CHARACTERIZING MICROELECTRONIC DEVICES VIA SCALABLE, CONFINEMENT-AWARE EQUIVARIANT NETWORKS

Krishna Bhattaram[1*], **Pratik Brahma**[1*], **Sayeef Salahuddin**[1,2]
[1] Department of Electrical Engineering and Computer Sciences, University of California Berkeley
[2] Materials Science Division, Lawrence Berkeley National Laboratory
{krishnabhattaram,pratik_brahma,sayeef}@berkeley.edu

ABSTRACT

Modern microelectronic devices are comprised of materials with critical dimensions of a few nanometers. At these sizes, material properties change in non-trivial ways due to quantum confinement and atomic-level variability, creating a multi-scale modeling challenge that requires atomistic simulations for accurate prediction. However, such simulations are often prohibitively slow or intractable, making highly expensive iterative rounds of experimentation the default option. To address this issue, we introduce EBFormer, a geometry-aware equivariant neural network that predicts electronic properties of nanostructures by jointly capturing atomistic interactions and geometric effects, achieving orders of magnitude speed-up over state-of-the-art physical simulators while preserving high accuracy. This is accomplished through the introduction of a boundary cross-attention mechanism, a scalable approach to augment local graph convolution with information of the nanostructure geometry. We validate EBFormer on nanowire and nanosheet transistors, representing advanced modern microelectronic architectures, and show superior in-distribution and out-of-distribution performance on both material property inference and downstream device characteristics compared to leading architectures. Combined with superior asymptotic scalability and data- and parameter-efficiency, our work paves a pathway to atomistic, automated, high-throughput and predictive nanoscale design that is otherwise not available today.

1 INTRODUCTION

Moore's Law and the constant demand for greater performance within fixed energy budgets and space have pushed material design into the nanoscale. Key technologies such as image sensors, LEDs, phase-change memories, and transistors rely on materials with one or more dimensions scaled to only a few nanometers. At these dimensions the wave-nature of electrons interacts with material boundaries, leading to discrete resonant effects similar to the harmonics of sound waves on a guitar string or surface of a drum Paras et al. (2023); Ekimov & Onushchenko (2023). This quantum confinement of the electronic wavefunction leads to distinct properties of nanostructures compared to bulk counterparts, and has been leveraged to engineer semiconductor bandgaps Nozik (2002), boost transistor currents by up to 20% Liao et al. (2022); Park et al. (2025), and even dictate optical properties such as LED brightness and color Xia et al. (2018). Confinement effects are also spatially heterogeneous; much like the variation of nodes and antinodes across a vibrating string, electronic properties vary across nanostructures, impacting electron distribution and other material properties Jiang et al. (2008).

In addition to geometry effects, atomic-scale variation in nanostructures also leads to outsized effects in material properties. The current-conducting channel of a modern transistor may only be tens to hundreds of atoms thick Park et al. (2025). Effects such as interfacial strain or surface defects can lead to significant variation in electrostatic Reddy et al. (1999) and transport properties Tsutsui et al. (2019), among others.

* These authors contributed equally to this work

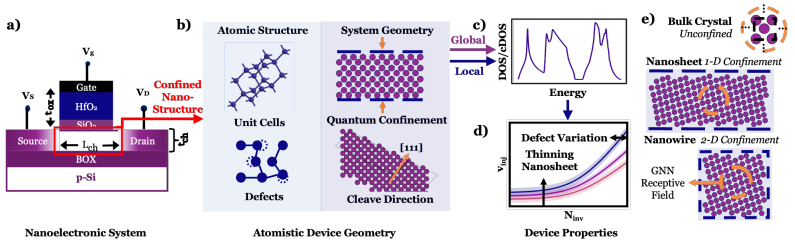

Figure 1: **Local and global structure effects in nanoscale device properties**: An example pipeline depicting inference for an SOI transistor. Transistor properties depend on the channel electronic behavior, a nanosheet (i.e. thin-film) or nanowire whose properties in turn depend on atomic variation and nanostructure geometry. Figure (e) highlights the insufficiency of local message-passing approaches for nanostructures; while bulk systems can be fully represented with a relatively small unit-cell, nanostructure critical dimensions can exceed model receptive fields, meaning the impact of geometry and cleaving orientation on local material properties cannot be resolved.

The combination of structural and atomic-level effects leads to a unique mesoscale problem in which a large number of atoms have to be simulated to jointly capture atomic structure and the variation of material properties across the nanostructure geometry. Atomistic quantum methods such as tight-binding or density functional theory remain state-of-the-art for high accuracy atomistic simulation of nanostructures Jancu et al. (1998); Stradi et al. (2019); Lee et al. (2021). However, the poor scaling characteristics of these methods render them impractically slow for realistic nanoelectronic devices, making highly expensive rounds of trial-and-error experimentation the default option Yuan (2022). Machine learning has recently shown significant promise in accelerating high-accuracy but computationally-expensive atomistic quantum simulation for small molecules and crystalline systems Unke et al. (2021); Batzner et al. (2022); Musaelian et al. (2023); Batatia et al. (2024); Xie & Grossman (2018); Schütt et al. (2017); Ward et al. (2016). Most architectures are based primarily on local graph convolution, neglecting global structural effects. Recent architectures such as DOSTransformer Lee et al. (2023), M3GNet Chen & Ong (2022), and Neural P^3M Wang et al. (2024b) capture system-level features of bulk unit cells, but either do not directly encode global geometry or have poor asymptotic scaling.

In this work, we present EBFormer, a neural network designed to address the mesoscale simulation challenges of nanostructures by encoding the influence of global geometry on local material properties while preserving linear scaling. EBFormer uses equivariant attention between atoms and material interfaces Fuchs et al. (2020), a physically and empirically-motivated sparse approach to propagate global geometry information through the material structure, augmenting local chemical information captured via conventional equivariant graph convolution Tan et al. (2025). We validate our approach by predicting material properties of one- and two-dimensionally confined nanostructures representative of channels in modern advanced transistors (SOI and GAAFETs Liao et al. (2022); Huang et al. (2017)), and demonstrate superior in- and out-of-distribution performance in inference of downstream device characteristics compared to leading architectures.

2 Background

2.1 Machine Learning for Material Simulation

Machine learning in materials science has predominantly focused on learning structure–property relationships from atomistic simulations. Among various approaches, graph neural networks (GNNs) have proven effective, as they naturally represent atomic systems. Examples of such architectures include CGCNN Xie & Grossman (2018), SchNet Schütt et al. (2017), ALIGNN Choudhary & DeCost (2021) and M3GNet Chen & Ong (2022), which have demonstrated strong performance on scalar property prediction of molecules and bulk crystal unit cells.

Building on this success, recent models have replaced local convolution with transformer-based attention mechanisms Vaswani et al. (2017) over local neighborhoods, leading to architectures such as Equiformer Liao & Smidt (2023), COMFormer Yan et al. (2024), and Matformer Yan et al. (2022) that further improve accuracy. Moreover, both GNN-based models (e.g., ALIGNN Choudhary

& DeCost (2021)) and transformer-based architectures (e.g., DOSTransformer Lee et al. (2023)) have demonstrated good performance on spectral properties such as the electronic DOS and phonon DOS. However, most of the above models primarily focus on learning local chemical environments through either graph convolution or attention, and largely overlook the role of global structural features (such as confinement, cleaving direction, and surface termination) on macroscopic properties. Architectures including M3GNet Chen & Ong (2022) incorporate system-level features such as pressure and temperature, and DOSTransformer Lee et al. (2023) employs global cross-attention between energy and atomic embeddings, but neither explicitly model global geometry effects. Approaches such as Ewald Summations or Neural P^3M incorporate long-range information through reciprocal space, but scale superlinearly and require tuning multiple additional hyperparameters Kosmala et al. (2023); Wang et al. (2024b). In contrast, our proposed architecture explicitly propagates geometry-induced global effects across the system using equivariant attention while preserving linear scaling and introducing only the added hyperparameter of boundary embedding size.

Finally, previous datasets have predominantly focused on molecular and bulk-crystal systems. In such systems, local message-passing approaches are ubiquitous due to the low receptive field required to collate geometry information over the entire atomic graph. JARVIS, MC2D, C2DB, and others Choudhary et al. (2023); Haastrup et al. (2018); Liu et al. (2025) include heterostructures, but primarily focus on 2D material mono- or bilayers, which have no long-range geometric structure. In this work, we introduce a dataset of geometrically-confined materials, and explicitly demonstrate the effect of confinement on spatial variation of electronic properties (LDOS and LcDOS). To the best of our knowledge, this is the first such dataset.

2.2 Predicted Material Properties

We focus on predicting the electronic properties of density of states (DOS) and current density of states (cDOS), used in nanoelectronic characterization to capture various downstream properties ranging from optical response to conductivity Choudhary & DeCost (2021); Lee et al. (2023). In this work, we leverage these two quantities to model the ballistic current through nanosheets and nanowires, which comprise the channels of modern transistor architectures Liao & Smidt (2023); Huang et al. (2017). Ballistic current represents the current through a transistor in the absence of electron scattering, and is a nonequilibrium quantity written as the product of injection velocity (v_{inj}, the average velocity of electrons entering the channel) and the inversion charge (N_{inv}, the density of electrons inside the channel at inversion) Rahman et al. (2003). The inversion charge and ballistic current can be written in turn as Fermi-Dirac integrals over DOS and cDOS, as outlined in Appendix B.

To resolve the spatial distribution of electronic states within the device, we compute atomic projections of the DOS (LDOS) and cDOS (LcDOS) using a projection operator Soriano & Palacios (2014), as detailed in Appendix C. Conceptually, the LDOS reflects the geometric distribution of states that electrons at a certain energy can occupy, while the LcDOS determines the current-carrying capacity of different parts of the nanostructure for electrons of given energies. The former impacts the electrostatic behavior of a transistor (i.e. at which voltages and how sharply the transistor turns on or off), while the latter determines how much current different parts of the device carry at a given voltage distribution. A visualization of the variation of LDOS and LcDOS over a nanowire is included in Appendix F.2.

3 Methods

3.1 Model Architecture

EBFormer is designed to capture the effects of both local atomic variation and quantum confinement on local material properties. Representing local atomic variation with graph convolution has been empirically well-motivated in literature Schütt et al. (2017); Batzner et al. (2022); Brahma et al. (2023). Numerical simulation has supported the theory of nearsightedness of electronic materials Prodan & Kohn (2005) by demonstrating exponential decay of electronic interaction terms with distance Xia et al. (2025), establishing that local convolution with a well-tuned cutoff (commonly between 0.4 to 1.0 nm) is sufficient to capture variation in neutral materials due to atomic defects Batatia et al. (2024). Replicating long-range quantum confinement effects with neural net-

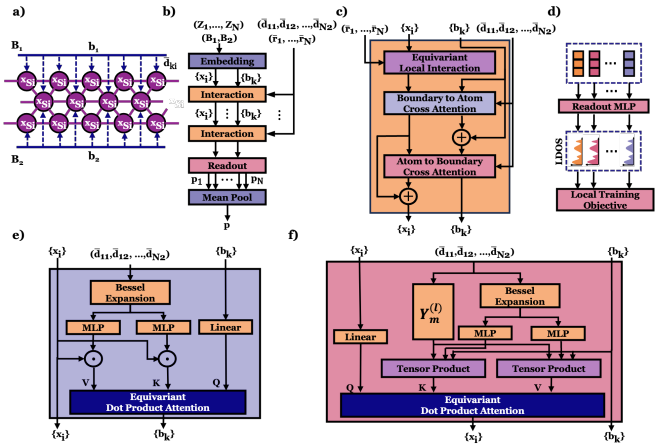

Figure 2: **EBFormer network architecture**: a) Nanosheet with material interfaces B_1 and B_2. b) Overall architecture. c) Interaction Block containing local convolution and global attention. d) Readout neural network predicting the atomistic projection of material property. e) Equivariant attention block to update boundary nodes. f) Equivariant attention block to update atomic nodes.

works, however, has limited precedence for atomic systems. Empirically, we observe confinement to impact local material properties at scales beyond 12.0 nm in crystalline silicon (Appendix F.1). Significant evidence and theory has demonstrated that simple increase of GNN cutoff radii is untenable due to cubically-scaling memory and data requirements, and graph over-squashing effects Kosmala et al. (2023); Alon & Yahav (2021); Batzner et al. (2022). Structural information must therefore be incorporated through an auxiliary mechanism.

Physical theory establishes that each atom needs to be able to resolve the distance and orientation of the boundaries relative to the crystal structure to capture their impact on local material properties Park et al. (2025); Yang et al. (2010). To accomplish this, EBFormer introduces a distance-dependent SE(3) equivariant atom-to-boundary cross-attention mechanism, enabling each atom to capture its distance to every boundary while resolving orientation through higher-order equivariant features. Our approach presents a linearly-scaling method for each atom to be aware of its local environment and its location within the nanostructure, permitting accurate prediction of local material properties. Further discussion and motivation that interaction with only boundaries is sufficient is included in Appendix Section A, where we present a comparison with an O(N^2) distance-dependent equivariant self-attention mechanism, an all-to-all connected network representing a relaxation of EBFormer's boundary-only inductive bias. We find that both architectures achieve similar errors, and that the learned self-attention matrices are sparse, focusing on local environments and atoms at the boundaries.

Overview: The pipeline from atomic structure to material predictions proceeds as follows. Input nanostructures are first mapped to a graph, with nodes placed at atomic positions ($\vec{r}_i$) and edges drawn between atoms within a radial cutoff. Boundary embeddings, representing interfacial planes between materials $\{\mathbf{B}_k\}$, are each initialized with the same learnable vector $\mathbf{b}_k^{(0)}$ to preserve rotational symmetry, while each atomic node is initialized with an embedding $\mathbf{x}_i^{(0)}$ derived from its atomic number Z_i. These embeddings are iteratively refined through local graph convolution and global boundary attention layers as described below, with a final readout layer used to predict local atomic and system-level properties.

Equivariant Local Graph Convolution: EBFormer updates atomic embeddings with local chemical information using SE(3) equivariant message-passing graph convolution introduced in NequIP Tan et al. (2025). Using an equivariant mechanism ensures preservation of the spatial symmetries of the nanostructure, which are important to ensure physical consistency of local property predictions.

Equivariant Global Boundary Attention: To encode global system structure and make each atom aware of its position within the global structure, EBFormer updates local atomic embeddings using cross-attention with boundary embeddings. Boundary embeddings are also refined through cross-attention with atomic nodes, allowing them to collect global geometric information and increase

model expressivity. Both updates employ distance-aware equivariant attention mechanisms implemented with E3NN Geiger et al. (2022), as described below.

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \sum_k \alpha_{ik} \left(\mathbf{b}_k^{(t)} \overset{\mathbf{W}_V^A}{\otimes} Y_J^{(l)}(\vec{\mathbf{d}}_{ik}) \right) \tag{1}$$

$$\mathbf{b}_k^{(t+1)} = \mathbf{b}_k^{(t)} + \sum_i \alpha_{ki} \left(\text{EquivLinear}_V^B \left(\mathbf{x}_i^{(t)} \middle| \text{MLP}_V^B \left(\mathbf{f}_{ik} \right) \right) \right) \tag{2}$$

$\text{EquivLinear}\left(\cdot|\mathbf{W}\right)$ denotes an E3NN equivariant linear layer parameterized by $\mathbf{W}$, used here to incorporate the distance of atom i to plane $\mathbf{B}_k$, represented as $\vec{\mathbf{d}}_{ik}$. The scalar distance is expanded over a trainable Bessel function $\mathbf{f}_{ik} = \mathbf{f}(|\vec{\mathbf{d}}_{ik}|)$, ensuring that boundary-to-atom features decay smoothly with distance so that atoms far from a boundary resemble bulk behavior. Equivariant features are combined with spherical harmonic expansions $Y_J^{(l)}$ of interatomic distance vectors using an E3NN fully-connected tensor product $\overset{\mathbf{W}}{\otimes}$ parameterized by $\mathbf{W}$. Attention scores α_{ki} are calculated using a softmax as follows:

$$\alpha_{ki} = \frac{\exp(\mathbf{q}_k \overset{\text{dot}}{\otimes} \mathbf{k}_i)}{\sum_j \exp\left(\mathbf{q}_k \overset{\text{dot}}{\otimes} \mathbf{k}_j \right)} \tag{3}$$

where $\overset{\text{dot}}{\otimes}$ is a parameterized fully-connected tensor product with scalar output Thomas et al. (2018). Query and keys are generated for boundary-to-atom and atom-to-boundary cross attention as:

$$\mathbf{q}_k = \text{EquivLinear}_Q^B \left(\mathbf{b}_k^{(t)} \right) \quad \mathbf{k}_i = \text{EquivLinear}_K^B \left(\mathbf{x}_i^{(t)} \middle| \text{MLP}_K^B \left(\mathbf{f}_{ik} \right) \right) \tag{4}$$

$$\mathbf{q}_i = \text{EquivLinear}_Q^A \left(\mathbf{x}_i^{(t)} \right) \quad \mathbf{k}_i = \mathbf{x}_i^{(t)} \overset{\mathbf{W}_K^A}{\otimes} Y_J^{(l)}(\mathbf{d}_{ik}) \tag{5}$$

where $\mathbf{W}_K^A = \text{MLP}_K\left(\mathbf{f}_{ik}\right)$ and $\mathbf{W}_V^A$, defined similarly, are the distance-dependent weights parameterizing the fully-connected tensor product.

