The Journal of Chemical Physics ARTICLE pubs.aip.org/aip/jcp

Enhancing the prediction of TADF emitter properties using Δ-machine learning: A hybrid semi-empirical and deep tensor neural network approach

Cite as: J. Chem. Phys. 162, 144103 (2025); doi: 10.1063/5.0263384
Submitted: 5 February 2025 • Accepted: 16 March 2025 •
Published Online: 8 April 2025

Nikhitha R. and Anirban Mondal[a)]

AFFILIATIONS

Department of Chemistry, Indian Institute of Technology, Gandhinagar, Gujarat 382355, India

[a)]Author to whom correspondence should be addressed: amondal@iitgn.ac.in

ABSTRACT

This study presents a machine learning (ML)-augmented framework for accurately predicting excited-state properties critical to thermally activated delayed fluorescence (TADF) emitters. By integrating the computational efficiency of semi-empirical PPP+CIS theory with a Δ-ML approach, the model overcomes the inherent limitations of PPP+CIS in predicting key properties, including singlet (S_1) and triplet (T_1) energies, singlet–triplet gaps (ΔE_{ST}), and oscillator strength (f). The model demonstrated exceptional accuracy across datasets of varying sizes and diverse molecular features, notably excelling in predicting oscillator strength and ΔE_{ST} values, including negative regions relevant to TADF molecules with inverted S_1–T_1 gaps. This work highlights the synergy between physics-inspired models and machine learning in accelerating the design of efficient TADF emitters, providing a foundation for future studies on complex systems and advanced functional materials.

I. INTRODUCTION

Advances in computational chemistry have revolutionized the numerical determination of molecular properties, offering unprecedented insights into chemical systems.[1,2] However, the vast chemical space remains challenging to explore due to the high computational costs associated with accurate quantum mechanical (QM) methods, particularly for large molecular systems.[3,4] While molecular mechanics offers computational efficiency, it lacks the accuracy and transferability required for electronic property predictions.[5] Semiempirical QM (SQM) methods, such as the Pariser–Parr–Pople (PPP) model,[6–8] strike a balance between accuracy and efficiency by approximating the interaction between electrons at a reduced computational cost.[9] The PPP model, for instance, provides a simplified yet insightful description of π electrons in planar conjugated molecules, accounting for interelectronic interactions within the zero differential overlap (ZDO) approximation.[9] These intermediate methods are particularly valuable for systems requiring electronic property predictions beyond the reach of molecular mechanics.

Machine learning (ML) has emerged as a transformative tool in this domain, enabling the efficient prediction of molecular properties with accuracy comparable with QM methods at a fraction of the cost.[10,11] Atomistic ML models are now integral to exploring complex biology, chemistry, and materials science systems.[5] These models facilitate large-scale dynamic simulations[12] and high-throughput screening,[13] significantly reducing the computational burden of conventional density functional theory (DFT) calculations. By training on representative datasets, ML models predict properties for new systems with remarkable efficiency, democratizing access to computationally intensive analyses.[14]

Thermally activated delayed fluorescence (TADF) emitters exemplify the need for accurate electronic property predictions. These materials achieve up to 100% internal quantum efficiencies by up-converting non-radiative triplet excitons into radiative singlet excitons through reverse intersystem crossing (RISC).[15,16] Efficient TADF emitters require a minimal singlet–triplet energy gap (ΔE_{ST}) to facilitate RISC, along with strong oscillator strength (f) for effective light emission.[17] A recently discovered class of TADF emitters,

known as INVEST (inverted singlet–triplet gap) molecules, challenges conventional paradigms by exhibiting a negative ΔE_{ST}, with the singlet state stabilized below the triplet state.[18] Accurately modeling such systems necessitates the inclusion of double-excitation configurations,[19–24] significantly increasing the computational cost of QM methods. This limitation underscores the critical need for efficient and accurate approaches to predict the electronic properties of TADF emitters.

The Δ-ML framework addresses this challenge by leveraging a lower-cost baseline, such as DFT or SQM methods, to calibrate predictions to the accuracy of high-level QM methods such as coupled-cluster or experimental results.[25,26] Recent studies have highlighted the potential of SQM methods as effective baselines for Δ-ML models.[26,27] For example, Jorner *et al.* demonstrated that a combination of configuration interaction singles (CIS) and dynamic spin polarization (DSP) could partially describe the inverted singlet–triplet gap in INVEST molecules.[28] Although the PPP+CIS+DSP theory has been shown to be useful for screening INVEST molecules, its poor correlation with Algebraic Diagrammatic Construction [ADC(2)] values for properties such as oscillator strength is primarily due to the method's inability to capture wavefunction-dependent quantities accurately.[28] This limitation arises because PPP+CIS only corrects energies to the second order without corresponding corrections to the wavefunction. In this work, we address this limitation by combining the computational efficiency of the PPP+CIS method with the established accuracy of ADC(2) predictions relative to higher-level methods through the Δ-ML framework, which effectively corrects systematic deficiencies in the low-level method to align with high-level reference results.

Building on this foundation, our study employs the PPP+CIS method as the baseline for predicting electronic properties relevant to TADF emitters. High-level QM calculations, specifically ADC(2), serve as the target data. Using the Δ-ML approach implemented in the deep tensor neural network (DTNN) framework, SchNet,[29] we validated four key properties of TADF emitters: singlet energies (S_1), triplet energies (T_1), singlet–triplet gap (ΔE_{ST}), and oscillator strength (f). Our results demonstrate that this hybrid approach, leveraging PPP+CIS as the baseline and SchNet for machine learning, achieves high accuracy for key TADF-relevant properties, including S_1, T_1, ΔE_{ST}, and f. In addition, we validate the pre-trained model on the benchmark INVEST15 dataset[30] to showcase the transferability and generalizability of the model. Furthermore, this framework can also be employed to screen TADF molecules efficiently, enabling the identification of TADF candidates with desirable properties. To our knowledge, this is the first application of such a framework to TADF-relevant properties in the context of OLED materials, paving the way for cost-effective and accurate predictions that were previously unattainable. This study not only validates the Δ-ML framework for modeling TADF emitters but also establishes a robust platform for the efficient screening and design of novel materials with desirable electronic characteristics.

