

Universal Ensemble-Embedding Graph Neural Network for Direct Prediction of Optical Spectra from Crystal Structures

Nguyen Tuan Hung,* Ryotaro Okabe, Abhijatmedhi Chotrattanapituk, and Mingda Li*

Optical properties in solids, such as refractive index and absorption, hold vast applications ranging from solar panels to sensors, photodetectors, and transparent displays. However, first-principles computation of optical properties from crystal structures is a complex task due to the high convergence criteria and computational cost. Recent progress in machine learning shows promise in predicting material properties, yet predicting optical properties from crystal structures remains challenging due to the lack of efficient atomic embeddings. Here, Graph Neural Network for Optical spectra prediction (GNNOpt) is introduced, an equivariant graph-neural-network architecture featuring universal embedding with automatic optimization. This enables high-quality optical predictions with a dataset of only 944 materials. GNNOpt predicts all optical properties based on the Kramers-Krönig relations, including absorption coefficient, complex dielectric function, complex refractive index, and reflectance. The trained model is applied to screen photovoltaic materials based on spectroscopic limited maximum efficiency and search for quantum materials based on quantum weight. First-principles calculations validate the efficacy of the GNNOpt model, demonstrating excellent agreement in predicting the optical spectra of unseen materials. The discovery of new quantum materials with high predicted quantum weight, such as SiOs, which host exotic quasiparticles with multifold nontrivial topology, demonstrates the potential of GNNOpt in predicting optical properties across a broad range of materials and applications.

circuits (PICs). These devices play an indispensable role in the current resurgence of the semiconductor industry.^[1–4] The linear optical responses, in particular, offer insights into fundamental parameters such as energy bandgaps, transparency, reflectivity, and refractive index, which are essential for controlling the light-matter interactions.^[5,6] Tremendous scientific and industrial interests have driven both experimental and computational efforts toward high-throughput screening of candidate materials for tailored optical applications. Current experimental techniques, including ellipsometry, UV–vis spectroscopy, and Fourier transform infrared spectroscopy (FTIS), are commonly used to obtain materials' optical spectra. However, they are limited to specific wavelength ranges and often require stringent sample conditions, making them not ideal for high-throughput material screening.^[7–10] On the other hand, first-principles calculations based on density functional theory (DFT) can compute optical spectra across all wavelength ranges.^[11] However, DFT requires dense k-point sampling for the convergence of complex dielectric functions or absorption coefficients. For instance, in graphite, the transition of excited electrons by light

1. Introduction

Understanding the optical properties of materials is crucial for designing and optimizing optoelectronic devices such as LEDs, solar cells, photodetectors, and photonic integrated

occurs near the Dirac point, necessitating a k-point sampling of over 100,000 k-points to accurately capture the optical transitions, resulting in a time-consuming calculation.^[12] Consequently, high-throughput DFT calculations for optical spectra are

N. T. Hung
Frontier Research Institute for Interdisciplinary Sciences
Tohoku University
Sendai 980-8578, Japan
E-mail: nguyen.tuan.hung.e4@tohoku.ac.jp
N. T. Hung, R. Okabe, A. Chotrattanapituk, M. Li
Quantum Measurement Group
MIT
Cambridge MA 02139-4307, USA

R. Okabe
Department of Chemistry
MIT
Cambridge MA 02139-4307, USA
A. Chotrattanapituk
Department of Electrical Engineering and Computer Science
MIT
Cambridge MA 02139-4307, USA
M. Li
Department of Nuclear Science and Engineering
MIT
Cambridge MA 02139-4307, USA
E-mail: mingda@mit.edu

© 2024 The Author(s). Advanced Materials published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/adma.202409175

largely limited to materials with a small number of atoms per unit cell.^[13]

Machine-learning methods are increasingly being adopted in materials research to accelerate materials discovery through high-throughput property prediction.^[14–16] One successful approach is the graph neural networks (GNNs), which enables a direct prediction of material properties from crystal structures.^[17–19] Chen et al.^[19] build a GNN model using an equivariant neural network with E(3)NN,^[20] which predicts the phonon density-of-states (DOS) by using only atomic species and positions as input parameters. The success of machine learning models in structure-property prediction suggests a potential for applying machine learning to optical spectra prediction. However, challenges remain in developing an effective predictive model, which should accurately predict optical properties from a small available database, and, moreover, extract useful information from complex relationships in optical spectra, such as the Kramers–Krönig (K-K) relations and the f -sum rule.^[21] Such a model could be instrumental in searching for photovoltaic materials for energy conversion or in understanding fundamental physics through the optical spectra of materials.

In this work, we develop GNNOpt, a GNN model that establishes a direct relationship between crystal structures and frequency-dependent optical properties. Rather than focusing on building a more complex neural network structure, we emphasize automatic embedding optimization, which could serve as a universal input layer in equivariant neural networks. This approach surpasses the commonly used fixed embedding schemes and goes beyond the feature selection and importance method by integrating different features through additional neural networks. This saves the effort of manually picking the embedding and, meanwhile, improves the prediction accuracy. As a result, it enables high-quality optical predictions with a small dataset of 944 materials.^[13,22] In a GNN model, converting crystal structure into machine-readable graph representations is essential. While the distance vector between an atom and its neighbor is a mandatory fixed embedding for applying the equivariant neural networks in E(3)NN,^[20] particularly as input parameters of the spherical harmonics in a tensor product, the feature embedding, or representation of an atom, serve as initial parameters and often relies on human intuition to select the best descriptors in a GNN. For phonon DOS predictions, atomic mass is a natural embedding.^[19,23] However, atomic mass is less relevant for optical properties, and identifying suitable atomic embeddings can be challenging since there are various physical, chemical, structural, and environmental atomic descriptors available. To address this challenge, we propose a universal ensemble embedding layer for the GNN model. The ensemble embedding layer assigns a learnable weight to each feature embedding. As a result, the GNN model can automatically identify the most important descriptors for specific physical properties from several selected descriptors. Combining the universal ensemble embedding and the equivariant neural networks, our GNN model can directly and accurately predict optical spectra using only crystal structure as input, which shows superior performance than any fixed embeddings. Additionally, thanks to the K-K relation, we were able to extract all frequency-dependent optical spectra, such as the absorption coefficient, the refractive index, and reflectivity, from only the real or imaginary part of the dielectric function. By pre-

dicting the optical spectra in 5,281 unseen crystal structures, we identify a list of high-performing materials for solar cell application, supported by additional first-principles calculations. Lastly, the f -sum rule is another universal constraint in the linear optical response, which determines the quantum characteristics of a material by integrating over an infinite spectral range. A recent theory shows that quantum weight, a parameter based on the variation of f -sum rule,^[24] is directly connected to the ground state quantum geometry and topology. Using quantum weight, our GNNOpt model successfully identifies several quantum materials with high quantum weight, such as SiOs, validated by first-principles calculations.

