se-vacancies can be strong or weak together. (c) Selected system used in the statistical study. Hf, O, W, and Se atoms have been hidden to highlight the Ni dopants spread through the a-HfO_2.

First, we consider c_{Ni}, the proportionality coefficient between the magnitude of hole doping of WSe_2 and the Ni doping rate. In the histogram in the second row of Fig. 2(b), we see c_{Ni} has a positive mean of approximately $9.7^{11.2}_{8.3}$. This implies a strong positive correlation between the Ni doping rate and the hole concentration in the interfaced WSe_2, since the posterior likelihood of the doping coefficient is entirely positive. Importantly, to the best of our knowledge, this is the first time variations in induced hole concentrations due to modulation doping have been accounted for at an atomistic level. Our studies provide strong statistical evidence of robust Ni-induced p-doping in interfaced WSe_2 and greatly extend previous work [5] to amorphous gate oxides and realistic doping rates.

Next, we focus on the effect of system density on the hole concentration in the WSe_2 layer, which is represented by c_ρ. Since the distribution of c_ρ is peaked at approximately $4.1^{6.2}_{2.0}$, and the distribution is almost entirely positive, this is strong evidence that greater system densities facilitate higher p-doping of the WSe_2 layer. This is strong evidence for the intuitive picture that lower densities lead to larger structural variations that can create trap states and increase the potential barrier through which the Ni electrons tunnel through. Both of these effects reduce the doping in the WSe_2 layer.

Finally, we consider the effect of Se vacancies on the hole doping, which is represented by $c_{\mathrm{V_{Se}}}$. The distribution of $c_{\mathrm{V_{Se}}}$ is peaked at $-5.0^{-2.2}_{-8.0}$, a negative value, suggesting that Se vacancies contribute negatively to p-doping in WSe_2. In other words, there is no compensating mechanism for the n-doping of $\mathrm{V_{Se}}$ from the gate dielectric that we find from our data. While this relationship is weaker than the p-doping due to Ni, we see that Se vacancies and Ni doping are competing variables in the p-doping of WSe_2.

We now focus on the interactions between the coefficients of the model, shown in the off-diagonal subplots in Fig. 2(b). The interactions between the residual variable σ and both c_{Ni} and $c_{\mathrm{V_{Se}}}$ suggest a stronger doping effect weakly corresponds to increased residual of the model, indicating that the linear model may need additional corrections in the strong doping regime. Very interestingly, we notice a strong correlation

in the joint posterior distribution of c_{Ni} and $c_{\mathrm{V_{Se}}}$. This suggests it is likely both the p-doping due to Ni and n-doping due to Se-vacancies can be strong or weak together, but it is very unlikely that one is strong while the other is weak. Investigation of this correlation using more involved numerical experiments is a promising avenue for further work. Once again, GEARS H makes this possible.