II. METHOD

To implement the Δ-ML framework for TADF-relevant properties, we utilized two distinct datasets, referred to as Dataset-I and Dataset-II. Dataset-I was derived from a more extensive collection of 68 695 rationally designed compounds, whose geometries were optimized at the B97-3c level, with excitation energies computed at the second-order algebraic diagrammatic construction, ADC(2)/cc-pVDZ level.[31] From this extensive dataset, we selected a subset of 12 000 molecules to form Dataset-I, ensuring a diverse representation of molecular structures and electronic properties. We chose this dataset size to reduce the computational cost, as the training was performed on a single NVIDIA RTX A6000 GPU. This ensured efficient utilization of the computational resources while also showcasing the framework's capabilities. Dataset-II, on the other hand, comprises 250 substituted azaphenalenes, with geometries optimized at the B3LYP/cc-pVDZ level and excitation energies calculated using ADC(2)/cc-pVDZ.[32] These two datasets provide a robust foundation for training and validating the machine-learning models in this study.

We adopted SchNet,[29] a deep learning architecture optimized for atomistic systems, to implement the Δ-ML framework. SchNet, as a variant of deep tensor neural networks (DTNNs), inherently respects the fundamental symmetries of atomistic systems, such as rotational and translational invariance, as well as invariance to the ordering of atom indices.[29] Its architecture utilizes continuous-filter convolutional layers to effectively capture both spatial and chemical interactions.[33] The model represents each atom as a feature vector $\mathbf{x}_i^{(l)} \in \mathbb{R}^D$, where D denotes the feature space dimension and l represents the layer in the network. Atomic interactions are iteratively updated T times through pairwise interactions between feature vectors of atoms within a defined cutoff distance, incorporating information about the chemical environment and complex many-body interactions. The continuous-filter convolution layers, facilitated by filter-generating networks, refine these feature representations. Finally, the model pools atom-wise updates to predict global molecular properties, ensuring an accurate and efficient mapping of structure–property relationships.[5]

The machine learning pipeline was built using a combination of well-established tools and frameworks. Data handling and preprocessing were conducted using NumPy[34] and Pandas,[35] which facilitated the organization and manipulation of molecular property datasets. For data splitting and cross-validation, we utilized scikit-learn's `train_test_split` and `KFold` modules, enabling robust performance validation.[36] PyTorch[37] served as the core deep learning framework, while SchNetPack[14,38] provided pre-built modules for atomic structure representation and property prediction. The Atomic Simulation Environment (ASE)[39] was employed to handle atomic configurations, including reading and manipulating XYZ files. Model performance was evaluated using TorchMetrics, which provided metrics such as mean absolute error (MAE) and root mean squared error (RMSE). These results were visualized and monitored through TensorBoard,[40] ensuring a comprehensive evaluation of the model's predictive accuracy and training progression.

The model was implemented using SchNetPack 2.0,[14,38] a framework specifically designed for atomistic simulations, to accurately predict key TADF descriptors: excited singlet energies (S_1), excited triplet energies (T_1), singlet–triplet gap (ΔE_{ST}), and oscillator strength (f). SchNet's continuous-filter convolutional layers were utilized to model atomic interactions and effectively capture the chemical environment. To ensure compatibility with SchNetPack, the molecular dataset and the corresponding target properties, calculated using the baseline PPP+CIS method, were converted

into the Atomic Simulation Environment database (`ASE db`) format.

The Δ-ML framework was designed to enhance property prediction accuracy by incorporating a correction factor that accounts for the discrepancy between the baseline PPP+CIS calculations and the higher-level ADC(2) results. This correction factor expressed mathematically in Eq. (1) served as the target property for the machine learning model. By learning this correction, the model bridges the gap between the computational efficiency of PPP+CIS and the predictive accuracy of ADC(2). SchNetPack was customized to predict this correction factor using molecular coordinates and atomic numbers as input features. The final property prediction was obtained by adding the predicted correction factor to the baseline PPP+CIS value, as shown in Eq. (2). This two-step process ensures that the computational efficiency of PPP+CIS is retained while achieving the precision of ADC(2) calculations,

$$\Delta E_{\text{correction-factor}} = E_{\text{ADC(2)}} - E_{\text{PPP+CIS}}, \tag{1}$$

$$E^{\text{pred}}_{\text{ADC(2)}} = E_{\text{PPP+CIS}} + \Delta E^{\text{pred}}_{\text{correction-factor}}. \tag{2}$$

The training process involved splitting the dataset into training, validation, and test sets, with fivefold cross-validation used to identify the best model and ensure robustness. Hyperparameter optimization was conducted using Optuna,[41] an efficient framework for automated hyperparameter tuning. Key hyperparameters were systematically optimized to achieve optimal performance, including learning rate, batch size, number of convolutional layers, radial basis functions, atom basis size, patience limits, and cutoff radius. An early stopping mechanism was implemented to monitor validation loss and halt training when overfitting was detected. The AdamW optimizer was employed for stable and efficient weight updates during training, with the optimization objective being to minimize the validation loss, thereby ensuring the model's generalizability. By integrating SchNetPack 2.0 with Optuna, the training process was tailored to predict all four properties accurately. This approach demonstrates the robustness and adaptability of the Δ-ML framework for modeling TADF-related electronic properties, combining computational efficiency with predictive precision.

III. RESULTS

A. Dataset overview and characteristics

To evaluate the robustness and adaptability of our approach, we analyzed two datasets of markedly different sizes, as shown in Fig. 1. Dataset-I contains 12 000 molecules, while Dataset-II comprises only 250 molecules. Despite this stark contrast, our method demonstrated consistent performance, highlighting its invariance to dataset size and composition (*vide infra*). Both datasets were curated to include potential INVEST candidates and consisted solely of carbon, hydrogen, and nitrogen atoms [Figs. 1(c)–1(g)]. However, notable differences in molecular characteristics emerge upon closer examination. The molecular weight distribution [Fig. 1(a)] in Dataset-I is broader and peaks at higher values (around 250–300 g/mol), whereas Dataset-II [Fig. 1(e)] exhibits a narrower range with molecular weights clustering around 180–200 g/mol. This suggests that molecules in Dataset-I are generally larger and more diverse in size compared to those in Dataset-II. The number of rings [Figs. 1(b)–1(f)] is remarkably similar across both datasets, with most molecules containing three rings. This similarity reflects a shared structural motif among the potential INVEST candidates. However, the hybridization patterns reveal a key distinction: Dataset-I is dominated by sp and sp^2 hybridized atoms

