Graph Transformer Model Integrating Physical Features for Projected Electronic Density of States Prediction

Jiahao Wu, Weipeng Lu, Jingwen Wu, Bing Zhang, and Yanchuan Guo*

Cite This: https://doi.org/10.1021/acs.jpca.5c03006

Read Online

ACCESS | 📊 Metrics & More | 📰 Article Recommendations | 🆘 Supporting Information

ABSTRACT: The prediction of the projected density of states (PDOS) in materials has traditionally relied on deep learning models based on graph convolutional networks (GCN) and Graph Attention Networks (GAT). In this study, utilizing PDOS data from the Materials Project, we demonstrate that the Graph Transformer (GT)—which employs multi-head scaled dot-product attention over each node and its local neighborhood—consistently outperforms both GAT and GCN models in prediction accuracy under identical energy-level-corrected conditions. Furthermore, by incorporating the valence electron counts of the s, p, d, and f orbitals for the entire structure as an additional feature into a GT-based model, we further enhance the model's predictive performance. The novel framework and physics-informed feature enhancements offer valuable insights into improving the prediction of PDOS and potentially other electronic properties.
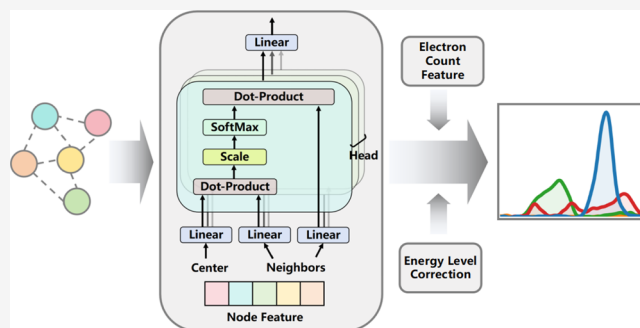
1. INTRODUCTION

The electronic density of states (DOS)[1] is a crucial indicator of material electronic structure and is increasingly used for characterizing the electronic properties of materials. It describes the number of electronic states per unit energy interval in a material, providing insights into the width of the valence band, the band gap, and the electron occupancy at the Fermi level. The projected electronic density of states (PDOS), obtained by projecting the DOS onto individual atomic orbitals, offers more detailed information, such as the energy levels of different atomic orbitals, bonding situations between orbitals, band centers, and band widths. The primary method for obtaining the PDOS is theoretical calculation. In the mid-20th century, scientists simplified the complex many-electron wave function equations into iteratively solvable Hartree—Fock equations[2−4] using the Born—Oppenheimer approximation, mean-field approximation, and Pauli exclusion principle, marking the dawn of material electronic structure calculations. In 1965, Kohn and Sham introduced the Kohn—Sham equations,[5] incorporating electron exchange-correlation energy terms and using electron density instead of wave functions, significantly reducing computational costs while maintaining high accuracy. This approach forms the core of the modern density functional theory (DFT). However, for complex materials, DFT calculations based on the Kohn—Sham self-consistent equations still involve high computational costs, often requiring thousands of seconds or more for conventional computing resources. This makes obtaining electronic structure information for a large batch of structures costly, hindering the goal of high-throughput

screening and the construction of databases through DFT calculations.

In recent years, with the rapid development of machine learning technology, materials science research has encountered new opportunities. Several studies have attempted to combine neural networks with high-throughput calculations, achieving high-accuracy predictions of materials' electronic structure information and significantly reducing the cost. Previous models that predict DOS (or PDOS), as well as other models that predict various electronic properties, generally utilize deep learning models baed on normal Graph Neural Networks (GNNs) frameworks, such as Graph Attention Networks (GAT)[6,7] and Graph Convolutional Networks (GCN),[8−10] as well as simpler models based on conventional neural network architectures.[11,12] GNNs are specifically designed for processing graph-structured data and addressing problems in non-Euclidean spaces. Traditional neural networks, such as Multi-layer Perceptron (MLP) and Convolutional Neural Networks (CNNs), are primarily suited for data with regular grid-like structures and struggle to effectively represent data characterized by invariances to rotation and translation, such as chemical molecules or crystal cells. In contrast, GNNs propagate
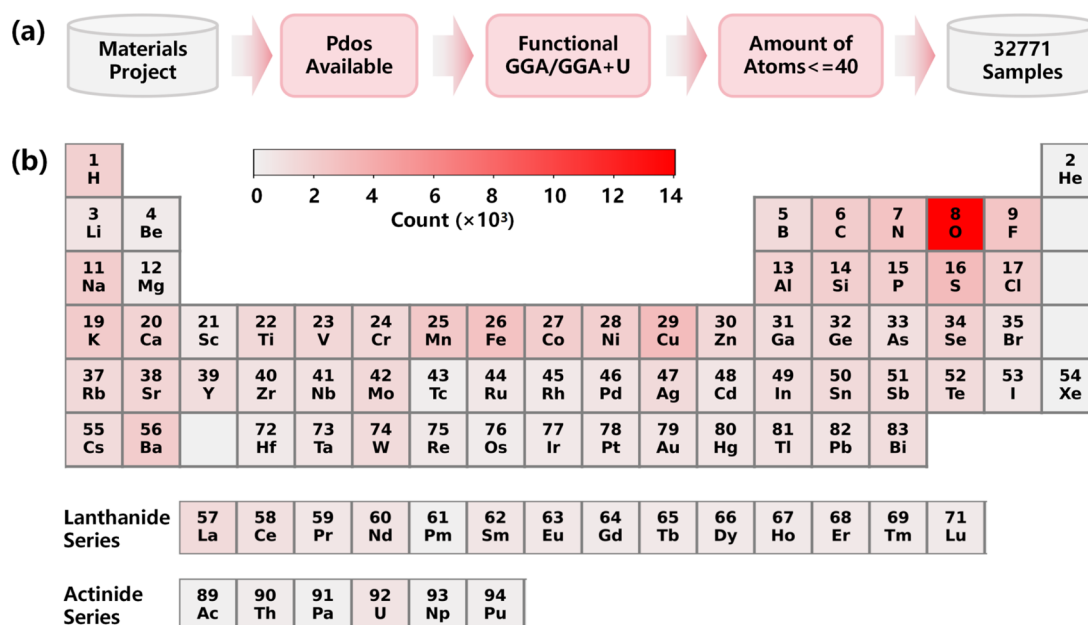
Figure 1. Overview of the data set. (a) Workflow for data set construction; (b) elemental composition distribution of the sample structures.