2.3 Application to diverse chemical systems and atomic environments

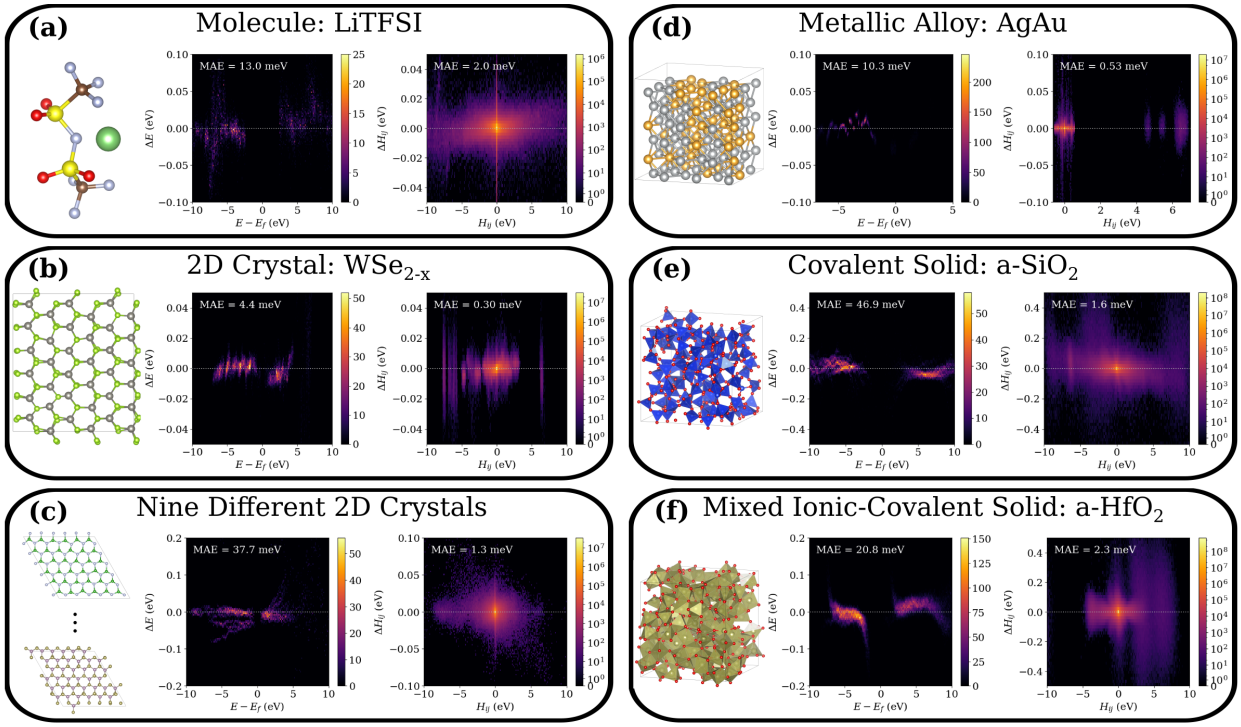

Figure 3: Sample structure from the training set and 2D histograms of eigenvalue and Hamiltonian matrix element errors of GEARS H models trained on a wide variety of materials chosen for their varied local environments. These include 1) Li-TFSI, a molecular system with 6 atomic species, 2) WSe_{2-x}, a 2D TMD, 3) a collection of 9 different 2D TMDs with 8 atomic species, 4) AgAu, a metallic alloy, 5) a-SiO_2, a covalent solid, and 6) a-HfO_2, a mixed ionic-covalent solid. Errors shown were calculated on the validation sets, and H_{ij} errors are shown with an `asinh` scale. Note that error scales differ between figures, due to the variation in the range of errors. GEARS H performs extremely well across all classes of materials.

In Fig. 3, we show the eigenvalue and Hamiltonian matrix element distribution of GEARS H models trained on a diverse set of systems chosen for their widely varying atomic environments. GEARS H can handle molecules, 2D materials, metallic alloys, amorphous solids, and systems combining these without extensive hyperparameter tuning. Notably, in addition to the 5 species system discussed in the previous section, we include a 6 species and an 8 species system here. Only `HamGNN` [18, 31] has report models with more species, and even then, they have not included defects or amorphous structures.

For each material, we show an example structure from the training dataset, the validation set eigenvalue error distribution MAE, and Hamiltonian matrix element error distribution and MAE. Both MAEs provided are averaged MAEs across the validation set structures, and the eigenvalue MAEs are computed using eigenvalues within $\pm 5\,\mathrm{eV}$ of the Fermi level. Additional details about the model hyperparameters along with the MAEs are shown in Table 1.

The first example system is LiTFSI, which is molecular system with 6 elements. This system presents a strong alchemical and structural challenge due to the large number of atomic species and large configurational space. GEARS H achieves $2\,\mathrm{meV}$ MAE on the Hamiltonian matrix elements.

The next set of systems we consider are 2D materials. First, we consider WSe_{2-x}. Se vacancies are commonly formed during synthesis, and GEARS H handles them with aplomb. While a direct comparison is not

possible with our dataset, the Hamiltonian matrix element MAE of $0.30\,\mathrm{meV}$ we achieved with Se vacancies is lower than that achieved by `DeePTB-E3`, `DeepH-E3`, and `HamGNN` on MoS_2 without defects [11, 12, 18] with 12%, 8.3%, and 3% the parameter count, respectively. As a combination of the many atomic species we showed in LiTFSI and the defect WSe_2, we consider a dataset of nine defect-free binary 2D crystals comprised of eight different atomic species. The 2D crystals included in this dataset are BN, GeS, GeSe, GeTe, MoS_2, $MoSe_2$, $MoTe_2$, WS_2, and WSe_2. The dataset includes only 16 snapshots of each 2D crystal in the training set, and 2 of each crystal in the validation set. Despite the wide range of atomic species and limited training data, GEARS H achieves $1.3\,\mathrm{meV}$ MAE on the Hamiltonian matrix elements.

Next, we consider bulk systems. $\mathrm{Ag}_x\,\mathrm{Au}_{1-x}$ forms a metallic solid solution with the atoms on an FCC lattice. Even with the wide range of compositions considered $(0.34 < x < 0.72)$, the Hamiltonian matrix element MAE is still sub-meV.