2. GNNOpt: A Machine Learning Model for Optical Spectra Prediction

The input and output of the GNNOpt are shown in Figure 1a, in which the crystal structure is the only input parameter for the GNNOpt, and the output is the optical spectra including complex dielectric function ($\epsilon_1(\omega)$ is the real part and $\epsilon_2(\omega)$ is the imaginary part), absorption coefficient ($\alpha(\omega)$), complex refractive index ($n(\omega)$ is the real part and $k(\omega)$ is the imaginary part), and reflectance ($R(\omega)$), where ω is the angular frequency of light and the photon energy is given by $E = \hbar\omega$, where \hbar is the reduced Planck constant. All optical spectra are related by the Kramers–Krönig relations, which will be discussed shortly. In Figure 1b, we show the input features of each atomic species (or each node), in which we adopt three features, including the atomic mass (x_0), dipole polarizability (x_1), and the effective covalent radius (x_2). All input features use a one-hot encoding to denote the atomic species. For example, the oxygen atom is encoded as $[0, \dots, x_0^O, \dots, 0]$ for the feature 0 (i.e., atomic mass) with a 118-element-long array of zeroes except for an element with an index = 8, which equals to the atomic number of the oxygen minus one. Similarly, we have $[0, \dots, x_1^O, \dots, 0]$ and $[0, \dots, x_2^O, \dots, 0]$ for the feature 1 (dipole polarizability) and the feature 2 (covalent radius), respectively. Thus, the one-hot representation of the oxygen atom with three features is $\mathbf{x}_{\text{in}}^O = ([0, \dots, x_0^O, \dots, 0]; [0, \dots, x_1^O, \dots, 0]; [0, \dots, x_2^O, \dots, 0])$. After that, all input features of each atom are automatically optimized for their embedding mixture by an ensemble embedding layer, as shown in Figure 1c. Then, the embedded features are passed through a sequence of equivariant graph convolution and gated nonlinear layers parameterized with the input parameter \mathbf{r}_{ij} . Here, \mathbf{r}_{ij} is a distance vector between i -th atom and neighbor j -th atoms (see Figure 1a) up to a radial cut-off value $r_{\text{cut}} = 6$ Å, which is selected based on the optimized parameters testing (see Figure S4, Supporting Information). Here, the periodic boundary condition is considered when constructing a graph. To achieve the equivariance, the convolutional filters are designed to be composed of learnable radial functions $R(|\mathbf{r}_{ij}|)$ and spherical harmonics $Y_{\ell}^m(\mathbf{r}_{ij}/|\mathbf{r}_{ij}|)$, where the indices ℓ and m indicate the degree and order of the function. Therefore, the geometric information and all crystallographic symmetries of input crystal structures are preserved in GNNOpt. After the final convolution layer, all resulting features are summed and passed through a processing layer, including activation (ReLU), to predict the optical spectra. The weights of the GNNOpt are optimized by minimizing the mean squared error (MSE) loss function between the predicted and ground truth spectra. Figure 1d shows the graph

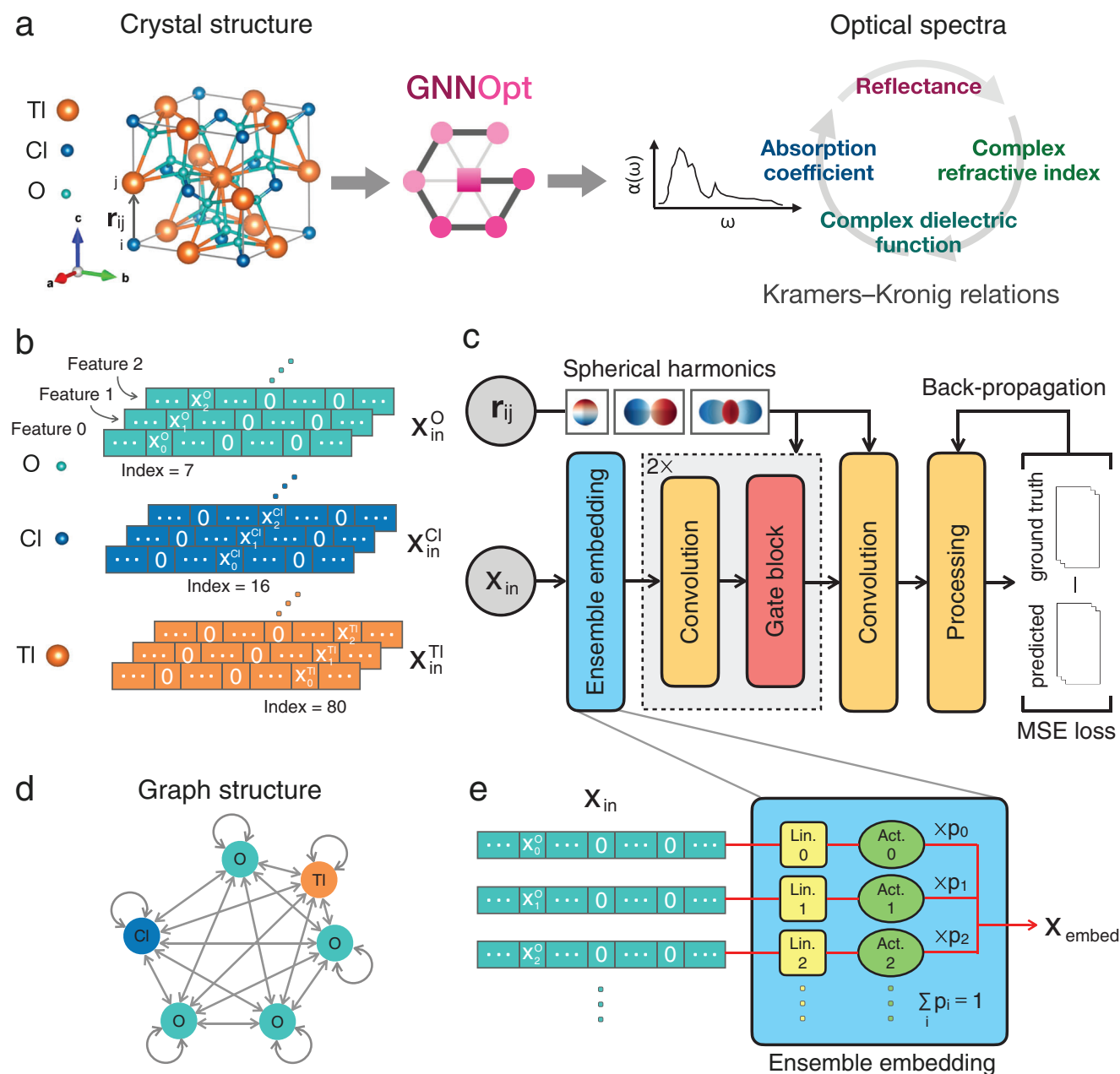


Figure 1. The GNNOpt for optical spectra prediction. a) The crystal structure is the only input parameter needed for the GNNOpt to predict the optical spectra, taking TiClO_4 as an example. r_{ij} is a distance vector between i -th and j -th atoms, and a number of frequency-dependent optical properties can be predicted. b) Input features of each atomic species (O, Cl, and Ti) in feature-weighted one-hot representation, which is a 118-element-long array of zeroes except an element with an index equal to the atomic number where the value is equal to the feature. c) Overview of the GNNOpt architecture. All atomic input features are automatically optimized by an ensemble embedding layer. Then, the embedded features are passed through a sequence of equivariant graph convolution and gated nonlinear layers parameterized with r_{ij} 's radial and spherical harmonic representations. After that, the result is passed to a post-process layer, including activation and aggregation, to generate the predicted output spectra. Finally, the network weights are trained by minimizing the mean-square error (MSE) loss function between the predicted and ground-truth spectra. d) Graph representation of periodic lattice of TiClO_4 , where the graph nodes represent the atoms inside a unit cell, and edges represent message passing directions of graph convolutions. e) In an ensemble embedding layer, each atomic feature is independently embedded with its linear layer (Lin. 0, Lin. 1,...) and activation layers (Act. 0, Act. 1,...). After that, all embedded features are weighted and averaged by learnable weight probability p_i . The same ensemble embedding layer's parameter values are used across all elements, allowing interpretability of the model feature importance.

representation of a unit-cell TiClO_4 , in which nodes represent atoms of a unit cell, and edges represent the message passing direction of graph convolutions layers. In Figure 1e, we show the detail of the universal ensemble embedding layer, which is the key ingredient toward performance improvement even without any neural network model change. For each atom, each feature is independently embedded with its linear and activation layers. After that, all embedded features are evaluated by a weighted average by learnable mixing probability p_i , in which p_i are normalized by $\sum_i p_i = 1$ (see Experimental Section for the full hyperparameters of the GNNOpt).