and aggregate information through graph structures, making them particularly suitable for capturing the complex relationships inherent in such domains. GCN aggregates the features of neighboring nodes and the connecting edges through operations such as averaging or summation, aiming to capture the underlying interaction patterns among nodes. However, traditional GCN assigns equal importance to all neighboring nodes during aggregation, which may overlook the varying significance of different neighbors. To address this limitation, GAT introduces dynamic weighting mechanisms that assign learnable, adaptive attention coefficients to each neighbor, thereby enabling the model to differentiate between neighbors with greater or lesser influence on the central node. While they have indeed played a significant role in predicting material properties, the rapid advancement of deep learning in computer science has led to the emergence of increasingly complex and promising new frameworks within the scientific community. Given the remarkable success and dominance of Transformer architectures[13] in natural language processing, their adaptation to graph-structured data (Graph Transformer, GT)[14,15] has attracted growing interest. However, the application of GT to PDOS prediction remains unexplored.

Moreover, most existing studies on models for predicting the electronic properties of materials primarily adapt general-purpose deep learning frameworks or algorithms with relatively limited efforts to enhance feature representations by leveraging the specific physical and chemical backgrounds of materials science. A recent study[16] demonstrated that principal component analysis (PCA, an unsupervised learning method) can efficiently identify favorable descriptors for machine learning by capturing the directions of greatest variance in the dataset. By employing PCA to extract electronic-structure descriptors, the study highlighted clear advantages in terms of predictive performance and interpretability. Notably, in more complex models such as GNNs, atomic-level descriptors—used as node features in graph-based representations—are commonly parametrized and optimized during training. In such cases, manually extracting principal components may offer limited benefits. To further enhance the model performance, it could be

advantageous to incorporate certain physically meaningful features alongside end-to-end learning, thereby helping the model to better capture the underlying physical principles.

In this work, we employ supervised learning on the Materials Project dataset's PDOS to demonstrate that GT outperforms conventional GCN and GAT under identical energy-level-corrected conditions. This superiority stems from GT's multi-head scaled dot-product attention mechanism, which utilizes dot-product-based interactions to capture the relationships between each node and its local neighborhood. Furthermore, we enhance the model by incorporating the valence electron counts of the s, p, d, and f orbitals for the entire structure as a physics-informed feature, which significantly improves the accuracy of the PDOS prediction. Our findings provide an effective strategy for advancing PDOS or other electronic property predictions in materials science.

2. METHODS

2.1. Dateset and Data Augmentation. All the structures and PDOS data were bulk downloaded from the Materials Project,[17] with samples calculated using GGA[18] or GGA+U[19] functionals and primitive structures containing no more than 40 atoms being selected. (Figure 1a) A total of 32,771 sample-label pairs were ultimately obtained, covering materials or molecules such as elemental substances, alloys, metallic compounds, and nonmetallic compounds. Their element distribution is shown in Figure 1b, covering elements with atomic numbers from 1 (hydrogen) to 83 (bismuth) and 89 (actinium) to 94 (plutonium). The elemental composition of these samples predominantly distributes across the first three periods and main groups of the large periods, consistent with natural elemental abundances and demonstrating a rational distribution pattern. In addition, Supporting Information Figure S1 shows the space group distribution of these materials, with a relatively higher frequency of popular space groups in the field of materials chemistry research. All PDOS data were shifted by subtracting the Fermi level with the purpose of setting the Fermi level as the zero energy reference point. The PDOS data for all materials are divided into four orbitals: s, p, d, and f. The energy range for each

orbital's data is from -10 to 10 eV, with a smearing of approximately 0.1 eV between data points. Therefore, the PDOS data for each sample form a 4×201 matrix. The PDOS data for all materials are normalized by dividing by the number of atoms in the corresponding structure. Additionally, the PDOS data for all materials are processed using a Gaussian filter to apply appropriate smoothing, thereby reducing the impact of sharp and insignificant small peaks on model training.

2.2. Model Hyperparameters and Model Details. For a model comparison, the GCN, GAT, and GT frameworks were all configured with three layers and employed residual connections. The input and output dimensions of the node features were both set to 256. In the process of collecting edge connections, a cutoff radius of 2.5 Å was employed. This radius is large enough to encompass the majority of chemical bonds within the system, yet sufficiently small to prevent the model from overfitting by including spurious connections between nodes that are not physically meaningful. Moreover, periodic boundary conditions (PBC) were incorporated during the edge collection process,[20] ensuring that the models consider the inherent periodicity of crystal structures. Consequently, these models are denoted as crystal graph convolutional network (CGCNN), crystal graph attention network (CGAT), and crystal graph transformer (CGT), respectively. Edge features were expanded by using a Gaussian function, resulting in an edge feature dimension of 256. After the graph layer, a global pooling operation is applied. Both sum pooling, max pooling, and average pooling are considered, and their respective outputs are concatenated before being passed into the fully connected layer. After pooling, each model concatenates the energy level correction feature encoding to correct for the differences in the original energy level data across different samples. For the CGAT and CGT models, the number of attention heads was set to 4.

2.3. Model Training and Evaluation. The data set was randomly divided into training, validation, and test sets with a ratio of 8:1:1. The process of model training and evaluation was mainly based on the PyTorch[21] and PyTorch Geometric libraries,[22] with hardware support provided by NVIDIA GeForce RTX 3080 Ti 12GB GPUs. The model training was based on the backpropagation algorithm,[23] conducted for 300 epochs. The training batch size was set to 256, and a fixed learning rate of 1×10^{-5} was used to promote stable training and convergence of the models. L2 regularization was applied with a weight decay factor set to 1×10^{-6}. The loss function employed was the MSE loss, and the AdamW optimizer was used. During model training, both the MSE loss and RMSE loss on the training and validation sets were recorded to monitor the training process and ensure the stability and convergence of the model (Supporting Information Figure S2). A denoising procedure was applied by setting the negative values of the model output to zero, followed by Gaussian filtering (implemented using the sklearn[24] library) for smoothing. Denoising was incorporated only during the forward propagation step for model evaluation and inference and was not involved in model training. Three metrics are used for test set evaluation: mean squared error (MSE), root mean squared error (RMSE), and mean absolute error (MAE).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{2}$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{3}$$

where y_i is the ith value in the actual PDOS, $\hat{y}_i$ is the ith value in the predicted PDOS, and n is the number of data points, which is $4 \times 201 = 804$.