Angular resolution of boundary planes is important to derive information such as cleaving planes, which has important impact on material properties Yang et al. (2010). To enable this resolution, $\ell > 0$ equivariance is included in the atom-to-boundary attention mechanism. However, readout uses only the $\ell = 0$ features, ensuring the network predictions remains rotationally and translationally invariant. Furthermore, as the attention mechanism is a weighted average, EBFormer is also invariant to supercell replication, enabling generalization to larger systems. Formal proofs of symmetry preservation are included in Appendix K. The boundary mechanism is visualized for the nanosheet and nanowire cases in Appendix E.1.

Readout: After generating the final atomic embeddings, material properties are generated per atom using a readout equivariant linear layer. For our application, we use two distinct readout heads to generate the atomic LDOS and LcDOS from the invariant features of the atomic embeddings. The system average DOS and cDOS are generated using mean-pooling on the predicted local quantities.

Model Training: We define two loss functions $\mathcal{L}_G$ and $\mathcal{L}_L$:

$$\mathcal{L}_G = \frac{1}{N} \sum_{j,E} \mathbf{w}_{DOS} \left|\left| \hat{D}^j(E) - D^j(E) \right|\right| + \mathbf{w}_{cDOS} \left|\left| \hat{J}_x^j(E) - J_x^j(E) \right|\right| \tag{6}$$

$$\mathcal{L}_L = \frac{1}{N} \sum_{j,a,E} \mathbf{w}_{DOS} \left|\left| \hat{D}^j(E,a) - D^j(E,a) \right|\right| + \mathbf{w}_{cDOS} \left|\left| \hat{J}_x^j(E,a) - J_x^j(E,a) \right|\right| \tag{7}$$

Here, $\mathcal{L}_G$ is the mean-absolute error over total DOS $D(E)$ and cDOS $J_x(E)$, while $\mathcal{L}_L$ is the MAE over their corresponding atomistic projections onto atom a. The weights $\mathbf{w}_{DOS}$ and $\mathbf{w}_{cDOS}$ balance the relative contribution of DOS and cDOS losses (a sensitivity analysis of their ratio is included in Appendix E). EBFormer is trained on either $\mathcal{L}_G$ or $\mathcal{L}_L$, with the latter explicitly encouraging the model to capture spatial variation of material properties over the nanostructures.

3.2 DATASET

Our dataset includes a variety of nanosheet and nanowire geometries, central building blocks of a variety of modern nanoelectronic devices Park et al. (2025); Liao & Smidt (2023); Huang et al. (2017). The nanosheet portion comprises 12,485 silicon and germanium structures, with cleaving planes over various orientations ($\langle 100 \rangle$, $\langle 110 \rangle$, $\langle 111 \rangle$, $\langle 211 \rangle$), nanosheet thickness scaled from 15 to 100 atomic layers, and various isotropic strains and point-defects introduced to the atomic structure. The nanowire dataset, demonstrating two dimensions of confinement compared to the one of nanosheets, includes 749 silicon and germanium structures with each lateral dimension ranging from 10-30 atomic layers and similarly varied isotropic strains. Dataset generation required $\sim$4 months computational wall time. Further details regarding generation, statistics, and visualizations are included in the Appendix C and D.

4 EXPERIMENTS AND RESULTS

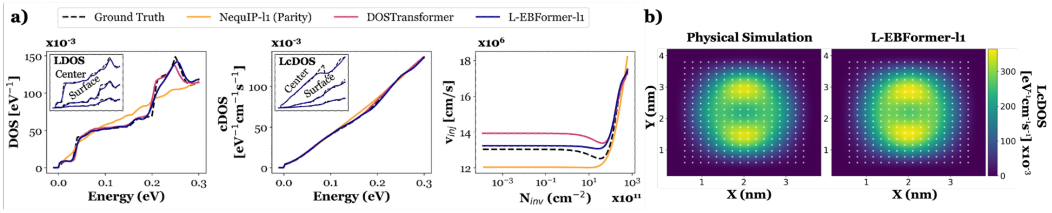

Figure 3: **Qualitative Visualization** on (a) the nanosheet interpolation task on a structure from the test set with 35 atomic layers of silicon cleaved along $\langle 100 \rangle$ and (b) the LcDOS from physical simulation and L-EBFormer-$\ell 1$ at $E = 0.22eV$ on a test structure from the nanowire interpolation task, with 16x18 atomic layers. White dots represent the location of silicon nuclei. LcDOS represents the geometric distribution of the wire's current-carrying capacity for electrons at a certain energy, while the LDOS (pictured for this system in Appendix F.2) represents how electrons distribute spatially.

We compare three different configurations of EBFormer on our datasets; an SE(3) invariant model trained to minimize $\mathcal{L}_G$ (EBFormer-$\ell 0$), an $\ell = 1$ equivariant model trained to minimize $\mathcal{L}_G$ (EBFormer-$\ell 1$), and an $\ell = 1$ equivariant model trained to minimize local loss $\mathcal{L}_L$ (L-EBFormer-$\ell 1$). As an ablation of the global attention mechanism, and a comparison against local graph networks, we benchmark against NequIP-$\ell 0$ and $\ell 1$, using identical configuration to the local convolution mechanism of EBFormer-$\ell 0$/$\ell 1$. We also benchmark against DOSTransformer, a state-of-the-art, linearly-scaling model for spectral quantity prediction that incorporates global information, but using a rotationally-invariant, geometry-agnostic attention mechanism. Finally, we include an MLP as a baseline. We maintain parity in the number of trainable parameters in all the models (except the ablation networks) by modifying the hidden dimension to reduce confounding effects from scaling Qu & Krishnapriyan (2024). All models are coded in PyTorch-CUDA and trained on an NVIDIA A100 GPU. Further details regarding model hyperparameters are included in Appendix E. Qualitative visualizations of experimental results are included in Appendix F.1.

Nanosheet Interpolation Task: We begin by evaluating the model in the in-distribution regime. We use an 80/10/10 train/validation/test split on the nanosheet dataset, with all models trained on $\mathcal{L}_G$ with a 3:1 ratio of $\mathbf{w}_{\text{DOS}} : \mathbf{w}_{\text{cDOS}}$, except L-EBFormer-l1 that was trained on $\mathcal{L}_L$ with the same ratio. The experiment results are summarized in Table 1.

Results: We make the following observations: **1)** Purely local models, such as the MLP and NequIP baseline approaches, show significantly worse performance in most downstream tasks compared to those that incorporate global information (DOSTransformer and EBFormer). However, NequIP shows that incorporation of local geometric information provides a significant improvement to a geometry-agnostic MLP approach. **2)** Compared to the ablation networks, the inclusion of global structural information through the boundary nodes provides a significant improvement in inference tasks. The weight-parity experiments show that simply increasing the network size in a local approach is insufficient to reach low MAEs. **3)** Comparing EBFormer-$\ell 0$ and -$\ell 1$, we see that the incorporation of higher-order geometric information improved performance, even when controlling for the number of parameters. We hypothesize that this is due to the ability to resolve cleaving

Model	Params [M]	DOS [$\times 10^{-3}$]		cDOS [$\times 10^{-3}$]		N_{inv} [$\times 10^{11}$]		I_{inv} [$\times 10^{-3}$]		v_{inj} [$\times 10^{6}$]	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MLP	9.25	14.82 (0.02)	23.03 (0.27)	12.07 (0.10)	20.22 (0.51)	2.69 (0.01)	6.21 (0.04)	7.26 (0.05)	19.99 (0.62)	1.63 (0.05)	2.63 (0.08)
NequIP-ℓ0 (Ablation)	1.39	5.59 (0.01)	9.14 (0.08)	2.40 (0.05)	4.51 (0.06)	0.45 (0.00)	1.35 (0.03)	1.43 (0.03)	4.26 (0.10)	0.87 (0.01)	1.97 (0.19)
NequIP-ℓ1 (Ablation)	1.18	5.50 (0.05)	8.92 (0.04)	2.31 (0.06)	4.31 (0.11)	0.44 (0.01)	**1.30** (0.04)	1.36 (0.04)	4.06 (0.10)	0.88 (0.02)	1.83 (0.08)
NequIP-ℓ0 (Parity)	9.98	5.53 (0.01)	9.17 (0.05)	2.29 (0.00)	4.39 (0.04)	0.44 (0.00)	1.36 (0.02)	1.37 (0.01)	4.18 (0.03)	0.91 (0.04)	2.24 (0.50)
NequIP-ℓ1 (Parity)	9.46	5.46 (0.06)	8.91 (0.05)	2.30 (0.08)	4.27 (0.07)	0.44 (0.01)	1.31 (0.05)	1.36 (0.06)	4.02 (0.10)	0.85 (0.02)	1.70 (0.31)
DOSTransformer	9.40	**2.80** (0.15)	6.59 (0.12)	1.55 (0.09)	3.80 (0.09)	0.37 (0.02)	1.33 (0.04)	0.90 (0.06)	3.51 (0.04)	0.80 (0.16)	2.34 (1.39)
EBFormer-ℓ0	9.48	3.27 (0.10)	6.89 (0.07)	1.32 (0.05)	3.93 (0.04)	0.33 (0.01)	1.38 (0.01)	0.78 (0.03)	3.75 (0.05)	0.61 (0.02)	1.25 (0.14)
EBFormer-ℓ1	9.35	3.17 (0.06)	6.74 (0.14)	**1.26** (0.03)	3.76 (0.16)	0.32 (0.01)	1.38 (0.05)	0.75 (0.02)	3.61 (0.11)	0.60 (0.01)	1.35 (0.28)
L-EBFormer-ℓ1	9.35	2.85 (0.06)	**6.21** (0.10)	**1.26** (0.04)	**3.37** (0.11)	**0.31** (0.00)	**1.30** (0.03)	**0.67** (0.02)	**3.24** (0.13)	**0.49** (0.01)	**0.90** (0.07)

Table 1: **Nanosheet Interpolation Errors** DOS (eV^{-1}), cDOS ($\text{eV}^{-1}\text{cm}^{-1}\text{s}^{-1}$), N_{inv} (cm^{-2}), I_{inv} (mA/μm), and injection velocity (cm/s) errors on the interpolation task are shown below. Errors are averaged over three seeds with standard deviation in parentheses. Bold figures denote the best (lowest) mean in each sub-column. L-EBFormer LDOS errors are $3.19(0.06)/7.67(0.14)$ MAE/RMSE, and LcDOS $1.99(0.04)/4.92(0.12)$ in the same units as DOS/cDOS.

orientation with $\ell = 1$ equivariance, as discussed in Section 3.1. **4)** The more complex task of predicting local DOS/cDOS *improves* downstream mean-pooled quantity errors. We hypothesize that the added information of spatial variation of quantities provides useful information regarding the impact of structure on LDOS/LcDOS, and therefore on the total nanostructure properties, a nontrivial result which to the best of our knowledge has not been observed before. Qualitatively, we see the L-EBFormer reproduces LDOS and LcDOS with high fidelity (inset panels in Figure 3), achieving similar order of error as DOS/cDOS predictions, which are spatial averages of the local quantities. Note that DOSTransformer is unable to predict local quantities, and therefore cannot leverage this information. **5)** EBFormer-ℓ1 shows slightly lower performance compared to DOSTransformer in DOS prediction, but comparable or improved performance in cDOS and downstream tasks, notably in injection velocity. Injection velocity is a ratio of quantities derived from DOS and cDOS, and thus depends on not only the accuracy but the mutual consistency of the two values. High performance on this downstream tasks indicates physically consistent covariance of EBFormer's predictions, and thus a more rich learned embedding. Analysis of correlation and consistency of these quantities are included in Appendix G. **6)** L-EBFormer-ℓ1 shows significantly improved performance on cDOS and all downstream tasks, while showing comparable performance to DOSTransformer on the DOS task. This demonstrates that incorporating geometric information in global attention improves material property prediction in structured materials.

4.1 Out-of-Distribution Nanosheet Inference

Compared to bulk materials in which a single unit-cell fully determines system properties, nanostructures incorporate a large number of atoms with unique local environments that define the system behavior. Simulating nanostructures involves very large unit-cells, and therefore is computationally difficult. Models that are capable of capturing the behavior of much larger systems from smaller, easier to simulate structures would therefore be of great value.

Nanosheet Extrapolation Task: To quantify extrapolative performance in the size of the system, we train the model on silicon nanosheets from 15 to 45 atomic layers thick with varied strain, cleaving, and point-defects. We test and validate on nanosheets from 46-100 layers. All test errors are generated using the model with the minimum validation loss during training.

Results: Errors on the extrapolation task are shown in Table 2. We notice first that the weight-parity local-convolution network is more robust to the distribution shift than DOSTransformer. This is likely because the GNN is unaware of the global structure, and therefore predicts similar local properties compared to smaller structures, which are good first-order approximations. DOSTransformer includes global information with geometry-agnostic attention, and therefore cannot leverage physical information regarding variation of properties due to geometry well. EBFormer demonstrates that our mechanism of symmetry-aware inclusion of geometry in global attention is better able to capture the physics of quantum confinement for application to out-of-distribution situations. Finally, we notice that L-EBFormer generally performs more poorly than the global prediction approach. We

Model	Params [M]	DOS $[\times 10^{-3}]$		cDOS $[\times 10^{-3}]$		N_{inv} $[\times 10^{11}]$		I_{inv} $[\times 10^{-3}]$		v_{inj} $[\times 10^{6}]$	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MLP	0.99	27.69 (5.88)	41.18 (9.80)	23.88 (1.56)	37.09 (1.72)	2.38 (0.28)	5.15 (0.20)	8.39 (0.43)	17.93 (1.72)	7.94 (0.12)	13.20 (1.35)
NequIP-ℓ1 (Parity)	1.04	8.82 (0.30)	11.70 (0.33)	4.39 (0.14)	6.05 (0.30)	1.22 (0.02)	2.44 (0.06)	2.68 (0.06)	5.64 (0.34)	2.37 (0.04)	2.98 (0.10)
DOSTransformer	0.95	11.88 (1.18)	16.89 (0.79)	7.11 (0.63)	10.16 (0.50)	1.42 (0.20)	3.21 (0.28)	4.20 (0.92)	8.57 (1.66)	4.99 (2.31)	7.34 (3.61)
EBFormer-ℓ0	0.94	9.12 (0.19)	12.42 (0.17)	4.37 (0.26)	7.41 (0.23)	1.35 (0.17)	3.04 (0.17)	2.42 (0.16)	6.61 (0.21)	1.00 (0.17)	1.32 (0.14)
EBFormer-ℓ1	0.91	7.97 (0.23)	10.89 (0.25)	**3.37** (0.23)	**5.39** (0.26)	**1.00** (0.09)	**2.29** (0.10)	**1.81** (0.16)	**4.69** (0.25)	**0.89** (0.09)	**1.24** (0.05)
L-EBFormer-ℓ1	0.91	**7.52** (0.24)	**10.55** (0.44)	4.58 (0.29)	6.98 (0.62)	1.03 (0.02)	2.48 (0.15)	2.76 (0.35)	6.60 (0.87)	0.97 (0.08)	1.35 (0.06)

Table 2: **Nanosheet Extrapolation Errors** Model errors on the nanosheet extrapolation task (training on 15-45 layers, test on 46-100 layers), in the same units as Table 1. Errors are averaged over three seeds with standard deviation in parentheses. L-EBFormer LDOS errors are $9.49(0.23)/14.53(0.74)$ MAE/RMSE, and LcDOS $7.01(0.41)/10.85(0.57)$.

hypothesize that this is due to the added variability of the local quantities compared to the system-level DOS and cDOS, which may lead to decreased robustness under distribution shift (discussed in Appendix I). Qualitative visualization of the extrapolation task is included in Appendix F.1.