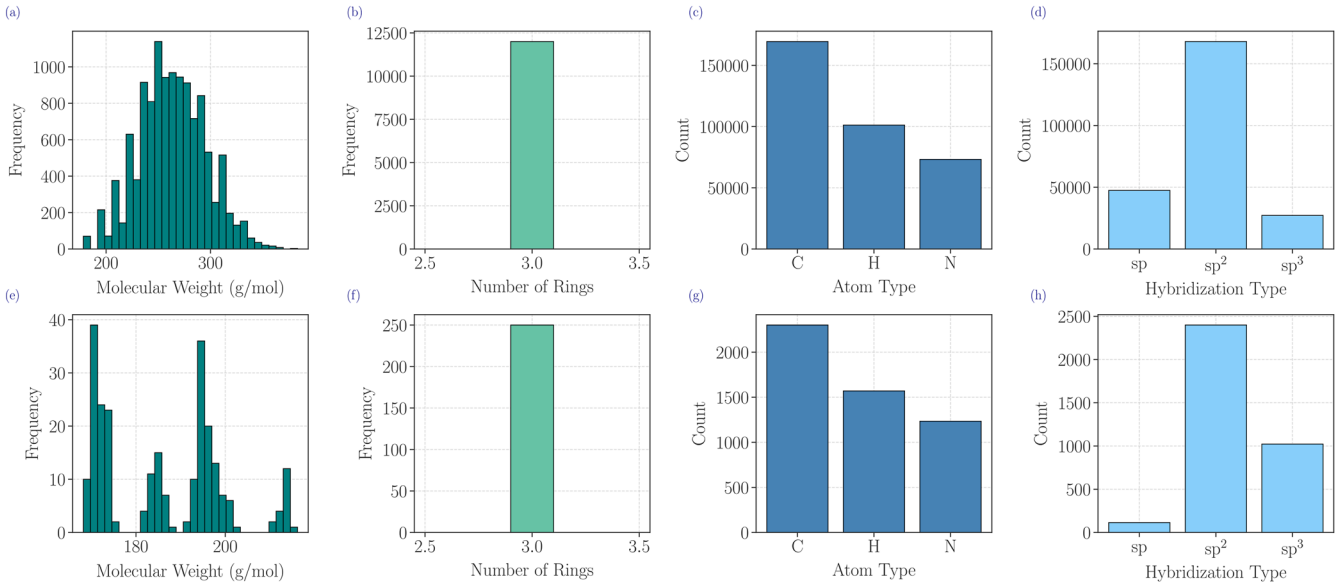

FIG. 1. Comparative analysis of Dataset-I (12 000 molecules) (Top row) and Dataset-II (250 molecules) (Bottom row) illustrates [(a) and (e)] molecular weight distribution, [(b) and (f)] number of rings, [(c) and (g)] atom types, and [(d) and (h)] hybridization types in both datasets.

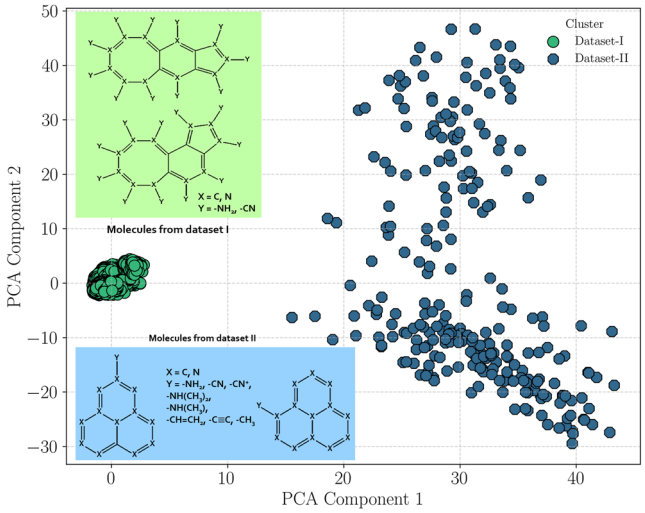

FIG. 2. Principal component analysis (PCA) plot illustrating the structural diversity of Dataset-I and Dataset-II. Representative molecular structures from each dataset are highlighted in the inset boxes.

[Fig. 1(d)], whereas Dataset-II primarily features sp^3 and sp^2 hybridized atoms [Fig. 1(h)], reflecting a greater presence of single bonds and increased structural diversity owing to varied substituents. These differences underscore the complementary nature of the two datasets. While Dataset-I provides a broader representation of larger molecules with fewer substituents, Dataset-II explores a narrower range of molecular sizes but a more diverse set of smaller molecules with varying substitutions. This comprehensive analysis demonstrates that our approach remains robust and effective across datasets with varying sizes and structural compositions.

The principal component analysis (PCA) depicted in Fig. 2 highlights significant differences in the structural diversity between Dataset-I and Dataset-II. Dataset-I exhibits a tightly clustered distribution, indicating a high degree of structural similarity among its molecules. This uniformity likely arises from the dataset's limited chemical composition, with molecules predominantly containing double and triple bonds and a narrow range of structural features. In contrast, Dataset-II demonstrates a broader distribution across the PCA components, reflecting its greater structural diversity. The molecules in this dataset incorporate a wider variety of bond types, including a larger number of single bonds, as well as more varied substitution patterns and molecular configurations. This increased heterogeneity underscores the complexity of Dataset-II, distinguishing it as a more diverse and chemically versatile collection compared to Dataset-I.

Jorner *et al.* demonstrated that while the PPP+CIS+DSP theory could be employed for screening INVEST molecules, it showed poor correlation with ADC(2) values, particularly for oscillator strengths.[28] This discrepancy is primarily due to the PPP+CIS method's lack of second-order corrections to the wavefunction, which is essential for accurate oscillator strength predictions. Similarly, our analysis reveals that PPP+CIS predictions for TADF properties exhibit weak correlations with ADC(2) values, as illustrated in Fig. 3, Figs. S1 and S2 of the supplementary material. In particular, the scatterplots highlight significant discrepancies across key properties: singlet–triplet energy gaps (ΔE_{ST}), singlet excitation energies (S_1), triplet excitation energies (T_1), and oscillator strengths (f). For instance, the R^2 values for ΔE_{ST}, f, S_1, and T_1 indicate that PPP+CIS significantly underperforms compared to ADC(2), with noticeable deviations in magnitudes. Particularly for ΔE_{ST}, PPP+CIS fails to capture the correct trend, consistently