For a material with covalent bonding, we choose a-SiO_2. This is a very challenging system to model. While every Si atom is tetrahedrally coordinated by O, there is enormous freedom in how the tetrahedra connect. While Hamiltonian matrix element MAE is relatively low, the errors have larger variance than the other systems considered thus far. The eigenvalues are sensitive to outlier errors in the Hamiltonian, leading to the larger MAE. Better performance on this material can be achieved with optimization of the training dataset and model hyperparameters, which we did not deem necessary for this demonstration.

Finally, we consider a-HfO_2, a mixed ionic and covalent solid. Despite the increase in the possible number of bonding environments for Hf relative to Si in SiO_2, the Hamiltonian matrix element MAE is still a low $2.3\,\mathrm{meV}$.

System	Train./Val. Split	N_{TD}	Optimizer	Eigenvalue MAE (meV)	H_{ij} MAE (meV)
LiTFSI	1200/200	2	`adan`	13.0	2.0
WSe_{2-x}	160/40	1	`lamb`	4.4	0.30
Nine 2D	144/36	2	`adan`	37.7	1.3
$\mathrm{Ag}_x\,\mathrm{Au}_{1-x}$	96/24	1	`adan`	10.3	0.53
a-SiO_2	140/20	1	`adan`	46.9	1.6
a-HfO_2	160/40	1	`adan`	20.8	2.3

Table 1: Model training and validation set sizes, number of `TensorDense` layers (N_{TD}), the optimizer used and MAEs for the example systems shown in Fig. 3.

3 Conclusion

We present GEARS H, a state-of-the-art MLH framework that can be applied to the widest range of chemical systems and atomic environments of any MLH framework reported in the literature.

Using GEARS H, we train a model on a Ni-doped a-HfO_2 gate oxide interfaced with WSe_2 and use the model to perform a statistical study on realistic, device-scale systems. We analyze the effect of the Ni doping rate, system density, and Se vacancy rate on the induced hole concentration in the WSe_2 layer. This is a direct example of first-principles simulations, atomistic deep learning, and statistical modeling being leveraged to guide future scientific and engineering exploration for novel semiconductor design. Without GEARS H, this kind of study would be impossible. GEARS H enables this work by bypassing lengthy self-consistent cycles required by Kohn-Sham DFT—the inference of the large-scale structures takes just $\sim 13\,\mathrm{s}$ on our hardware.

We further demonstrate the remarkable flexibility of GEARS H by training models on molecular, 2D, metallic alloy, amorphous covalent solids, and mixed ionic-covalent solid systems. In all cases, the MAE of the Hamiltonian matrix elements is smaller than $2.4\,\mathrm{meV}$.

With GEARS H, MLH frameworks are now production-ready tools for next-generation device modeling.

4 Methods

4.1 Training and validation data generation

4.1.1 Structure generation

To generate the defect WSe_2 structures, We use the `mx2` build module in `ASE` [32] with a 6×3 orthogonal supercell of WSe_2. The primitive cell is generated with a lattice constant of 3.32 Å, and a thickness (vertical spacing between Se atoms) of 3.2 Å, with 2.3 Å of vacuum (for subsequent stacking of Ni-doped HfO_2). The precise values the lattice constants and thickness are not optimized, since the structures are annealed and relaxed later, and the underlying variation in strain is an intended variability in the dataset. A uniform random diagonal strain of $\pm 2\%$ is applied to the lattice constants. A random number of Se vacancies is incorporate. A Poisson distributed with a mean of 1.0 is used.

The a-HfO_2 training snapshots were generated using Packmol and HfO_2 "molecules". Target densities were randomly chosen from a uniform random distribution between $7.0 \pm 1.0\ \mathrm{g\,cm^{-3}}$. The HfO_2 geometries were then optimized using the LBFGS optimizer in `ASE` to a maximum force of $0.2\ \mathrm{eV\,Å^{-1}}$, followed by a piecewise-constant-temperature anneal from 800 K to 400 K using the Bussi velocity rescaling thermostat in the NVT ensemble, as implemented in ASE. A timestep of 2.0 fs was used, with a 100 fs thermostat coupling constant. The snapshots were finally optimized using LBFGS to a maximum force of $0.1\ \mathrm{eV\,Å^{-1}}$. We found that amorphous structures generated using conventional melt-quench methods provide similar structures for amorphous HfO_2.