While there is no standard metric for required accuracy, architectural decisions between nodes typically yield 15-60% impact in quantities such as power, current, and electron mobility Park et al. (2025). Resolving the impact of these decisions requires less than around 10% errors in predictions. DOSTransformer, the strongest model after EBFormer, demonstrated 17% nRMSE in-distribution and 52% in out-of-distribution nanosheet inference of injection velocity, while EBFormer demonstrated errors of 6% in-distribution and *9%* in the OOD case (statistics from Appendix D).

4.2 In-Distribution Nanowire Inference

Model	Params [M]	DOS $[\times 10^{-3}]$		cDOS $[\times 10^{-3}]$		N_{inv} $[\times 10^{11}]$		I_{inv} $[\times 10^{-3}]$		v_{inj} $[\times 10^{6}]$	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MLP	0.99	27.73 (1.71)	43.41 (0.54)	12.76 (1.11)	19.22 (0.63)	0.29 (0.01)	0.85 (0.04)	6.78 (0.60)	17.81 (0.46)	2.14 (0.31)	2.75 (0.40)
NequIP-l1 (Parity)	1.04	20.60 (0.04)	35.91 (0.12)	2.96 (0.09)	4.66 (0.14)	0.14 (0.00)	0.33 (0.01)	0.93 (0.06)	2.53 (0.37)	1.68 (0.02)	2.17 (0.05)
DOSTransformer	0.95	20.69 (2.54)	34.73 (2.25)	6.18 (2.21)	8.23 (2.62)	0.14 (0.02)	0.32 (0.10)	3.91 (1.93)	7.18 (3.27)	2.89 (1.11)	3.34 (1.11)
EBFormer-l0	0.94	16.66 (0.29)	31.37 (0.66)	1.59 (0.21)	2.90 (0.55)	**0.10** (0.01)	0.25 (0.01)	**0.51** (0.11)	**1.54** (0.52)	1.52 (0.02)	1.93 (0.02)
EBFormer-l1	0.91	16.07 (1.20)	30.64 (1.14)	**1.50** (0.35)	**2.62** (0.66)	**0.10** (0.00)	**0.24** (0.01)	**0.51** (0.15)	**1.54** (0.52)	1.53 (0.05)	1.96 (0.06)
L-EBFormer-l1	0.91	**14.34** (0.13)	**29.30** (0.24)	1.70 (0.06)	2.93 (0.18)	**0.10** (0.00)	**0.24** (0.01)	0.59 (0.05)	1.69 (0.23)	**1.44** (0.02)	**1.84** (0.02)

Table 3: **Nanowire Interpolation Errors** Model errors on the nanowire interpolation task, in the same units as Table 1. Errors are averaged over three seeds with standard deviation in parentheses. L-EBFormer LDOS errors are $12.89(0.09)/35.68(0.14)$ MAE/RMSE, and LcDOS $4.05(0.08)/9.37(0.21)$.

Finally, we benchmark EBFormer against other architectures on the nanowire dataset (Table 3). Nanowires introduce an additional dimension of confinement, leading to distinct physical effects compared to the nanosheet case Huang et al. (2017); Jiang et al. (2008). We adopt the same 80/10/10 train/validation/test split and loss functions as in the nanosheet interpolation experiment. We observe that the local convolution network outperforms DOSTransformer despite having no nonlocal structural information. We attribute this to the flexible inductive bias of DOSTransformer relative to NequIP, which in this data-sparse regime likely leads to greater model variance and overfitting. In contrast, both the global and local variants of EBFormer generalize well to the nanowire task, achieving significantly better performance than other architectures. L-EBFormer also quantitatively shows good reproduction of spatial variation of local quantities (pictured for cDOS in Figure 3). Further discussion and numerical analysis of local inference is included in Appendix I.

4.3 Model Parameterization and Learning Curves

Figure 4 shows learning curves and parameter efficiency of EBFormer compared to NequIP and DOSTransformer on the nanosheet interpolation task. We notice that EBFormer achieves low losses in cDOS and injection velocity with a relatively small number of parameters, implying the model

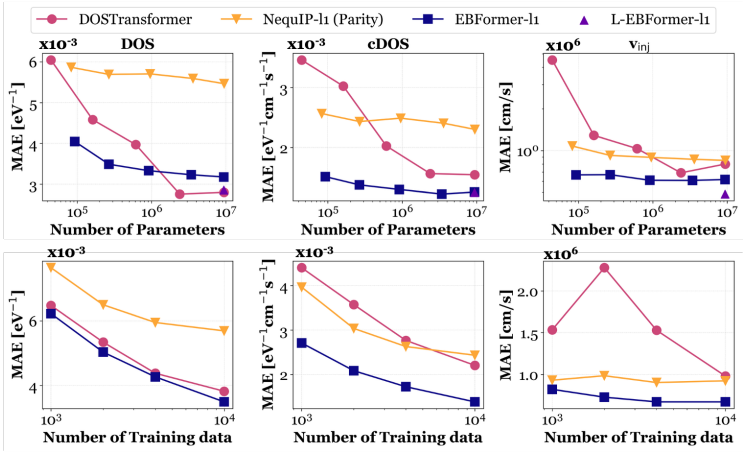

Figure 4: **Parameterization & Learning Curves**: Various model configurations are tested on the nanosheet interpolation task. Learning curves models maintain weight parity at ∼1M parameters. We observe superior performance of EBFormer in the low-data regime, and both that EBFormer performance continues to scale with more parameters while DOSTransformer plateaus, as well as superior performance in the low parameter regime, indicating the suitability of the inductive bias.

inductive bias is well-suited to the inference task. We also see that training on local projections of DOS and cDOS (L-EBFormer-$\ell 1$) shows marked improvement in predictions with the same number of model parameters (further discussed in Appendix I). While DOSTransformer achieves superior DOS losses at high parameterization, we notice that EBFormer continues to improve with scaling, and achieves superior downstream injection velocity performance. Improved injection velocity performance is also apparent in the learning curves; we see that EBFormer achieves better injection velocity losses with 1k training data than DOSTransformer with 10k. In contrast, NequIP, a purely local convolutional model, saturates with few data. By augmenting local convolution with global structural context, EBFormer demonstrates a significant shift in learning efficiency and predictive accuracy, surpassing both NequIP and DOSTransformer. Future steps for our work include demonstrating generalization of the boundary mechanism on a more comprehensive material dataset (we provide a first step through a transfer learning task in Appendix J), and reducing the computational cost of the boundary mechanism and E3NN SO(3) equivariant-operations through SO(2) or bespoke $\ell = 0$ and $\ell = 1$ equivariant operations (discussed further in Appendix H).

5 Conclusion

In this work, we propose EBFormer, a physically-motivated architecture for atomistic property prediction in confined microelectronic systems. Through equivariant geometry-aware cross attention between atom nodes and boundary nodes representing interfacial planes, we demonstrate high performance in in-distribution atomistic material property predictions and downstream device-level tasks in silicon and germanium nanostructures representing modern transistor channels. We also show high-fidelity reproduction of spatial variations of material property within systems with different dimensionalities, and prediction of geometry-induced effects in electronic properties. Finally, we demonstrate superlative performance compared to state-of-the-art models in the data-limited and extrapolative regimes, demonstrating the strength of our model's inductive bias. EBFormer preserves linear scaling in system size, and achieves 4–5 orders of magnitude acceleration over conventional atomistic simulations (Figure 5). Through a combination of scalability, predictive accuracy, and the capacity to model geometric effects at the nanoscale regime, EBFormer presents a concrete framework for the rapid design and modeling of next-generation nanoscale devices.

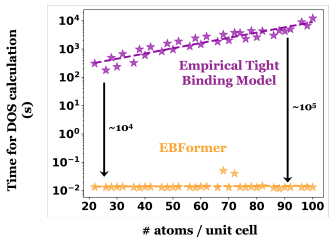

Figure 5: **Acceleration**

6 ETHICS STATEMENT

We acknowledge that we have carefully reviewed the ICLR ethics statement, and assure that all studies in this work adhere to the outlined guidelines. EBFormer is a neural network architecture designed to capture quantum confinement effects in nanostructures. Nanostructures are prevalent across various domains, including materials science and semiconductor device engineering. By enabling accelerated and accurate modeling of these systems, EBFormer can significantly enhance the design cycle for next-generation, energy-efficient devices. This has the potential to positively impact society by promoting sustainable technologies. However, training such large neural networks, particularly those incorporating equivariant features and attention operations, can be very computationally intensive and energy demanding. As global energy consumption from machine learning continues to rise, such models risk contributing to environmental degradation if not paired with sustainable computing practices.

7 REPRODUCIBILITY

The code is available at https://github.com/EBFormer/EBFormer, containing the neural network configuration files, training scripts, and test scripts to reproduce our results. Appendix sections C, D, and E also contain the dataset splits and the implementation details about the neural network architecture. The datasets for all tasks are available at Zenodo.

8 ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE 2146752. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Katsumi Abe, Toshio Kamiya, and Hideo Hosono. Quantum confinement effects in amorphous in–ga–zn–o thin-film transistors with quantum well channel. *Applied Physics Letters*, 121(22): 222104, 11 2022. ISSN 0003-6951. doi: 10.1063/5.0132431. URL https://doi.org/10.1063/5.0132431.

Uri Alon and Eran Yahav. On the bottleneck of graph neural networks and its practical implications. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=i80OPhOCVH2.

Ilyes Batatia, Philipp Benner, Yuan Chiang, Alin M. Elena, and Dávid P. Kovács et al. A foundation model for atomistic materials chemistry, 2024. URL https://arxiv.org/abs/2401.00096.

Simon Batzner, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nature Communications*, 13, 2022.

Pratik Brahma, Krishnakumar Bhattaram, and Sayeef Salahuddin. Accelerated modelling of interfaces for electronic devices using graph neural networks, 2023.

Rich Caruana. Multitask Learning. *Machine Learning*, 28(1):41–75, July 1997. ISSN 1573-0565. doi: 10.1023/A:1007379606734. URL https://doi.org/10.1023/A:1007379606734.

Chi Chen and Shyue Ping Ong. A universal graph deep learning interatomic potential for the periodic table. *Nature Computational Science*, 2, 2022.

Kamal Choudhary and Brian DeCost. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials*, 7, 2021.

Kamal Choudhary, Kevin F. Garrity, Steven T. Hartman, Ghanshyam Pilania, and Francesca Tavazza. Efficient computational design of two-dimensional van der waals heterostructures: Band alignment, lattice mismatch, and machine learning. *Physical Review Materials*, 7(1):014009, 2023. doi: 10.1103/PhysRevMaterials.7.014009.

Robert Cameron Colwell. Xxvi. the vibrations of membranes and plates. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 12(76):320–328, 1931.

Supriyo Datta. *Quantum Transport: Atom to Transistor*. Cambridge University Press, 2005.

A. I. Ekimov and A. A. Onushchenko. Quantum size effect in three-dimensional microscopic semiconductor crystals. *JETP Letters*, 118, 2023.

Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks, 2020.

Mario Geiger, Tess Smidt, Alby M., Benjamin Kurt Miller, Wouter Boomsma, Bradley Dice, Kostiantyn Lapchevskyi, Maurice Weiler, Michał Tyszkiewicz, Simon Batzner, Dylan Madisetti, Martin Uhrin, Jes Frellsen, Nuri Jung, Sophia Sanborn, Mingjian Wen, Josh Rackers, Marcel Rød, and Michael Bailey. Euclidean neural networks: e3nn, 2022.

Sten Haastrup, Mikkel Strange, Mohnish Pandey, Thorsten Deilmann, Per S. Schmidt, Nicki F. Hinsche, Morten N. Gjerding, Daniele Torelli, Peter M. Larsen, Anders C. Riis-Jensen, Jakob Gath, Karsten W. Jacobsen, Jens Jørgen Mortensen, Thomas Olsen, and Kristian S. Thygesen. The computational 2d materials database: high-throughput modeling and discovery of atomically thin crystals. *2D Materials*, 5(4):042002, 2018. doi: 10.1088/2053-1583/aacfc1.

Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. STRATEGIES FOR PRE-TRAINING GRAPH NEURAL NETWORKS. 2020.

Ya-Chi Huang, Meng-Hsueh Chiang, Shui-Jinn Wang, and Jerry G. Fossum. Gaafet versus pragmatic finfet at the 5nm si-based cmos technology node. *IEEE Journal of the Electron Devices Society*, 5(3):164–169, 2017. doi: 10.1109/JEDS.2017.2689738.

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, 1998.

Xiangwei Jiang, Hui-Xiong Deng, Jun-Wei Luo, Shu-Shen Li, Jian-Bai Xia, and Lin-Wang Wang. A fully 3d atomistic quantum mechanical study on random dopant induced effects in 25nm mosfets. *IEEE Transactions on Electron Devices - IEEE TRANS ELECTRON DEVICES*, 55, 2008.

Arthur Kosmala, Johannes Gasteiger, Nicholas Gao, and Stephan Günnemann. Ewald-based long-range message passing for molecular graphs, 2023.

Dávid Péter Kovács, J. Harry Moore, Nicholas J. Browning, Ilyes Batatia, Joshua T. Horton, Yixuan Pu, Venkat Kapil, William C. Witt, Ioan-Bogdan Magdău, Daniel J. Cole, and Gábor Csányi. Mace-off: Transferable short range machine learning force fields for organic molecules, 2025. URL `https://arxiv.org/abs/2312.15211`.

Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, and Zhuowen Tu. Deeply-Supervised Nets. In Guy Lebanon and S. V. N. Vishwanathan (eds.), *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics*, volume 38 of *Proceedings of Machine Learning Research*, pp. 562–570, San Diego, California, USA, 09–12 May 2015. PMLR. URL `https://proceedings.mlr.press/v38/lee15a.html`.

Hyeongu Lee, Yucheol Cho, Seonghyeok Jeon, and Mincheol Shin. First-principles-based quantum transport simulations of interfacial point defect effects on inas nanowire tunnel fets. *IEEE Transactions on Electron Devices*, 68, 2021.

Namkyeong Lee, Heewoong Noh, Sungwon Kim, Dongmin Hyun, Gyoung S. Na, and Chanyoung Park. Density of states prediction of crystalline materials via prompt-guided multi-modal transformer. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

Yunhai Li, Zhen Zhan, Xueheng Kuang, Yonggang Li, and Shengjun Yuan. Tbplas: A tight-binding package for large-scale simulation. *Computer Physics Communications*, 285:108632, 2023. ISSN 0010-4655. doi: https://doi.org/10.1016/j.cpc.2022.108632. URL https://www.sciencedirect.com/science/article/pii/S0010465522003514.

Yi-Lun Liao and Tess Smidt. Equiformer: Equivariant graph attention transformer for 3d atomistic graphs. In *The Eleventh International Conference on Learning Representations*, 2023.

Yu-Hung Liao, Khandker Akif Aabrar, Wriddhi Chakraborty, Wenshen Li, Suman Datta, and Sayeef Salahuddin. Large injection velocities in highly scaled, fully depleted silicon on insulator transistors. *IEEE Electron Device Letters*, 2022.

Zhongwei Liu, Zhimin Zhang, Xuwei Liu, Mingjia Yao, Xin He, Yuanhui Sun, Xin Chen, and Lijun Zhang. Monolayer two-dimensional materials database (ML2DDB) and applications. *arXiv preprint*, 2025. doi: 10.48550/arXiv.2507.00584. URL https://arxiv.org/abs/2507.00584.

Shengjie Luo, Tianlang Chen, and Aditi S. Krishnapriyan. Enabling efficient equivariant operations in the fourier basis via gaunt tensor products. In *The Twelfth International Conference on Learning Representations*, 2024.

Albert Musaelian, Simon Batzner, Anders Johansson, Lixin Sun, Cameron J. Owen, Mordechai Kornbluth, and Boris Kozinsky. Learning local equivariant representations for large-scale atomistic dynamics. *Nature Communications*, 14, 2023.