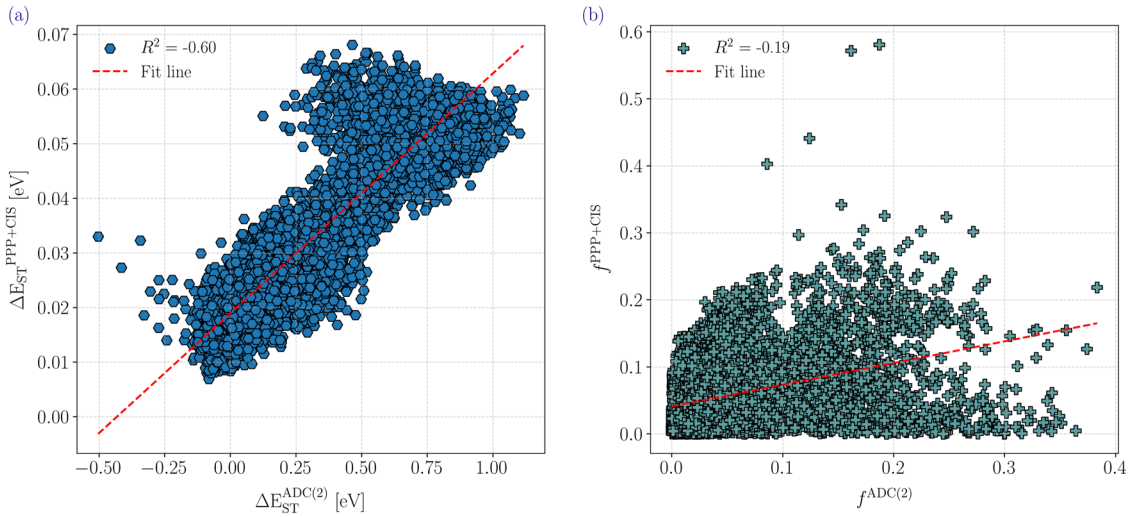

FIG. 3. Comparison of TADF excited-state properties calculated using PPP+CIS and ADC(2) for Dataset-I. The scatterplots show correlations for (a) singlet–triplet energy gap (ΔE_{ST}) and (b) oscillator strength (f). The weak correlations, as indicated by the R^2 values, highlight the limitations of PPP+CIS in accurately predicting TADF properties. The comparison for triplet energies (T_1) and singlet energies (S_1) are shown in Fig. S1. A similar comparison for Dataset-II is provided in Fig. S2.

predicting positive ΔE_{ST} values even for INVEST molecules. These findings underscore the inherent limitations of PPP+CIS in capturing the nuanced electronic structure features required for accurate TADF property predictions. The gap between PPP+CIS and ADC(2) values serves as a critical motivation for developing a Δ-ML framework. By directly learning the corrections required to bridge this gap, Δ-ML offers a robust and computationally efficient approach to improve the predictive accuracy of PPP+CIS, enabling reliable screening of TADF molecules. This approach not only addresses the shortcomings of PPP+CIS but also provides a scalable solution for large-scale molecular screening, paving the way for a more accurate and efficient design of TADF emitters.

TABLE I. Performance metrics for the machine learning model trained on Dataset-I. RMSE and MAE are reported for both the training and test datasets across the predicted properties: ΔE_{ST}, S_1, T_1, and f. The low error values indicate high model accuracy and generalizability.

Target property	RMSE (train)	MAE (train)	RMSE (test)	MAE (test)
S_1	0.0034	0.0442	0.0042	0.0488
T_1	0.0021	0.0307	0.0028	0.0407
ΔE_{ST}	0.0028	0.0419	0.0036	0.0455
f	0.0006	0.0167	0.0007	0.0162

B. Dataset I: Model training and testing outcomes

For Dataset-I, consisting of 12 000 molecules, the train–validation–test split was carefully designed to ensure a robust evaluation of the model. The training set comprised 7680 molecules; the validation set included 1920 molecules; and the remaining 2400 molecules were reserved for testing. This split was chosen to balance the need for sufficient training data with adequate samples for validation and testing, thereby enabling a thorough assessment of the model's generalization capabilities.

The dataset spans a wide range of ΔE_{ST} from −0.4 to 1.0 eV, encompassing molecules with both positive and negative singlet–triplet gaps. This diversity is critical for training a model capable of predicting ΔE_{ST} across different chemical environments. The model demonstrated high predictive accuracy, achieving an R^2 value of 0.95 on the test set, as shown in Fig. 4(b). Notably, the model performs consistently well for molecules with positive ΔE_{ST} values and those near zero and INVEST candidates, which are particularly interesting for thermally activated delayed fluorescence applications. However, a slight deviation is observed for molecules with extreme negative ΔE_{ST} values (less than −0.2 eV). This discrepancy may be attributed to the limited representation of such molecules in the training set, highlighting the potential need for targeted data augmentation in future work.

The loss metrics further underscore the model's robustness (see Table I). The RMSE for the train, validation, and test sets were 0.0028, 0.0029, and 0.0036, respectively, while the corresponding MAE values were 0.0419, 0.0414, and 0.0455. These values indicate minimal overfitting and excellent generalization, with only a slight increase in loss from the training to the testing phase. The training curves depicted in Fig. S3 show a smooth convergence for RMSE and MAE, confirming the model's stability during training. The model exhibits a strong predictive performance across the dataset, accurately capturing the trends in ΔE_{ST} for the majority of

