DFT Hamiltonian Neural Network Training with Semi-supervised Learning

Yucheol Cho* **Guenseok Choi*** **Gyeongdo Ham**
Mincheol Shin† **Dae-Shik Kim†**

School of Electrical Engineering,
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, 34141, South Korea
{yc_cho, ieggiegg, rudeh6185, mshin, daeshik}@kaist.ac.kr

Abstract

Recent efforts have focused on training neural networks to replace density functional theory (DFT) calculations. However, prior neural network training methods required an extensive number of DFT simulations to obtain the ground truth (Hamiltonians). Conversely, when working with limited training data, deep learning models often exhibit increased errors in predicting Hamiltonians and band structures for testing data. This phenomenon carries the potential risk of yielding inaccurate physical interpretations, including the emergence of unphysical branches within band structures. To address this challenge, we introduce a novel deep learning-based method for calculating DFT Hamiltonians, specifically designed to generate accurate results with limited training data. Our framework not only employs supervised learning with the calculated Hamiltonian but also generates pseudo Hamiltonians (targets for unlabeled data) and trains the neural networks on unlabeled data. We compare our results with those obtained using the state-of-the-art method, which trains neural networks using atomic structures as inputs and DFT Hamiltonians as targets. We demonstrate the superior performance of our framework compared to the previous approach on various datasets, such as MoS_2, Bi_2Te_3, HfO_2, and InGaAs.

1 Introduction

Over the past years, there have been tremendous efforts to train a neural network, to replace density functional theory (DFT) calculation [14, 24, 8, 23, 11, 25]. While previous research has focused on individual physical properties [6, 2, 22, 7], such as charge density and band structure, applying deep learning to the DFT Hamiltonian itself is a critical and challenging task [8]. When attempting to use neural networks to model the relationship between material structure and the DFT Hamiltonian for large-scale material systems, challenges arise due to the exponential growth in the number of independent variables and the dimensions of the Hamiltonian matrix. To address this issue, a comprehensive deep learning approach for DFT Hamiltonians was introduced under the name DeepH [14]. This framework was specifically designed to investigate crystalline materials by employing a message-passing neural network. The complex challenges associated with the inherently large dimensions and covariance concerns of the DFT Hamiltonian matrix were effectively resolved by incorporating locality principles. These included the utilization of local coordinates, localized basis transformations, and orbitals localized as basis functions. The DeepH framework consistently

*Equal contribution.
†Equal correspondence.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

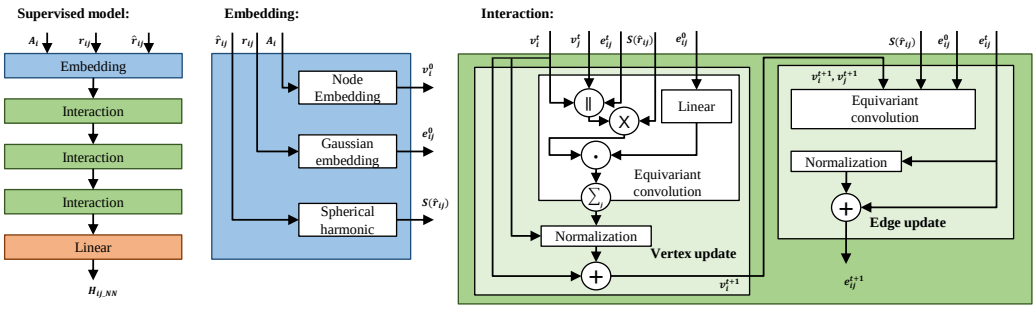

Figure 1: The neural network structure consisting of blocks such as embedding and interaction.

exhibited exceptional accuracy, not only in constructing the DFT Hamiltonian but also in computing various physical properties related to band structures and wavefunctions.

Despite these efforts in previous studies, it is still overlooked that more than hundreds of DFT simulation results are required to train neural networks. When the model learns from an insufficient quantity of training data, it exhibits elevated errors in Hamiltonian and band structure predictions. In such instances, there exists a potential hazard of yielding erroneous physical interpretations, such as the formation of unphysical branches. This phenomenon carries notable implications, not only for the accuracy of DFT predictions but also for their application in transport simulations. In this work, we propose a novel framework for training a neural network with a limited number of training data (DFT simulation results). To achieve greater precision with a smaller training dataset, we introduce a method that incorporates unlabeled training data into the learning process. Our approach entails the generation of pseudo Hamiltonians for various atomic structures, employing them as targets for unlabeled data during training. This method offers the prospect of significantly reducing the expenses associated with obtaining training data.