A.J Nozik. Quantum dot solar cells. *Physica E: Low-dimensional Systems and Nanostructures*, 14(1):115–120, 2002. ISSN 1386-9477. doi: https://doi.org/10.1016/S1386-9477(02)00374-0. URL https://www.sciencedirect.com/science/article/pii/S1386947702003740.

Paras, Kushal Yadav, Prashant Kumar, Dharmasanam Ravi Teja, Sudipto Chakraborty, Monojit Chakraborty, Soumya Sanjeeb Mohapatra, Abanti Sahoo, Mitch M. C. Chou, Chi-Te Liang, and Da-Ren Hang. A review on low-dimensional nanomaterials: Nanofabrication, characterization and applications. *Nanomaterials*, 13, 2023.

Jong Ho Park, Nirmaan Shanker, Suraj Cheema, Shang-Lin Hsu, Aditya Varma, Chia-Chun Lee, Chirag Garg, Urmita Siker, Li-Chen Wang, Chenming Hu, and Sayeef Salahuddin. Demonstration of high transconductance gate-all-around transistors using negative capacitance 'super high-k' gate stack. *IEEE Electron Device Letters*, 46, 2025.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library, 2019.

E. Prodan and W. Kohn. Nearsightedness of electronic matter. *Proceedings of the National Academy of Sciences*, 102(33):11635–11638, 2005. doi: 10.1073/pnas.0505436102. URL https://www.pnas.org/doi/abs/10.1073/pnas.0505436102.

Eric Qu and Aditi S. Krishnapriyan. The importance of being scalable: Improving the speed and accuracy of neural network interatomic potentials across chemical domains. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

Anisur Rahman, Jing Guo, S. Datta, and M.S Lundstrom. Theory of ballistic nanotransistors. *Electron Devices, IEEE Transactions on*, 50, 2003.

A.J. Reddy, J.V. Chan, T.A. Burr, R. Mo, C.P. Wade, C.E.D. Chidsey, J. Michel, and L.C. Kimerling. Defect states at silicon surfaces. *Physica B: Condensed Matter*, 273-274:468–472, 1999. ISSN 0921-4526. doi: https://doi.org/10.1016/S0921-4526(99)00527-X. URL https://www.sciencedirect.com/science/article/pii/S092145269900527X.

Victor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. *CoRR*, abs/2102.09844, 2021. URL https://arxiv.org/abs/2102.09844.

Kristof Schütt, Pieter-Jan Kindermans, Huziel Enoc Sauceda Felix, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

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, oct 2019.

M. Soriano and J. J. Palacios. Theory of projections with nonorthogonal basis sets: Partitioning techniques and effective hamiltonians. *Phys. Rev. B*, 90, 2014.

Daniele Stradi, Ulrik G. Vej-Hansen, Petr A. Khomyakov, Maeng-Eun Lee, Gabriele Penazzi, Anders Blom, Jess Wellendorff, Søren Smidstrup, and Kurt Stokbro. Atomistic modeling of nanoscale ferroelectric capacitors using a density functional theory and non-equilibrium green's-function method. In *2019 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2019.

Chuin Wei Tan, Marc L. Descoteaux, Mit Kotak, Gabriel de Miranda Nascimento, Seán R. Kavanagh, Laura Zichi, Menghang Wang, Aadit Saluja, Yizhong R. Hu, Tess Smidt, Anders Johansson, William C. Witt, Boris Kozinsky, and Albert Musaelian. High-performance training and inference for deep equivariant interatomic potentials, 2025.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation- and translation-equivariant neural networks for 3d point clouds, 2018.

G. Tsutsui, S. Mochizuki, N. Loubet, S. W. Bedell, and D. K. Sadana. Strain engineering in functional materials. *AIP Advances*, 9(3):030701, 03 2019. ISSN 2158-3226. doi: 10.1063/1.5075637. URL https://doi.org/10.1063/1.5075637.

Oliver T. Unke, Stefan Chmiela, Michael Gastegger, Kristof T. Schütt, Huziel E. Sauceda, and Klaus-Robert Müller. Spookynet: Learning force fields with electronic degrees of freedom and nonlocal effects. *Nature Communications*, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

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. ISSN 0022-3697. doi: https://doi.org/10.1016/0022-3697(83)90064-1. URL https://www.sciencedirect.com/science/article/pii/0022369783900641.

Tong Wang, Xinheng He, Mingyu Li, Yatao Li, Ran Bi, Yusong Wang, Chaoran Cheng, Xiangzhen Shen, Jiawei Meng, He Zhang, Haiguang Liu, Zun Wang, Shaoning Li, Bin Shao, and Tie-Yan Liu. Ab initio characterization of protein molecular dynamics with ai2bmd. *Nature*, 635 (8040):1019–1027, 2024a. doi: 10.1038/s41586-024-08127-z. URL https://doi.org/10.1038/s41586-024-08127-z.

Yusong Wang, Chaoran Cheng, Shaoning Li, Yuxuan Ren, Bin Shao, Ge Liu, Pheng-Ann Heng, and Nanning Zheng. Neural p^3m: A long-range interaction modeling enhancer for geometric GNNs. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024b. URL https://openreview.net/forum?id=ncqauwSyl5.

Yusong Wang, Tong Wang, Shaoning Li, Xinheng He, Mingyu Li, Zun Wang, Nanning Zheng, Bin Shao, and Tie-Yan Liu. Enhancing geometric representations for molecules with equivariant vector-scalar interactive message passing. *Nature Communications*, 15(1):313, January 2024c.

Logan Ward, Ankit Agrawal, Alok Choudhary, and Christopher Wolverton. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials*, 2, 2016.

Chen Hao Xia, Manasa Kaniselvan, Alexandros Nikolaos Ziogas, Marko Mladenović, Rayen Mahjoub, Alexander Maeder, and Mathieu Luisier. Learning the electronic hamiltonian of large atomic structures. In *Forty-second International Conference on Machine Learning*, 2025. URL https://openreview.net/forum?id=WGejWCgrpD.

Chenghui Xia, Weiwei Wu, Ting Yu, Xiaobin Xie, Christina van Oversteeg, Hans C. Gerritsen, and Celso de Mello Donega. Size-dependent band-gap and molar absorption coefficients of colloidal cuins2 quantum dots. *ACS Nano*, 12(8), 2018.

Tian Xie and Jeffrey C. Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.*, 120, 2018.

Keqiang Yan, Yi Liu, Yuchao Lin, and Shuiwang Ji. Periodic graph transformers for crystal material property prediction. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022.

Keqiang Yan, Cong Fu, Xiaofeng Qian, Xiaoning Qian, and Shuiwang Ji. Complete and efficient graph transformers for crystal material property prediction. In *The Twelfth International Conference on Learning Representations*, 2024.

Peizhen Yang, W.S. Lau, Seow Wei Lai, V.L. Lo, S.Y. Siah, and L. Chan. Effects of switching from $\langle 110 \rangle$ to $\langle 100 \rangle$ channel orientation and tensile stress on n-channel and p-channel metal–oxide-semiconductor transistors. *Solid-State Electronics*, 54, 2010.

Lipen Yuan. What is dtco?: An introduction to design-technology co-optimization - taiwan semiconductor manufacturing company limited, Jun 2022. URL https://www.tsmc.com/english/news-events/blog-article-20220615.

APPENDIX

A ARCHITECTURAL MOTIVATION

Our work presents a generalizable framework to incorporate the effects of atomic interactions with boundary planes on local and global electronic structural variation. Specifically, we focus on structured materials with long-range order such as those dictating the electronic properties of modern nanoelectronic devices, which can be comprised of hundreds of thousands of atoms. In such contexts, favorable asymptotic scaling is essential, and warrants careful engineering of the model inductive

bias. The novelty of our approach lies in explicitly introducing boundaries into atomic graph networks, and introducing a framework to communicate requisite long-range information to and from atomic nodes via equivariant attention mechanisms to capture geometric confinement. Crucially, we do this in a manner preserving O(N) asymptotic scaling, and demonstrate high accuracy and extensibility to nanosheets and nanowires, including in OOD situations. In the following paragraphs, we motivate the main architectural decisions of EBFormer as both necessary and sufficient, and emphasize the suitability of its inductive bias to our task of modeling atomic interactions with boundaries.

Necessity of our Architectural Components The local electronic properties of an atom in a nanostructure depend on the local atomic environment as well as the global material structure. For example, the local density of states in a crystalline nanosheet is determined by local properties such as the system lattice constant, elemental composition of the unit-cell, local atomic defects, and others. To capture such local effects, we motivated the utilization of a GNN with a local cutoff.

In addition to local effects, nanostructure critical dimensions can be significantly smaller than electronic de Broglie wavelengths and phase decay lengths, which leads to quantum confinement Abe et al. (2022). To capture the long-range information of confinement, we introduced boundary nodes, vector embeddings to collect and communicate information regarding geometric confinement. We utilized a cross-attention mechanism to exchange the heterogeneous information between the boundary embeddings and atom embedding while incorporating atom-boundary distance, a crucial factor that determines confinement. Furthermore, we utilize Bessel expansions over the boundary-atom distance, ensuring that interactions asymptotically decay as boundaries move far from atoms, leading local material properties to approach bulk behavior. In our main text, we showed that communication with the boundary planes was necessary to achieve high accuracy with our ablations with NequIP, the underlying local GNN. Furthermore, we demonstrated the necessity of an equivariant interaction with the boundary planes through another ablation with EBFormer-l0, which enables encoding angular information to determine cleaving direction (the relative orientation of the boundary plane with the crystal structure).

Sufficiency of our Architectural Components Limiting attention to only boundaries, however, is a strong inductive bias. Currently, we only include local and boundary information. It is possible that this is insufficient, and information about all atoms in the structure are required to be communicated (for example with a mechanism such as in Neural P^3M Wang et al. (2024b)). To answer this question, we relaxed the inductive bias of boundary-only attention to permit all atom-to-atom interactions, replacing the boundary interaction step with a fully trainable, equivariant, distance-conditioned self-attention mechanism identical to the boundary node interaction mechanism but between all atoms instead. In this way, we could visualize the attention matrices between atoms to determine if the interatomic interactions are dense or have some pattern or sparsity.

Model	Params (M)	DOS MAE	DOS RMSE	cDOS MAE	cDOS RMSE
NequIP-l1 (Parity)	0.27	5.69	9.19	2.43	4.48
DOSTransformer	0.27	3.82	7.58	2.20	4.63
Self-Attention-l1	0.27	3.67	6.96	**1.34**	**3.33**
EBFormer-l1	0.27	**3.49**	**6.94**	1.38	3.77

Table 4: Model size and errors for DOS and cDOS for the self-attention network compared to NequIP (which doesn't include nonlocality), DOSTransformer (nonlocality without geometry information), and EBFormer, which only includes boundary attention

Table 4 presents the errors of the self-attention architecture on the interpolation nanosheet task compared to the purely local NequIP, the cross-attention globally-aware mechanism of DOSTransformer, and EBFormer-l1, which uses boundary attention. We note that EBFormer and the self-attention network are significantly superior to the other two methods, while both achieve similar performance to the other, providing motivation that the mechanism we use is sufficient. This is reinforced by the attention masks (visualized in Figure 6). We utilized four layers of self-attention for the nanosheet, identical to EBFormer, and present the top 5% of attention scores sorted by score below for a 97-layer silicon nanosheet cleaved along $\langle 100 \rangle$. Atomic indices are ordered by z-axis position, from

least to most, along the nanosheet. Atoms 0-1 and 98-99 are passivating hydrogens masked when predicting total system electronic properties. For the nanowire, due to the larger size of the structures, we could only fit three layers of self-attention within our GPU, visualized in panel b) for a 26x26 silicon nanowire cleaved along $\langle 100 \rangle$. The architecture is highlighted in panel c).

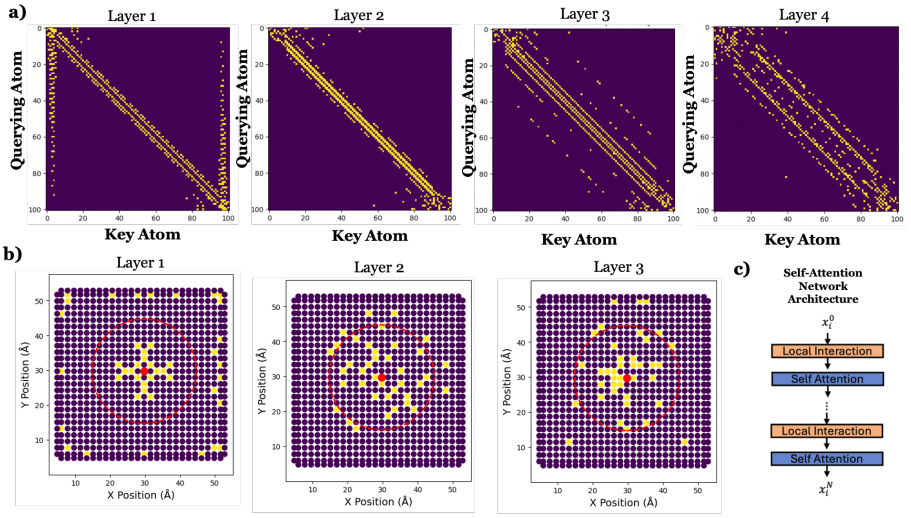

Figure 6: **Visualization of the Attention Matrices**: (a) Top 5% of attention scores (four layers of self-attention) for a silicon nanosheet with 97 layers. The attention pattern is sparse, with a strong local (on-diagonal) component, and nonlocal "legs" representing attention with the boundaries. (b) Top 5% of attention scores (three layers of self-attention) of the central atom in a 26x26 layer silicon nanowire (represented in real space as the cross-section of the nanowire, where dots are atoms). Red circle represents the convolutional layer receptive field. Attention scores appear to focus locally (within the receptive field of the convolution) or the boundaries. (c) Architecture of the self-attention network. Intersperses local graph convolution with nonlocal, fully-connected self-attention using an identical attention formulation as EBFormer

Visually, the attention mask for the nanosheet has the form of an "N", with a diagonal component (which is local), and two strong nonlocal "legs" which show that each atom has a high attention score with at least one of the atoms at the boundaries. Masks 2-4 primarily demonstrate local attention, with strong scores mainly along the diagonal. Similarly, the nanowire attention scores (visualized over the 2D face of the nanowire for the central atom) show that top 5% scores are either on the boundaries (in layer 1), or primarily within the receptive field (layers 2-3). Leveraging the sparsity of these attention matrices, we motivate that including nonlocal interactions only with the boundaries of a structure (representing the "legs" of the "N" for the nanosheet, or the boundary scores in the nanowire) is *sufficient* for our systems of interest. Rather than connecting all atoms with all surface atoms, we use a single vector embedding to collect boundary information, which ensures preservation of O(N) scaling for arbitrary geometries. Finally, augmenting the nonlocal information with local convolution covers the diagonal of the "N" in the attention mask of the nanosheet, or the scores within the circle of the nanowire.

With the preceding discussion, we have motivated the architectural approaches taken with EBFormer as required and physically-motivated but frugal, with only the requisite interactions included to preserve favorable asymptotic scaling.

Empirical Validation To conclude, we note that our architecture demonstrates superior in-distribution and OOD when compared to a global, linearly scaling approach DOSTransformer which does not carefully consider the inductive biases of our domain. Furthermore, we show the suitability of the inductive bias of our approach by demonstrating high performance with parameter and data sparsity. We also demonstrate the improvement of our predictions in the OOD thickness case with inclusion of larger structures from another elemental species in the transfer learning task in Appendix J, demonstrating the ability of our architecture to leverage structure information separately from local information. Finally, we also demonstrate our method's applicability to nanowires such

as those in GAAFETs which face an additional dimension of confinement. These results serve to further empirically motivate that the EBFormer architecture is well-suited to our task of capturing boundary effects with efficient scaling and preservation of a local basis.

We note that the matrices from the self-attention experiment presented above were our initial motivation to engineer EBFormer.