The band center $\varepsilon_{\text{center}}$ and bandwidth $\varepsilon_{\text{width}}$ are calculated using the following formula.[25,26] The lower and upper limits of integration are -5 and 5 eV, respectively.
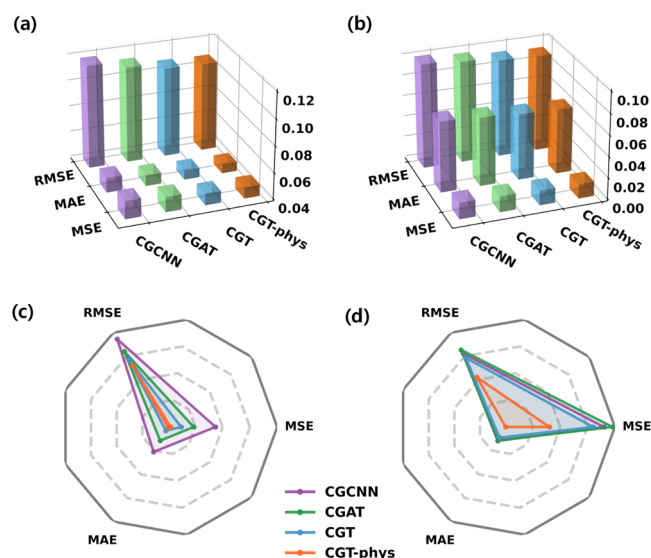
$$\varepsilon_{\text{center}} = \frac{\int_{-\infty}^{\infty} n(\varepsilon) \varepsilon d\varepsilon}{\int_{-\infty}^{\infty} n(\varepsilon) d\varepsilon} \tag{4}$$

$$\varepsilon_{\text{width}} = \frac{\int_{-\infty}^{\infty} n(\varepsilon) |\varepsilon - \varepsilon_{\text{center}}| d\varepsilon}{\int_{-\infty}^{\infty} n(\varepsilon) d\varepsilon} \tag{5}$$

where $n(\varepsilon)$ is the density of states and ε is the energy level.

3. RESULTS AND DISCUSSION

3.1. Model Comparison. We trained the CGCNN, CGAT, and CGT models using the same hyperparameters and subsequently evaluated their performance on the test set. As shown in Figure 2a and Table 1, the prediction accuracy for

Figure 2. Evaluation metrics of each model on the test set. (a) Bar chart: prediction error metrics for the PDOS; (b) bar chart: error metrics are for the p-band; (c) radar chart: error metrics for the p-band center. The innermost layer represents 0.45, and the outermost layer represents 0.80; (d) Radar chart: Error metrics for the d-band center. The innermost layer represents 0.60, and the outermost layer represents 1.20.

PDOS follows the following order: CGT > CGAT > CGCNN, with the distribution of RMSE values for these models' prediction errors shown in Supporting Information Figure S3. This suggests that the CGT model outperforms the other two in terms of accuracy. Notably, almost all materials have the p-band, and recent studies have highlighted that the properties of an increasing number of materials are closely linked to the p-band.

Table 1. Detailed Numerical Values of the Prediction Error Metrics for PDOS and p-Band on the Test Set

	PDOS			p-band		
	MSE	RMSE	MAE	MSE	RMSE	MAE
CGCNN	0.0536	0.1185	0.0512	0.0155	0.0995	0.0671
CGAT	0.0515	0.1133	0.0483	0.0150	0.0969	0.0652
CGT	0.0489	0.1085	0.0471	0.0140	0.0932	0.0628
CGT-phys	0.0486	0.1078	0.0469	0.0137	0.0922	0.0618

Table 2. Detailed Numerical Values of the Error Metrics for p-Band Center and d-Band Center of Predicted PDOS on the Test Set

	p-band center			d-band center		
	MSE	RMSE	MAE	MSE	RMSE	MAE
CGCNN	0.6023	0.7761	0.5424	1.1401	1.0678	0.6882
CGAT	0.5325	0.7297	0.5006	1.1878	1.0899	0.6848
CGT	0.4919	0.7014	0.4641	1.0813	1.0399	0.6675
CGT-phys	0.4583	0.6770	0.4506	0.8399	0.9165	0.6000
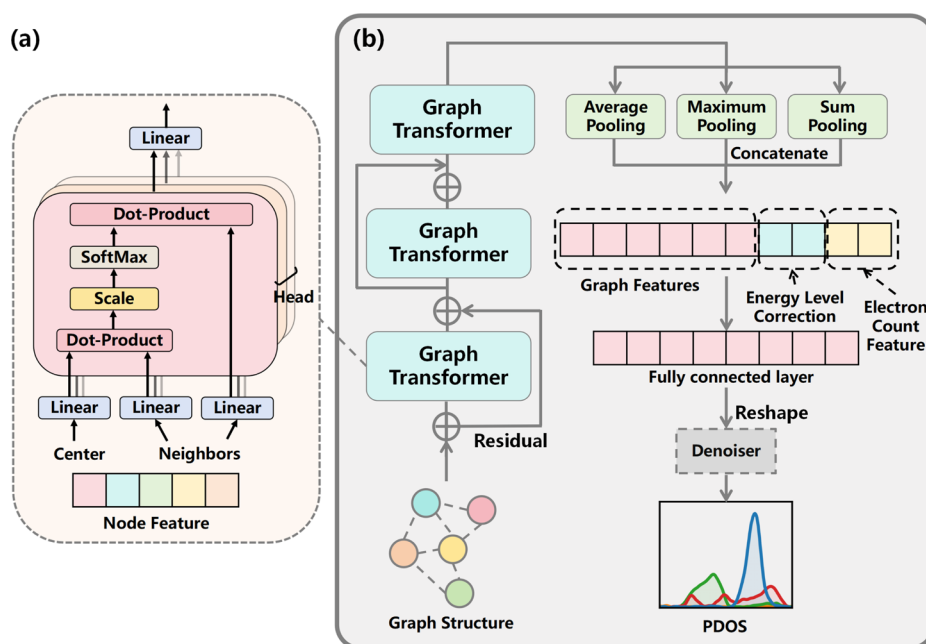
Figure 3. Model Schematic Diagram of CGT-phys. (a) Multi-head scaled dot-product attention (MHSDPA) mechanism and (b) main architecture of the model.