Next, we apply GNNOpt to predict the optical spectra given the limited training data with 944 materials, which are calculated by DFT within the independent-particle approximation (IPA).^[13] The database is obtained from the Materials Project.^[22] The details for the data preparation and statistics are given in Section 7 and Figure S1 (Supporting Information). We note that the linear optical properties are not independent but follow the K-K relations:^[21,25]

$$\epsilon_1(\omega) = 1 + \frac{2}{\pi} P \int_0^\infty \frac{\epsilon_2(\omega')\omega'}{\omega'^2 - \omega^2} d(\omega') \quad (1)$$

$$\epsilon_2(\omega) = -\frac{2\omega}{\pi} P \int_0^\infty \frac{\epsilon_2(\omega')\omega'}{\omega'^2 - \omega^2} d(\omega') \quad (2)$$

$$\alpha(\omega) = \frac{\sqrt{2}\omega}{c} \sqrt{\sqrt{\epsilon_1^2(\omega) + \epsilon_2^2(\omega)} - \epsilon_1(\omega)} \quad (3)$$

$$n(\omega) = \sqrt{\frac{1}{2} \left(\sqrt{\epsilon_1^2(\omega) + \epsilon_2^2(\omega)} + \epsilon_1(\omega) \right)} \quad (4)$$

$$k(\omega) = \sqrt{\frac{1}{2} \left(\sqrt{\epsilon_1^2(\omega) + \epsilon_2^2(\omega)} - \epsilon_1(\omega) \right)} \quad (5)$$

$$R(\omega) = \frac{(n(\omega) - 1)^2 + k^2(\omega)}{(n(\omega) + 1)^2 + k^2(\omega)} \quad (6)$$

where P in Equations (1) and (2) denotes the Cauchy principal value. The database, which originally only included ϵ_1 , ϵ_2 , and α , is expanded for n , k , and R by using Equations (4), (5), and (6), respectively. It is noted that only the averaged optical values, i.e., $\alpha = (\alpha_{xx} + \alpha_{yy} + \alpha_{zz})/3$, are available in the database.^[13]

Equations (1)–(6) are considered for the 3D materials. For a 2D system, when the thickness L is smaller than the Fermi wavelength λ_F , the quantum size effect will play an important role in the optical properties.^[26,27] For common semiconductors like Si, GaAs, InAs, and InP, λ_F typically ranges from about 10 to 40 nm, depending on the doping level.^[27] Thus, the optical prediction in the present study is limited to the 2D thin film case with $L \gg \lambda_F$.

To compare the predicted spectra with the DFT ground-truth spectra, we calculate the weighted mean, \overline{W} , of the optical spectra, $W(\omega)$, by

$$\overline{W} = \frac{\int d\omega W(\omega)\omega}{\int d\omega W(\omega)} \quad (7)$$

where $W(\omega)$ denotes the optical spectra, such as $\epsilon_1(\omega)$, $\epsilon_2(\omega)$, $\alpha(\omega)$, $n(\omega)$, $k(\omega)$, or $R(\omega)$. The corrections between the GNNOpt-predicted, \overline{W} , and DFT ground-truth spectra, \overline{W}^* are plotted in Figure 2a–f. Here, the coefficients of determination, R^2 , of the test set are 0.72, 0.74, 0.93, 0.51, 0.86, and 0.55 for ϵ_1 , ϵ_2 , α , n , k , and R , respectively, which show excellent agreement for the cases of α and k . On the other hand, the relative error $|\overline{W} - \overline{W}^*|/\overline{W}^*$ below 10% for the test set (see the inset figures) of ϵ_1 , ϵ_2 , α , n , k , and R are 79%, 61%, 79%, 95%, 82%, and 62%, respectively, which shows the high performance of our model to predict ϵ_1 , α , n , and k . To visualize the model performance, in Figure 2g, we also plot $\alpha(\omega)$ of the 24 randomly selected materials from the test set in each mean-squared-error (MSE) quartile, in which the color lines are the GNNOpt-predicted spectra and the black lines are the DFT ground-truth spectra. In Figure 2g, the 1st to 4th rows in the right figure correspond to the 1st to 4th error quartiles in the left figure by the same color. The overlap between the predicted $\alpha(\omega)$ (color lines) and the DFT calculation (black line) in Figure 2g suggests that our model can accurately predict optical spectra, in which all spectrum peaks are reproduced from the DFT calculation even for the 4th row. In the 3rd and 4th rows, some noise can be found in the predicted $\alpha(\omega)$ of NaTi_2Bi or TiNiF_3 . This is because of the origin of the Dirac delta function in the dielectric function formula, which is often solved by adopting a broadening parameter or increasing the number of k-point samples in the DFT calculation.^[11] For the GNNOpt model, a uniform filter of the SciPy library^[28] can be applied instead of the broadening parameter in the DFT to reduce the noise of the spectrum. The MSE loss and the compared full spectra of the training, validation, and test datasets for ϵ_1 , ϵ_2 , α , n , k , and R are given in Figures S5–S11 (Supporting Information), respectively.

In Table 1, we compare the R^2 values of the test set with the ensemble embedding layer and fixed embeddings. In this present study, we use three feature embeddings: atomic mass, dipole polarizability, and covalent radius, since these features are provided for all elements in the periodic table. Nevertheless, the ensemble embedding layer can work with any other added features. Table 1 shows that the GNNOpt with the ensemble embedding layer is comparable with the best-trained fixed embedding model. We note that R^2 of the ϵ_2 prediction is more sensitive to the train-test split sets than that of the α prediction because of the narrow peaks in the ϵ_2 spectra compared with the α spectra (see Figures S7–S8, Supporting Information). The value of p_i of the ensemble embedding layer can also tell us which feature is the main contribution. For the GNNOpt, the mixing probability shows that the covalent radius (feature 2) is almost completely dominant with $p_2 \approx 1$. We note that the covalent radius is a measure of the size of an atom that forms part of one covalent bond.^[29] On the other hand, the dipole moment, which is related to the optical properties, is defined as the product of the total amount of charge and the covalent bond length. This relationship might explain why the covalent radius is the most dominant feature in the GNNOpt.

To evaluate the scalability of the GNNOpt model for unseen materials, we perform the GNNOpt with the test set containing larger atomic site numbers N_{test} than the training and validation sets. There are three cases of the test set as $N_{\text{test}} = 8$, 9 (case I), $N_{\text{test}} = 7$, 8, 9 (case II), and $N_{\text{test}} = 6$, 7, 8, 9 (case III). As shown