2 Preliminaries

Neural Network Hamiltonian. In materials science, graph neural networks (GNNs), especially the Message-passing neural network (MPNN) variant, model molecular properties by graphically representing atomic connections and facilitating atomic interaction through node-based information aggregation. Exploiting MPNNs' benefits, they have been utilized in neural network potentials for MD simulations [4, 3], predicting electron charge density in DFT [10], and neural network Hamiltonians (NNH) in DFT [14]. The NNH provides an efficient alternative to traditional DFT Hamiltonians, sidestepping the Kohn-Sham equations, which facilitates more streamlined evaluations in areas like non-equilibrium green's function method [20] for semiconductor devices and comprehensive atomic structure analysis.

Semi-supervised Learning. Deep neural networks often excel through supervised learning, relying on labeled datasets. However, the benefits of using larger datasets come with significant costs due to the human effort required for labeling. Semi-supervised learning (SSL) addresses this challenge, reducing reliance on labeled data by leveraging unlabeled data. Obtaining unlabeled data is typically less labor-intensive, making SSL cost-effective. This has led to various SSL techniques tailored for deep neural networks [19, 1, 9]. Among these techniques, we utilize the Pseudo-label method [13], which incorporates pseudo labels for unlabeled data during training. Inspired by these previous efforts, we integrate SSL into neural network training by generating pseudo Hamiltonians to leverage unlabeled data (input data for which we have not performed DFT simulations). There have been attempts to apply SSL in the field of molecular dynamics [15], but as far as we know, there are no cases of applying SSL to neural network training that predicts the DFT Hamiltonian that includes information about various physical properties. Therefore, this new approach provides a novel viewpoint in the field of DFT Hamiltonian neural network training.

Algorithm 1 Overall process of our framework

Input: Labeled data x and their target Hamiltonians H, unlabeled data u, model F, initial step for semi-supervised learning I, pseudo Hamiltonian generation step $s \in (I, \ldots, S)$, trainging weight α, number of epoch $e \in (1, \ldots, E)$ and mean squared error loss l.

```
for e ← 1 to E do
    h = F(x_e)                                  ▷ predict Hamiltonian for labeled data
    if e = s then
        F(u_e) → Ĥ                              ▷ pseudo Hamiltonian generation for unlabeled data
        continue
    end if
    if e ≧ I then
        h' = F(u_e)                             ▷ predict Hamiltonian for unlabeled data
    end if
end for
L_S = l(H, h)                                   ▷ calculate supervised loss
L_U = l(Ĥ, h')                                  ▷ calculate unsupervised loss
return L_S + αL_U                               ▷ calculate the total loss L from x and u
```

3 Methods

Data Preparation. The data preparation involves three steps: atomic structure generation, DFT Hamiltonian computation, and Hamiltonian transformations. The first two steps utilize the Vienna ab initio simulation package [12] and OpenMX software package [16], incorporating PBE exchange-correlation energy functional [17] and norm-conserving pseudopotentials [21]. Using the Wigner D-matrix, Hamiltonian matrices are transformed covariantly [14]. Datasets were secured for four materials: MoS2 (sourced from [5]; 500 configurations of 75 atoms each after 300K relaxation; s3p2d2 orbital basis for Mo and s2p2d1 for S), Bi2Te3 (acquired through [5]; 256 configurations of 90 atoms each; s3p2d2 orbital basis for Bi and Te), HfO2 (monoclinic structure with 96 atoms; 500 configurations from 1500K to 300K; s2p2d1 orbital basis for Hf and s2p2 for O), and InGaAs (initial trigonal structure of 108 atoms; 420 configurations from 600K to 300K; s2p2d2 orbital basis for In, Ga, and As).