For Ni-doped a-HfO_2 interfaced with defect WSe_2, we start with the same initial structures as used in the defect WSe_2 and a-HfO_2 structures discussed above. Hf is substitutionally doped with Ni with a Poisson-distributed concentration with mean of 3% of the Hf count. We then stack the defect WSe_2 2.3 Å above the initial Ni doped a-HfO_2. We performed molecular dynamics with a piecewise constant annealing schedule using the Bussi thermostat [33] in `ASE` with a timestep of $2.0\,\mathrm{fs}$ and a thermostat coupling constant of $\tau = 100\,\mathrm{fs}$, using the MACE MPA-0 foundation potential [28, 34, 35]. Geometries were first optimized to 0.2 $\mathrm{eV\,Å^{-1}}$ using the LBFGS optimizer in ASE. They were then annealed down from 800 K to 400 K in steps of -100 K, running for 2000 steps to 300 steps, in steps of -400 steps. A final optimization using LBFGS down to a maximum force of $0.1\ \mathrm{eV\,Å^{-1}}$ was performed.

To generate the large scale structures for the statistical study, we generate structures in the same manner as we generated the training and validation set above.

LiTFSI training snapshots were generated starting from a single conformer of TFSI, replacing the H with Li. The geometries were optimized to $0.05\ \mathrm{eV\,Å^{-1}}$, followed a 2 ps molecular dynamics at 100 K. The `Bussi` thermostat in `ASE` was used, with a couple time constant of 50 fs. Both the optimization and molecular dynamics were performed using The MACE-MPA-0 [35] foundation potential including DFT-D3 dispersion correction [36, 37]. The training/validation split was 1200/200 structures, owing to the small amount of data per snapshot for a single molecule.

The nine 2D system dataset includes 18 structures each of BN, GeS, GeSe, GeTe, MoS_2, $MoSe_2$, $MoTe_2$, WS_2, and WSe_2.

The $\mathrm{Ag}_x\,\mathrm{Au}_{1-x}$ $(0.34 < x < 0.72)$ training snapshots were generated by first generating bulk silver $3 \times 3 \times 3$ supercells and then replacing N atoms of silver with gold, where $N \sim Poisson(\lambda = N_{atoms}/2)$. The AgAu geometries were then optimized using the LBFGS optimizer in `ASE` to a maximum force of $0.5\,\mathrm{eV\,Å^{-1}}$, including cell, but maintaining cell shape. This was followed by a piecewise-constant-temperature anneal from 1600 K to 700 K using the Bussi velocity rescaling thermostat in the NVT ensemble, as implemented in ASE. An adaptive timestep was used based on temperature-dependent heuristic to make sure atoms almost never move beyond a given distance (0.08 Å per time step). A 100 fs thermostant coupling constant was used. The snapshots were finally optimized using LBFGS to a maximum force of $0.1\ \mathrm{eV\,Å^{-1}}$.

The a-SiO_2 dataset snapshots were generated by randomly scattering SiO_2 trimers (to ensure the local stoichiometry was correct) using `Packmol`. The densities of the generated structures ranged from $2.025\,\mathrm{g/cm^3}$ to $2.4\,\mathrm{g/cm^3}$ The structures were then pre-relaxed until $F_{\max} \leq 5\,\mathrm{eV/Å}$. Next, the structures were annealed from $2300\,\mathrm{K}$ to $1000\,\mathrm{K}$ in steps of $100\,\mathrm{K}$ for $3\,\mathrm{ps}$ per step. Finally, the structure was relaxed until

$F_{\max} \leq 1 \times 10^{-2}\,\mathrm{eV/Å}$. Both relaxations used LBFGS and the MACE-MP-0a large model, while the annealing used the MACE-MP-0a medium model. 160 structures were generated in total.

4.1.2 LCAO DFT calculations

The LCAO DFT GPAW calculations were done using the szp basis sets included with GPAW. At the time of writing, GEARS only supports Γ-point datasets, so after the electronic structures were converged, a non-self-consistent calculation was done using the converged density to extract the Hamiltonian and S-matrix at the Γ-point only. We use the generalized gradient approximation for the exchange-correlation functional [38], grid spacing of $h = 0.2$ for all datasets except the WSe_2 dataset, where we used $h = 0.25$. Calculations were converged to a maximum change in the electron density smaller than 0.001 electrons per valence electron. To reduce the extent of the basis functions and thereby reduce the size of the training structures and number of neighbors in the training data, the Ag, Au, and Hf basis functions were confined until the atomic eigenstates shifted up by $0.3\,\mathrm{eV}$. This is a strong confinement, but we expect minimal effects due to the large number of basis functions available in the bulk. For the Ni atoms, we place the $3p$ electrons in the core and use an effective on-site interaction interaction of $U_{\text{eff}} = 4.5\,\mathrm{eV}$ on the $3d$ electrons. For Ag, we froze the $4p$ electrons in the core.