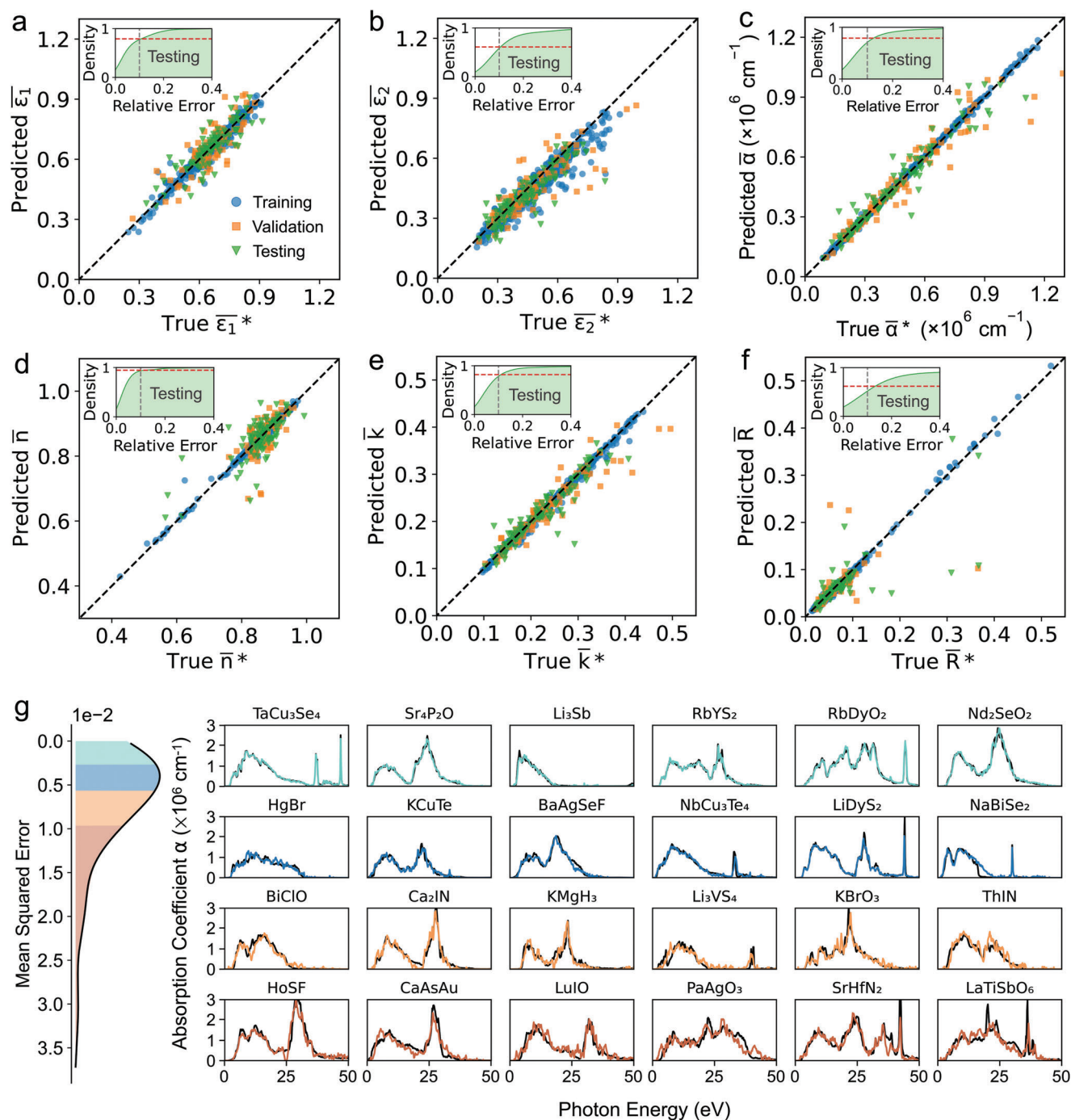


Figure 2. Performance of the GNNOpt in predicting different optical properties. a–f) Weighted average magnitudes of a few optical properties predicted from the GNNOpt model, \bar{W} , compared with the ground-truth (true), \bar{W}^* , with the $\bar{W} = \bar{\epsilon}_1, \bar{\epsilon}_2, \bar{\alpha}, \bar{n}, \bar{k}$, and \bar{R} , for complex dielectric function, absorption coefficient, complex refractive index, and reflectance, respectively. The data points of circle (blue), square (orange), and triangle (green) denote the training, validation, and test datasets, respectively. The inset shows the cumulative kernel-density-estimator (KDE) plot of the relative error, $|\bar{W} - \bar{W}^*|/\bar{W}^*$, of the only test dataset. g) Mean squared error (MSE) distribution and 24 randomly selected materials in the test dataset corresponding to each error quartile in the MSE distribution for absorption coefficient $\alpha(\omega)$. Color lines are the GNNOpt-predicted spectra, and black lines are the DFT ground-truth spectra.

Table 1. The coefficient of determination, R^2 , of the test set by the models with fixed and ensemble feature embeddings. The atomic descriptors (atomic mass, dipole polarizability, and covalent radius) are individually embedded in different fixed feature embedding models, while, in ensemble feature embedding, all atomic descriptors' feature embeddings are combined with trainable probability weights (p_0 , p_1 , and p_2 , respectively.) The table shows trained probability weights that give R^2 of the ensemble feature embedding model.

Optic	Fixed embedding			Ensemble embedding	
	Atomic mass	Dipole polarizability	Covalent radius	Ensemble	Probability (p_0 , p_1 , p_2)
ϵ_1	0.69	0.67	0.72	0.72	0.0021, 0.0017, 0.9962
ϵ_2	0.62	0.63	0.73	0.74	0.0026, 0.0002, 0.9972
α	0.93	0.93	0.91	0.93	0.0001, 0.0008, 0.9991
n	0.48	0.51	0.48	0.51	0.0011, 0.0013, 0.9976
k	0.85	0.86	0.85	0.86	0.0001, 0.0002, 0.9997
R	0.61	0.48	0.57	0.55	0.0033, 0.0018, 0.9949

in Figure S12a (Supporting Information), R^2 for the test set are 0.83, 0.84, and 0.86 for the cases I, II, and III, respectively. Moreover, the training set contains only 443 materials with $2 \leq N_{\text{train}} \leq 5$ in case III. However, the prediction optical spectra show a good overlap with the DFT ground truth for the test set with $6 \leq N_{\text{test}} \leq 9$ (see Figure S11b, Supporting Information). Thus, it suggests that the GNNOpt can accurately predict the optical properties of unseen complex materials with larger systems than the training set.

3. The Kramers-Krönig Relations

To evaluate the role that the Kramers-Krönig relations play in predicted optical properties, first, we use the complex dielectric function $\tilde{\epsilon}_1$ and $\tilde{\epsilon}_2$, which are predicted by the GNNOpt, to calculate other optical properties by using Equations (1)–(6). It is noted that $\tilde{\epsilon}$ denotes the predicted dielectric function while $\bar{\epsilon}$ denotes the weighted mean of the dielectric function (see Equation (7)). Then, we calculate the R^2 coefficients for the test set for each of the optical properties. Finally, we compare these R^2 values from the K-K relations with R^2 of the corresponding optical properties from GNNOpt, as shown in Figure 3. It notes that we use the same training, validation, and test sets for all cases. The highest R^2 value in the 1st row suggests that directly predicting the optical properties is better than calculating the optical properties from the predicted $\tilde{\epsilon}_1$ (or $\tilde{\epsilon}_2$). In other words, the K-K relations must be applied before the GNNOpt model to predict optical spectra, i.e., $\epsilon_1(\epsilon_2) \rightarrow \text{K-K relations} \rightarrow W \rightarrow \text{GNNOpt} \rightarrow \text{predicted-}W$. In the 2nd and 3rd rows of Figure 3, R^2 of $\alpha/\tilde{\epsilon}_2$ (0.71) and $k/\tilde{\epsilon}_2$ (0.74) are larger than that of $\alpha/\tilde{\epsilon}_1$ (0.14) and $k/\tilde{\epsilon}_1$ (0.41). This implies that α and k are more dependent on ϵ_2 than ϵ_1 , which is clearly explained by Equations (3) and (4), $\alpha(\omega) = 2\omega k(\omega)/c \propto \sqrt{|\epsilon_2|}$. In contrast, R^2 of $n/\tilde{\epsilon}_1$ (0.50) and $R/\tilde{\epsilon}_1$ (0.23) are larger than that of $n/\tilde{\epsilon}_2$ (0.18) and $R/\tilde{\epsilon}_2$ (0.00). That means that n and R are more dependent on ϵ_1 than ϵ_2 . R^2 of $\epsilon_1/\tilde{\epsilon}_1$ (0.72) and $\epsilon_2/\tilde{\epsilon}_2$ (0.74) are same with that of GNNOpt since $\epsilon_1 = \tilde{\epsilon}_1$ and $\epsilon_2 = \tilde{\epsilon}_2$. By evaluating the K-K relations before and after training the GNNOpt model, we found that the K-K relations should be applied before training the model. That means that the K-K relations are used to construct the ground truth database for n , k , and R based on ϵ_1 and ϵ_2 , as shown in Equations (4)–(6). After that, these databases are used to train the GNNOpt models for n , k , and R predictions.

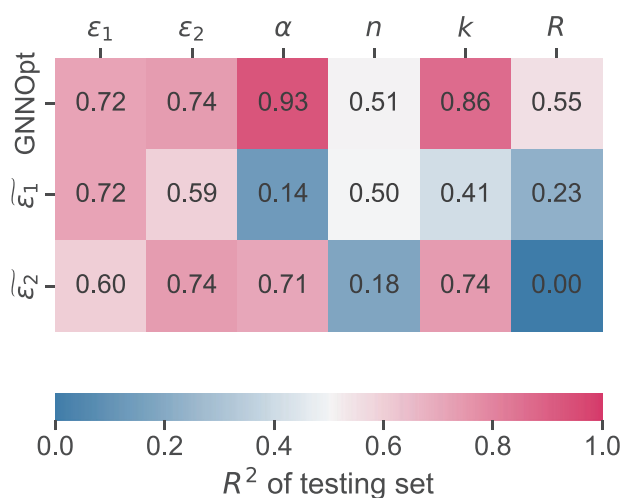


Figure 3. Coefficient of determination R^2 of the test set for the optical spectra $W = \epsilon_1$, ϵ_2 , α , n , k , and R . For the 1st row, W is directly predicted by using GNNOpt. For the 2nd and 3rd rows, W is calculated from the predicted $\tilde{\epsilon}_1$ and $\tilde{\epsilon}_2$ by using the Kramers-Krönig relations, respectively.