Message-passing Neural Network. In the MPNN, vertices v_i denote atoms and edges e_{ij} represent atom pairs. The initial values for these vertices and edges are derived from the embeddings of the atomic number A_i and distances $|r_{ij}|$, respectively. This initialization is defined in Eq. (2), where c_n and δ act as parameters governing the Gaussian basis [18]. Following this initialization, the vertices and edges are updated: the tensor product of vertices v_i^t, v_j^t (t is the number of update), and edges e_{ij}^t, coupled with spherical harmonics $S(\hat{r}_{ij})$, generates a message m_i^{t+1} through message function M_t. This message, encapsulating neighboring atom information, is then integrated with existing vertex or edge data via the update function U_t, as described in Eq. (1).

$$v_i^{t+1} = U_t(v_i^t, m_i^{t+1}), \quad m_i^{t+1} = \sum_j M_t(v_i^t || v_j^t || e_{ij}^t, S(\hat{r}_{ij})) \tag{1}$$

$$v_i^0 = Node\ Embedding(A_i), \quad e_{ij}^0 = exp(-\frac{(|r_{ij}| - c_n)^2}{\delta^2}) \tag{2}$$

Following multiple updates through multilayer layers, Hamiltonian elements are computed using the Wigner-Eckart layer and symmetry-aligned features are managed with an Equivariant Neural Network (ENN), transforming them into vector components. Our framework has been applied to the state-of-the-art model, DeepH-E3 [5], which incorporates the methodologies previously discussed. The neural network architecture used for learning is as shown in Figure 1. We used SILU or Sigmoid as the activation functions, Adam as the optimizer, starting with a learning rate of 0.005.

Semi-supervised Learning. As illustrated in Algorithm 1, we utilize both labeled data x and unlabeled data u to train the neural network. Before the initial step I, only the supervised loss $\mathcal{L}_\mathcal{S}$ is applied. For each pseudo Hamiltonian generation step s, we create pseudo Hamiltonians to use as targets for the unlabeled data. As the learning process advances, we re-generate increasingly accurate

	Hamiltonian error ($\times 10^{-5}$ eV^2)						Band structure error ($\times 10^{-2}$ eV)					
# of labeled data	10		20		30		10		20		30	
Method	baseline	**ours**	baseline	**ours**	baseline	**ours**	baseline	**ours**	baseline	**ours**	baseline	**ours**
MoS_2	14.2	**3.94**	5.40	**2.36**	2.97	**1.64**	2.52	**1.39**	1.89	**1.19**	1.49	**1.20**
Bi_2Te_3	9.43	**3.20**	3.99	**1.97**	2.54	**1.48**	4.21	**2.00**	2.17	**1.61**	1.68	**1.27**
HfO_2	9.80	**4.79**	3.99	**2.91**	2.54	**2.36**	8.06	**5.79**	3.97	**3.80**	2.03	**2.02**
InGaAs	11.0	**6.53**	6.49	**5.25**	5.50	**4.99**	4.59	**2.43**	**1.04**	1.29	1.62	**1.49**

Table 1: Average error of the (left) Hamiltonian and (right) band structure. We compared our method with DeepH-E3 [5]. The best results are indicated in bold for each case. We used unlabeled data equal to 10 times the number of labeled data for neural network training.

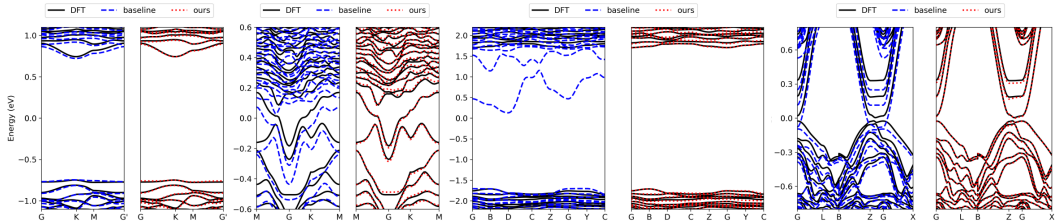

Figure 2: Band structures of MoS_2, Bi_2Te_3, HfO_2, and InGaAs (from left to right) calculated by the baseline (blue dashed lines), our method (red doted lines), and DFT results (balck solid lines).

pseudo Hamiltonians. After the initial step I, we introduce unsupervised loss $\mathcal{L}_\mathcal{U}$ by computing the loss between the pseudo Hamiltonian $\hat{H}$ and the predicted Hamiltonian h' for unlabeled data. However, pseudo Hamiltonians generated in the early epochs, which have not been sufficiently learned yet, contain a significant amount of uncertainty. Therefore, we set α relatively low in the early epochs and increase it as learning progresses. We use the total loss as follows:

$$\mathcal{L} = \mathcal{L}_\mathcal{S} + \alpha \mathcal{L}_\mathcal{U}. \tag{3}$$

