

Neural-Network-Based Effective-Mass Hamiltonian for Quantum Transport Simulations of Disordered Devices

Guenseok Choi*, Jooseok Kim*, Gyeongwon Roh, and Mincheol Shin

(* Equal Contribution)

School of Electrical Engineering, Korea Advanced Institute of Science and Technology

Daejeon 34141, Rep. of Korea

E-mail: mshin@kaist.ac.kr

Abstract—Although first-principles methods accurately describe semiconductor electronic structures, their direct application to device-level quantum transport remains computationally expensive. We propose NNH-EMA, a neural-network framework that constructs an effective-mass Hamiltonian by fitting first-principles band structures. Focusing on the dispersion near the conduction-band edge, the model generates a Hamiltonian for transport calculations. The framework is validated for c-IGZO with composition disorder, and it reproduces the band structures, transmission characteristics, and mobility behavior obtained from first-principles calculations. These results show that the method is suitable for large-scale quantum transport simulations and provides a link between first-principles calculations and device-level transport analysis.

Index Terms—Effective-mass Hamiltonian, Machine learning, First-principles Band Structure, Quantum Transport, and Graph Neural Network

I. INTRODUCTION

First-principles electronic-structure calculations are widely used to study semiconductor materials and devices, particularly when interfaces, heterostructures, compositional disorder, or amorphous regions strongly influence the electronic structure and transport properties [1], [2]. However, density functional theory (DFT) is computationally expensive for device-level quantum transport because both Hamiltonian construction and transport calculations with large matrices are costly. Neural network Hamiltonian (NNH) methods can learn DFT Hamiltonians, enabling direct prediction of first-principles Hamiltonians for electronic-structure calculations [3], [4] and accelerating atomistic simulations such as molecular dynamics [5]. However, transport calculations remain costly because the learned Hamiltonians retain the full localized basis and remain large. The computational cost of non-equilibrium Green's function (NEGF)-based transport calculations increases rapidly with matrix size. Therefore, it is more important to reduce the size of the Hamiltonian itself than to simply reduce the time required to construct it.

As a method for constructing a compact transport Hamiltonian, the effective-mass approximation Hamiltonian (EMA Hamiltonian) has been widely used. The EMA Hamiltonian reduces the atomistic complexity of first-principles Hamiltonians into a compact representation, enabling efficient quantum transport simulations while retaining the essential band-edge properties relevant to carrier transport. Based on this advan-

tage, previous studies have modeled cation-disorder-induced transport in crystalline indium gallium zinc oxide (c-IGZO) by extracting the electrostatic potential from DFT calculations and incorporating it into an EMA Hamiltonian [6]. In this approach, the random distribution of Ga and Zn cations is treated as a disorder potential that functions as an electron-scattering source in c-IGZO [7]. However, the parameters for the model disorder potential were calibrated to reproduce the spatially averaged disorder effect, rather than fully resolved atomistic configurations. Therefore, a general and transferable route for constructing a transport-ready EMA Hamiltonian directly from first-principles band structures of complex, disordered device materials has remained lacking.

We address this with the Neural Network Hamiltonian Effective Mass Approach (NNH-EMA), which learns the spatially varying potential of a fixed-mass EMA Hamiltonian from first-principles band structures. A graph neural network learns this disorder potential from the local cation environment, so the EMA Hamiltonian captures configuration-specific scattering while remaining a compact, single-band real-space operator whose reduced matrix size makes NEGF transport tractable. On compositionally disordered c-IGZO, NNH-EMA reproduces the band structures, transmission characteristics, and mobility behavior obtained from first-principles calculations.

II. METHOD

To reproduce the target conduction band, we use an EMA Hamiltonian of the form

$$\hat{H}_{\text{EMA}} = -\frac{\hbar^2}{2m^*} \nabla^2 + V_s(\mathbf{r}) \quad (1)$$

with $m^* = 0.2 m_0$ for c-IGZO. The neural network is trained to predict the spatially varying potential $V_s(\mathbf{r})$ within this fixed-mass EMA framework. In this framework the effective mass is fixed by the bulk band curvature, so the network is left to capture the entire effect of cation disorder through the potential alone. The model is implemented using a message-passing neural network adapted from the code of Refs. [8]–[10]. The network operates on this graph by passing messages along the edges, alternately updating the node and