4. Screening Solar Cell Materials

The first application of the GNNOpt is to identify potential solar cell materials with high-performance energy conversion. For this task, Shockley-Queisser (SQ) limit is a key factor to estimate the upper limit for solar-energy conversion efficiency.^[30] However, the SQ limit assumes that the absorption coefficient is a step function or the infinity thickness of the absorber material. An alternative parameter that is more suitable for real PV materials and devices is introduced by Yu and Zunger,^[31] which is the spectroscopic limited maximum efficiency (SLME). For the SLME method, the efficiency of the energy conversion, η , is defined as the ratio between the maximum output power density, P_{out} , and the incident solar power density, P_{solar} ,

$$\eta = \frac{P_{\text{max}}}{P_{\text{solar}}} \quad (8)$$

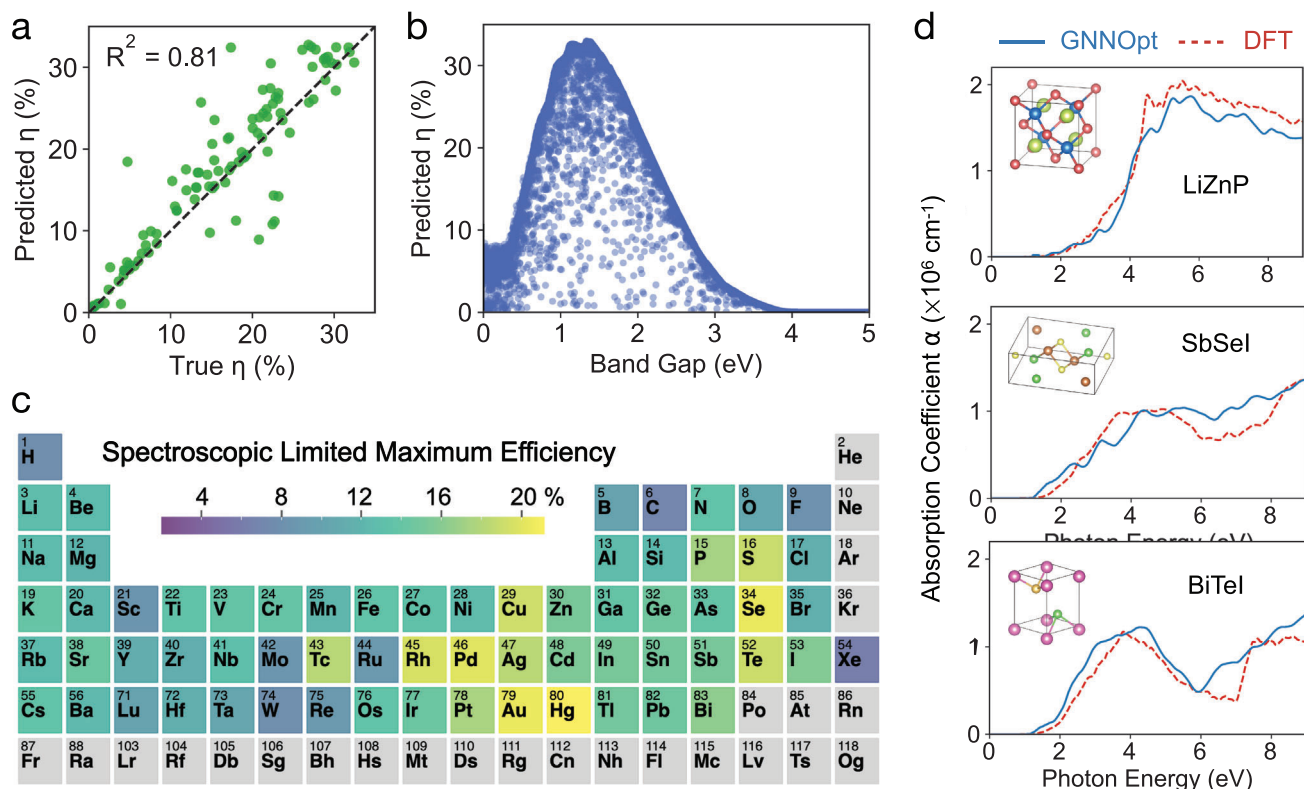


Figure 4. Searching material candidates for solar energy harvesting applications with GNNOpt. a) Comparison between predicted and true efficiencies η , obtained by the GNNOpt and the DFT ground truth, respectively, for the test set. Here, η is estimated by the spectroscopic limited maximum efficiency (SLME). b) η -predicted by the GNNOpt of the 5,281 stable materials from the Materials Project, which is plotted as a function of the energy bandgap. c) The periodic table is colored by the SLME of materials containing each element. d) Absorption coefficients $\alpha(E)$ are plotted as a function of photon energy E by using the GNNOpt (solid lines) and the DFT (dashed lines) for the unseen materials, including LiZnP, SbSeI, and BiTeI. The inset figures show the crystal structures of these materials.

P_{\max} and P_{solar} can be written in terms of $J - V$ characteristic of the solar cell and the solar spectrum, respectively:^[31]

$$P_{\max} = \max\{JV\}_V = \max\{[J_{\text{sc}} - J_0(e^{qV/k_B T} - 1)]V\}_V \quad (9)$$

and

$$P_{\text{solar}} = \int_0^\infty EI_{\text{solar}}(E)dE \quad (10)$$

respectively, where J is the total current density, V is potential over the absorber layer, k_B is the Boltzmann constant, T is the temperature of the solar device, q is the elementary charge. J_{sc} and J_0 are the short-circuit current density and the reverse saturation current density, respectively, which depend on the absorption coefficient, $\alpha(E)$, and the thickness, L , of the material as follows:^[31]

$$J_{\text{sc}} = q \int_0^\infty [1 - e^{-2\alpha(E)L}] I_{\text{solar}}(E)dE \quad (11)$$

and

$$J_0 = \frac{J_r}{f_r} = \frac{q\pi}{f_r} \int_0^\infty [1 - e^{-2\alpha(E)L}] I_{\text{bb}}(E, T)dE \quad (12)$$

where $I_{\text{solar}}(E)$ is the AM1.5G solar spectrum, $I_{\text{bb}}(E, T)$ is the black-body spectrum, and f_r is the radiative recombination current density. From Equations (8)–(12), the material property-related inputs $\alpha(E)$, L , T , and f_r are required to calculate the SLME. In this work, we assume the $L = 500$ nm, $T = 300$ K, and the radiative recombination is the only recombination process, i.e., $f_r = 1$, which is a good approximation for the materials where radiative recombination dominates, such as GaAs.^[32] We note that $L = 500$ nm satisfies the condition $L > \lambda_F$ for most semiconductors.^[27] Since the solar spectrum for solar energy harvesting ranges from 200 – 2500 nm (i.e., 0.5 – 6.0 eV), we thus retrain the GNNOpt model with the photon energy < 10 eV (see Figures S14–S15, Supporting Information), which is sufficient for the SLME calculation.

In Figure 4a, we compare the predicted and true efficiencies for the test set, in which $R^2 = 0.81$ shows a high prediction accuracy of the GNNOpt for the SLME. Using the trained GNNOpt model, we predict the η values of 5,281 unseen crystal structures from the Materials Project without ground-truth optical spectra.^[22] Here, we select only the stable insulators with an energy band gap E_g , suitable for solar-energy harvesting applications ($0 \leq E_g \leq 5.0$ eV). Statistical plots of the unseen materials with the number of materials as a function of E_g , N , and lattice constant are shown in Figure S13 (Supporting Information). In Figure 4b, we plot the

predicted- η as a function of E_g . We observe that the maximum η is about 32% at $E_g \approx 1.3$ eV, which is consistent with the SQ limit. However, the SLME is a more stringent selection parameter than the SQ limit for the solar cell materials because the SLME shows η values over a wide range for materials with similar bandgaps, indicating the significant contribution of $\alpha(E)$ to η .