B BALLISTIC TRANSPORT THEORY

In this work, we focus on inference of the density of states and current density-of-states (DOS and cDOS, respectively), which in the case of transistors are crucial quantities that describe the final current-carrying and electrostatic capabilities of the device. The ballistic current is related to the cDOS and DOS as follows Rahman et al. (2003):

$$N_{inv} = \int dE D(E) f(E - E_f) \quad v_{inj} = \int dE \frac{J_x(E)}{N_{inv}} f(E - E_f) \qquad I_{inv} = q N_{inv} v_{inj} \tag{8}$$

Here, $D(E)$ denotes the DOS , $J_x(E)$ is the cDOS along the transport direction, f is the Fermi-Dirac distribution, and N_{inv} represents the inversion charge density inside the nanostructure. The injection velocity v_{inj} is a key figure of merit for transistors representing the average velocity of electrons entering the channel. I_{inv} is the final ballistic current flowing through the transistor. Using the projection operator as detailed in Appendix C.2, we can derive the atom-local quantities of LDOS and LcDOS, which represent the distribution of the DOS and cDOS in space. The above DOS and cDOS integrals will then be modified to include a spatial integral over the cross-section of the nanostructure, which can incorporate the varying potential profile over the nanostructure at a given drain and gate voltage. These atom-resolved projections are crucial for capturing the microscopic distribution of charge and current within the semiconductor channel, which in turn influence the electrostatics and device characteristics.

C DATA GENERATION

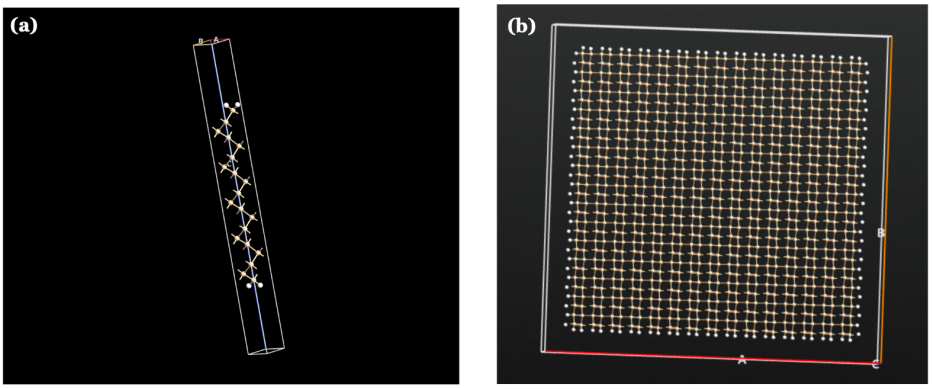

Figure 7: **Visualization of the Structures**: (a) Silicon $\langle 100 \rangle$ nanosheet visualization (b) Silicon 30x30 layer nanowire visualization. Tan atoms are silicon, and the white atoms are hydrogen.

Table 5: Counts of cleaving directions for silicon and germanium nanosheets

Material	$\langle 1\,0\,0 \rangle$	$\langle 1\,1\,0 \rangle$	$\langle 1\,1\,1 \rangle$	$\langle 2\,1\,1 \rangle$	Total
Silicon	3,786	2,353	1,126	481	7,746
Germanium	2,458	1,394	444	443	4,739
Total	6,244	3747	1570	924	12,485

Our nanosheet dataset consists of 12,485 confined, hydrogen-passivated nanosheets representing the channels of SOI transistors Park et al. (2025). Table 5 shows the distribution over cleaving orientation and elemental species. The nanowire dataset consists of 749 confined, hydrogen-passivated silicon and germanium nanowires representing the channels of GAAFETs Huang et al. (2017). Density of states and current density of states Rahman et al. (2003) are derived for each structure using an empirical tight-binding simulation Jancu et al. (1998). Our structure generation and simulation are both conducted using QuantumATK Smidstrup et al. (2019), with details presented in the following sections. All simulations were conducted on Intel Xeon Gold 6330 processors parallelized across 16 to 64 processes.

C.1 Structure Generation

We begin with a unit-cell of either silicon or germanium material in the face-centered diamond cubic crystal configuration. The system lattice constant is scaled by a multiplicative constant in the range of (0.9, 1.035) for silicon, and (0.9, 1.02) for germanium to simulate the effect of strain. The strained unit cell is then cleaved to form a nanosheet such that the normal direction of the sheet is oriented along one of four orientations: $\langle 100 \rangle$, $\langle 110 \rangle$, $\langle 111 \rangle$, or $\langle 211 \rangle$, or to form a nanowire, with orientation selected along $\langle 100 \rangle$. Final nanosheet atomic layer count is varied between 15 to 100 layers for all cleaving directions, or 10-30 layers in each lateral direction for the nanowire. Point-defects are introduced for a subset of silicon nanosheet structures, with displacement amplitude chosen in the range of (0.05 Å, 0.15 Å). Finally, the system is passivated with hydrogen using the QuantumATK passivation tool to reduce surface defects due to dangling bonds. The displacement amplitude and strain ranges were chosen to ensure preservation of a band gap to retain the semiconductor nature of the nanostructures.

C.2 Electronic Structure Simulation

Using the generated atomic structures, electronic properties are derived using an empirical tight-binding simulation (commonly used for atomistic property simulation for semiconductors Vogl et al. (1983)) with QuantumATK. Silicon structures are simulated using the Bassani.SiH basis set, while germanium-containing structures use Bassani.GeH Jancu et al. (1998). LDOS is generated using tetrahedral expansion over energies calculated at a 150x150 Monkhorst-Pack grid over k_x and k_y for nanosheets, and 150 points on k_z for nanowires. cDOS is calculated using Gaussian broadening of 0.01 eV over the band structure sampled over the same grid. Projection for LDOS calculation is conducted using a projection operator $\hat{\mathbf{P}}_a$ onto atom a Soriano & Palacios (2014), while LcDOS is approximated as follows:

$$D(E,a) = \sum_{\vec{k}} \delta(E - E_{\vec{k}}) \langle \psi_{\vec{k}} | \hat{\mathbf{P}}_a | \psi_{\vec{k}} \rangle ; \quad J_x(E,a) = J_x(E) \times \frac{D(E,a)}{D(E)} \tag{9}$$

Both DOS/LDOS and cDOS/LcDOS are calculated over 100 energy bins from E_c - 0.01 eV to E_c + 0.3 eV, where E_c is the conduction band minimum. Injection velocity and inversion charge and current calculations are conducted by varying the fermi level over 100 energies in the range (-0.3 eV, 0.3 eV) around E_c Rahman et al. (2003).

D Description of Datasets

D.1 In-Distribution Nanosheet Dataset

The statistics and distributions of quantities in the dataset for the nanosheet interpolation task are presented below. The task uses a 80/10/10 randomized split over all structures in the dataset. Figure 8 depicts the distributions of DOS, cDOS, and injection velocity values across all energy bins and fermi levels in the test, training, and validation sets. Units and scales are the same as presented in all error tables.

Table 6 includes the means and standard deviations across all energy bins and fermi levels of DOS, cDOS, and downstream derived quantities over the test, train, and validation datasets. From the histograms and tables, we can draw the conclusion that the training data is representative of the test task.

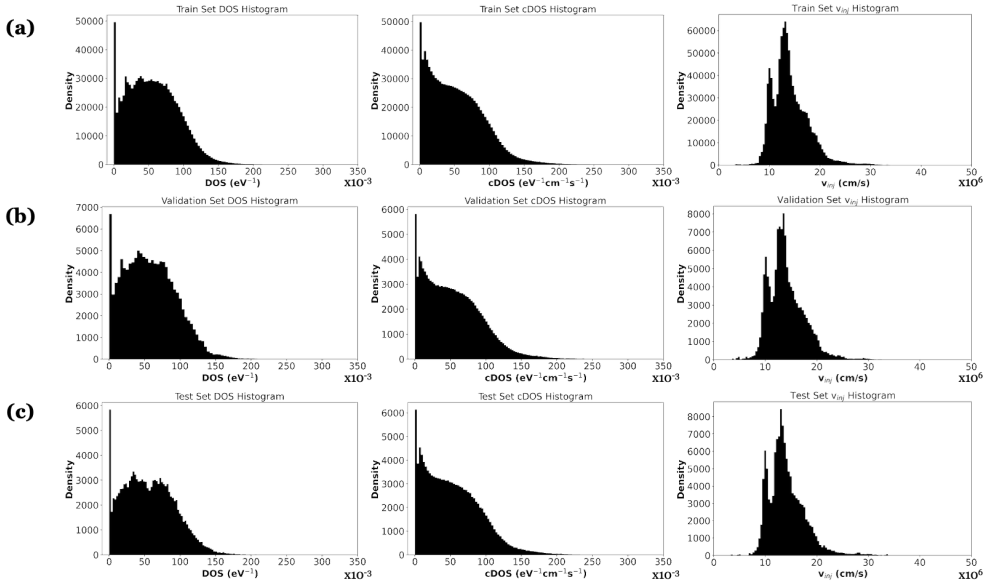

Figure 8: **Statistics of the In-Distribution Nanosheet Dataset**: (a) Training Data Histograms for DOS, cDOS, and injection velocity (b) Validation Data Histograms (c) Test Data Histograms

Table 6: **In-Distribution nanosheet statistics**: DOS (eV^{-1}), cDOS ($eV^{-1}cm^{-1}s^{-1}$), N_{inv} (cm^{-2}), I_{inv} (mA/μm), and injection velocity (cm/s) means over the train, validation, and test set for the in-distribution task are included below. Standard deviations are presented in parentheses below

Set	Count	DOS $[\times 10^{-3}]$	cDOS $[\times 10^{-3}]$	N_{inv} $[\times 10^{11}]$	I_{inv} $[\times 10^{-3}]$	v_{inj} $[\times 10^{6}]$
Train Stats	9991	56.35 (35.75)	52.80 (38.60)	10.29 (15.98)	28.67 (47.08)	14.10 (3.60)
Validation Stats	1246	57.52 (36.21)	54.10 (39.55)	10.49 (16.22)	29.48 (48.56)	14.06 (3.49)
Test Stats	1248	56.75 (35.65)	53.46 (39.04)	10.32 (15.98)	29.11 (48.00)	14.12 (3.58)

D.2 Out-of-Distribution Nanosheet Dataset

The statistics and distributions of quantities in the dataset for the zero-shot out-of-distribution task are presented below. The extrapolation task uses nanosheets 15-45 atomic layers thick of all strains, cleaving directions, and point-defective silicon nanosheets as the training data, and validates/tests on 46-100 layer structures. The validation and test datasets are a 50-50 randomized distribution of the 46-100 layer nanosheets. Figure 9 depicts the distributions of DOS, cDOS, and injection velocity values across all energy bins and fermi levels in the test, training, and validation sets. Units and scales are the same as presented in all error tables.

Table 7: **Out-of-Distribution nanosheet statistics**: DOS (eV^{-1}), cDOS ($eV^{-1}cm^{-1}s^{-1}$), N_{inv} (cm^{-2}), I_{inv} (mA/μm), and injection velocity (cm/s) means over the train, validation, and test set for the out-of-distribution task are included below. Standard deviations are presented in parentheses below

Set	Count	DOS $[\times 10^{-3}]$	cDOS $[\times 10^{-3}]$	N_{inv} $[\times 10^{11}]$	I_{inv} $[\times 10^{-3}]$	v_{inj} $[\times 10^{6}]$
Train Stats	3291	60.24 (37.03)	57.65 (41.24)	11.42 (17.23)	32.63 (54.60)	13.62 (3.25)
Validation Stats	2225	53.79 (31.41)	47.44 (32.25)	10.00 (15.24)	25.45 (39.75)	13.19 (2.64)
Test Stats	2227	53.88 (31.49)	47.53 (32.30)	10.06 (15.35)	25.49 (39.81)	13.24 (2.61)

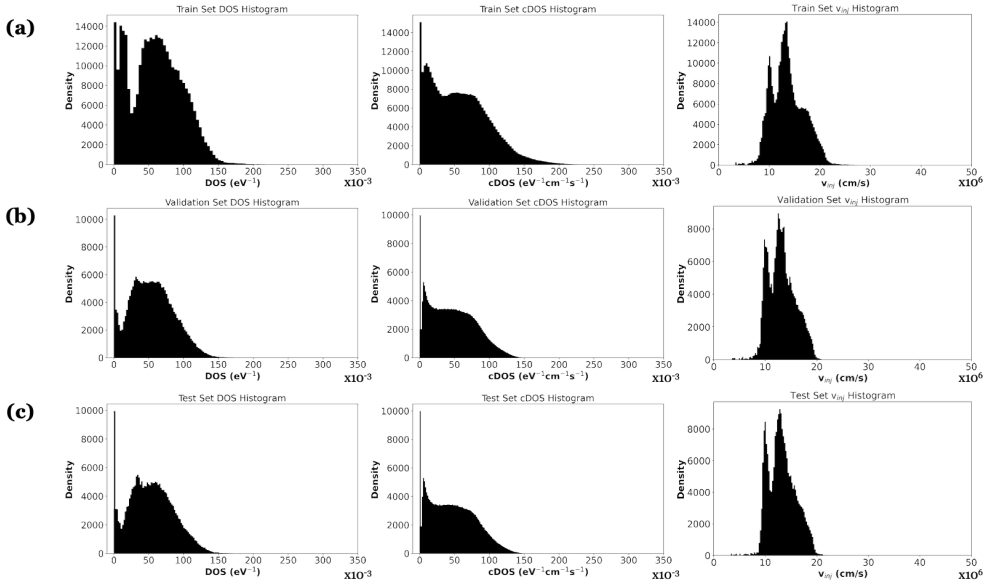

Figure 9: **Statistics of the Out-of-Distribution Nanosheet Dataset**: (a) Training Data Histograms for DOS, cDOS, and injection velocity. These are silicon nanosheets with atomic layers from 15-45 inclusive. (b) Validation Data Histograms (c) Test Data Histograms. Validation and test set are over silicon structure with 46-100 layers inclusive. A distribution shift is apparent between the training and test data

Table 7 includes the means and standard deviations across all energy bins and fermi levels of DOS, cDOS, and downstream derived quantities over the test, train, and validation datasets for the extrapolation task. We qualitatively see from the histograms that the training and test distributions are dissimilar.

D.3 In-Distribution Nanowire Experiment

The statistics and distributions of quantities in the dataset for the nanowire interpolation task are presented below. The dataset is comprised of 380 silicon and 369 germanium structures. Training uses a 80-10-10 randomized split over all structures in the dataset. Figure 10 depicts the distributions of DOS, cDOS, and injection velocity values across all energy bins and fermi levels in the test, training, and validation sets. Units and scales are the same as presented in the error tables.

Table 8 includes the means and standard deviations across all energy bins and fermi levels of DOS, cDOS, and downstream derived quantities over the test, train, and validation datasets.