GEARS H requires that each atom must have unique neighbors—that is, each atom cannot have an interaction with an atom and the periodic images of the same atom. This sets a minimum cell size of $2\times$ the longest basis function in the system. For WSe_2, HfO_2, and the combined HfO_2:Ni + WSe_2 datasets, the longest cutoff length was $8.0\,$Å. For SiO_2, since we did not confine the Si basis functions, the longest cutoff length was $8.4\,$Å. With the confinement of the Ag and Au basis functions, the cutoff for the AgAu dataset was $6.1\,$Å.

4.2 Hole concentration calculation

Using the GEARS H model detailed in Fig. 1, we infer the Hamiltonians of the 72 large-scale structures generated as discussed above. To get the eigenvalues, we require an S-matrix, which we compute for each structure. The S-matrix is computed pairwise across atoms and is therefore not computationally intensive to generate. Using the S-matrix and inferred Hamiltonian, we solve the generalized eigenvalue problem to get the eigenvalues, which we then shift such that the Fermi level is at $0\,\mathrm{eV}$. We then species-project the density of states (DOS) and then smooth the projected DOSes using 20.000 points with Gaussians of width $0.05\,\mathrm{eV}$.

To get the hole concentration from the projected, smoothed DOSes, we integrate the W- and Se-projected DOS from the Fermi level to $0.2\,\mathrm{eV}$ above the Fermi level.

4.3 Bayesian analysis

We consider weakly regularizing priors for the coefficients and residual. The model is sampled over 8 chains, each for 4000 samples, with 2000 samples of burn-in using `PyMC` v5.23.0 [39] with the default No-U-Turn sampler.

4.4 Machine-learned H model

There is very little variation in the model hyperparameters used to train the models presented in this work. A full configuration (broken into its distinct pieces) is provided in the Supplement (to be added). Here, we only briefly discuss the most important hyperparameters used and which were varied between models.

In the `data` section, `n_train`, `n_valid`, `atoms_pad_multiple`, and `nl_pad_multiple` are changed depending on the dataset. The first two control the number of training and validation structures, while the last two control the number of recompilations of the model (through the maximum amount of padding a neighbor list array and an atomic species array is permitted) that the user is willing to allow. The total number of training and validation structures used for each model is shown in Table 1.

The only change made in the `atom_centered` section across models is the number of `TensorDenses`, which was set to 1 or 2 for all models shown in this work. The radial basis is an input to the atom-centered descriptor and is included as a subsection of `atom_centered`. The only change made between models in `radial_basis` is to adjust the cutoff radius to the maximum basis set cutoff across species in each dataset. Basis set cutoffs are discussed in Section 4.1.2.

Similarly, for the `bond_centered` section, all hyperparameters were left unchanged except for the cutoff, which was set to the largest cutoff across species in the dataset (see Section 4.1.2).

The residual dense layer (controlled by the `mlp` section of the config) was left unchanged across all models trained in this work. Three layers were used with output feature sizes of 32, 16, and 32, in that order. We use a `bent_identity` nonlinear activation function between layers.

The only changes made in the `optimizer` section was to switch between the `adan` [40] and `lamb` [41] optimizers, depending on which resulted in a lower loss model. `adan` was best for all models except the defect WSe_2 model. Which optimizer was used for each model is shown in 1. The only learning rate schedule changes made were to adjust the `accumulation_size` parameter to make the `reduce_on_plateau` scheduler check if the loss had plateaued only once per epoch.

Loss parameters were left unchanged across all models.

References

[1] Richard Van Noorden. "These Are the Most-Cited Research Papers of All Time". In: *Nature* 640.8059 (Apr. 17, 2025), pp. 591–591. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/d41586-025-01124-w. URL: https://www.nature.com/articles/d41586-025-01124-w (visited on 06/05/2025).

[2] Krithika Dhananjay et al. "Monolithic 3D Integrated Circuits: Recent Trends and Future Prospects". In: *IEEE Transactions on Circuits and Systems II: Express Briefs* 68.3 (Mar. 2021), pp. 837–843. ISSN: 1549-7747, 1558-3791. DOI: 10.1109/TCSII.2021.3051250. URL: https://ieeexplore.ieee.org/document/9321494/ (visited on 06/05/2025).

[3] Senfeng Zeng, Chunsen Liu, and Peng Zhou. "Transistor Engineering Based on 2D Materials in the Post-Silicon Era". In: *Nature Reviews Electrical Engineering* 1.5 (Apr. 30, 2024), pp. 335–348. ISSN: 2948-1201. DOI: 10.1038/s44287-024-00045-6. URL: https://www.nature.com/articles/s44287-024-00045-6 (visited on 06/09/2025).