Additionally, knowing which elements from the periodic table contribute most to high-efficiency solar cell materials can serve as an initial guideline for material design. As presented in Figure 4c, the transition metals such as Tc, Rh, Pd, Pt, Cu, Ag, Au, and Hg, and chalcogenides including S, Se, and Te, are the main constituent elements of the solar cell materials. This finding is in agreement with widely known solar cell materials, such as Cu-rich chalcopyrite,^[33] Pb-based perovskites,^[34,35] or CdTe.^[36] In Table S1 (Supporting Information), we identified 246 materials with η -SLME greater than 32%. To validate the predicted SLME of the unseen materials, we selected three examples from the highest-SLME materials list: LiZnP, SbSeI, and BiTeI. These materials are not present in the DFT database. We conducted DFT calculations to determine $\alpha(E)$ in these examples. The results, depicted in Figure 4d, show excellent agreement between the DFT calculations (dashed lines) and the GNNOpt-predicted α values (solid lines). This indicates that GNNOpt could be an effective materials screening tool at a much lower computational cost. We note that for a huge database, the GNNOpt can be accelerated by combining with the genetic algorithms (GA) to speed up the search of candidate materials.^[37]

5. Probing Quantum Materials

In this section, we provide another application scope of GNNOpt to probe quantum geometry and topology in quantum materials. Since the dipole moment matrix element for optical transition is closely related to the interband Berry connection,^[38] recent studies have established the relationship between the quantum geometry and the optical properties.^[39–41] In particular, Onishi and Fu^[24] theoretically showed that the generalized quantum weight, a concept that can be derived from optical spectra, is a direct measure of ground state quantum geometry and topology. The quantum weight K_{xx} is given by a modification of f -sum rule weighted by the inverse frequency as follows:^[24]

$$K_{xx} = \frac{2\hbar}{e^2} \int_0^\infty \frac{\text{Re}[\sigma_{xx}(\omega)]}{\omega} d\omega = \frac{2\hbar}{e^2} \int_0^\infty \epsilon_2(\omega) d\omega \quad (13)$$

where $\text{Re}[\sigma_{xx}(\omega)]$ is the real part of optical conductivity over the entire frequency range (i.e., $0 < \omega < \infty$).

The GNNOpt for prediction $\epsilon_2(\omega)$ with $0 < \omega < 50$ eV (see Figure S6, Supporting Information) is used to obtain K_{xx} . It is noted that for $\omega > 50$ eV, $\epsilon_2(\omega) \rightarrow 0$ for all materials in the training set. In Figure 5a, we compare the predicted and true K_{xx} in the unit of \hbar/e^2 for the test set. $R^2 = 0.73$ shows a good prediction between the GNNOpt and the DFT, particularly for $K_{xx} < 25$. The GNNOpt is thus used to predict K_{xx} of the 5,281 unseen insulator materials, as shown in Figure 5b. For simplicity, we consider the quantum weight of the well-known topology insulator Bi_2Te_3 ($K_{xx} = 28.87$) as a threshold for classifying quantum materials, in which the material with $K_{xx} > 28.87$ is considered as the high- K_{xx} materials. We identified 297 high- K_{xx} materials, listed in Table

S2 (Supporting Information), in which several materials, such as ZrTe_5 ($K_{xx} = 33.90$), TaAs_2 ($K_{xx} = 37.66$), FeSi ($K_{xx} = 48.74$), and NbP ($K_{xx} = 35.58$), etc., have been reported as quantum materials with anomalous hall effect,^[42] large magnetoresistance,^[43] topological Fermi arcs,^[44] and quantum oscillations.^[45]

To further validate the quantum characteristics from the predicted high- K_{xx} materials, we carry out additional DFT calculations for SiOs, which have exceedingly high quantum weight ($K_{xx} = 46.52$) but have not been well studied. As shown in Figure 5c, SiOs is found to host the threefold and double-Weyl fermions at the Γ and R points, respectively. Since both the threefold and double-Weyl fermions are located above the Fermi level (at 1.15 and 0.7 eV, respectively), it would not be possible to detect them using the angular resolved photoemission spectroscopy (ARPES) technique. Nevertheless, irrespective of the energy position of the manifold band crossing points, we could expect the associated Fermi surface arcs on the surface Brillouin zone.^[46,47] We thus calculated the surface band structure of the (001) SiOs by using the maximally-localized Wannier functions and the Green's function approach (see Experimental Section). As shown in Figure 5d, we find the surface state near the Fermi level form self-enclosed loops, indicating the triviality of these states, in contrast to the open Fermi arcs. This closed-loop shrinks in size when the surface state moves away from the Fermi energy (see Figure S17a, Supporting Information). However, the nontrivial topological Fermi arcs associated with manifold fermions can be found at 1.3 eV above the Fermi level (see Figure S17b, Supporting Information). These signatures suggest the ultra-quantum properties of SiOs.

6. Discussion

In this work, we present GNNOpt, a graph neural network model with engineered input embedding that directly predicts all linear optical spectra from crystal structures. The ensemble embedding layer for automatic embedding optimization improves prediction accuracy without modifying the neural network structures. Furthermore, such an embedding layer with auto-optimization capability is expected to be universally applicable as a universal input layer for GNNs. By integrating with the equivariant neural network, GNNOpt achieves high-quality predictions using a small data set of 944 materials. Furthermore, applying the Kramers-Kronig relation before training the GNNOpt model is shown to have better-predicted optical spectra, as observed by comparing optical properties before and after training. With GNNOpt, we were able to identify over 200 materials with over 32% conversion efficiency for solar energy harvesting applications. Additionally, thanks to the recent connection between optical property and ground-state topology, we were able to screen quantum materials carrying multiple nontrivial topologies, including SiOs.

The direct ensemble structure input in GNNOpt opens several promising future research directions. First, the atomic embedding with multiple features greatly facilitates the encoding of defect structures, which could be encoded by perturbing one or more atomic features. Given the significant impact of defects on optical properties, predicting optical properties with defects is highly desirable (see Figure S16, Supporting Information). This approach could enable tuning absorption spectra through new defect levels or identifying non-radiative recombination centers