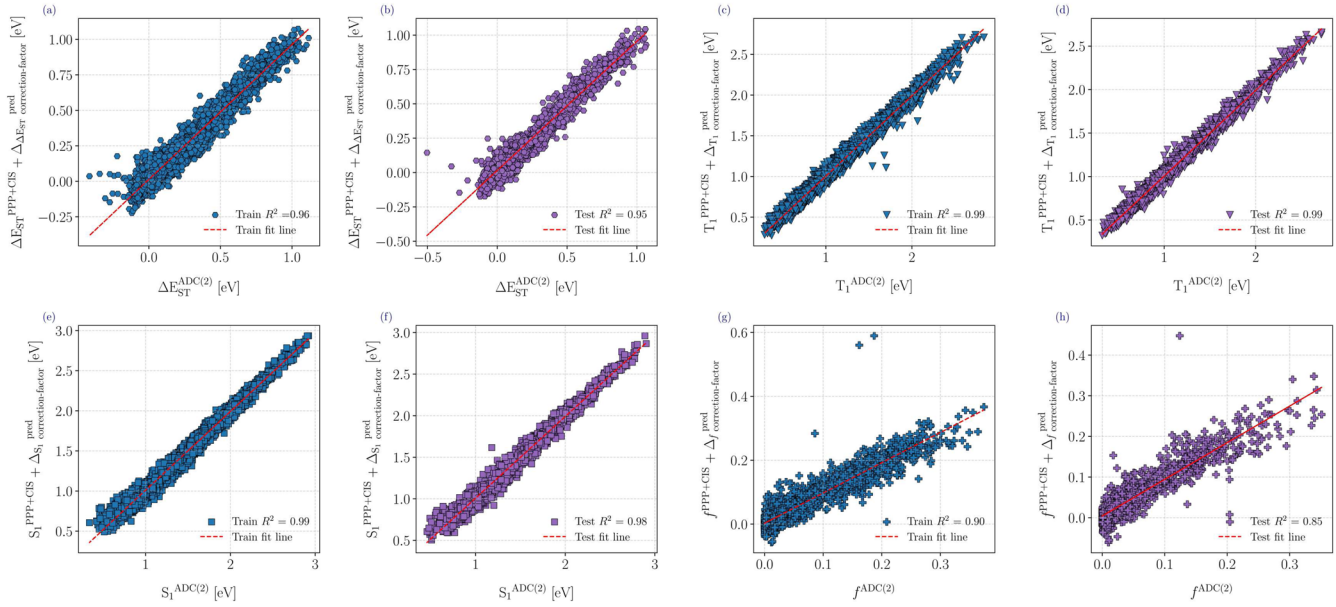

FIG. 4. Performance evaluation of the neural network model for predicting [(a) and (b)] ΔE_{ST}, [(c) and (d)] T_1 energies, [(e) and (f)] S_1 energies, and [(g) and (h)] f in the Dataset-I. Predicted correction factor added to the PPP+CIS values, compared against the true ADC(2) value, for the training set (Blue) and test set (Purple), with a consistent fit across datasets.

molecules while maintaining low error metrics. The observed deviations in extreme cases suggest opportunities for improvement by including more diverse training samples or applying transfer learning techniques to handle under-represented regions of the dataset better.

The triplet energies (T_1) range from 0.28 to 2.8 eV, covering the infrared to blue region. The model achieved exceptional accuracy in predicting triplet energies, with an R^2 value of 0.99 for both the training and test sets, as shown in Figs. 4(c) and 4(d). The predictions are particularly reliable in the visible region, a critical range for organic light-emitting diode applications. The RMSE values for the training, validation, and test sets were 0.0021, 0.0029, and 0.0028, respectively, while the MAE values were 0.0307, 0.0413, and 0.0407, as depicted in Table I. These metrics highlight the model's substantial predictive accuracy and capability to generalize effectively across the dataset.

Similarly, the singlet energies (S_1) in Dataset-I range from 0.5 to 3 eV, covering emissions from the infrared to the blue region. The model demonstrated excellent predictive performance, achieving an R^2 value of 0.99 and 0.98 for the training and test sets, respectively, as illustrated in Figs. 4(e) and 4(f). The model reliably predicts singlet energies across a broad range, including the 1.8–3 eV region, which is particularly relevant for OLED applications due to its alignment with visible-range emissions. The loss metrics are represented in Table I, where RMSE values for the training, validation, and test sets were 0.0034, 0.0043, and 0.0042, respectively, while the corresponding MAE values were 0.0442, 0.0484, and 0.0488. These results underscore the model's robustness and ability to generalize effectively, with minimal discrepancies between training and test losses, confirming the absence of overfitting.

The dataset also exhibits a wide range of oscillator strengths (f), varying from near zero to 0.35. The SchNet model was initially employed to predict oscillator strengths, achieving an R^2 value of 0.77 (Fig. S4) for the test set. Recognizing the direct correlation between oscillator strength and dipole moments, the model was refined using the polarizable atom interaction neural network (PaiNN),[42] which explicitly incorporates dipole moment calculations. This refinement led to a marked improvement in performance, with the R^2 value for the test set increasing to 0.85, as depicted in Figs. 4(g) and 4(h). Excluding significant outliers further enhanced the R^2 value to 0.87. The RMSE values for the training, validation, and test sets were 0.0006, 0.0005, and 0.0007, respectively, while the MAE values were 0.0167, 0.0157, and 0.0162, as depicted in Table I. Additional correction factor plots for ΔE_{ST}, S_1, T_1, and f, along with RMSE and MAE loss trends, are provided in Fig. S3. The R^2 values for the actual vs predicted correction factors for all properties across train and test sets are listed in Table S1. This progression of results demonstrates the model's capability to accurately predict singlet and triplet energies, as well as oscillator strengths, thereby showcasing its utility in designing and screening molecules for OLED applications.

C. Dataset II: Assessing robustness and consistency

To further validate the robustness of our approach and its adaptability to varying dataset sizes, we tested our model on a smaller dataset comprising 250 molecules (Dataset-II). This dataset primarily included substituted azaphenalenes with varied substitutions, offering a diverse yet compact set of molecular structures. The dataset was partitioned into 160 molecules for training, 40 for

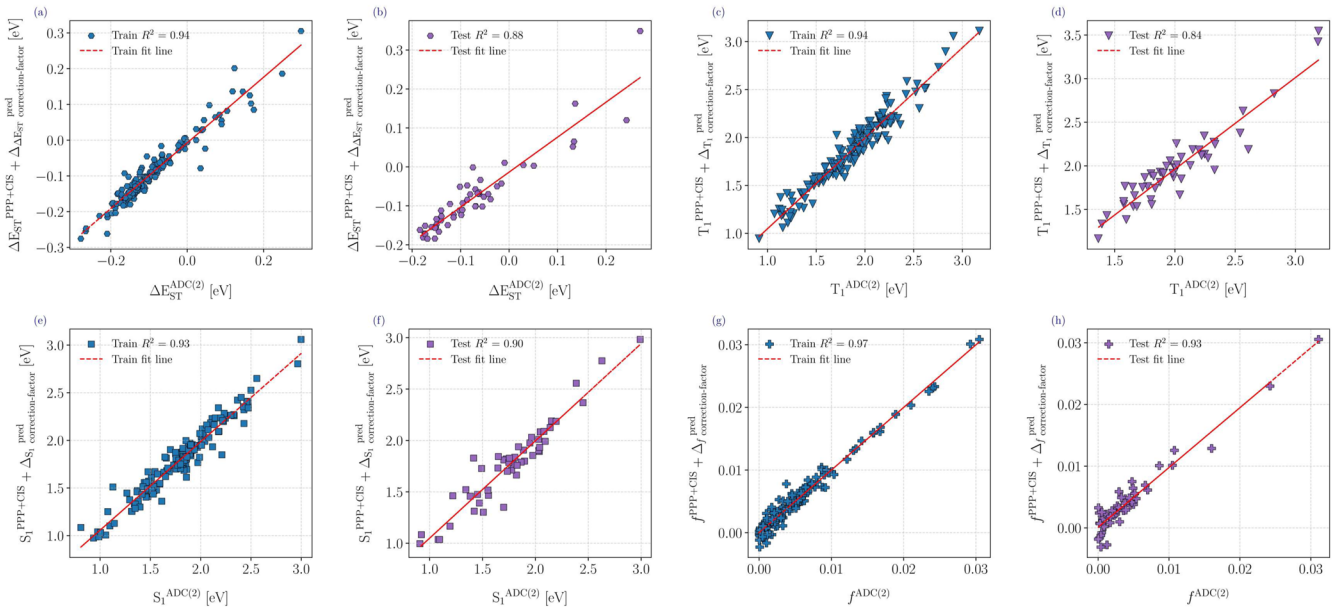

FIG. 5. Comparison of predicted and true values for key molecular properties [(a) and (b)] ΔE_{ST}, [(c) and (d)] T_1, [(e) and (f)] S_1, and [(g) and (h)] f in the Dataset-II. The R^2 values for ΔE_{ST}, T_1, S_1, and f are 0.88, 0.84, 0.90, and 0.93, respectively, indicating strong predictive performance across all properties.