In light of this, we additionally chose to extract the prediction results for the p-band specifically and calculated the corresponding model evaluation metrics. As shown in Figure 2b and Table 1, the prediction accuracy for the p-band DOS follows a consistent ranking: CGT > CGAT > CGCNN. Additionally, Figure 2c,d presents radar plots illustrating the prediction errors for the p-band and d-band centers, respectively, with detailed numerical values available in Table 2. These data further support the observed performance trend, although the comparison of MSE and RMSE for the d-band center between CGCNN and CGAT shows a slight deviation from the expected pattern. The above results indicate that the GAT-based model with learnable weight aggregation outperforms the GCN-based model, which lacks such a mechanism, in predicting the PDOS. Notably, the CGT model, which integrates a multi-head scaled dot-product attention (MHSDPA) mechanism,[13] consistently shows superior performance over GAT, especially in the evaluation of band center errors of the predicted PDOS. The mechanism described above originates from the multi-head self-attention mechanism

in natural language processing (NLP), where it performs parallel computations of keys, queries, and values across multiple heads using a scaled dot-product attention function. This approach has achieved remarkable success in various NLP tasks[27−29] and serves as a core component of the Transformer, a well-established architecture in machine learning. Recognizing the successes of Transformer in NLP, researchers have extended this architecture to graph-structured data, resulting in the GT.[14,15] However, unlike the traditional multi-head self-attention mechanism, the MHSDPA in the GT model functions as a localized dot-product attention rather than a global attention mechanism. It (Figure 3a) utilizes the center node features, neighbor node features, and neighbor node features as queries, keys, and values, respectively. It computes the query, key, and value matrices in parallel across multiple attention heads. The dot product of the query and key matrices is scaled and passed through the Softmax function to generate attention scores. These scores are then used to weight the value matrix, and the weighted values are combined to produce the final output. The
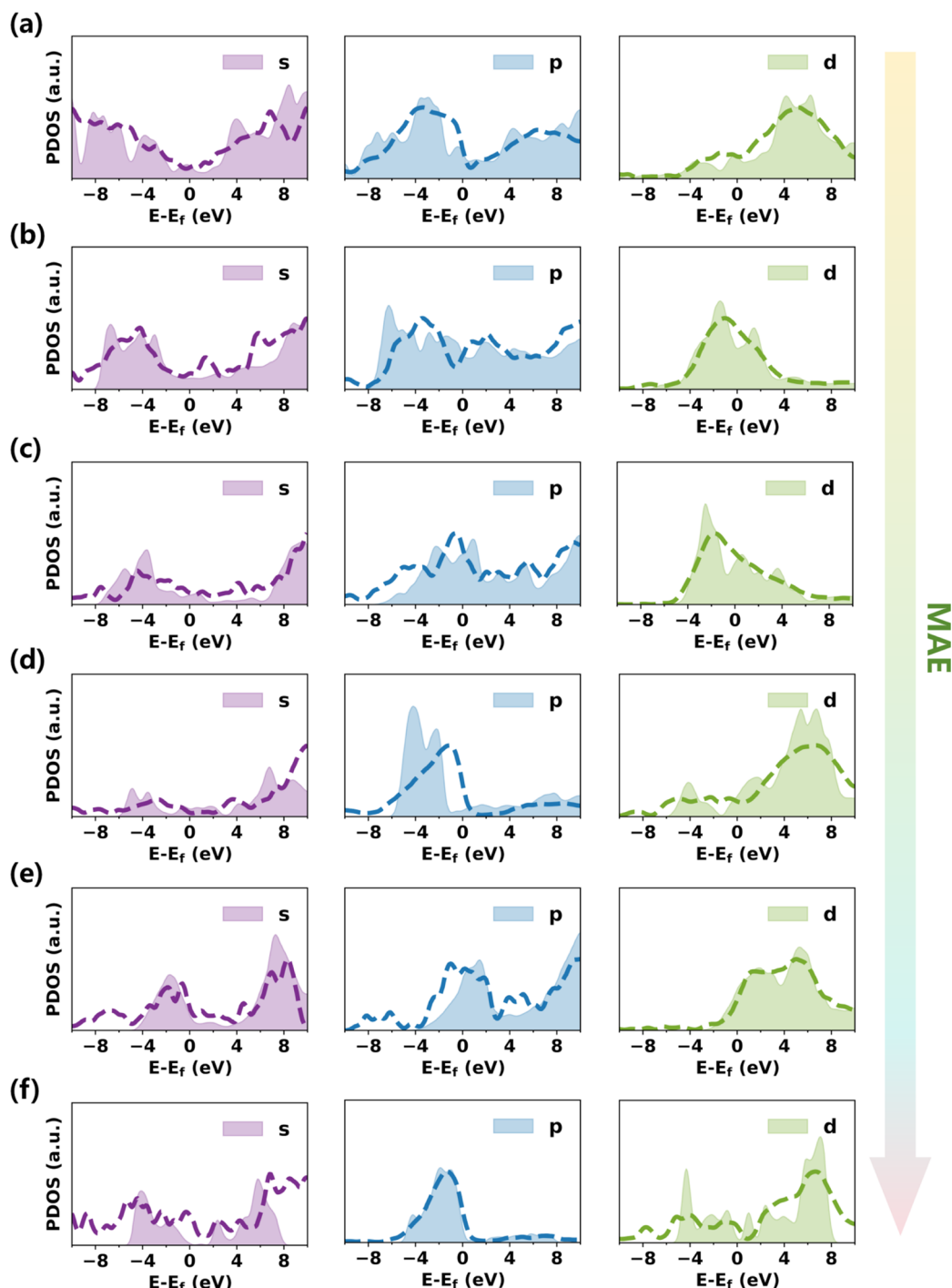
Figure 4. Prediction examples of CGT-phys on the test set. (a) mp-1189682, formula: LuB_2C, MAE: 0.0166; (b) mp-1220094, formula: $NdSmFe_{17}N_3$, MAE: 0.0324; (c) mp-1215264, formula: $ZrScVNi_3$, MAE: 0.0418; (d) mp-1187277, formula: $Tb_2Y_2O_5$, MAE: 0.0486; (e) mp-1184402, formula: Eu_3Dy, MAE: 0.0576; (f) mp-754261, formula: La_4FeO_8, MAE: 0.0710.