[4] Arnab Pal et al. "Three-Dimensional Transistors with Two-Dimensional Semiconductors for Future CMOS Scaling". In: *Nature Electronics* 7.12 (Dec. 16, 2024), pp. 1147–1157. ISSN: 2520-1131. DOI: 10.1038/s41928-024-01289-8. URL: https://www.nature.com/articles/s41928-024-01289-8 (visited on 06/09/2025).

[5] Nikhil Sivadas and Yongwoo Shin. *Modulation Doping and Control the Carrier Concentration in 2-Dimensional Transition Metal Dichalcogenides*. Apr. 18, 2025. DOI: 10.48550/arXiv.2504.14031. arXiv: 2504.14031 [cond-mat]. URL: http://arxiv.org/abs/2504.14031 (visited on 06/05/2025). Pre-published.

[6] Amil Merchant et al. "Scaling Deep Learning for Materials Discovery". In: *Nature* 624.7990 (Dec. 7, 2023), pp. 80–85. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-023-06735-9. URL: https://www.nature.com/articles/s41586-023-06735-9 (visited on 06/06/2025).

[7] Aaron D. Kaplan et al. *A Foundational Potential Energy Surface Dataset for Materials*. Version 1. 2025. DOI: 10.48550/ARXIV.2503.04070. URL: https://arxiv.org/abs/2503.04070 (visited on 06/06/2025). Pre-published.

[8] Han Yang et al. *MatterSim: A Deep Learning Atomistic Model Across Elements, Temperatures and Pressures*. May 10, 2024. DOI: 10.48550/arXiv.2405.04967. arXiv: 2405.04967 [cond-mat]. URL: http://arxiv.org/abs/2405.04967 (visited on 06/06/2025). Pre-published.

[9] Ganesh Hegde and R. Chris Bowen. "Machine-Learned Approximations to Density Functional Theory Hamiltonians". In: *Scientific Reports* 7.1 (Feb. 15, 2017), p. 42669. ISSN: 2045-2322. DOI: 10.1038/srep42669. URL: https://www.nature.com/articles/srep42669 (visited on 11/08/2024).

[10] K. T. Schütt et al. "Unifying Machine Learning and Quantum Chemistry with a Deep Neural Network for Molecular Wavefunctions". In: *Nature Communications* 10.1 (Nov. 15, 2019), p. 5024. ISSN: 2041-1723. DOI: 10.1038/s41467-019-12875-2. URL: https://www.nature.com/articles/s41467-019-12875-2 (visited on 11/11/2024).

[11] He Li et al. "Deep-Learning Density Functional Theory Hamiltonian for Efficient Ab Initio Electronic-Structure Calculation". In: *Nature Computational Science* 2.6 (June 23, 2022), pp. 367–377. ISSN: 2662-8457. DOI: 10.1038/s43588-022-00265-6. URL: https://www.nature.com/articles/s43588-022-00265-6 (visited on 08/03/2023).

[12] Xiaoxun Gong et al. "General Framework for E(3)-Equivariant Neural Network Representation of Density Functional Theory Hamiltonian". In: *Nature Communications* 14.1 (May 18, 2023), p. 2848. ISSN: 2041-1723. DOI: 10.1038/s41467-023-38468-8. URL: https://www.nature.com/articles/s41467-023-38468-8 (visited on 09/03/2024).

[13] Oliver T. Unke et al. *SE(3)-Equivariant Prediction of Molecular Wavefunctions and Electronic Densities*. Oct. 20, 2021. arXiv: 2106.02347 [physics]. URL: http://arxiv.org/abs/2106.02347 (visited on 11/11/2024). Pre-published.

[14] Jigyasa Nigam, Michael Willatt, and Michele Ceriotti. "Equivariant Representations for Molecular Hamiltonians and N-center Atomic-Scale Properties". In: *The Journal of Chemical Physics* 156.1 (Jan. 7, 2022), p. 014115. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/5.0072784. arXiv: 2109.12083 [physics]. URL: http://arxiv.org/abs/2109.12083 (visited on 11/11/2024).

[15] Mario Geiger and Tess Smidt. *E3nn: Euclidean Neural Networks*. July 18, 2022. DOI: 10.48550/arXiv.2207.09453. arXiv: 2207.09453 [cs]. URL: http://arxiv.org/abs/2207.09453 (visited on 06/05/2025). Pre-published.

[16] Qiangqiang Gu et al. "Deep Learning Tight-Binding Approach for Large-Scale Electronic Simulations at Finite Temperatures with Ab Initio Accuracy". In: *Nature Communications* 15.1 (Aug. 8, 2024), p. 6772. ISSN: 2041-1723. DOI: 10.1038/s41467-024-51006-4. URL: https://www.nature.com/articles/s41467-024-51006-4 (visited on 06/05/2025).