validation, and 50 for testing. To ensure consistency and reliability across different data partitions, we employed fivefold cross-validation. This approach provided a comprehensive assessment of the model's performance while mitigating the impact of data splitting variability.

We also applied scikit-learn's[36] `StandardScaler` to normalize the data, particularly addressing the small numerical values associated with properties such as the singlet–triplet gap (ΔE_{ST}) and oscillator strength (f). This preprocessing step improved numerical stability and enhanced model performance. The model demonstrated strong predictive accuracy across all properties, achieving an R^2 value of 0.88 for ΔE_{ST} for the test set [Fig. 5(b)]. For other key properties, namely, S_1, T_1, and f, the R^2 values for the test set were 0.90, 0.84, and 0.93, respectively, as shown in Fig. 5. These results underscore the model's robustness and predictive power across diverse molecular properties.

The correction factor predictions for ΔE_{ST}, S_1, T_1, and f, depicted in Fig. S5, further illustrate the model's ability to generalize across different molecular features. The RMSE and MAE values for both the train and test sets, summarized in Table II, highlight the low prediction errors, reinforcing the reliability of the model. Notably, for ΔE_{ST}, the RMSE and MAE values for the test set were 0.0351 and 0.0254, respectively, indicating precise predictions for the singlet–triplet gap. Similarly, for S_1 and T_1, the test set RMSE values were 0.0523 and 0.1558, respectively, while for f, the RMSE was as low as 0.0015. The R^2 values for the actual vs predicted correction factors for all properties across the train and test sets are listed in Table S2. Overall, these results demonstrate the robustness and scalability of our approach, even when applied to a smaller dataset. The consistency of the predictive performance across different dataset sizes and molecular properties underscores the generalizability of the correction factor model, paving the way for its application to a broader range of molecular systems.

D. Transferability on the benchmark dataset

We utilized the trained model, initially developed for predicting ΔE_{ST} in the azaphenalene dataset, to make predictions on the benchmark INVEST15 dataset (see Fig. S6)[30] for the same property, as depicted in Fig. 6. The molecules in the INVEST15 dataset were optimized in their ground state using the B97-3c functional, and excited-state calculations were performed at the LR-CC2/aug-cc-pVTZ level of theory. Molecules containing elements such as boron and phosphorus were excluded from the analysis, as the pre-trained model does not account for these atoms. In addition, a few molecules resembling those in the training set were omitted to maintain evaluation integrity. The model delivered an impressive predictive performance, achieving an R^2 value of 0.86, and demonstrated exceptional classification ability by accurately distinguishing INVEST molecules from those with a positive ΔE_{ST}. This successful application on a different dataset exemplifies effective domain adaptation, where the model has transferred knowledge from the azaphenalene training set to predict properties in a new but related dataset.[43] These results underscore the model's potential as a powerful tool for both efficient molecular screening and high-accuracy QM-level property prediction.

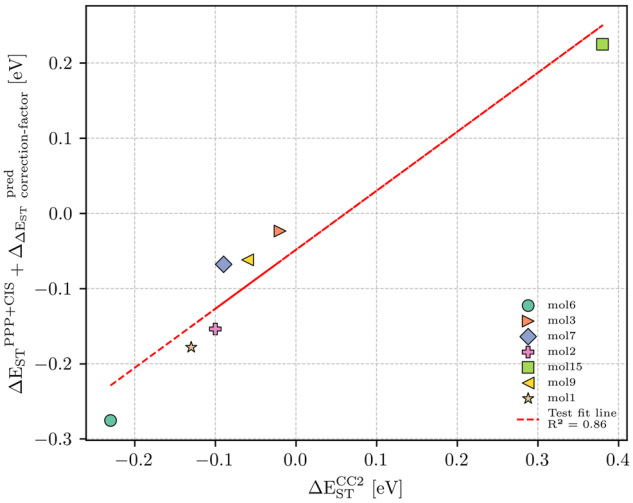

FIG. 6. Performance of the pre-trained model on a few molecules from the benchmark dataset INVEST15.[30] The plot depicts the predicted vs actual ΔE_{ST} values. The model achieves an impressive R^2 = 0.86, highlighting its strong transferability and ability to generalize well to similar molecules.

TABLE II. Performance metrics for the machine learning model trained on Dataset-II. RMSE and MAE are reported for both the training and test datasets across the predicted properties: ΔE_{ST}, S_1, T_1, and f. The low error values indicate high model accuracy and generalizability.

Target property	RMSE (train)	MAE (train)	RMSE (test)	MAE (test)
S_1	0.0316	0.129	0.0523	0.1685
T_1	0.0957	0.0752	0.1558	0.1153
ΔE_{ST}	0.0239	0.0156	0.0351	0.0254
f	0.0009	0.0006	0.0015	0.0011

E. Hyperparameter tuning—Balancing accuracy and robustness

In optimizing SchNetPack models using Optuna, several hyperparameters were identified as critical for modeling different molecular properties. For the singlet–triplet gap (ΔE_{ST}), the learning rate, batch size, and $n_{interactions}$ emerged as key factors. The learning rate dictates how the model adjusts its weights during training, striking a balance between faster convergence and avoiding overshooting. Batch size, which determines the amount of data processed before updating the model weights, influences both stability and generalization. Notably, $n_{interactions}$—the number of interaction layers in the model—proved especially important for accurately predicting ΔE_{ST}. These layers refine atomic representations by iteratively capturing complex interatomic interactions, essential for modeling the subtle electronic effects underlying the singlet–triplet gap.