Table 8: **In-Distribution Nanowire statistics**: DOS (eV^{-1}), cDOS (eV^{-1}cm^{-1}s^{-1}), N$_{inv}$ (cm^{-2}), I$_{inv}$ (mA/μm), and injection velocity (cm/s) means over the train, validation, and test set for the in-distribution nanowire task are included below. Standard deviations are presented in parentheses below

Set	Count	DOS $[\times 10^{-3}]$	cDOS $[\times 10^{-3}]$	N_{inv} $[\times 10^{10}]$	I_{inv} $[\times 10^{-3}]$	v_{inj} $[\times 10^{6}]$
Train Stats	603	62.36 (45.64)	57.93 (33.36)	1.23 (2.19)	35.66 (54.20)	12.36 (3.76)
Validation Stats	72	57.95 (41.17)	53.22 (30.78)	0.98 (1.69)	32.02 (47.92)	12.46 (3.68)
Test Stats	74	63.25 (46.79)	58.31 (34.69)	1.17 (2.03)	35.77 (55.15)	12.39 (3.58)

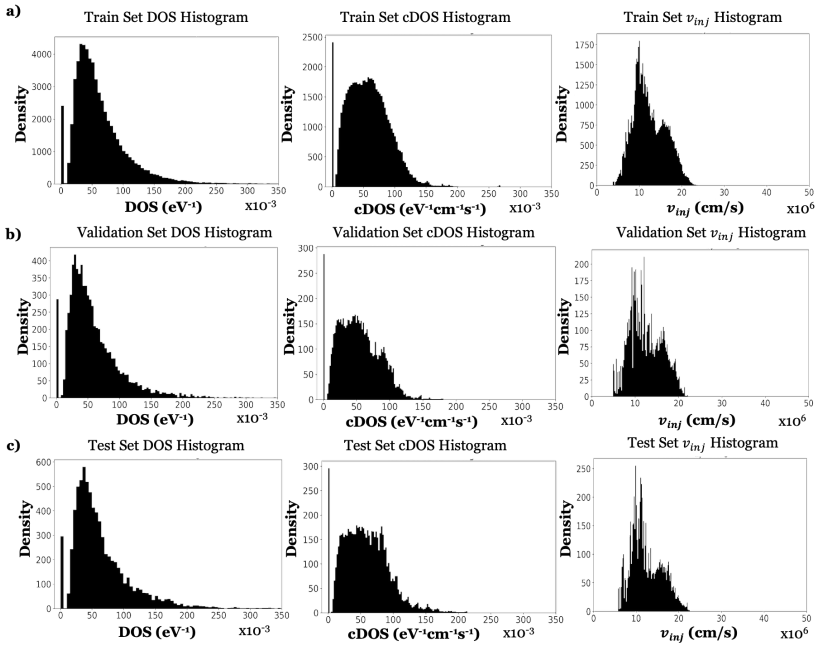

Figure 10: **Statistics of the In-Distribution Nanowire Experiment**: (a) Training Data Histograms for DOS, cDOS, and injection velocity. (b) Validation Data Histograms (c) Test Data Histograms. An 80/10/10 split is used over the entirety of the nanowire dataset. The training distribution qualitatively appears to be representative of the validation and test distributions

Table 9: Hyperparameter specifications of EBFormer

Hyperparameters	**In-Distribution Nanosheet**		**OOD Nanosheet & Nanowire**	
	EBFormer-l0	EBFormer-l1	EBFormer-l0	EBFormer-l1
Message Passing Layers	4	4	4	4
Boundary Attention Heads	1	1	1	1
$\mathbf{w}_{\text{DOS}}/\mathbf{w}_{\text{cDOS}}$	3	3	3	3
Cutoff Radius (Å)	5	5	5	5
Hidden Dimension	400	215	122	64
Learning Rate	0.005	0.005	0.005	0.005
Batch Size	5	5	5	5

E Implementation Details

We compare four architectures over a variety of configurations on the in-distribution and out-of-distribution tasks. EBFormer is characterized over a variety of training modes, equivariance degrees, and loss functions. The hyperparameters used for the invariant and l=1 equivariant approaches are included in Table 9. As a baseline, we include an MLP with a similar configuration as DOSTransformer Lee et al. (2023), which uses initial atomic embeddings as input to derive DOS and cDOS using a joint readout. The architecture has three hidden layers before the two readout heads, with hidden dimensions of 1000 for the interpolation task, and 300 for extrapolation to maintain weight parity.

Our GNNs are based off NequIP equivariant architectures Batzner et al. (2022). The ablation implementations are identical to the embedding layers in EBFormer, with the same shared hidden dimension, layer count, and cutoff as presented in Table 9. The invariant weight parity models use hidden dimension 1100 and 122 for the in-distribution and out-of-distribution tasks, respectively. The equivariant weight-parity models use hidden dimension 630 and 201 for the in-distribution and out-of-distribution tasks.

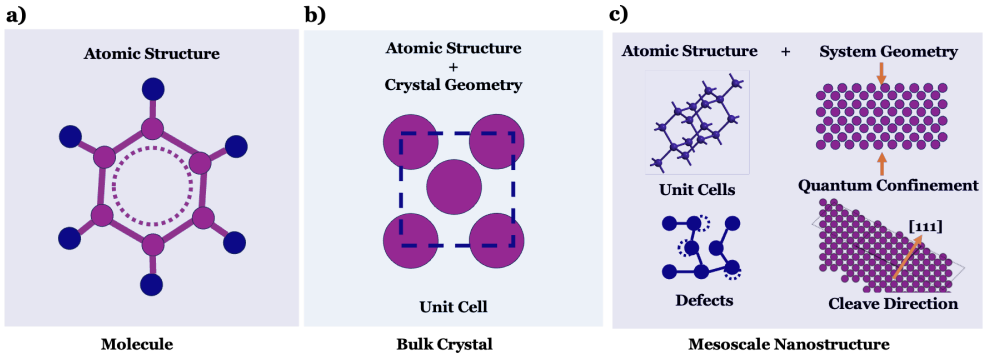

Figure 11: Visualization of (a) atomic (b) bulk crystal and (c) mesoscale nanostructure systems, which EBFormer focuses on

DOSTransformer hyperparameters are selected from the default configuration released with the code Lee et al. (2023). The readout layer is branched to two readout heads to facilitate joint training to fairly compare DOSTransformer with EBFormer, as we noted joint training led to a significant improvement in performance for our model. Hidden dimension is 256 for the in-distribution task, and 80 for the out-of-distribution task.

All models optimize over the same L1-loss functions $\mathcal{L}_{L/G}$ defined in Section 3.1 with $\mathbf{w}_{\text{DOS}}/\mathbf{w}_{\text{cDOS}} = 3.0$. We attempt to maintain weight parity across compared networks to avoid confounding effects of scaling Qu & Krishnapriyan (2024), which we accomplish by changing model hidden dimensions. This has the effect of scaling most of the main weight matrices involved in the network, including attention and convolution mechanisms. All the neural network code is written in PyTorch-CUDA Paszke et al. (2019). Training and inference were performed on an NVIDIA A100 GPU. Result tables with MAE and RMSE losses include mean (and standard deviation in parentheses below) over three runs of each model, each with a different seed. Seeds for NequIP-based approaches and EBFormer are selected from {123, 246, 369} for the model parameters, and {456, 912, 368} for dataset-related randomization, including data shuffling. The MLP and DOSTransformer use seeds {0, 1, 2}. Models on the result tables were selected based on minimum validation loss over 500 epochs of model training.

E.1 Visualization of Boundary Embeddings

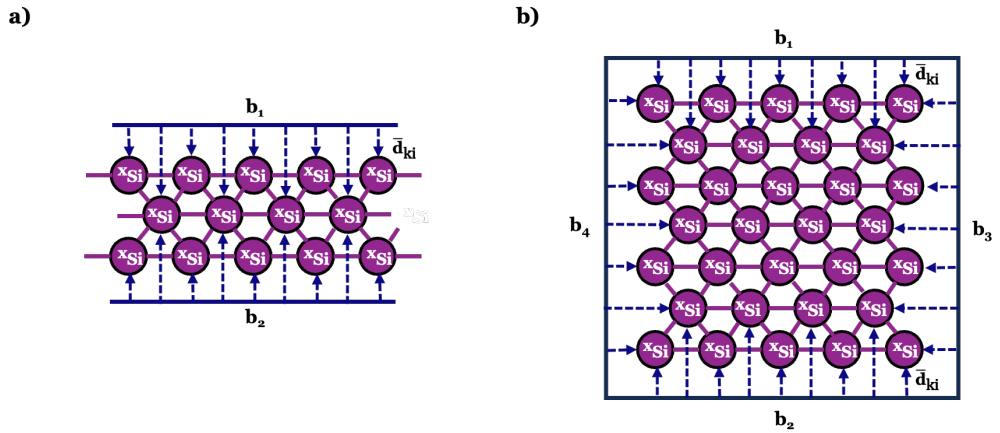

Figure 12: Visualization of the position and count of the boundary embeddings for the nanosheet (a) and nanowire (b) structures

Figure 11 includes a visualization of molecular, bulk crystal, and mesoscale nanostructure systems, the last of which is the focus of EBFormer. We note that in the first two cases, local message-passing approaches are sufficient to capture the entire system geometry (notice the relatively few atoms required to define the entire system). However, mesoscale effects such as quantum confinement and crystal cleaving orientation lead to variation in material properties which require global augmentation. When applied to materials without boundary planes (such as molecular systems or bulk crystals), our architecture reduces to the GNN we use for local convolution – in our case, a modified version of NequIP.

Figure 12 shows a visualization of the boundary embeddings for the nanowire and nanosheet structures. For the nanosheet, there are two boundaries; one at the top surface and one at the bottom of the nanosheet, in this case between the silicon and vacuum. In a full SOI gate-stack, this interface could be between the silicon and the bottom and gate oxides. For the nanowire, we have four boundary embeddings; one for each face of the nanostructure, again between the silicon and vacuum. Each boundary embedding interacts with all atoms in the boundary to atom cross-attention, while each atom collects information from both or all four of the boundaries for the nanosheet or the nanowire case, respectively. The interaction mechanisms are described in the main text in Equation 4.

E.2 Sensitivity Analysis

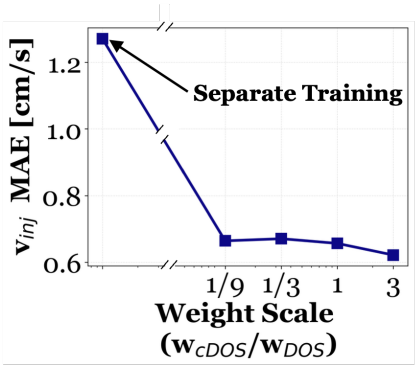

Figure 13: Sensitivity analysis of relative weighting of DOS and cDOS for the in-distribution nanosheet experiment for EBFormer

We conduct a sensitivity analysis on the relative weights of cDOS and DOS in the loss function, and also train separate models on each quantity individually, combining their predictions for injection velocity. We see that joint training yields a substantial improvement in injection velocity predictions, reinforcing the importance of physical consistency between DOS and cDOS for accurate downstream performance. Moreover, EBFormer maintains stable accuracy across a wide range of weight ratios, demonstrating robustness to a range of hyperparameter choices.

F Qualitative Visualization

F.1 Nanosheet Inference

Figure 14 compares model predictions on DOS, cDOS, and injection velocity on a structure from the test set. The ground-truth DOS show distinctive step-like features characteristic in confined 2D materials Datta (2005), and depends on the film thickness. NequIP-l1 is unable to capture these features since it is unaware of global geometry. DOSTransformer and EBFormer show comparable performances on DOS/cDOS, with EBFormer demonstrating improved injection velocity accuracy. The inset panel also shows high-fidelity reproduction of local atomic properties by EBFormer. Further characterization of LDOS/LcDOS prediction errors can be found in Appendix Section I.

Extrapolation performance, visualized in Figure 15, qualitatively shows improved DOS predictions by EBFormer with approximate reproduction of confinement features in global and local DOS/cDOS. Note that this structure (at 77 layers) is significantly larger than the largest in the

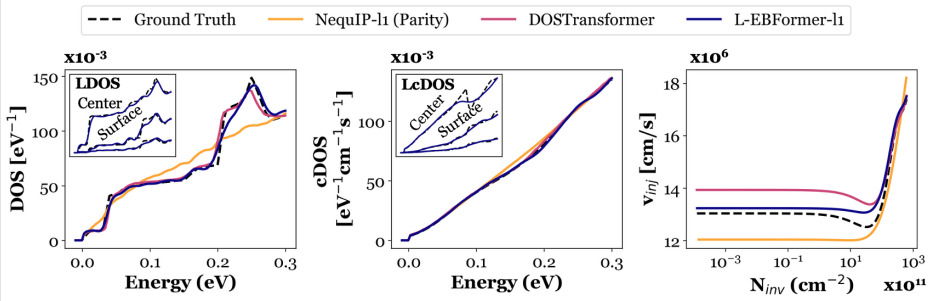

Figure 14: **Interpolation Qualitative Analysis** on a structure from the test set, 35 atomic layers of silicon cleaved along ⟨100⟩

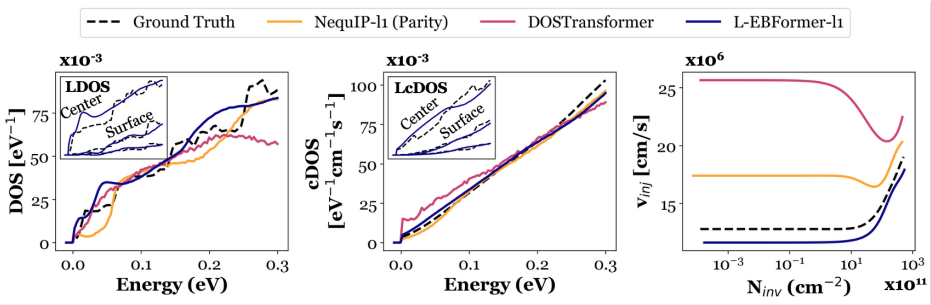

Figure 15: **Extrapolation Qualitative Analysis** on a structure from the test set with many more layers than in the training set, 77 atomic layers of silicon cleaved along ⟨100⟩

training set (45 layers). We also observe significantly improved injection velocity prediction in extrapolation by EBFormer compared to the two other architectures.

F.2 Nanowire Inference

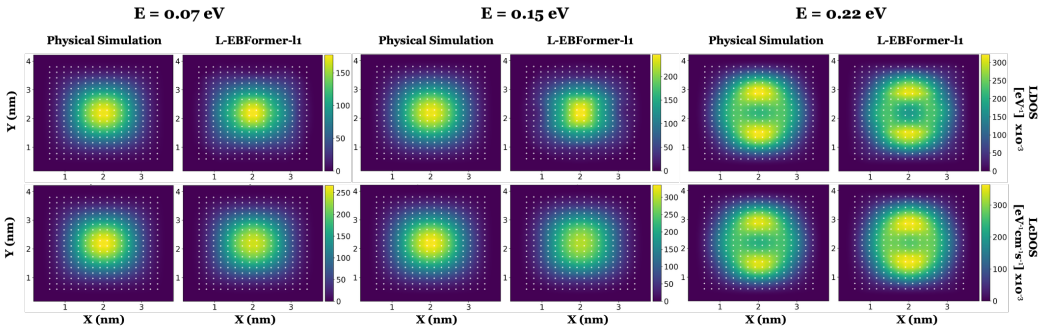

Figure 16: **Prediction of Local Quantities on a Nanowire** on a structure from the test set, 16x18 atomic layers of silicon cleaved along ⟨100⟩ with lattice scaling 0.981. Atom-local predicted values are Gaussian-broadened across real-space with width 0.1 nm. The white dots correspond to the locations of the silicon nuclei in the nanowire. Energies are chosen uniformly over the inference range. LcDOS reflects which areas of the wire are carrying current, and affects the device behavior under different applied voltage profiles, while LDOS determines the form and number of electrons in the wire based on an applied voltage.

Finally, we complete the visualization in the main paper by visualizing the Gaussian-smeared LDOS and LcDOS as predicted by L-EBFormer-l1 on a structure from the nanowire test set in Figure 16. The visualizations are included over three energies chosen to uniformly sample the energy range of the DOS/cDOS. We notice that different energy ranges correspond to different geometric distributions over the nanowire, which correspond to different subbands (or eigenmodes) of the wavefunc-

tion over the nanowire. This is conceptually similar to incrementing over different harmonics of the wave equation on a rectangular plate (such as a Chladni plate Colwell (1931)), though the form of the Schrodinger Equation differs in detail. The variation in the LDOS determines the distribution of electrons over the nanowire, while LcDOS corresponds to the geometric distribution of the current-carrying capacity of the various locations of the nanowire (i.e. which parts of the wire carry how much current for electrons injected at a certain energy). We see that EBFormer is able to reproduce the variation of local features quite successfully, capturing the quantum-mechanical feature of low current/electron states near the surface commonly employed as a first-order correction for device simulators as well as more complicated geometrical features at higher harmonics.

G Error Analysis of V_{inj}

Table 10: Error statistics and DOS–J_x correlation quality. Units are as in the main text, with best values in bold.