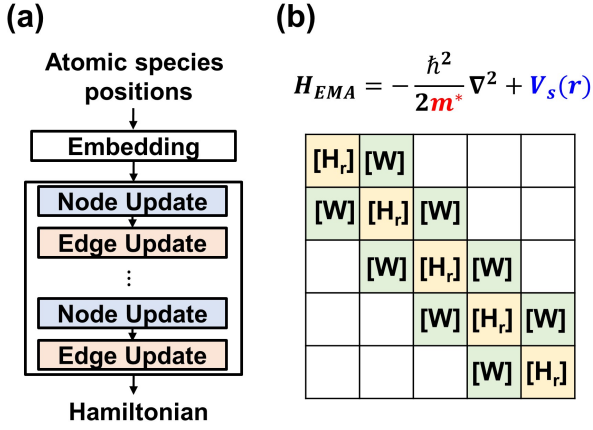


Fig. 1. Overview of the proposed method. (a) An e3nn-based neural network with alternating node-update and edge-update blocks is used for learning. (b) The EMA finite-difference Hamiltonian matrix is shown.

edge features so that each grid node aggregates information from its surrounding atomic environment. Stacking several such message-passing layers progressively expands the spatial range over which the predicted potential responds to the local cation arrangement. Edges are defined from their positional relationships, and the outputs are interpreted as the EMA potential.

A loss function based solely on band energies is insufficient to determine the local shape of the band dispersion. The energies fix the band only at the sampled k -points, leaving the slope and curvature of the dispersion between them unconstrained. Therefore, the total loss is defined to incorporate the band energies and their first and second derivatives, as

$$\mathcal{L}_{\text{tot}} = \lambda_E \mathcal{L}_E + \lambda_{\partial k} \mathcal{L}_{\partial k} + \lambda_{\partial^2 k} \mathcal{L}_{\partial^2 k}, \quad (2)$$

$$\mathcal{L}_E = \frac{1}{N_k} \sum_{i=1}^{N_k} \omega_i |E_n^{\text{pred}}(k_i) - E_n^{\text{DFT}}(k_i)|^2, \quad (3)$$

$$\mathcal{L}_{\partial k} = \frac{1}{N'_k} \sum_{i=1}^{N'_k} \omega_i \left| \frac{dE_n^{\text{pred}}}{dk} \Big|_{k_i} - \frac{dE_n^{\text{DFT}}}{dk} \Big|_{k_i} \right|^2, \quad (4)$$

$$\mathcal{L}_{\partial^2 k} = \frac{1}{N'_k} \sum_{i=1}^{N'_k} \omega_i \left| \frac{d^2 E_n^{\text{pred}}}{dk^2} \Big|_{k_i} - \frac{d^2 E_n^{\text{DFT}}}{dk^2} \Big|_{k_i} \right|^2, \quad (5)$$

Here, $E_n(k_i)$ is the target conduction-band energy at the i th k -point, where n labels the conduction band targeted for transport, and the fit is restricted to this single band. N_k and N'_k denote the numbers of sampled k -points used for the energy and derivative terms, respectively. The coefficients λ_E , $\lambda_{\partial k}$, and $\lambda_{\partial^2 k}$ set the relative weight of the energy, slope, and curvature contributions. The weight ω_i down-weights k -points

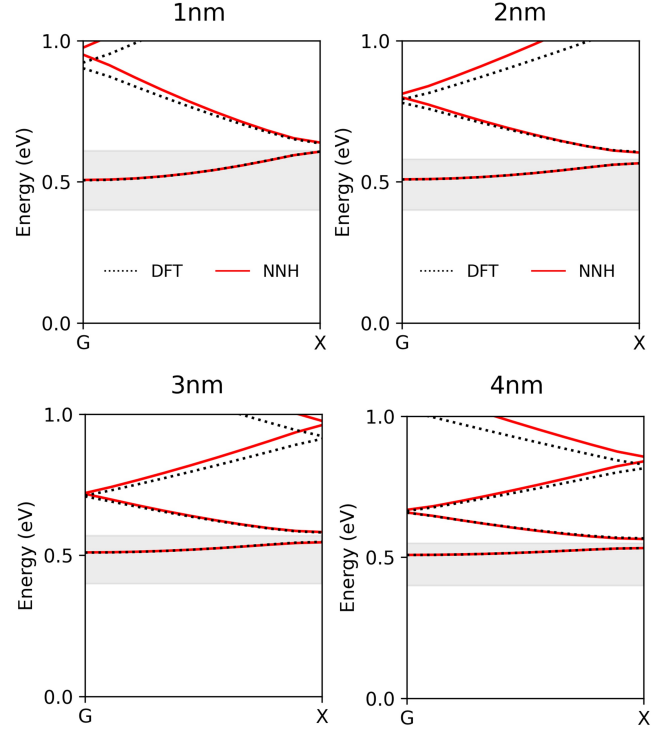


Fig. 2. Comparison of DFT and NNH-EMA band structures for c-IGZO supercells with 1, 2, 3, and 4 nm channel length. Trained using only the first conduction band, the model achieves band-edge errors of 0.5 meV, 0.3 meV, 0.7 meV, and 0.2 meV respectively.

whose energies lie outside the target window, so that the fit concentrates on the band-edge region relevant for low-field transport rather than on states far above it.