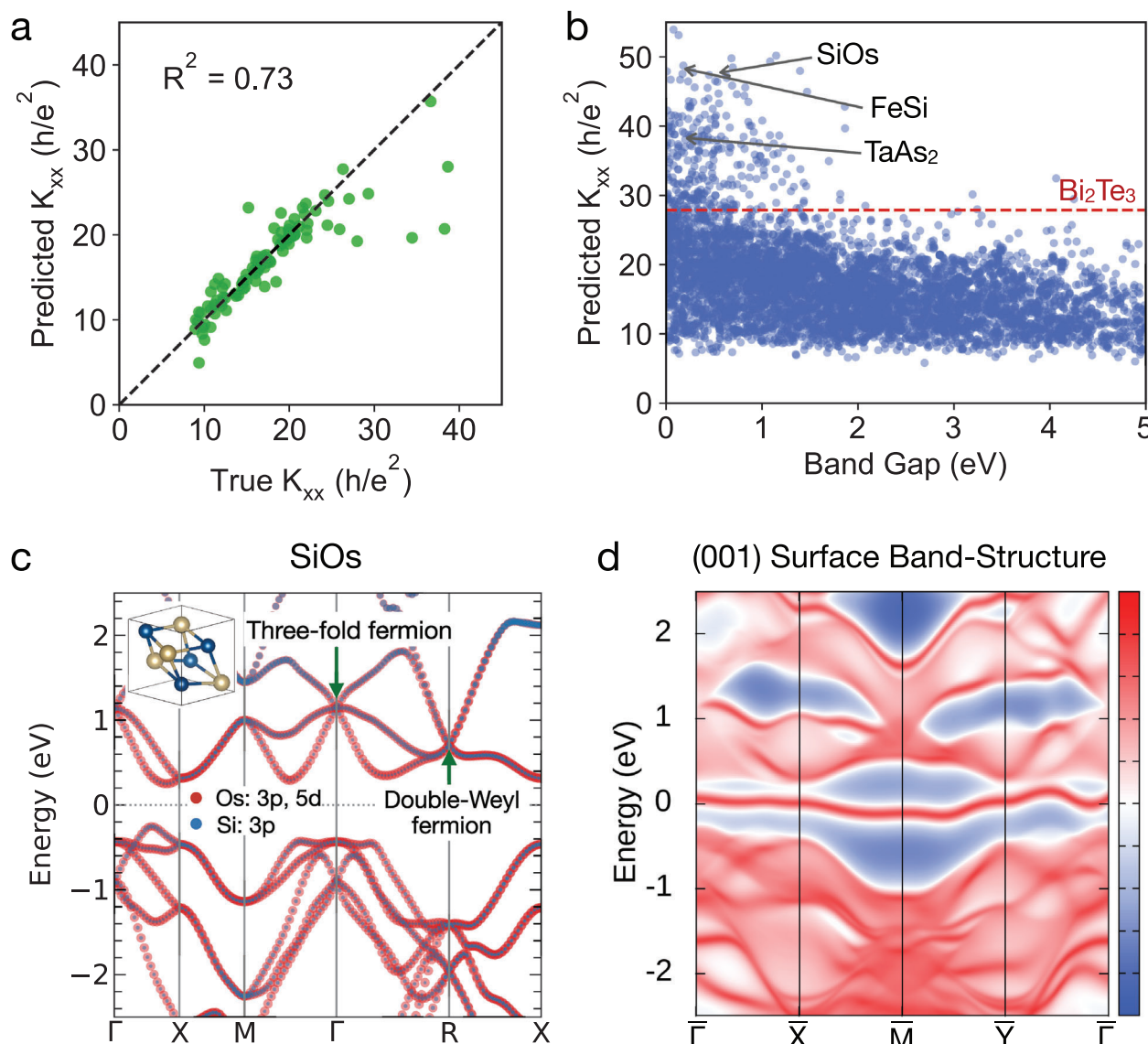


Figure 5. Searching for quantum materials with high quantum weight K_{xx} using GNNOpt. a) Comparison between predicted and true K_{xx} , obtained by the GNNOpt and the DFT ground truth, respectively, for the test set. b) K_{xx} -predicted by the GNNOpt of the 5,281 stable materials from the Materials Project, which is plotted as a function of the energy bandgap. The dashed line indicates $K_{xx} = 27.87$ of Bi_2Te_3 . c) Electronic band structure with atomic orbital projections in the color of the high- K_{xx} material, SiOs. The arrows indicate the double-Weyl and threefold fermions at the Γ and R points, respectively. d) The surface band structure along the high-symmetry lines on the (001) surface of SiOs. The color bar scale is in arbitrary units.

that reduce solar cell efficiency. Second, building on the successful prediction of linear optical properties, extending this method to second and third-order optical responses, such as second harmonic generation and Raman spectroscopy, offers valuable opportunities. Third, since the current ground truth database is DFT-based, future studies incorporating many-body effects, such as the GW approximation, could provide more accurate predictions, particularly for excitons. The absorption coefficient of the magnetic material depends on the dielectric function and magnetic permeability.^[48] Therefore, a future goal should be to predict the permeability spectra using GNNOpt, which requires a permeability database, in order to obtain the absorption coefficient for the magnetic material. Lastly, since the ensemble em-

bedding integrates multiple embedding types to enhance overall model performance and goes beyond feature selection, it is expected to be broadly applicable to other GNN models.

7. Experimental Section

Data Preparation for Optical Spectra: The GNNOptic was trained and tested with the 944 crystalline solids calculated from the density functional theory (DFT) by Yang et al.^[13]. The databases were obtained from the Materials Project^[22] using the API. The optical spectra in the database included the frequency-dependent dielectric function and the corresponding absorption coefficient, which was performed within the independent-particle approximation (IPA). Yang et al.^[13] showed that the IPA was

sufficient in most cases to reproduce the experiment spectra. The 944 materials in the database were selected based on the energy bandgap between $0 \leq E_g \leq 5.0$ eV and the number of atomic sites $N < 10$ in each unit cell. The static plot of the number of materials as a function of E_g , N , and lattice constants are given in Figure S1 (Supporting Information).

The entire dataset was randomly split into 80%, 10%, and 10% for the training (733 materials), validation (97 materials), and test (110 materials) sets, respectively (see Figure S2, Supporting Information). Given limited training data, the high photon energy ω resolution (with 2001 points) requests a large number of parameters, such as radius cut-off and the number of layers, in the neural network to fit the output dimension. Therefore, to ensure a balanced output dimension and the calculation cost, the smoothed spectrum was interpolated in the range $0 \leq \hbar\omega \leq 50$ eV to 251 points, which was suitable to reproduce all of the spectra peaks (see Figure S3, Supporting Information). To enhance the performance of the training model, the optical spectra are normalized was also applied by using the median value. The predictive model on 5,281 stable insulators from the Materials Project^[22] with $0 \leq E_g \leq 5.0$ eV with atomic site number $N < 20$ in each unit cell. The static plot of the number of materials as a function of E_g , N , and lattice constants for this database are given in Figure S12 (Supporting Information).

Hyperparameters Optimization: The set of optimized parameters to get the best results for GNNOpt is given as follows: the maximum cutoff radius $r_{\text{cut}} = 6$ Å, the maximum of spherical harmonics $l_{\text{max}} = 2$, the length of the embedding feature vector was 64, the multiplicity of irreducible representation was 32, the number of pointwise convolution layer $n_{\text{conv}} = 2$, the AdamW optimizer learning rate was $(5 \times 10^{-3}) \times 0.96^k$ with k being the epoch number, and the AdamW optimizer weight decay coefficient is 0.05.

First-Principles Calculations: The Quantum ESPRESSO package was used for the DFT calculations to obtain the absorption coefficient $\alpha(E)$ of LiZnP, SbSeI, and BiTeI within independent particle approximation.^[11,49] The crystal structures of these materials were taken from the Materials Project to be consistent with that in the GNNOpt prediction. The optimized norm-conserving Vanderbilt pseudopotentials^[50] with the Perdew–Burke–Ernzerhof exchange–correlation functional^[51] were selected for all atoms. A cutoff energy of 60 Ry for plane wave was used for all materials. For self-consistent field (SCF) calculations, the k-points meshes were $6 \times 6 \times 6$, $4 \times 8 \times 3$, and $9 \times 9 \times 5$ for LiZnP, SbSeI, and BiTeI, respectively, while the dense meshes of $20 \times 20 \times 20$, $14 \times 30 \times 12$, and $27 \times 27 \times 15$, respectively, were used for non-SCF calculations in order to achieve convergence in $\alpha(E)$.

To calculate the electronic band structure of SiOs, the tight-binding model was obtained by the maximally-localised Wannier functions using the Wannier90 code,^[52] in which Si 3p and Os 3p, 5d orbitals were selected as the basis. Then, the WannierTools code^[53] with the Green function approach was used for the analysis of the surface band structure.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

N.T.H. acknowledged the Researcher, Young Leaders Overseas Program and financial support from the Frontier Research Institute for Interdisciplinary Sciences from Tohoku University. R.O. acknowledged the support from the U.S. Department of Energy (DOE), Office of Science (SC), Basic Energy Sciences (BES), Award No. DE-SC0021940. A.C. thanks National Science Foundation (NSF) Designing Materials to Revolutionize and Engineer the Future (DMREF) Program with Award No. DMR-2118448. M.L. was partially supported by NSF ITE-2345084, the Class of 1947 Career Development Chair and the support from R. Wachnik.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are openly available on GitHub at <https://github.com/nguyen-group/GNNOpt>.