application of this dot-product attention mechanism in the MHSDPA enables the GT model to more effectively capture the interactions between the center node and its neighboring nodes, thus allowing it to learn the chemical structure features with greater efficiency and precision.

Building on this, we further aimed to optimize the model by incorporating richer physical features. After the graph layer output and pooling steps, we added a concatenation of the encoded features representing the electron counts of the s, p, d, and f orbitals across the entire structure (Figure 3b, electron

count feature). This step was intended to explicitly enhance the model's understanding of the structural electronic features. The resulting model, referred to as CGT-phys (Figure 3), was evaluated to assess the impact of these physical features on the model's performance. As shown in Figure 2, Tables 1 and 2, we observed that CGT-phys, which integrates physical features, slightly outperforms the version of the model without these additional physical features. This finding suggests that the inclusion of physical features is not only meaningful but also

improves the performance of the existing model framework, especially in specific application scenarios.

Finally, we extended our evaluation to assess the performance of these four models on the training set (Supporting Information Figures S4, S5, Tables S1 and S2) and validation set (Supporting Information Tables S3 and S4). The results on these sets followed the same trend as observed in the test set: CGT-phys > CGT > CGAT > CGCNN. This consistency across different data sets further supports the robustness of the CGT-phys model and its enhanced performance over the others.

3.2. Crystal Graph Transformer Model (Integrating Physical Features). Figure 3 shows the schematic diagram of the CGT-phys model. The model hyperparameters and details are described in Section 2.2. The difference from the CGT model is that the penultimate fully connected layer has an additional electron number feature embedded (the light yellow one) connected. Projection weights of query, key, and value in the MHSDPA (Figure 3a) module of the first Graph Transformer layer are shown in Supporting Information Figure S6. The computational logic of MHSDPA is as follows:

$$\mathbf{Q}_i^{(h)} = \mathbf{W}_Q^{(h)}\mathbf{x}_i$$

$$\mathbf{K}_j^{(h)} = \mathbf{W}_K^{(h)}\mathbf{x}_j + \mathbf{W}_E^{(h)}\mathbf{e}_{ij}$$

$$\mathbf{V}_j^{(h)} = \mathbf{W}_V^{(h)}\mathbf{x}_j + \mathbf{W}_E^{(h)}\mathbf{e}_{ij}$$

$$\alpha_{ij}^{(h)} = \text{softmax}_j\left(\frac{\mathbf{Q}_i^{(h)}\mathbf{K}_j^{(h)T}}{\sqrt{C}}\right)$$

$$\mathbf{m}_i^{(h)} = \sum_{j \in \mathcal{N}(i)} \alpha_{ij}^{(h)}\mathbf{V}_j^{(h)}$$

$$\mathbf{m}_i = \frac{1}{H}\sum_{h=1}^{H}\mathbf{m}_i^{(h)}$$

$$\mathbf{x}_i' = \mathbf{W}_S\mathbf{x}_i + \mathbf{m}_i$$

Definition of variables:

$x_i \in R^{256}$ (Input feature of center node i)

$x_j \in R^{256}$ (Input feature of neighbor node j)

$e_{ij} \in R^{256}$ (Input edge feature between nodes i and j)

$Q_i^{(h)}, K_j^{(h)}, V_j^{(h)} \in R^{256}$ (Projected query, key, value for head h)

$\alpha_{ij}^{(h)} \in R$ (Attention coefficient between node i and j for head h)

$m_i^{(h)} \in R^{256}$ (Aggregated message for head h)

$m_i \in R^{256}$ (Mean of all head messages)

$W_Q^{(h)}, W_K^{(h)}, W_V^{(h)}, W_E^{(h)} \in R^{256 \times 256}$
     (Trainable linear transformation matrices per head)

$W_S \in R^{256 \times 256}$ (Skip connection weight matrix)

3.3. Model Prediction Examples. We randomly selected six samples from the test set of the CGT-phys model, ordered by ascending PDOS prediction error (MAE), and sequentially displayed their DFT prediction results for a clearer comparison (with the filled regions representing DFT-calculated data, and the dashed lines indicating model predictions). Each of these samples has particular attention paid to the s, p, and d orbitals. Among the selected samples, those with better prediction performance (as shown in Figure 4a,b) exhibit good consistency between the predicted and actual values for the peak positions and heights across all three orbitals. These results underscore the model's potential for accurately predicting the PDOS when the underlying electronic structure is well understood.

However, in Figure 4c,d, a noticeable decline in prediction accuracy can be observed, particularly with respect to the peak heights of the d-band and peak positions of the p-band. Turning to Figure 4e,f, the prediction metrics for these two samples significantly deviate from the average MAE. In these cases, the model erroneously predicts the presence of peaks that do not exist in reality, such as the s orbitals in both samples and the d orbitals in La_4FeO_8.

The above results suggest that the interactions between different space groups and elements are complex, particularly with the more intricate electronic states found in higher-period elements, which often involve more sophisticated interactions and are prone to producing smaller, less pronounced peaks. This added complexity inherently challenges the model's ability to generalize across a broader range of materials, as it may struggle to accurately differentiate and predict these minor peaks. The discrepancies in these cases could also be attributed to the complexity of the computational parameters used in the DFT calculations, including the choice of the basis set and the application of DFT+U, which can introduce additional challenges. Nevertheless, the predictions for peak heights and the shapes of the main peaks remain within a reasonable range, demonstrating that the model is still capable of capturing key features of the electronic structure, even when finer details become difficult to predict.