For training, c-IGZO is described by DFT calculations performed using SIESTA on a $1 \times 4 \times 2$ k -point grid. For transport, the real-space EMA Hamiltonian is reduced with a supercell mode-space approach [11], which projects it onto a compact basis of supercell Bloch states retained within a narrow energy window above the conduction-band edge. Constructing this transformation in block-diagonal form lets the reduction accommodate the spatially varying disorder potential while keeping the number of modes in each region small, and it reproduces the real-space transmission to within a few percent. Because the cost of a NEGF calculation grows roughly as the cube of the Hamiltonian size, this reduction makes device-scale transport tractable. The transmission is then evaluated within the NEGF formalism, and the mobility is extracted using the dR/dL method from the disorder-averaged Landauer conductance [12], [13].

III. RESULT

For c-IGZO, we use 40 relaxed supercells with random Ga/Zn arrangements taken from Ref. [14]. These supercells span channel lengths from 1 to 4 nm at the same cation composition. To assess transferability across length, we train

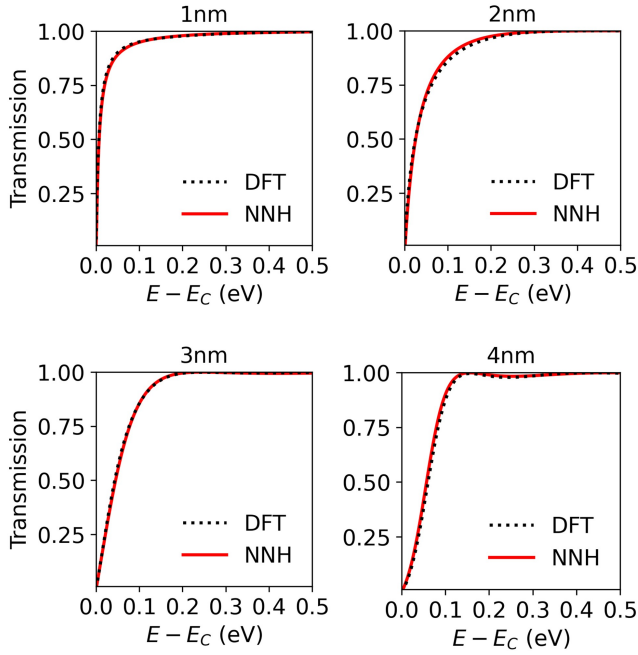


Fig. 3. Comparison of the transmissions obtained from DFT and NNH-EMA for c-IGZO. For each channel length, the transmissions are averaged over five test structures. NNH-EMA reproduces the transmission onset and overall spectral shape over the energy range considered.

only on the 20 structures with 1 and 2 nm channels and evaluate on the 3 and 4 nm channels. This split is demanding, as the network must extrapolate the disorder potential beyond any length seen in training rather than interpolate within a fixed device. This test thus distinguishes whether the network has learned a local, transferable mapping from the cation environment to the potential, or one specific to a particular supercell size.

As shown in Fig. 2, the predicted band structures agree closely with the DFT references across the full set of channel lengths. The average band-edge error remains below 1 meV at every channel length, with values of 0.5 and 0.3 meV for the 1 and 2 nm training channels and 0.7 and 0.2 meV for the 3 and 4 nm channels, so accuracy is preserved even outside the training range. This generalization to channel lengths beyond the training set indicates that the model learns a transferable mapping rather than interpolating within a fixed device size. Because each site's potential depends on the cations within the network's finite receptive field, this mapping extends to channels longer than those seen in training. This band-edge accuracy supports the reproduction of the transport quantities discussed below, since the transmission and mobility are largely governed by the disorder-induced scattering near the conduction-band edge.

The transmission and mobility trends in Fig. 3 and Fig. 4 are likewise well reproduced. The transmission curves obtained from the NNH-EMA Hamiltonian follow the DFT references. Because the dR/dL extraction requires a linear resistance-

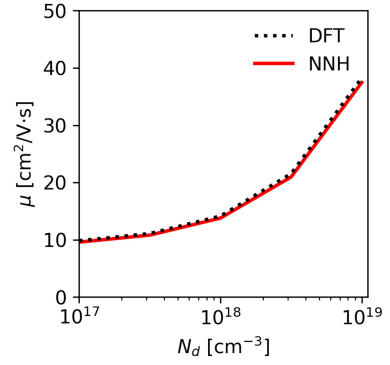


Fig. 4. Comparison of the electron mobility μ as a function of donor concentration N_d for c-IGZO. The mobility increases with N_d , and the NNH-EMA results closely follow the corresponding DFT trend over the concentration range considered.

length relation, we confirmed this linearity holds for the disorder-averaged NNH-EMA conductance before extracting mobility. The resulting mobility increases with donor concentration N_d in close agreement with the DFT trend for both the 1 and 2 nm channels. The rise of mobility with N_d is the defining transport signature of c-IGZO [14].