[17] Haiyang Yu et al. *Efficient and Equivariant Graph Networks for Predicting Quantum Hamiltonian*. Nov. 8, 2023. arXiv: 2306.04922 [physics]. URL: http://arxiv.org/abs/2306.04922 (visited on 09/03/2024). Pre-published.

[18] Yang Zhong et al. "Transferable Equivariant Graph Neural Networks for the Hamiltonians of Molecules and Solids". In: *npj Computational Materials* 9.1 (Oct. 6, 2023), p. 182. ISSN: 2057-3960. DOI: 10.1038/s41524-023-01130-4. URL: https://www.nature.com/articles/s41524-023-01130-4 (visited on 06/05/2025).

[19] Oliver T. Unke and Hartmut Maennel. *E3x:* $\mathrm{E}(3)$*-Equivariant Deep Learning Made Easy*. Jan. 17, 2024. arXiv: 2401.07595 [physics]. URL: http://arxiv.org/abs/2401.07595 (visited on 09/03/2024). Pre-published.

[20] Jens Jørgen Mortensen et al. "GPAW: An Open Python Package for Electronic Structure Calculations". In: *The Journal of Chemical Physics* 160.9 (Mar. 7, 2024), p. 092503. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/5.0182685. URL: https://pubs.aip.org/jcp/article/160/9/092503/3269902/GPAW-An-open-Python-package-for-electronic (visited on 12/16/2024).

[21] A. H. Larsen et al. "Localized Atomic Basis Set in the Projector Augmented Wave Method". In: *Physical Review B* 80.19 (Nov. 18, 2009), p. 195112. ISSN: 1098-0121, 1550-235X. DOI: 10.1103/PhysRevB.80.195112. URL: https://link.aps.org/doi/10.1103/PhysRevB.80.195112 (visited on 12/16/2024).

[22] Kaiming He et al. "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, June 2016, pp. 770–778. ISBN: 978-1-4673-8851-1. DOI: 10.1109/CVPR.2016.90. URL: http://ieeexplore.ieee.org/document/7780459/ (visited on 06/09/2025).

[23] Felix Musil et al. "Physics-Inspired Structural Representations for Molecules and Materials". In: *Chemical Reviews* 121.16 (Aug. 25, 2021), pp. 9759–9815. ISSN: 0009-2665, 1520-6890. DOI: 10.1021/acs.chemrev.1c00021. URL: https://pubs.acs.org/doi/10.1021/acs.chemrev.1c00021 (visited on 11/08/2024).

[24] Alexander V. Shapeev. "Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials". In: *Multiscale Modeling & Simulation* 14.3 (Jan. 2016), pp. 1153–1173. ISSN: 1540-3459, 1540-3467. DOI: 10.1137/15M1054183. URL: http://epubs.siam.org/doi/10.1137/15M1054183 (visited on 09/03/2024).

[25] Ralf Drautz. "Atomic Cluster Expansion for Accurate and Transferable Interatomic Potentials". In: *Physical Review B* 99.1 (Jan. 8, 2019), p. 014104. ISSN: 2469-9950, 2469-9969. DOI: 10.1103/PhysRevB.99.014104. URL: https://link.aps.org/doi/10.1103/PhysRevB.99.014104 (visited on 09/03/2024).

[26] Jigyasa Nigam et al. "Unified Theory of Atom-Centered Representations and Message-Passing Machine-Learning Schemes". In: *The Journal of Chemical Physics* 156.20 (May 28, 2022), p. 204115. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/5.0087042. URL: https://pubs.aip.org/jcp/article/156/20/204115/2841327/Unified-theory-of-atom-centered-representations (visited on 09/03/2024).

[27] James P. Darby, James R. Kermode, and Gábor Csányi. "Compressing Local Atomic Neighbourhood Descriptors". In: *npj Computational Materials* 8.1 (Aug. 11, 2022), p. 166. ISSN: 2057-3960. DOI: 10.1038/s41524-022-00847-y. URL: https://www.nature.com/articles/s41524-022-00847-y (visited on 09/03/2024).

[28] Ilyes Batatia et al. *MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields*. Jan. 26, 2023. DOI: 10.48550/arXiv.2206.07697. arXiv: 2206.07697 [stat]. URL: http://arxiv.org/abs/2206.07697 (visited on 12/09/2024). Pre-published.

[29] Zhangting Wu et al. "Defects as a Factor Limiting Carrier Mobility in WSe2: A Spectroscopic Investigation". In: *Nano Research* 9.12 (Dec. 2016), pp. 3622–3631. ISSN: 1998-0124, 1998-0000. DOI: 10.1007/s12274-016-1232-5. URL: http://link.springer.com/10.1007/s12274-016-1232-5 (visited on 06/09/2025).