Keywords

equivariant neural networks, kramers–krönig relations, machine learning, optical spectra, photovoltaic materials, quantum materials

Received: June 27, 2024

Revised: August 15, 2024

Published online: September 12, 2024

- [1] D. Marpaung, J. Yao, J. Capmany, *Nat. Photonics* **2019**, *13*, 80.
- [2] W. Bogaerts, D. Pérez, J. Capmany, D. A. Miller, J. Poon, D. Englund, F. Morichetti, A. Melloni, *Nature* **2020**, *586*, 207.
- [3] Y. Meng, Y. Chen, L. Lu, Y. Ding, A. Cusano, J. A. Fan, Q. Hu, K. Wang, Z. Xie, Z. Liu, Y. Yang, Q. Liu, M. Gong, Q. Xiao, S. Sun, M. Zhang, X. Yuan, X. Ni, *Light Sci. Appl.* **2021**, *10*, 1.
- [4] F. Priolo, T. Gregorkiewicz, M. Galli, T. F. Krauss, *Nat. Nanotechnol.* **2014**, *9*, 19.
- [5] H.-L. Liu, B. D. Annawati, N. T. Hung, D. P. Gulo, P. Solís-Fernández, K. Kawahara, H. Ago, R. Saito, *Phys. Rev. B* **2023**, *107*, 165421.
- [6] J. B. Khurgin, *ACS Photonics* **2022**, *9*, 743.
- [7] D. P. Gulo, N. T. Hung, R. Sankar, R. Saito, H.-L. Liu, *Phys. Rev. Mater.* **2023**, *7*, 044001.
- [8] R. Saito, N. T. Hung, T. Yang, J. Huang, H.-L. Liu, D. P. Gulo, S. Han, L. Tong, *Small* **2024**, *2308558*.
- [9] B. C. Smith, *Fundamentals of Fourier transform infrared spectroscopy*, 2nd ed., CRC Press, Boca Raton **2011**.
- [10] H.-H. Perkampus, *UV-VIS Spectroscopy and its Applications*, Springer Berlin, Heidelberg **1992**.
- [11] N. T. Hung, A. R. T. Nugraha, R. Saito, *Quantum ESPRESSO course for solid-state physics*, Jenny Stanford Publishing, Singapore **2022**.
- [12] D. P. Gulo, N. T. Hung, T.-J. Yang, G.-J. Shu, R. Saito, H.-L. Liu, *Carbon* **2022**, *197*, 485.
- [13] R. X. Yang, M. K. Horton, J. Munro, K. A. Persson, *arXiv preprint arXiv:2209.02918* **2022**.
- [14] Y. Liu, T. Zhao, W. Ju, S. Shi, *J. Materiomics* **2017**, *3*, 159.
- [15] G. Genty, L. Salmela, J. M. Dudley, D. Brunner, A. Kokhanovskiy, S. Kobtsev, S. K. Turitsyn, *Nat. Photonics* **2021**, *15*, 91.
- [16] R. Pollice, G. dos Passos Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao, A. Aspuru-Guzik, *Acc. Chem. Res.* **2021**, *54*, 849.
- [17] T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301.
- [18] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, G. Monfardini, *IEEE Trans. Neural Netw. Learn. Syst.* **2008**, *20*, 61.
- [19] Z. Chen, N. Andrejevic, T. Smidt, Z. Ding, Q. Xu, Y.-T. Chi, Q. T. Nguyen, A. Alatas, J. Kong, M. Li, *Adv. Sci.* **2021**, *8*, 2004214.
- [20] M. Geiger, T. Smidt, *arXiv preprint arXiv:2207.09453* **2022**.
- [21] V. Lucarini, J. J. Saarinen, K.-E. Peiponen, E. M. Vartiainen, *Kramers–Kronig relations in optical materials research*, vol. 110, Springer Berlin, Heidelberg **2005**.
- [22] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, *1*, 1.

- [23] R. Okabe, A. Chotrattanapituk, A. Boonkird, N. Andrejevic, X. Fu, T. S. Jaakkola, Q. Song, T. Nguyen, N. Drucker, S. Mu, Y. Wang, B. Liao, Y. Cheng, M. Li, *arXiv preprint arXiv:2301.02197* **2023**.
- [24] Y. Onishi, L. Fu, *Phys. Rev. X* **2024**, 14, 011052.
- [25] D. C. Hutchings, M. Sheik-Bahae, D. J. Hagan, E. W. Van Stryland, *Opt. Quantum Electron.* **1992**, 24, 1.
- [26] V. P. Kurbatsky, V. V. Pogosov, *Phys. Rev. B* **2010**, 81, 155404.
- [27] N. Mancini, G. Giaquinta, A. Pennisi, *J. Opt. Soc. Am.* **1981**, 71, 383.
- [28] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. v. der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, I. Polat, Y. Feng, E. W. Moore, J. VanderPlas, *Nat. Methods* **2020**, 17, 261.
- [29] P. Pyykkö, M. Atsumi, *Chem. - Eur. J.* **2009**, 15, 186.
- [30] W. Shockley, H. J. Queisser, *J. Appl. Phys.* **1961**, 32, 510.
- [31] L. Yu, A. Zunger, *Phys. Rev. Lett.* **2012**, 108, 068701.
- [32] I. Schnitzer, E. Yablonovitch, C. Caneau, A. Ersen, T. Gmitter, in *LEOS'92 Conference Proceedings*, IEEE, Piscataway, NJ **1992**, pp. 127–128.
- [33] J. Jiang, R. Giridharagopal, E. Jedlicka, K. Sun, S. Yu, S. Wu, Y. Gong, W. Yan, D. S. Ginger, M. A. Green, X. Hao, W. Huang, H. Xin, *Nano Energy* **2020**, 69, 104438.
- [34] Q. Tai, K.-C. Tang, F. Yan, *Energy Environ. Sci.* **2019**, 12, 2375.
- [35] W. Ke, C. C. Stoumpos, M. G. Kanatzidis, *Adv. Mater.* **2019**, 31, 1803230.
- [36] C. S. Ferekides, D. Marinskiy, V. Viswanathan, B. Tetali, V. Palekis, P. Selvaraj, D. Morel, *Thin Solid Films* **2000**, 361, 520.
- [37] S. Zhou, J. Qiu, C. Zhang, Y. Guo, Q. Pan, Q. Zhou, Y. Shuai, *Int. J. Therm. Sci.* **2022**, 179, 107680.
- [38] C. Aversa, J. E. Sipe, *Phys. Rev. B* **1995**, 52, 14636.
- [39] J. Ahn, G.-Y. Guo, N. Nagaosa, *Phys. Rev. X* **2020**, 10, 041041.
- [40] T. Holder, D. Kaplan, B. Yan, *Phys. Rev. Res.* **2020**, 2, 033100.
- [41] P. Bhalla, K. Das, D. Culcer, A. Agarwal, *Phys. Rev. Lett.* **2022**, 129, 227401.
- [42] T. Liang, J. Lin, Q. Gibson, S. Kushwaha, M. Liu, W. Wang, H. Xiong, J. A. Sobota, M. Hashimoto, P. S. Kirchmann, Z.-X. Shen, R. J. Cava, N. P. Ong, *Nat. Phys.* **2018**, 14, 451.
- [43] Y.-Y. Wang, Q.-H. Yu, P.-J. Guo, K. Liu, T.-L. Xia, *Phys. Rev. B* **2016**, 94, 041103.
- [44] S. Changdar, S. Aswartham, A. Bose, Y. Kushnirenko, G. Shipunov, N. Plumb, M. Shi, A. Narayan, B. Büchner, S. Thirupathaiah, *Phys. Rev. B* **2020**, 101, 235105.
- [45] J. Klotz, S.-C. Wu, C. Shekhar, Y. Sun, M. Schmidt, M. Nicklas, M. Baenitz, M. Uhlarz, J. Wosnitza, C. Felser, B. Yan, *Phys. Rev. B* **2016**, 93, 121105.
- [46] B. Bradlyn, J. Cano, Z. Wang, M. Vergniory, C. Felser, R. J. Cava, B. A. Bernevig, *Science* **2016**, 353, aaf5037.
- [47] Y. Sun, S.-C. Wu, M. N. Ali, C. Felser, B. Yan, *Phys. Rev. B* **2015**, 92, 161107.
- [48] D. Wood, N. Ashcroft, *Phys. Rev. B* **1982**, 25, 6255.
- [49] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, et al., *J. Phys. Condens. Matter* **2009**, 21, 395502.
- [50] D. Hamann, *Phys. Rev. B* **2013**, 88, 085117.
- [51] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, 77, 3865.
- [52] A. A. Mostofi, J. R. Yates, Y. S. Lee, I. Souza, D. Vanderbilt, N. Marzari, *Comp. Phys. Comm.* **2008**, 178, 685.
- [53] Q. Wu, S. Zhang, H.-F. Song, M. Troyer, A. A. Soluyanov, *Comput. Phys. Commun.* **2018**, 224, 405.