Looking ahead, we believe that further refinement of the model's structural descriptors or the incorporation of additional local structural features may enhance its predictive performance. These adjustments might help the model better learn the geometrical structure, potentially improving its ability to generalize across a broader range of compounds and configurations.

Overall, the prediction examples from the test set suggest that the CGT-phys model demonstrates a reasonable degree of accuracy in predicting the PDOS for most samples. Compared to traditional DFT methods (see Supporting Information Figure S8), the CGT-phys model significantly improves the efficiency of PDOS acquisition, particularly for large data sets, representing a meaningful research direction. The ability to accelerate PDOS predictions while maintaining a good level of accuracy is a significant step forward and could lead to broader applications in material science.

3.4. Downstream Information of PDOS Predicted by the Model. After obtaining the PDOS of the materials, it is common practice to calculate downstream physical descriptors, such as the d band center,[25,26] d-bandwidth,[30] p-band center[31,32] and p-bandwidth to evaluate or interpret adsorption properties,[33] electrical conductivity,[34] catalytic activity,[35,36] and optical properties.[37] In this study, we used the CGT-phys model to predict the PDOS for all samples in the test set. Subsequently, we calculated the d-band center and d-bandwidth for samples containing transition metals and the p-band center and p-
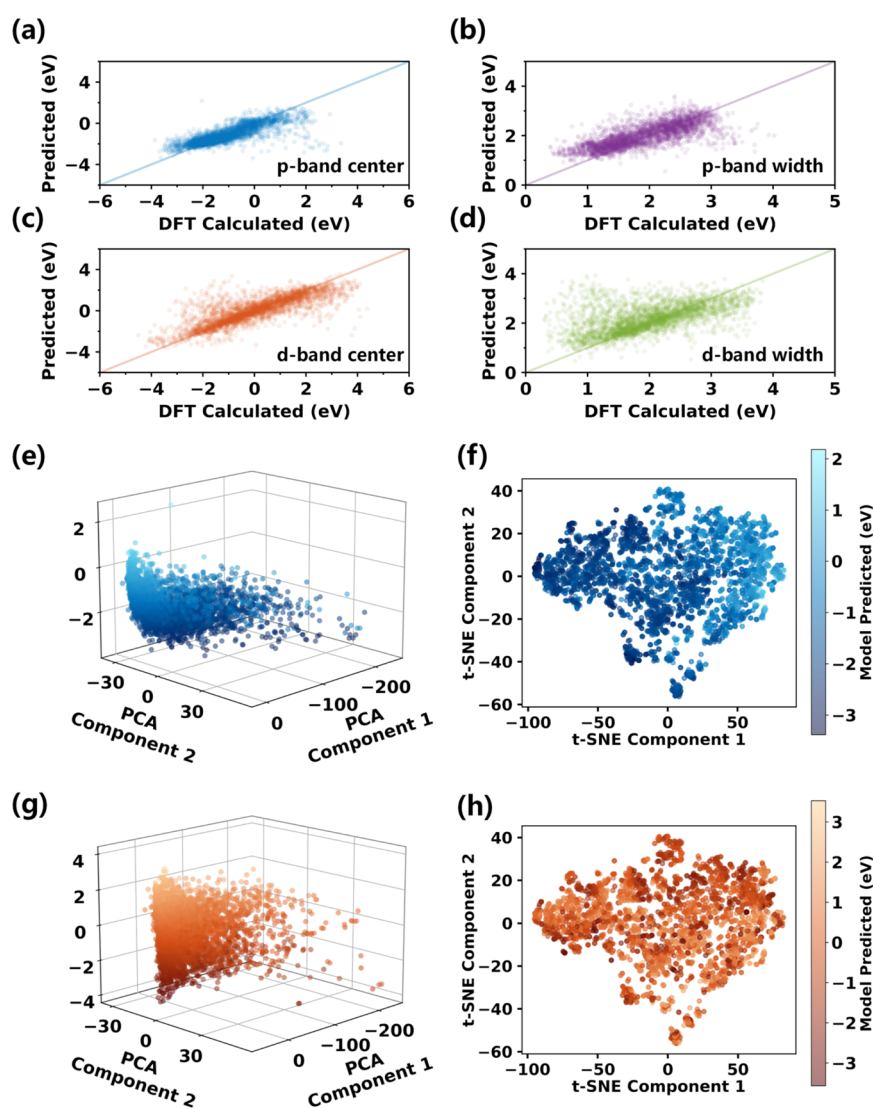
Figure 5. Scatter Plots PCA Analysis and t-SNE Analysis of Downstream Properties for PDOS Predicted by the CGT-phys, on the Test Set. (a) Predicted vs theoretical p-band center for all test set samples; (b) predicted vs theoretical p-bandwidth for all test set samples; (c) predicted vs theoretical d-band center for test set samples containing transition metal elements; (d) predicted vs theoretical d-bandwidth for test set samples containing transition metal elements; (e) PCA visualization of fully connected layer features, colored by the predicted p-band center; (f) t-SNE visualization of fully connected layer features, colored by the predicted p-band center; (g) PCA visualization of fully connected layer features, colored by the predicted d-band center; and (h) t-SNE visualization of fully connected layer features, colored by the predicted d-band center.

bandwidth for all samples. The values computed from the predicted PDOS were then systematically compared with those from theoretical calculations, with the average error shown in Supporting Information Figure S7.

Figure 5a,c present scatter plots illustrating the relationship between the band centers of the predicted PDOS and the theoretical references. The results show that within the energy range of −4 to 0 eV, the predicted p-band centers exhibit a good linear correlation with the theoretical values. However, at energy levels above 0 eV, the model tends to underestimate the p-band centers to a certain extent. Regarding the d-band centers, although most samples align well with the ideal linear trend, noticeable deviations are observed with both overestimations and underestimations present. These deviations may be attributed to the inherent complexity of high-period transition metal electronic structures. Figure 5b,d presents scatter plots comparing the predicted PDOS bandwidths to their theoretical values. It can be observed that the model demonstrates relatively

limited generalization performance for PDOS bandwidth prediction. For the p-band, both overestimation and underestimation are significant. Moving on to the d-band, the predicted bandwidths are mostly overestimated, which may be attributed to the model incorrectly predicting additional spurious d-band peaks that do not exist in the ground truth. In future work, incorporating constraints on both band center and bandwidth in model design could be a promising direction.