Model	Params [M]	DOS [$\times 10^{-3}$]		cDOS [$\times 10^{-3}$]		N_{inv} [$\times 10^{11}$]		I_{inv} [$\times 10^{-3}$]		v_{inj} [$\times 10^{6}$]		$R^2_{\text{N}_{inv}-\text{I}_{inv}}$
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	
MLP	9.25	14.82 (0.02)	23.03 (0.27)	12.07 (0.10)	20.22 (0.51)	2.69 (0.01)	6.21 (0.04)	7.26 (0.05)	19.99 (0.62)	1.63 (0.05)	2.63 (0.08)	-0.229 (0.212)
NequIP-l0 (Ablation)	1.39	5.59 (0.01)	9.14 (0.07)	2.40 (0.04)	4.51 (0.05)	0.45 (0.00)	1.35 (0.03)	1.43 (0.02)	4.26 (0.08)	0.87 (0.00)	1.97 (0.16)	0.677 (0.039)
NequIP-l1 (Ablation)	1.18	5.50 (0.04)	8.92 (0.03)	2.31 (0.06)	4.31 (0.10)	0.44 (0.01)	**1.30** (0.03)	1.36 (0.03)	4.06 (0.08)	0.88 (0.01)	1.83 (0.07)	0.569 (0.050)
NequIP-l0 (Parity)	9.98	5.53 (0.01)	9.17 (0.04)	2.29 (0.00)	4.39 (0.03)	0.44 (0.00)	1.36 (0.01)	1.37 (0.01)	4.18 (0.03)	0.91 (0.03)	2.24 (0.42)	0.632 (0.039)
NequIP-l1 (Parity)	9.46	5.46 (0.05)	8.91 (0.04)	2.30 (0.07)	4.27 (0.06)	0.44 (0.01)	1.31 (0.04)	1.36 (0.05)	4.02 (0.09)	0.85 (0.02)	1.70 (0.26)	0.602 (0.075)
DOSTransformer	9.40	**2.80** (0.15)	6.59 (0.12)	1.55 (0.09)	3.80 (0.09)	0.37 (0.02)	1.33 (0.04)	0.90 (0.06)	3.51 (0.04)	0.80 (0.16)	2.34 (1.39)	0.512 (0.133)
EBFormer-l0	9.48	3.27 (0.09)	6.89 (0.06)	1.32 (0.04)	3.93 (0.04)	0.33 (0.01)	1.38 (0.01)	0.78 (0.03)	3.75 (0.04)	0.61 (0.02)	1.25 (0.12)	0.724 (0.035)
EBFormer-l1	9.35	3.17 (0.05)	6.74 (0.12)	**1.26** (0.03)	3.76 (0.13)	0.32 (0.01)	1.38 (0.04)	0.75 (0.01)	3.61 (0.09)	0.60 (0.01)	1.35 (0.24)	0.704 (0.017)
L-EBFormer-l1	9.35	2.85 (0.05)	**6.21** (0.09)	**1.26** (0.03)	**3.37** (0.09)	**0.31** (0.00)	**1.30** (0.03)	**0.67** (0.02)	**3.24** (0.11)	**0.49** (0.01)	**0.90** (0.06)	**0.756** (0.012)

Injection velocity is a downstream function of inversion charge and current. These in turn are weighted integrals over DOS and cDOS in energy space, respectively, as defined in Appendix B. Errors propagate from the inversion charge and currents as follows:

$$v_{inj} = \frac{I_{inv}}{N_{inv}} \implies \frac{\Delta v_{inj}}{v_{inj}} = \frac{\Delta I_{inv}}{I_{inv}} - \frac{\Delta N_{inv}}{N_{inv}}$$

This implies that low injection velocity error requires (1) accurate N_{inv} and I_{inv} predictions and (2) similar magnitudes of N_{inv} and I_{inv} relative errors. Consistent relative errors in the inversion charge and current imply that these errors fall along the y=x line when plotted against each other. We quantify this error by calculating the coefficient of determination of the relative errors in N_{inv} and I_{inv} using the y=x line as the best fit ($\text{R}^2_{\text{N}_{inv}-\text{I}_{inv}}$). We calculate the errors over the in-distribution nanosheet dataset and present the results in Table 10.

We see a strong correlation between the closeness of fit as given by the coefficient of determination and the accuracy in injection velocity, and note that EBFormer achieves the highest degree of covariance of N_{inv} and I_{inv}. These quantities are derived from DOS and cDOS, which implies that these neural-network predictions also have a high covariance. Therefore, this shows that EBFormer's high success in injection velocity performance is due to accurate predictions as well as having learned the physical covariance of DOS and cDOS better than the benchmarked models.

H Timing and Memory Consumption

Figure 5 in the main text compares runtimes over a variety of structures for the empirical tight-binding method and EBFormer. Tight-binding is timed on an Intel Xenon Gold 6330 over 16-64

parallel processes. EBFormer inference is timed on an NVIDIA A100 GPU. The prediction time of the neural network only includes the inference time, and excluded data movement and formation. We see that EBFormer achieves 4-5 orders of magnitude speed-up compared to the baseline simulation method. Furthermore, while tight-binding methods scale as $O(N^3)$ Li et al. (2023) where N is the number of atoms in a unit-cell, EBFormer scales linearly, permitting simulation of significantly larger systems.

H.1 Timing and Memory Consumption

Timing and memory comparison and throughput are presented in Table 11 for the models in our study.

Model	Parameters (M)	Inference Speed (Samples / s)	Inference Memory (Peak / Sample) MB	Training Speed (Samples / s)	Training Memory (Peak / Sample) GB
MLP	9.25	502.2	53	77.7	0.16
ablation-l0	1.39	89.7	25	26.8	0.14
wp-l0	9.98	95.5	79	28.0	0.54
ablation-l1	1.18	64.8	58	15.3	0.38
wp-l1	9.46	64.6	176	18.0	1.68
DOSTransformer	9.40	124.0	55	27.1	0.20
EBFormer-l0	9.48	30.5	126	8.3	3.89
EBFormer-l1	9.35	16.4	109	4.3	3.84
L-EBFormer-l1	9.35	18.3	109	4.4	3.84

Table 11: Model size, speed, and memory during inference and training.

As noted in our limitations, using SE(3) equivariant attention involving Clebsch-Gordon tensor products is generally an expensive operation Batzner et al. (2022); Geiger et al. (2022); Luo et al. (2024). This is reflected in the table above in terms of memory and time usage of NequIP and EBFormer, which are equivariant architectures built on e3nn. However, we note that there have been many works that implement equivariant message-passing with higher-order tensorial features in an efficient manner through bespoke engineering and limited order of equivariance Wang et al. (2024c); Batatia et al. (2024); Wang et al. (2024a). Furthermore, our application demonstrates equivariant features only up to $l = 1$, meaning we operate with only scalar and vector-type objects. Interactions of these quantities require implementation only of vector-vector dot and cross products, scalar-vector multiplication, and scalar operations. While beyond the purview of our current work, replacing the complete machinery of e3nn and its generic treatment of irreps with a more tailor-made mechanism is a valuable and attainable next step.

Finally, we also observe that local training leads to minimal additional cost in memory or speed in both inference and training due to the fact that the intermediate values of LDOS/LcDOS are generated regardless of if they are used as a training signal. The improvement in downstream performance indicates that including these signals is an effective and low-cost way to improve performance given data availability.

I Local DOS/cDOS Errors

Tables 12 and 13 show the MAE and RMSE errors of LDOS and LcDOS along with all other model predictions on the nanosheet tasks, while Table 14 includes the same for the nanowire interpolation task. LDOS and LcDOS are the atom-wise predicted contributions to DOS/cDOS prior to mean pooling that are used to generate the final mean DOS/cDOS. These quantities are only meaningful for architectures which make predictions on a per-atom basis – MLP and DOSTransformer do not generate local predictions, and therefore cannot generate or leverage the information in locally-projected quantities.

Atom-local quantities such as LDOS and LcDOS can vary significantly over nanostructures. Furthermore, capturing this spatial variation of material properties can have significant impact on downstream device properties Jiang et al. (2008). For this reason, EBFormer is designed to be able to predict atom-local quantities. Table 12 demonstrates the high accuracy of L-EBFormer-l1 on the inference task of LDOS/LcDOS. In addition, we see that training on local quantities leads to *improvement* in performance in in-distribution regimes. This indicates that information regarding spatial variation of physical quantities is helpful to accurately reproduce physical quantities in systems with structure.

Table 12: **Nanosheet Interpolation Including LDOS/LcDOS Errors**: Interpolation accuracy including LDOS/LcDOS errors. Note that only L-EBFormer-l1 was trained on the local task. Models which cannot project predictions to a local basis (MLP and DOSTransformer) have no errors listed for the local tasks

Model	Params [M]	DOS [$\times 10^{-3}$]		LDOS [$\times 10^{-3}$]		cDOS [$\times 10^{-3}$]		LcDOS [$\times 10^{-3}$]		N_{inv} [$\times 10^{11}$]		I_{inv} [$\times 10^{-3}$]		v_{inj} [$\times 10^{6}$]	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MLP	9.25	14.82 (0.02)	23.03 (0.27)	–	–	12.07 (0.10)	20.22 (0.51)	–	–	2.69 (0.01)	6.21 (0.04)	7.26 (0.05)	19.99 (0.62)	1.63 (0.05)	2.63 (0.08)
NequIP-l0 (Ablation)	1.39	5.59 (0.01)	9.14 (0.08)	1055.06 (9.03)	1710.27 (5.08)	2.40 (0.05)	4.51 (0.06)	1058.15 (27.16)	1754.87 (21.45)	0.45 (0.00)	1.35 (0.03)	1.43 (0.03)	4.26 (0.10)	0.87 (0.01)	1.97 (0.19)
NequIP-l1 (Ablation)	1.18	5.50 (0.05)	8.92 (0.04)	1029.29 (49.49)	1741.76 (74.85)	2.31 (0.06)	4.31 (0.11)	974.32 (51.10)	1773.18 (92.41)	0.44 (0.01)	**1.30** (0.04)	1.36 (0.04)	4.06 (0.10)	0.88 (0.02)	1.83 (0.08)
NequIP-l0 (Parity)	9.98	5.53 (0.01)	9.17 (0.05)	993.23 (14.01)	1600.48 (24.00)	2.29 (0.00)	4.39 (0.04)	985.88 (22.61)	1625.68 (15.52)	0.44 (0.00)	1.36 (0.02)	1.37 (0.01)	4.18 (0.03)	0.91 (0.04)	2.24 (0.50)
NequIP-l1 (Parity)	9.46	5.46 (0.06)	8.91 (0.05)	981.23 (35.68)	1693.83 (70.94)	2.30 (0.08)	4.27 (0.07)	947.03 (34.15)	1779.91 (81.19)	0.44 (0.01)	1.31 (0.05)	1.36 (0.06)	4.02 (0.10)	0.85 (0.02)	1.70 (0.31)
DOSTransformer	9.40	**2.80** (0.15)	6.59 (0.12)	–	–	1.55 (0.09)	3.80 (0.09)	–	–	0.37 (0.02)	1.33 (0.04)	0.90 (0.06)	3.51 (0.04)	0.80 (0.16)	2.34 (1.39)
EBFormer-l0	9.48	3.27 (0.10)	6.89 (0.07)	1005.10 (10.87)	2101.54 (62.78)	1.32 (0.05)	3.93 (0.04)	853.81 (34.00)	1880.41 (94.95)	0.33 (0.01)	1.38 (0.01)	0.78 (0.03)	3.75 (0.05)	0.61 (0.02)	1.25 (0.14)
EBFormer-l1	9.35	3.17 (0.06)	6.74 (0.14)	1014.96 (14.18)	2108.01 (53.59)	**1.26** (0.03)	3.76 (0.16)	819.26 (11.23)	1703.02 (21.65)	0.32 (0.01)	1.38 (0.05)	0.75 (0.02)	3.61 (0.11)	0.60 (0.01)	1.35 (0.28)
L-EBFormer-l1	9.35	2.85 (0.06)	**6.21** (0.10)	**3.19** (0.06)	**7.67** (0.14)	**1.26** (0.04)	**3.37** (0.11)	**1.99** (0.04)	**4.92** (0.12)	**0.31** (0.00)	**1.30** (0.03)	**0.67** (0.02)	**3.24** (0.13)	**0.49** (0.01)	**0.90** (0.07)

Table 13: **Nanosheet Extrapolation Including LDOS/LcDOS Errors**: Extrapolation accuracy including LDOS/LcDOS errors. Note that only L-EBFormer-l1 was trained on the local task. Models which cannot project predictions to a local basis (MLP and DOSTransformer) have no errors listed for the local tasks

Model	Params [M]	DOS [$\times 10^{-3}$]		LDOS [$\times 10^{-3}$]		cDOS [$\times 10^{-3}$]		LcDOS [$\times 10^{-3}$]		N_{inv} [$\times 10^{11}$]		I_{inv} [$\times 10^{-3}$]		v_{inj} [$\times 10^{6}$]	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
MLP	0.99	27.69 (5.88)	41.18 (9.80)	–	–	23.88 (1.56)	37.09 (1.72)	–	–	2.38 (0.28)	5.15 (0.20)	8.39 (0.43)	17.93 (1.72)	7.94 (0.12)	13.20 (1.35)
NequIP-l1 (Parity)	1.04	8.82 (0.30)	11.70 (0.33)	975.21 (141.96)	1574.29 (268.93)	4.39 (0.14)	6.05 (0.30)	913.48 (161.69)	1511.96 (336.40)	1.22 (0.02)	2.44 (0.06)	2.68 (0.06)	5.64 (0.34)	2.37 (0.04)	2.98 (0.10)
DOSTransformer	0.95	11.88 (1.18)	16.89 (0.79)	–	–	7.11 (0.63)	10.16 (0.50)	–	–	1.42 (0.20)	3.21 (0.28)	4.20 (0.92)	8.57 (1.66)	4.99 (2.31)	7.34 (3.61)
EBFormer-l0	0.94	9.12 (0.19)	12.42 (0.17)	940.02 (22.33)	1694.37 (34.84)	4.37 (0.26)	7.41 (0.23)	797.94 (33.99)	1220.09 (24.82)	1.35 (0.17)	3.04 (0.17)	2.42 (0.16)	6.61 (0.21)	1.00 (0.17)	1.32 (0.14)
EBFormer-l1	0.91	7.97 (0.23)	10.89 (0.25)	831.37 (17.69)	1497.16 (121.50)	**3.37** (0.23)	**5.39** (0.26)	721.85 (113.18)	1043.88 (181.63)	**1.00** (0.09)	**2.29** (0.10)	**1.81** (0.16)	**4.69** (0.25)	**0.89** (0.09)	**1.24** (0.05)
L-EBFormer-l1	0.91	**7.52** (0.24)	**10.55** (0.44)	**9.49** (0.23)	**14.53** (0.74)	4.58 (0.29)	6.98 (0.62)	**7.01** (0.41)	**10.85** (0.57)	1.03 (0.02)	2.48 (0.15)	2.76 (0.35)	6.60 (0.87)	0.97 (0.08)	1.35 (0.06)

Table 14: **Nanowire Interpolation Including LDOS/LcDOS Errors**: Nanowire interpolation accuracy including LDOS/LcDOS errors. Note that only L-EBFormer-l1 was trained on the local task. Models which cannot project predictions to a local basis (MLP and DOSTransformer) have no errors listed for the local tasks

Model	Params [M]	DOS [$\times 10^{-3}$]		LDOS [$\times 10^{-3}$]		cDOS [$\times 10^{-3}$]		LcDOS [$\times 10^{-3}$]		N_{inv} [$\times 10^{11}$]		I_{inv} [$\times 10^{-3}$]		v_{inj} [$\times 10^{6}$]	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE				
MLP	0.99	27.73 (1.71)	43.41 (0.54)	–	–	12.76 (1.11)	19.22 (0.63)	–	–	0.29 (0.01)	0.85 (0.04)	6.78 (0.60)	17.81 (0.46)	2.14 (0.31)	2.75 (0.40)
NequIP-l1 (Parity)	1.04	20.60 (0.04)	35.91 (0.12)	1350.25 (51.65)	3076.54 (484.69)	2.96 (0.09)	4.66 (0.14)	1921.92 (77.02)	4970.71 (1142.77)	0.14 (0.00)	0.33 (0.01)	0.93 (0.06)	2.53 (0.37)	1.68 (0.02)	2.17 (0.05)
DOSTransformer	0.95	20.69 (2.54)	34.73 (2.25)	–	–	6.18 (2.21)	8.23 (2.62)	–	–	0.14 (0.02)	0.32 (0.10)	3.91 (1.93)	7.18 (3.27)	2.89 (1.11)	3.34 (1.11)
EBFormer-l0	0.94	16.66 (0.29)	31.37 (0.66)	1104.04 (33.82)	1703.17 (99.40)	1.59 (0.21)	2.90 (0.55)	1527.16 (147.21)	2257.03 (272.30)	0.10 (0.01)	0.25 (0.01)	**0.51** (0.11)	1.54 (0.52)	1.52 (0.02)	1.93 (0.02)
EBFormer-l1	0.91	16.07 (1.20)	30.64 (1.14)	1483.64 (120.16)	2693.52 (121.83)	**1.50** (0.35)	**2.62** (0.66)	1303.90 (70.34)	2150.89 (137.52)	0.10 (0.00)	**0.24** (0.01)	0.51 (0.15)	**1.54** (0.52)	1.53 (0.05)	1.96 (0.06)
L-EBFormer-l1	0.91	**14.34** (0.13)	**29.30** (0.24)	**12.89** (0.09)	**35.68** (0.14)	1.70 (0.06)	2.93 (0.18)	**4.05** (0.08)	**9.37** (0.21)	**0.10** (0.00)	0.24 (0.01)	0.59 (0.05)	1.69 (0.23)	**1.44** (0.02)	**1.84** (0.02)

Note that out of all the models presented in the tables, only L-EBFormer-l1 was trained to optimize local targets. All other networks were trained to minimize error on the DOS/cDOS after mean-pooling. Therefore, the graph networks presented in these tables have high losses.