For singlet energy (S_1), the interplay of n_{rbf} and $n_{interactions}$ was particularly significant. n_{rbf}, which defines the number of radial basis functions used to model atomic distances, plays a fundamental role in representing the continuous nature of interatomic interactions. This precise distance representation is crucial for energy predictions, as molecular energy is inherently tied to the arrangement

and interactions of atoms. Complementarily, $n_{\text{interactions}}$ enhances the model's ability to capture deeper, more complex atomic relationships, thereby improving the accuracy of singlet energy predictions. Triplet energy (T_1) predictions, on the other hand, were predominantly influenced by $n_{\text{interactions}}$. This underscores the importance of capturing intricate interatomic relationships and environmental interactions for modeling triplet states, which involve subtle spin-related effects. These effects depend on the delicate balance of electronic configurations, making them particularly challenging to predict accurately without sufficient model depth.

For oscillator strength (f), a measure of the probability of electronic transitions between energy levels, both n_{rbf} and $n_{\text{interactions}}$ were pivotal. The precise representation of atomic distances through n_{rbf} enables the model to capture short-range spatial relationships, which are critical for accurately modeling electronic structure and transition probabilities. Meanwhile, $n_{\text{interactions}}$ enhances the model's ability to account for the complex interatomic relationships that govern these transitions, thereby improving the prediction accuracy for oscillator strength.

These findings highlight the overarching importance of $n_{\text{interactions}}$ across all tasks, as it enables the model to capture nuanced atomic interactions and refine feature representations. While $n_{\text{interactions}}$ consistently plays a critical role, other hyperparameters, such as n_{rbf} (for distance representation), batch size, and learning rate (for optimization), exhibit varying levels of importance depending on the specific property being modeled. Interestingly, the hyperparameters were found to be largely independent of one another, meaning they could be optimized individually. This independence simplifies the optimization process, allowing for a focused approach to tuning each parameter without concern for cross-dependencies. Consequently, this property-specific yet independent tuning framework enables efficient hyperparameter optimization tailored to the unique demands of each predictive task.

IV. CONCLUSION

This work advances the capabilities of the SchNetPack framework by developing a tailored and efficient methodology for accurately predicting and screening molecular excited-state properties critical to thermally activated delayed fluorescence emitters. By integrating the computational efficiency of the PPP+CIS theory with the accuracy enhancements provided by the Δ-ML component, the model effectively addresses the inherent limitations of the PPP+CIS approach in predicting precise property values. Hyperparameter optimization techniques, facilitated through Optuna, played a pivotal role in fine-tuning the model, ensuring high predictive performance while minimizing the risk of overfitting.

This study leveraged two varying-sized datasets, encompassing molecules with diverse structural and electronic features, as confirmed by principal component analysis. This diversity validated the model's robustness and generalizability across a wide chemical space. The model demonstrated exceptional accuracy in predicting key TADF properties, including singlet (S_1) and triplet (T_1) energies, the singlet–triplet gap (ΔE_{ST}), and oscillator strength (f). Its performance in predicting oscillator strength was of particular significance—a property for which PPP+CIS theory alone shows weak correlations with high-level ADC(2) values. The inclusion of the PaiNN model, capable of accurately calculating dipole moments, proved instrumental, as oscillator strength is directly linked to the spatial distribution and dynamics of dipole moments. The pre-trained model was also validated on the benchmark INVEST15 dataset,[30] which features molecules similar to Dataset-II. The model's impressive performance highlights its transferability and generalization capabilities. Furthermore, the model's ability to predict ΔE_{ST} with high accuracy, including in the negative region, highlights its potential for identifying and screening INVEST molecules. This capability underscores the model's applicability to a broader range of TADF emitters, including those with unconventional electronic configurations. Hyperparameter optimization revealed key insights into the model's architecture: while $n_{\text{interactions}}$ emerged as a consistently critical parameter across all properties, others such as n_{rbf}, batch size, and learning rate exhibited property-specific importance. The independence of these hyperparameters further streamlined the optimization process, enabling efficient and targeted tuning to maximize performance.

In conclusion, the model represents a significant step forward in computational design and prediction of TADF properties, demonstrating exceptional accuracy, versatility, and efficiency. This approach highlights the potential for integrating physics-inspired models and machine learning to advance the understanding and design of functional materials, paving the way for future studies to incorporate more complex features and diverse datasets.

SUPPLEMENTARY MATERIAL

The supplementary material contains a comparison of excited-state properties computed using PPP+CIS and ADC(2), a representation of the benchmark set INVEST15, and a performance evaluation of the neural network model for Dataset-I and Dataset-II.

ACKNOWLEDGMENTS

The authors acknowledge the Indian Institute of Technology Gandhinagar, India, for providing research facilities and financial support. N.R. and A.M. thank PARAM Ananta for computational resources.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

R. Nikhitha: Conceptualization (supporting); Data curation (lead); Formal analysis (lead); Methodology (equal); Validation (lead); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Anirban Mondal**: Conceptualization (lead); Formal analysis (supporting); Funding acquisition (lead); Investigation (equal); Methodology (equal); Project administration (lead); Resources (lead); Supervision (lead); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal).

DATA AVAILABILITY

The data and code to reproduce the results are available in our GitHub repository (https://github.com/nikhitha79/Delta-ML).

REFERENCES

[1] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, "Comparing molecules and solids across structural and alchemical space," Phys. Chem. Chem. Phys. **18**, 13754–13769 (2016).

[2] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, "Machine learning unifies the modeling of materials and molecules," Sci. Adv. **3**, e1701816 (2017).

[3] P. Zheng, R. Zubatyuk, W. Wu, O. Isayev, and P. O. Dral, "Artificial intelligence-enhanced quantum chemical method with broad applicability," Nat. Commun. **12**, 7022 (2021).

[4] G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Müller, and O. Anatole von Lilienfeld, "Machine learning of molecular electronic properties in chemical compound space," New J. Phys. **15**, 095003 (2013).

[5] M. Stöhr, L. Medrano Sandonas, and A. Tkatchenko, "Accurate many-body repulsive potentials for density-functional tight binding from deep tensor neural networks," J. Phys. Chem. Lett. **11**, 6835–6843 (2020).

[6] R. Pariser and R. G. Parr, "A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. I," J. Chem. Phys. **21**, 466–471 (1953).

[7] R. Pariser and R. G. Parr, "A semi-empirical theory of the electronic spectra and electronic structure of complex unsaturated molecules. II," J. Chem. Phys. **21**, 767–776 (1953).

[8] J. A. Pople, "Electron interaction in unsaturated hydrocarbons," Trans. Faraday Soc. **49**, 1375–1385 (1953).

[9] M. Bedogni, D. Giavazzi, F. Di Maiolo, and A. Painelli, "Shining light on inverted singlet–triplet emitters," J. Chem. Theory Comput. **20**, 902–913 (2024).