To further investigate the relationship between the model's internal features and the d-band centers and p-band centers, we conducted principal component analysis (PCA)[38] and t-distributed stochastic neighbor embedding (t-SNE)[39] on the feature vectors extracted from the penultimate fully connected layer of the CGT-phys model (as shown in Figure 3b, which combines structural features, energy-level correction features, and electron count features). This layer aggregates the richest and most comprehensive set of features within the model and thus serves as a representative layer for analyzing the features

learned by the model. PCA identifies directions of maximum variance through orthogonal linear transformation, and we computed the first and second principal components. The results (Figure 5e) indicate that the p-band center exhibits a positive correlation with the first principal component: as the first principal component increases, the p-band center shifts upward. The correlation between the p-band center and the second principal component is relatively weak, showing a slightly negative correlation. Furthermore, the t-SNE visualization of the fully connected layer features projected into two dimensions (Figure 5f) reveals that, after dimensionality reduction, samples with similar p-band center values form localized clusters. t-SNE is a nonlinear dimensionality reduction technique that preserves the local structure by mapping high-dimensional data into a low-dimensional space, which makes it effective for visualizing potential cluster patterns or local neighborhood structures in the data. Specifically, samples with lower p-band centers (darker colors) are predominantly located on the left, while those with higher values (lighter colors) cluster on the right. These observations suggest that the extracted features effectively capture the underlying variations in the p-band center.

In contrast, the analysis results for the d-band centers differ significantly. As shown in Figure 5g, the d-band center exhibits no significant correlation with the first principal component of PCA and only a weak negative correlation with the second principal component. Similarly, the t-SNE visualization (Figure 5h) does not reveal a strong clustering pattern, although a slight aggregation of darker regions toward the upper part of the plot can be observed. This phenomenon likely arises from the complexity of d-band electronic structures in transition metals: samples with similar d-band centers may possess distinct d-band shapes, resulting in significant variations in the corresponding fully connected layer features. Conversely, the p-band center is mainly attributed to oxygen and certain simple nontransition metal elements, which possess fewer electrons, limited orbital types, and relatively simple chemical bonding configurations. As a result, their electronic structures are inherently less diverse and complex. Under the same p-band center, the shape of the p-band and the corresponding model features tend to be more similar.

4. CONCLUSIONS

In recent years, the rapid advancement of deep learning technologies has opened new research avenues for predicting electronic structures in materials science. This study demonstrates the superior performance of the Graph Transformer (GT) framework over conventional graph neural networks (GCN and GAT) in predicting the PDOS of materials under identical energy-level-corrected conditions. The key advantage of GT lies in its multi-head scaled dot-product attention mechanism, which utilizes dot-product-based interactions to capture the relationships between the center node and its neighboring nodes. Furthermore, by integrating physics-informed features—the valence electron counts of s, p, d, and f orbitals for the entire structure—the model achieves significantly improved prediction of PDOS and its downstream information accuracy.

This work pioneers the application of the GT architecture to PDOS prediction and establishes its superiority over mainstream GNNs, offering a novel technical pathway for predicting electronic properties in materials science. It also highlights the potential of integrating physicochemical features (e.g., electronic orbital characteristics) into deep learning frameworks.

These strategies not only enhance model interpretability but also overcome performance bottlenecks of purely data-driven approaches in complex scenarios. Future research should focus on embedding richer physical principles into the framework design to advance next-generation models that balance high accuracy, generalizability, and efficiency.

■ ASSOCIATED CONTENT

Data Availability Statement

All PDOS and structural data are from the open-source database Materials Project. The organized data set and complete model codes for this study is available at the following GitHub repository: https://github.com/jiahao-codes/GNNs-PDOS.

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.5c03006.

> Space group distribution of the structures in the data set; loss function value during the model training; RMSE distribution for PDOS prediction errors on test set; visualization of the Q, K, V projection weights in the first GT layer's MHSDPA; evaluation metrics of each model on the training set; RMSE distribution for PDOS prediction errors on training set; errors of the band centers and band widths of PDOS predicted; comparison of the time required for DFT calculations and model; and detailed numerical values of the prediction error on the training set and validation set (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Yanchuan Guo — *Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100190, China; University of Chinese Academy of Sciences, Beijing 100190, China*; Phone: +86-10-82543583; Email: yanchuanguo@mail.ipc.ac.cn

Authors

Jiahao Wu — *Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100190, China; University of Chinese Academy of Sciences, Beijing 100190, China;* ⓞ orcid.org/0009-0000-7124-7236

Weipeng Lu — *Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100190, China*

Jingwen Wu — *Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100190, China*

Bing Zhang — *Technical Institute of Physics and Chemistry, Chinese Academy of Sciences, Beijing 100190, China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jpca.5c03006

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Kittel, C.; McEuen, P. *Introduction to solid state physics*; John Wiley & Sons: 2018.

(2) Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Zeitschrift für Physik* **1930**, *61*, 126−148.

(3) Hartree, D. R. The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods. In *Mathematical Proceedings of the Cambridge Philosophical Society*; Cambridge University Press: 1928; Vol. *24*, pp 89−110.

(4) Hartree, D. R. The wave mechanics of an atom with a non-coulomb central field. Part II. Some results and discussion. In *Mathematical Proceedings of the Cambridge Philosophical Society*; Cambridge University Press: 1928; Vol. *24*, pp 111−132.

(5) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133.

(6) Kong, S.; Ricci, F.; Guevarra, D.; Neaton, J. B.; Gomes, C. P.; Gregoire, J. M. Density of states prediction for materials discovery via contrastive learning from probabilistic embeddings. *Nat. Commun.* **2022**, *13* (1), 949.