Beyond reproducing these quantities, NNH-EMA enables rapid construction of an effective Hamiltonian for transport, substantially reducing the transmission cost relative to DFT. For c-IGZO, the per-block Hamiltonian dimension is reduced from 3852×3852 in the reference DFT to 54×54 , and the corresponding per-configuration runtime drops from about two days to about one minute. This reduction makes disorder-averaged transport practical, since obtaining a converged mobility requires evaluating many random cation configurations, and a per-configuration cost of minutes rather than days brings the full ensemble within reach.

IV. CONCLUSION

In this work, we presented NNH-EMA, a neural-network-based method for learning effective-mass Hamiltonians from first-principles band structures. For c-IGZO with composition disorder, it accurately reproduces band structures and the corresponding transmission and mobility trends. It also shows transferability to longer c-IGZO channels beyond those used in training, while completing band and transmission evaluation in about one minute per configuration. These results demonstrate a practical route to rapid quantum transport analysis of larger and more complex device structures.

ACKNOWLEDGMENT

This work was supported by K-CHIPS (Korea Collaborative & High-tech Initiative for Prospective Semiconductor Research) (RS-2024-00406007, 2410010283, 00406007) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea)

REFERENCES

- [1] M. Brandbyge, J.-L. Mozos, P. Ordejón, J. Taylor, and K. Stokbro, "Density-functional method for nonequilibrium electron transport," *Physical Review B*, vol. 65, no. 16, p. 165401, 2002.
- [2] M. Shin, W. J. Jeong, and J. Lee, "Density functional theory based simulations of silicon nanowire field effect transistors," *Journal of Applied Physics*, vol. 119, no. 15, p. 154505, 2016.
- [3] X. Gong, H. Li, N. Zou, R. Xu, W. Duan, et al., "General framework for E(3)-equivariant neural network representation of density functional theory Hamiltonian," *Nature Communications*, vol. 14, no. 2848, 2023.
- [4] M. Kaniselman, B. K. Miller, M. Gao, J. Nam, and D. S. Levine, "Learning from the electronic structure of molecules across the periodic table," arXiv preprint arXiv:2510.00224, 2025.
- [5] G. Choi and M. Shin, "Efficient Quantum-Mechanics-Based Molecular Dynamics Through Machine Learning-Driven Density Functional Theory Hamiltonian," in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, Grenoble, France, 2025.
- [6] D. Seo et al., "Transport Properties of Crystalline IGZO Channel Devices: Effects of Cation Disorders, Composition and Dimensions," in *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2024.
- [7] Y. Kang, Y. Cho, and S. Han, "Cation disorder as the major electron scattering source in crystalline InGaZnO," *Applied Physics Letters*, vol. 102, no. 15, p. 152104, 2013.
- [8] C. H. Xia, M. Kaniselman, A. N. Ziogas, M. Mladenović, R. Mahjoub, et al., "Learning the Electronic Hamiltonian of Large Atomic Structures," in *Proceedings of the International Conference on Machine Learning (ICML)*, Vancouver, Canada, PMLR, 2025.
- [9] Y.-L. Liao, B. Wood, A. Das, and T. Smidt, "EquiformerV2: Improved Equivariant Transformer for Scaling to Higher-Degree Representations," in *Proceedings of the International Conference on Learning Representations (ICLR)*, Vienna, Austria, 2024.
- [10] S. Passaro and C. L. Zitnick, "Reducing SO(3) Convolutions to SO(2) for Efficient Equivariant GNNs," in *Proceedings of the International Conference on Machine Learning (ICML)*, Honolulu, Hawaii, USA, 2023.
- [11] M. Shin, "Large-scale Quantum Simulations Enabled By Real-space-equivalent SuperMS Method," in *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2025.
- [12] T. Markussen, M. Palsgaard, D. Stradi, T. Gunst, M. Brandbyge, and K. Stokbro, "Electron-phonon scattering from Green's function transport combined with molecular dynamics: Applications to mobility predictions," *Physical Review B*, vol. 95, no. 24, p. 245210, 2017.
- [13] K. Rim, S. Narasimha, M. Longstreet, A. Mocuta, and J. Cai, "Low field mobility characteristics of sub-100 nm unstrained and strained Si MOSFETs," in *IEEE International Electron Devices Meeting (IEDM)*, San Francisco, CA, USA, 2002.
- [14] S. H. Han, D. Seo, J.-H. Cha, S. Kim, and M. Shin, "Cation Disorder Limited IGZO Mobility Calculation Based on the Density Functional Theory," in *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, San Jose, USA, 2024.