[10] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross, "How to represent crystal structures for machine learning: Towards fast prediction of electronic properties," Phys. Rev. B **89**, 205118 (2014).

[11] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, "Prediction errors of molecular machine learning models lower than hybrid DFT error," J. Chem. Theory Comput. **13**, 5255–5264 (2017).

[12] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, and K.-R. Müller, "Machine learning force fields," Chem. Rev. **121**, 10142–10186 (2021).

[13] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, and E. D. Cubuk, "Scaling deep learning for materials discovery," Nature **624**, 80–85 (2023).

[14] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller, "SchNetPack: A deep learning toolbox for atomistic systems," J. Chem. Theory Comput. **15**, 448–455 (2019).

[15] H. Uoyama, K. Goushi, K. Shizu, H. Nomura, and C. Adachi, "Highly efficient organic light-emitting diodes from delayed fluorescence," Nature **492**, 234–238 (2012).

[16] H. Kaji *et al.*, "Purely organic electroluminescent material realizing 100% conversion from electricity to light," Nat. Commun. **6**, 8476 (2015).

[17] D. Zhang, L. Duan, C. Li, Y. Li, H. Li, D. Zhang, and Y. Qiu, "High-efficiency fluorescent organic light-emitting devices using sensitizing hosts with a small singlet–triplet exchange energy," Adv. Mater. **26**, 5050–5055 (2014).

[18] N. Aizawa *et al.*, "Delayed fluorescence from inverted singlet and triplet excited states," Nature **609**, 502–506 (2022).

[19] M. H. Garner, J. T. Blaskovits, and C. Corminboeuf, "Enhanced inverted singlet–triplet gaps in azaphenalenes and non-alternant hydrocarbons," Chem. Commun. **60**, 2070–2073 (2024).

[20] Ö. H. Omar, X. Xie, A. Troisi, and D. Padula, "Identification of unknown inverted singlet–triplet cores by high-throughput virtual screening," J. Am. Chem. Soc. **145**, 19790–19799 (2023).

[21] Y.-J. Pu, D. Valverde, J. C. Sancho-García, and Y. Olivier, "Computational design of multiple resonance-type BN molecules for inverted singlet and triplet excited states," J. Phys. Chem. A **127**, 10189–10196 (2023).

[22] S. Ghosh and K. Bhattacharyya, "Origin of the failure of density functional theories in predicting inverted singlet–triplet gaps," J. Phys. Chem. A **126**, 1378–1385 (2022).

[23] Sanyam, R. Khatua, and A. Mondal, "Cost-effective approach for modeling of multiresonant thermally activated delayed fluorescence emitters," J. Chem. Theory Comput. **19**, 9290–9301 (2023).

[24] Sanyam, P. Sorout, and A. Mondal, "Rational design of organic emitters with inverted singlet–triplet gaps for enhanced exciton management," J. Phys. Chem. A **128**, 7114–7123 (2024).

[25] Y. Chen, W. Yan, Z. Wang, J. Wu, and X. Xu, "Constructing accurate and efficient general-purpose atomistic machine learning model with transferable accuracy for quantum chemistry," J. Chem. Theory Comput. **20**, 9500–9511 (2024).

[26] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Big data meets quantum chemistry approximations: The Δ-machine learning approach," J. Chem. Theory Comput. **11**, 2087–2096 (2015).

[27] K. Atz, C. Isert, M. N. A. Böcker, J. Jiménez-Luna, and G. Schneider, "Δ-Quantum machine-learning for medicinal chemistry," Phys. Chem. Chem. Phys. **24**, 10775–10783 (2022).

[28] K. Jorner, R. Pollice, C. Lavigne, and A. Aspuru-Guzik, "Ultrafast computational screening of molecules with inverted singlet–triplet energy gaps using the Pariser–Parr–Pople semiempirical quantum chemistry method," J. Phys. Chem. A **128**, 2445–2456 (2024).

[29] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, "SchNet—A deep learning architecture for molecules and materials," J. Chem. Phys. **148**, 241722 (2018).

[30] L. Kunze, T. Froitzheim, A. Hansen, S. Grimme, and J.-M. Mewes, "ΔDFT predicts inverted singlet–triplet gaps with chemical accuracy at a fraction of the cost of wave function-based approaches," J. Phys. Chem. Lett. **15**, 8065–8077 (2024).

[31] R. Pollice, B. Ding, and A. Aspuru-Guzik, "Rational design of organic molecules with inverted gaps between the first excited singlet and triplet," Matter **7**, 1161–1186 (2024).

[32] R. Pollice, P. Friederich, C. Lavigne, G. D. P. Gomes, and A. Aspuru-Guzik, "Organic molecules with inverted gaps between first excited singlet and triplet states and appreciable fluorescence rates," Matter **4**, 1654–1682 (2021).

[33] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, and K.-R. Müller, "SchNet: A continuous-filter convolutional neural network for modeling quantum interactions," arXiv:1706.08566 (2017).

[34] C. R. Harris *et al.*, "Array programming with NumPy," Nature **585**, 357–362 (2020).

[35] Pandas Development Team (2024). "T. pandas-dev/pandas: Pandas," Zenodo, https://doi.org/10.5281/zenodo.13819579

[36] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," J. Mach. Learn. Res. **12**, 2825–2830 (2011).

[37] A. Paszke *et al.*, *Advances in Neural Information Processing Systems 32* (Curran Associates, Inc., 2019), pp. 8024–8035.

[38] K. T. Schütt, S. S. P. Hessmann, N. W. A. Gebauer, J. Lederer, and M. Gastegger, "SchNetPack 2.0: A neural network toolbox for atomistic machine learning," J. Chem. Phys. **158**, 144801 (2023).

[39] A. H. Larsen *et al.*, "The atomic simulation environment—A Python library for working with atoms," J. Phys.: Condens. Matter **29**, 273002 (2017).

[40] M. Abadi *et al.*, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015, https://www.tensorflow.org/,Softwareavailablefromtensorflow.org.

[41] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. O. Koyama, "A next-generation hyperparameter optimization framework," in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Association for Computing Machinery, New York, NY, 2019), pp. 2623–2631.

[42] K. Schütt, O. Unke, and M. Gastegger, "Equivariant message passing for the prediction of tensorial properties and molecular spectra," in *Proceedings of the 38th International Conference on Machine Learning* (PMLR, 2021), pp. 9377–9388.

[43] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. March, and V. Lempitsky, "Domain-adversarial training of neural networks," J. Mach. Learn. Res. **17**, 1–35 (2016).