(7) Kotobi, A.; Singh, K.; Höche, D.; Bari, S.; Meißner, R. H.; Bande, A. Integrating explainability into graph neural network models for the prediction of X-ray absorption spectra. *J. Am. Chem. Soc.* **2023**, *145* (41), 22584−22598.

(8) Fung, V.; Ganesh, P.; Sumpter, B. G. Physically informed machine learning prediction of electronic density of states. *Chem. Mater.* **2022**, *34* (11), 4848−4855.

(9) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters* **2018**, *120* (14), No. 145301.

(10) Zhao, G.; Chung, Y. G. PACMAN: A Robust Partial Atomic Charge Predicter for Nanoporous Materials Based on Crystal Graph Convolution Networks. *J. Chem. Theory Comput.* **2024**, *20*, 5368.

(11) del Rio, B. G.; Kuenneth, C.; Tran, H. D.; Ramprasad, R. An Efficient Deep Learning Scheme To Predict the Electronic Structure of Materials and Molecules: The Example of Graphene-Derived Allotropes. *J. Phys. Chem. A* **2020**, *124* (45), 9496−9502.

(12) Kiyohara, S.; Hinuma, Y.; Oba, F. Band Alignment of Oxides by Learnable Structural-Descriptor-Aided Neural Network and Transfer Learning. *J. Am. Chem. Soc.* **2024**, *146* (14), 9697−9708.

(13) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In *Advances in neural information processing systems*; Curran Associates Inc.: 2017; Vol. *30*.

(14) Yun, S.; Jeong, M.; Kim, R.; Kang, J.; Kim, H. J. Graph transformer networks. In *Advances in neural information processing systems*; Curran Associates Inc.: 2019; Vol *32*.

(15) Shi, Y.; Huang, Z.; Feng, S.; Zhong, H.; Wang, W.; Sun, Y. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509* 2020.

(16) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Uncovering electronic and geometric descriptors of chemical activity for metal alloys and oxides using unsupervised machine learning. *Chem. Catalysis* **2021**, *1* (4), 923−940.

(17) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**.

(18) Perdew, n.; Burke, n.; Ernzerhof, n. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(19) Jain, A.; Hautier, G.; Ong, S. P.; Moore, C. J.; Fischer, C. C.; Persson, K. A.; Ceder, G. Formation enthalpies by mixing GGA and GGA+ U calculations. *Phys. Rev. B* **2011**, *84* (4), No. 045115.

(20) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314−319.

(21) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*; Curran Associates Inc.: 2019; Vol. *32*.

(22) Fey, M.; Lenssen, J. E. Fast graph representation learning with PyTorch Geometric. *arXiv preprint arXiv:1903.02428* 2019.

(23) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. *nature* **1986**, *323* (6088), 533−536.

(24) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.

(25) Bligaard, T.; Nørskov, J. K.; Lundqvist, B. Understanding heterogeneous catalysis from the fundamentals. *Handbook of Surface Science* **2008**, *3*, 269−340.

(26) Nørskov, J. K.; Studt, F.; Abild-Pedersen, F.; Bligaard, T. *Fundamental concepts in heterogeneous catalysis*; John Wiley & Sons: 2014.

(27) Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*; ACL: 2019; pp 4171−4186.

(28) Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* 2019.

(29) Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. *Improving language understanding by generative pre-training*; 2018.

(30) Nørskov, J.; Studt, F.; Abild-Pedersen, F.; Bligaard, T.The electronic factor in heterogeneous catalysis. In *Fundamental concepts in heterogeneous catalysis*; Wiley: 2014; pp 114−137.

(31) Grimaud, A.; May, K. J.; Carlton, C. E.; Lee, Y.-L.; Risch, M.; Hong, W. T.; Zhou, J.; Shao-Horn, Y. Double perovskites as a family of highly active catalysts for oxygen evolution in alkaline solution. *Nat. Commun.* **2013**, *4* (1), 2439.

(32) Xiao, K.; Wang, Y.; Wu, P.; Hou, L.; Liu, Z. Q. Activating lattice oxygen in spinel ZnCo2O4 through filling oxygen vacancies with fluorine for electrocatalytic oxygen evolution. *Angew. Chem., Int. Ed.* **2023**, *62* (24), No. e202301408.

(33) Hammer, B.; Nørskov, J. K. Theoretical surface science and catalysis—calculations and concepts. In *Advances in catalysis*; Elsevier: 2000; Vol. *45*, pp 71−129.

(34) Greeley, J.; Nørskov, J. K. Combinatorial density functional theory-based screening of surface alloys for the oxygen reduction reaction. *J. Phys. Chem. C* **2009**, *113* (12), 4932−4939.

(35) Shi, Y.; Bai, J.; Ma, A.; Zou, Z.; Liu, Y.; Lin, H.; Chen, D.; Zhang, R.; Lai, J.; Wang, L. Ordered 2D/3D CoNi-LDH/FeSe electrode for efficient and durable seawater oxidation. *Chemical Engineering Journal* **2024**, *498*, No. 155834.

(36) Wu, J.; Zhang, X.; Yu, F.; Tian, M.; Wang, Y.; Wu, J.; Lu, W.; Guo, Y. Exploration of cobalt-based spinel oxide nanocatalysts MCo2O4 (M = Mn, Fe, Co, Ni, Cu, Zn) for glucose electrochemical sensing: NiCo2O4 exhibits largest Faradaic current. *Chemical Engineering Journal* **2024**, *499*, No. 156011.

(37) Li, Y.-F.; Liu, Z.-P.; Liu, L.; Gao, W. Mechanism and activity of photocatalytic oxygen evolution on titania anatase in aqueous surroundings. *J. Am. Chem. Soc.* **2010**, *132* (37), 13008−13015.

(38) Maćkiewicz, A.; Ratajczak, W. Principal components analysis (PCA). *Computers & Geosciences* **1993**, *19* (3), 303−342.

(39) Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9* (2605), 2579.

■ NOTE ADDED AFTER ASAP PUBLICATION

This paper was published ASAP on June 12, 2025, with errors in Tables 1 and 2 and the Supporting Information. The corrected version was reposted on June 16, 2025.