4 Results

For each of the systems, we assessed the Hamiltonian and band structure errors by comparing our method to the state-of-the-art method, DeepH-E3 [5]. Table 1 demonstrates that our framework outperforms the baseline method on various material datasets, particularly when a smaller number of labeled data is used for training. Empirically, a Hamiltonian error close to 3×10^{-5} is deemed accurate. However, as depicted in Figure 2, there are instances where further error reduction is necessary, even within this range. Due to varying standards for accurate Hamiltonians across datasets, our emphasis was on minimizing Hamiltonian error by leveraging unlabeled data alongside labeled data. As depicted in Figure 2, when limited labeled data is employed, the baseline method exhibit inaccurate band structures or even generate unphysical branches, potentially leading to erroneous physical interpretations. In contrast, our framework, which effectively incorporates unlabeled data in neural network training, addresses this issue. Consequently, our framework can be employed in scenarios where neural networks must be trained with limited data.

We conducted ablation studies, varying the hyperparameters used for neural network training. Figure 3.(a) shows that the Hamiltonian error decreases as more unlabeled data are used. This supports our argument that unlabeled data should be used to generate more accurate Hamiltonians. Figure 3.(b) shows that the Hamiltonian error decreases as the initial step I that generates the pseudo Hamiltonian for the first time becomes smaller and the pseudo Hamiltonian is updated more frequently. We also verified the effect of the training weight α, which determines how much the influence of unlabeled data is reflected in training, in Figure 3.(c). The results indicate that in earlier epochs, when the neural network has not been sufficiently trained, it is more likely to generate a less accurate pseudo Hamiltonian. Therefore, it is best to start with a small α and gradually increase it. We verified that gradually increasing α from 0.1 to 0.4 yields better performance compared to a constant α on Bi_2Te_3 dataset.

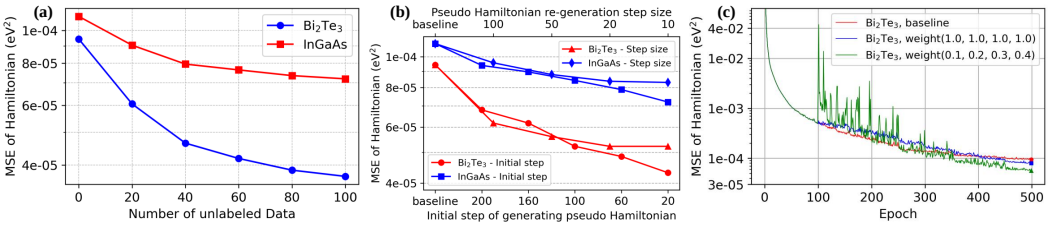

Figure 3: Hamiltonian error measured while varying (a) the number of unlabeled data used in neural network training (b) the initial pseudo Hamiltonian generation step I and the re-generation step size (c) the training weight α.

5 Discussion

Current neural network training approaches to replace DFT calculations are based on the assumption that a substantial number of DFT simulation results must be acquired for learning. However, when dealing with a limited amount of training data, there exists a risk of obtaining distorted results in the subsequent physical analysis. To address this challenge, we introduce a framework that mitigates the limitations arising from insufficient training data. We achieve this by incorporating semi-supervised learning techniques into neural network training. These methods have already demonstrated significant advancements in computer vision tasks, and our framework has also exhibited enhanced performance across various datasets.

6 Broader Impact

The exploration of neural networks for predicting physics-based simulations or experimental results remains a subject of ongoing research. However, there are scenarios where access to extensive simulations or conducting experiments can be limited due to various constraints, including resource limitations or high costs. In such situations, it becomes imperative to achieve reliable and meaningful results using a limited amount of training data. Our framework offers a versatile solution that can be applied effectively to a wide range of examples in these circumstances.

References

[1] David Berthelot, Nicholas Carlini, Ian Goodfellow, Nicolas Papernot, Avital Oliver, and Colin A Raffel. Mixmatch: A holistic approach to semi-supervised learning. *Advances in neural information processing systems*, 32, 2019.

[2] Anand Chandrasekaran, Deepak Kamal, Rohit Batra, Chiho Kim, Lihua Chen, and Rampi Ramprasad. Solving the electronic structure problem with machine learning. *npj Computational Materials*, 5(1):22, 2019.

[3] Johannes Gasteiger, Florian Becker, and Stephan Günnemann. Gemnet: Universal directional graph neural networks for molecules. *NeurIPS*, 2021.