[30] Yury Yu. Illarionov et al. "Ultrathin Calcium Fluoride Insulators for Two-Dimensional Field-Effect Transistors". In: *Nature Electronics* 2.6 (June 17, 2019), pp. 230–235. ISSN: 2520-1131. DOI: 10.1038/s41928-019-0256-8. URL: https://www.nature.com/articles/s41928-019-0256-8 (visited on 06/05/2025).

[31] Yang Zhong et al. "Universal Machine Learning Kohn–Sham Hamiltonian for Materials". In: *Chinese Physics Letters* 41.7 (June 1, 2024), p. 077103. ISSN: 0256-307X, 1741-3540. DOI: 10.1088/0256-307X/41/7/077103. URL: https://iopscience.iop.org/article/10.1088/0256-307X/41/7/077103 (visited on 06/09/2025).

[32] Ask Hjorth Larsen et al. "The Atomic Simulation Environment—a Python Library for Working with Atoms". In: *Journal of Physics: Condensed Matter* 29.27 (July 12, 2017), p. 273002. ISSN: 0953-8984, 1361-648X. DOI: 10.1088/1361-648X/aa680e. URL: https://iopscience.iop.org/article/10.1088/1361-648X/aa680e (visited on 12/16/2024).

[33] Giovanni Bussi, Davide Donadio, and Michele Parrinello. "Canonical Sampling through Velocity Rescaling". In: *The Journal of Chemical Physics* 126.1 (Jan. 7, 2007), p. 014101. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.2408420. URL: https://pubs.aip.org/jcp/article/126/1/014101/186581/Canonical-sampling-through-velocity-rescaling (visited on 12/16/2024).

[34] Ilyes Batatia et al. *The Design Space of E(3)-Equivariant Atom-Centered Interatomic Potentials*. Nov. 24, 2022. DOI: 10.48550/arXiv.2205.06643. arXiv: 2205.06643 [stat]. URL: http://arxiv.org/abs/2205.06643 (visited on 12/09/2024). Pre-published.

[35] Ilyes Batatia et al. *A Foundation Model for Atomistic Materials Chemistry*. Mar. 1, 2024. DOI: 10.48550/arXiv.2401.00096. arXiv: 2401.00096 [physics]. URL: http://arxiv.org/abs/2401.00096 (visited on 06/09/2025). Pre-published.

[36] Stefan Grimme et al. "A Consistent and Accurate *Ab Initio* Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu". In: *The Journal of Chemical Physics* 132.15 (Apr. 21, 2010), p. 154104. ISSN: 0021-9606, 1089-7690. DOI: 10.1063/1.3382344. URL: https://pubs.aip.org/jcp/article/132/15/154104/926936/A-consistent-and-accurate-ab-initio (visited on 06/09/2025).

[37] So Takamoto et al. "Towards Universal Neural Network Potential for Material Discovery Applicable to Arbitrary Combination of 45 Elements". In: *Nature Communications* 13.1 (May 30, 2022), p. 2991. ISSN: 2041-1723. DOI: 10.1038/s41467-022-30687-9. arXiv: 2106.14583 [cond-mat]. URL: http://arxiv.org/abs/2106.14583 (visited on 06/09/2025).

[38] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. "Generalized Gradient Approximation Made Simple". In: *Physical Review Letters* 77.18 (Oct. 28, 1996), pp. 3865–3868. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.77.3865. URL: https://link.aps.org/doi/10.1103/PhysRevLett.77.3865 (visited on 12/18/2024).

[39] Oriol Abril-Pla et al. "PyMC: A Modern, and Comprehensive Probabilistic Programming Framework in Python". In: *PeerJ Computer Science* 9 (Sept. 1, 2023), e1516. ISSN: 2376-5992. DOI: 10.7717/peerj-cs.1516. URL: https://peerj.com/articles/cs-1516 (visited on 06/09/2025).

[40] Xingyu Xie et al. *Adan: Adaptive Nesterov Momentum Algorithm for Faster Optimizing Deep Models*. Nov. 29, 2024. DOI: 10.48550/arXiv.2208.06677. arXiv: 2208.06677 [cs]. URL: http://arxiv.org/abs/2208.06677 (visited on 06/09/2025). Pre-published.

[41] Yang You et al. *Large Batch Optimization for Deep Learning: Training BERT in 76 Minutes*. Jan. 3, 2020. DOI: 10.48550/arXiv.1904.00962. arXiv: 1904.00962 [cs]. URL: http://arxiv.org/abs/1904.00962 (visited on 06/09/2025). Pre-published.