I.1 Local vs. Global Training

In the preceding sections, we observe the local variant of EBFormer (L-EBFormer) *outperforming* the global version on inference of global material properties including DOS, cDOS, and downstream derived properties, notably on in-distribution inference tasks. In other cases such as the OOD task, the global-variant (EBFormer) seems to have lower errors across most metrics. This section provides a brief analysis across data- and parameter-sparse regimes to motivate a qualitative understanding of when local training may be advantageous.

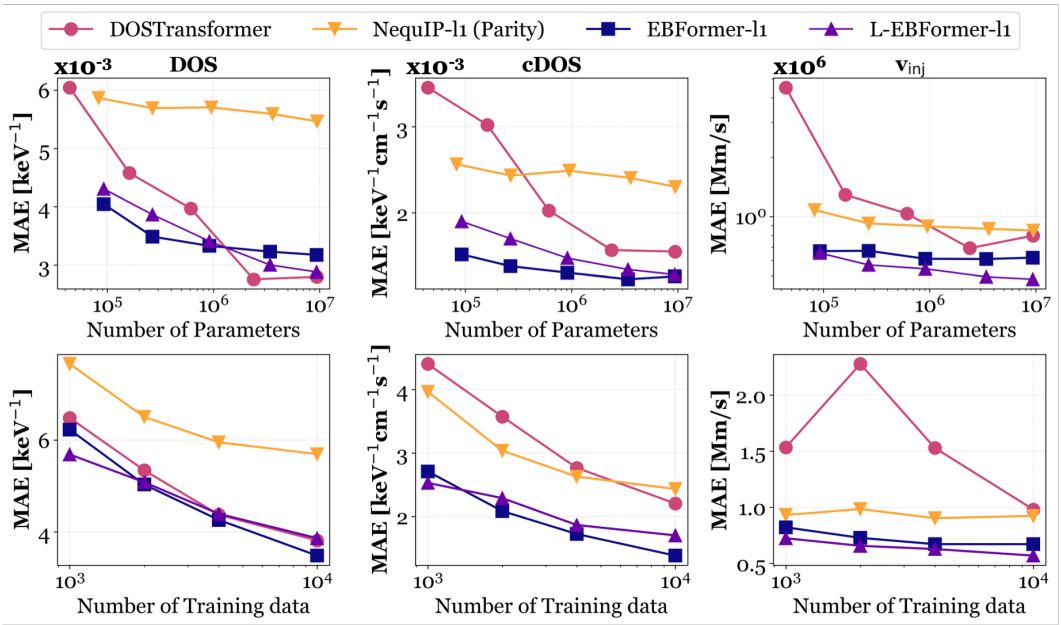

Figure 17: **Parameterization & Learning Curves with L-EBFormer**: Reproduction of Figure 4 with full L-EBFormer data included. The parameterization curves demonstrate that the local variant achieves comparable performance to the global variant only with sufficient parameters. The learning curve compares various models all scaled to 10^6 parameters, demonstrating mediocre local performance.

Figure 17 reproduces Figure 4 (errors on the nanosheet dataset with varied number of training data and model parameters) with the full L-EBFormer curve included. The parameterization curves (first row) demonstrate L-EBFormer underperforming EBFormer until the local model is sufficiently large on the training targets of DOS and cDOS. Capturing local feature variation is a more complex task with significantly more input data; the model must learn a DOS/cDOS spectral vector *per every atom*, with each energy bin scaled by the corresponding energy eigenfunction. The requirement for more parameters is therefore plausible, as the model must spend representational capacity on reproducing local property variation.

However, with sufficient parameterization, the local model has lower error than the global variant. This implies that inclusion of geometric variation of material properties provides the model useful additional information that permits more accurate pooled predictions. From a materials perspective, a similar phenomenon has been observed by providing forces as an additional training target to energy-prediction models, providing significantly lower losses even with orders-of-magnitude fewer data Schütt et al. (2017). This is a specific instance of multitask learning (MTL), an example of inductive transfer in which related auxiliary tasks improve model generalization Caruana (1997). From a graph-theory approach, Hu et al. (2020) empirically observed that training graph- and node-level features simultaneously improved downstream model prediction on molecular and protein inference tasks. Finally, deeply-supervised networks Lee et al. (2015) demonstrate that enforcing training objectives on the discriminativeness of hidden layers can improve downstream predictions and allow better gradient propagation. In our case, the ground-truth of the atom-local quantities are known directly from physical theory, rather than using surrogate losses to improve the hidden feature information.

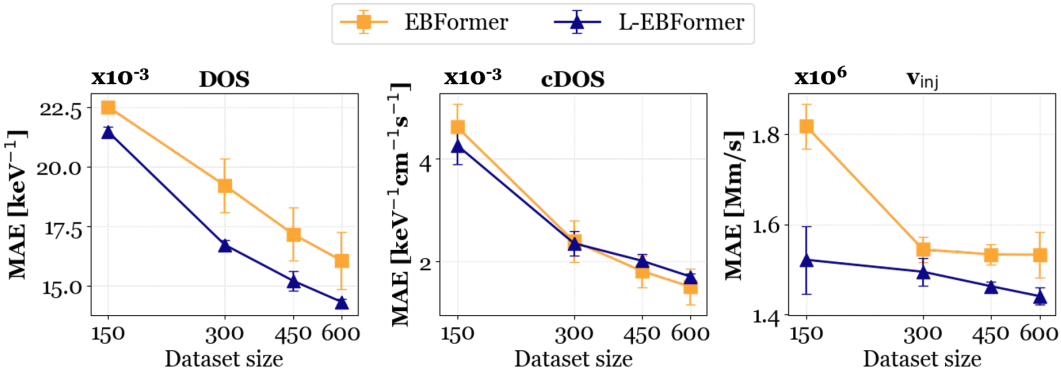

Figure 18: **Nanowire Learning Curves** L-EBFormer performs similarly on DOS/cDOS inference in the data-limited regime, and significantly better in the downstream property of injection velocity. Model parameterization is kept at ∼1M parameters. Further discussion of consistency of error in model predictions providing better downstream performance is provided in Appendix G

Figure 18 and the second row of Figure 17 demonstrate that while L-EBFormer underperforms in the parameter-limited regime, it performs similarly or better to the pooled, global variant in the data-sparse regime, demonstrated on the interpolation tasks for both the nanosheet and nanowire cases. This demonstrates that inclusion of local information is a generally useful auxiliary signal with high task *relatedness*, in the parlance of MTL.

In summary, L-EBFormer can be interpreted as a multitask-style or deeply supervised variant of EBFormer, where atom-local LDOS/LcDOS prediction acts as an intermediate supervision signal. This supervision imposes a stronger inductive bias on the model by constraining the spatially decomposed DOS/cDOS (LDOS/LcDOS), thereby improving in-distribution global and downstream predictions. However, as in the general paradigm of MTL, the benefit depends on task alignment; under stronger distribution shift or higher local-field complexity, the auxiliary local task can introduce additional variance or unfavorable bias that reduces performance. This is likely the case in the OOD study, where the local property distribution of smaller structures was significantly different from those of larger systems. We finally note that the local variant has steeper scaling with increased model parameterization, demonstrating that the model capabilities have not plateaued with the model sizes chosen in this work.

J Transfer Learning

The out-of-distribution task tested the zero-shot ability of EBFormer to make predictions on systems with confined dimensions larger than those in the training set. This was accomplished by training the model on 15-45 atomic layers of silicon of all cleaving orientations and strain, and testing on silicon nanosheets of 46-100 atomic layers. Simulation of systems with large number of atoms rapidly becomes computationally infeasible, which makes extrapolation performance an important metric to quantify.

However, as the paradigm of foundation models becomes increasingly central to materials informatics Batatia et al. (2024); Kovács et al. (2025), understanding the ability of our model to transfer information across chemical systems is also of interest. It is conceivable that a researcher exploring the properties of a novel material systems such as III-V semiconductor nanosheets may use a pretrained model with knowledge of confinement physics over a large range of nanostructures of traditional materials, and finetune with a few, small, cheaply generated examples of their new system. They may then desire to use the model to infer the behavior of nanostructures of various geometries comprised of the novel material.

We conduct an additional experiment to study EBFormer's potential in this setting by quantifying its ability to transfer geometric knowledge across chemical species. We achieve this by augmenting the out-of-distribution training dataset with germanium nanosheets of all cleaving orientations and strains with thickness between 15-100 atomic layers. This has the effect of including geometric information of larger (germanium) nanoslabs to the test set of small (silicon) nanoslabs. We then test

Table 15: **Extrapolation and Transfer Learning Errors** DOS (eV^{-1}), cDOS ($eV^{-1}cm^{-1}s^{-1}$), N_{inv} (cm^{-2}), I_{inv} (mA/μm), and injection velocity (cm/s) errors on the nanosheet extrapolation task are shown below. In addition, performance on the same silicon test-task of 46-100 atomic layers is shown on models trained with 15-100 atomic layers of germanium added to the 15-45 atomic-layer silicon dataset

Model	Params [M]	DOS $[\times 10^{-3}]$		cDOS $[\times 10^{-3}]$		N_{inv} $[\times 10^{11}]$		I_{inv} $[\times 10^{-3}]$		v_{inj} $[\times 10^{6}]$	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
EBFormer-l0	0.94	9.12 (0.19)	12.42 (0.17)	4.37 (0.26)	7.41 (0.23)	1.35 (0.17)	3.04 (0.17)	2.42 (0.16)	6.61 (0.21)	1.00 (0.17)	1.32 (0.14)
EBFormer-l1	0.91	7.97 (0.23)	10.89 (0.25)	**3.37** (0.23)	**5.39** (0.26)	1.00 (0.09)	2.29 (0.10)	**1.81** (0.16)	**4.69** (0.25)	0.89 (0.09)	1.24 (0.05)
L-EBFormer-l1	0.91	7.52 (0.24)	10.55 (0.44)	4.58 (0.29)	6.98 (0.62)	1.03 (0.02)	2.48 (0.15)	2.76 (0.35)	6.60 (0.87)	0.97 (0.08)	1.35 (0.06)
Ge+EBFormer-l0	0.94	8.24	11.55	3.91	6.89	1.13	2.74	2.20	6.22	0.77	1.09
Ge+EBFormer-l1	0.91	8.02	10.98	3.96	5.91	1.18	2.54	2.22	5.28	0.76	1.17
Ge+L-EBFormer-l1	0.91	**6.51**	**9.35**	3.44	5.53	**0.80**	**2.21**	1.86	4.88	**0.71**	**1.07**

performance on the same test set of large silicon nanoslabs as studied in the out-of-distribution evaluation in Section 4.2, which becomes in-distribution in terms of chemical composition and geometry individually, but not both simultaneously. The results are appended to the losses in the extrapolation task, with models trained with germanium labeled with the prefix **Ge+**. Losses are summarized in Table 15.

We notice minor improvements or comparable performance between EBFormer-l1 and the same model when augmented with germanium information. However, we see significant improvement in performance for the L-EBFormer model, indicating that EBFormer is able to leverage geometric information using spatial variation of atom-local quantities to improve global and local predictions.

K Proof of Symmetry Preservation

In this section, we prove that EBFormer is $SO(3)$ Equivariant, and identical under supercell replication. This implies that the updated embeddings of atomic and boundary nodes $\left(\mathbf{x_i}^{(t+1)}, \mathbf{b_k}^{(t+1)}\right)$ remain equivariant under spatial transforms and invariant under supercell replication after the message-passing and update phases. Formally, equivariance implies for any orthogonal matrix $Q \in \mathbb{R}^{3\times 3}$, EBFormer satisfies

$$Q\mathbf{x_i}^{(t+1)}, Q\mathbf{b_k}^{(t+1)} = \text{EBFormer}\left(Q\mathbf{x_i}^{(t)}, Q\mathbf{b_k}^{(t)}\right) \tag{10}$$

EBFormer contains two stages: (i) The local equivariant convolution, and (ii) The global equivariant attention between atom and the boundary nodes. If each stage preserves the requisite symmetries, then their composition is equivariant. The equivariance and supercell invariance of the local convolution follows from prior work Batzner et al. (2022); Satorras et al. (2021); Xie & Grossman (2018). We therefore focus on the cross-attention between boundary and atom nodes, specifically Eqs. (4) and (5) in the main text. If these two equations are equivariant and preserve symmetry under replication, then the entirety of EBFormer does as well.

K.1 Atom nodes

Assume $\mathbf{x}_i^{(t)}$ and $\mathbf{b}_k^{(t)}$ transform equivariantly under $SO(3)$. We show that $\mathbf{x}_i^{(t+1)}$ is also equivariant:

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + \sum_k \alpha_{ik} \left(\mathbf{b}_k^{(t)} \overset{\mathbf{W}_V^A}{\otimes} Y_J^{(l)}(\vec{\mathbf{d}}_{ik}) \right) \tag{11}$$

Here, $Y_J^{(\ell)}(\vec{\mathbf{d}}_{ik})$ denotes the spherical-harmonic expansion of the relative vector from atom i to boundary plane k, which is equivariant by construction. The tensor product with $\mathbf{b}_k^{(t)}$ uses Clebsch–Gordan coefficients Thomas et al. (2018), thus preserving equivariance (similar to standard

equivariant graph convolutions). The attention weights α_{ik} are *scalars* computed from queries $\mathbf{q}_k$ and keys $\mathbf{k}_i$ via equivariant linear maps applied independently to each irrep, followed by tensor products and extraction of scalar (invariant) components. Hence α_{ik} are invariant under rotations. Multiplying invariant scalars with equivariant features preserves equivariance, and summing over k messages preserves equivariance by closure under addition. Therefore, $\mathbf{x}_i^{(t+1)}$ is equivariant given $\mathbf{x}_i^{(t)}$ and $\mathbf{b}_k^{(t)}$ are equivariant.

In addition, under systems that are periodic in one or more directions, we know that atomic embeddings remain identical from local convolutions Xie & Grossman (2018); Batzner et al. (2022). Therefore, attention scores are also identical to the boundary nodes (assuming inductively that the boundary node embeddings are also the same, shown in the next section). Since the number of boundary nodes and distance remain identical, the attention mechanism of this step remains unchanged by supercell replication.

K.2 BOUNDARY NODES

Similarly, assume $\mathbf{x}_i^{(t)}$ and $\mathbf{b}_k^{(t)}$ are $SO(3)$-equivariant. We show $\mathbf{b}_k^{(t+1)}$ is equivariant:

$$\mathbf{b}_k^{(t+1)} = \mathbf{b}_k^{(t)} + \sum_i \alpha_{ki} \left(\text{EquivLinear}_V^B \left(\mathbf{x}_i^{(t)} \middle| \text{MLP}_V^B \left(\mathbf{f}_{ik} \right) \right) \right) \tag{12}$$

The features $\mathbf{f}_{ik}$ are invariant, as they are functions of scalar distance quantities. The map EquivLinear_V^B acts as an irrep-wise linear operator and therefore preserves equivariance. The attention weights α_{ki} are scalar invariants constructed as in the atom node case. Products of invariant scalars with equivariant features are equivariant, and summation over i preserves equivariance. Hence $\mathbf{b}_k^{(t+1)}$ is equivariant given $\mathbf{x}_i^{(t)}$ and $\mathbf{b}_k^{(t)}$ are equivariant.

Finally, assuming inductively that the atomic embeddings and boundary nodes are identical in a supercell from the previous message passing layers, which is true in the initialization step as initial atomic embeddings are by element and boundaries are all identical, then the attention scores to the replicated system are the same as the original unit-cell. The weighted average of attention then provides the same update to the boundary node, thereby preserving invariance to supercell replication.