[4] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *ICLR*, 2020.

[5] Xiaoxun Gong, He Li, Nianlong Zou, Runzhang Xu, Wenhui Duan, and Yong Xu. General framework for e (3)-equivariant neural network representation of density functional theory hamiltonian. *Nature Communications*, 14(1):2848, 2023.

[6] Andrea Grisafi, Alberto Fabrizio, Benjamin Meyer, David M Wilkins, Clemence Corminboeuf, and Michele Ceriotti. Transferable machine-learning model of the electron density. *ACS central science*, 5(1):57–64, 2018.

[7] Andrea Grisafi, David M Wilkins, Gábor Csányi, and Michele Ceriotti. Symmetry-adapted machine learning for tensorial properties of atomistic systems. *Physical review letters*, 120(3):036002, 2018.

[8] Qiangqiang Gu, Linfeng Zhang, and Ji Feng. Neural network representation of electronic structure from ab initio molecular dynamics. *Science Bulletin*, 67(1):29–37, 2022.

[9] Gyeongdo Ham, Yucheol Cho, Jae-Hyeok Lee, and Daeshik Kim. P-pseudolabel: Enhanced pseudo-labeling framework with network pruning in semi-supervised learning. *IEEE Access*, 10:115652–115662, 2022.

[10] Peter Bjørn Jørgensen and Arghya Bhowmik. Deepdft: Neural message passing network for accurate charge density prediction. *NeurIPS*, 2020.

[11] James Kirkpatrick, Brendan McMorrow, David HP Turban, Alexander L Gaunt, James S Spencer, Alexander GDG Matthews, Annette Obika, Louis Thiry, Meire Fortunato, David Pfau, et al. Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, 374(6573):1385–1389, 2021.

[12] G. Kresse and J. Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B*, 54:11169–11186, Oct 1996.

[13] Dong-Hyun Lee et al. Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In *Workshop on challenges in representation learning, ICML*, volume 3, page 896. Atlanta, 2013.

[14] He Li, Zun Wang, Nianlong Zou, Meng Ye, Runzhang Xu, Xiaoxun Gong, Wenhui Duan, and Yong Xu. Deep-learning density functional theory hamiltonian for efficient ab initio electronic-structure calculation. *Nature Computational Science*, 2(6):367–377, 2022.

[15] Hai Nguyen, Shin-ichi Maeda, and Kenta Oono. Semi-supervised learning of hierarchical representations of molecules using neural message passing. *arXiv preprint arXiv:1711.10168*, 2017.

[16] T. Ozaki. Variationally optimized atomic orbitals for large-scale electronic structures. *Phys. Rev. B*, 67:155108, Apr 2003.

[17] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[18] Kristof T Schütt, Huziel E Sauceda, P-J Kindermans, Alexandre Tkatchenko, and K-R Müller. Schnet–a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24), 2018.

[19] Kihyuk Sohn, David Berthelot, Nicholas Carlini, Zizhao Zhang, Han Zhang, Colin A Raffel, Ekin Dogus Cubuk, Alexey Kurakin, and Chun-Liang Li. Fixmatch: Simplifying semi-supervised learning with consistency and confidence. *Advances in neural information processing systems*, 33:596–608, 2020.

[20] Jeremy Taylor, Hong Guo, and Jian Wang. Ab initio modeling of quantum transport properties of molecular electronic devices. *Phys. Rev. B*, 63:245407, Jun 2001.

[21] N. Troullier and José Luís Martins. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B*, 43:1993–2006, Jan 1991.

[22] Masashi Tsubaki and Teruyasu Mizoguchi. Quantum deep field: data-driven wave function, electron density generation, and atomization energy prediction and extrapolation with machine learning. *Physical Review Letters*, 125(20):206401, 2020.

[23] Oliver Unke, Mihail Bogojeski, Michael Gastegger, Mario Geiger, Tess Smidt, and Klaus-Robert Müller. Se (3)-equivariant prediction of molecular wavefunctions and electronic densities. *Advances in Neural Information Processing Systems*, 34:14434–14447, 2021.

[24] 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*, 12(1):7273, 2021.

[25] Tetiana Zubatiuk and Olexandr Isayev. Development of multimodal machine learning potentials: Toward a physics-aware artificial intelligence. *Accounts of Chemical Research*, 54(7):1575–1585, 2021.