THE JOURNAL OF PHYSICAL CHEMISTRY A
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCA

Article

Two-Step Machine Learning Approach for Charge-Transfer Coupling with Structurally Diverse Data

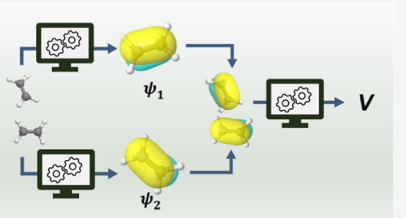
Hung-Hsuan Lin, Chun-I Wang, Chou-Hsun Yang, Muhammad Khari Secario, and Chao-Ping Hsu*

Cite This: *J. Phys. Chem. A* 2024, 128, 271−280

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Electronic coupling is important in determining charge-transfer rates and dynamics. Coupling strength is sensitive to both intermolecular, e.g., orientation or distance, and intramolecular degrees of freedom. Hence, it is challenging to build an accurate machine learning model to predict electronic coupling of molecular pairs, especially for those derived from the amorphous phase, for which intermolecular configurations are much more diverse than those derived from crystals. In this work, we devise a new prediction algorithm that employs two consecutive KRR models. The first model predicts molecular orbitals (MOs) from structural variation for each fragment, and coupling is further predicted by using the overlap integral included in a second model. With our two-step procedure, we achieved mean absolute errors of 0.27 meV for an ethylene dimer and 1.99 meV for a naphthalene pair, much improved accuracy amounting to 14-fold and 3-fold error reductions, respectively. In addition, MOs from the first model can also be the starting point to obtain other quantum chemical properties from atomistic structures. This approach is also compatible with a MO predictor with sufficient accuracy.



INTRODUCTION

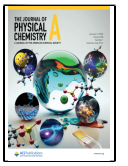
Organic semiconductors have attracted considerable attention owing to their potential applications in electronics, light-emitting diodes, and photovoltaics. This interest reflects several advantages, such as low mass, mechanical flexibility, and tunable functionality. Yet, compared to inorganic semiconductors, i.e., Si-based materials, their relatively low carrier mobility, which results in charge recombination, limits their applications.[1] Moreover, efficiency for electro-optical conversion is also determined by factors that are difficult to control, such as microscopic packing structure[1] and thermal fluctuations.[2] To advance the development of materials, computational simulation and modeling can offer valuable insights into working mechanisms and can reveal design principles.[2−6]

The difficulty in investigating organic semiconductors theoretically is our incomplete understanding of their charge carrier dynamics. Conventionally, charge transfer is modeled either by Marcus theory,[7] in which electronic coupling between molecules is small enough to be treated as a perturbation of nuclear dynamics, or band theory,[8] in which thermal fluctuations are relatively small, treated as perturbations of electronic degrees of freedom. For organic semiconductors, intermolecular electronic couplings can lead to delocalized polarons, but electron−phonon couplings, exhibiting similar orders of magnitude, can cause them to deteriorate. A direct dynamic simulation of both electron and nuclear degrees of freedom is needed to properly describe such nonadiabatic dynamics. Several approaches have been developed based on mean-field Ehrenfest[9] or fewest switch surface hopping molecular dynamics (MD)[10] for charge transfer in organic semiconductors.[11−17]

Establishing an effective electronic Hamiltonian in the site basis representation is the fundamental requirement for simulating dynamics with electron−phonon couplings. In principle, one can build a high-quality Hamiltonian using an all-first-principles calculation, but the computational cost can be prohibitive. This issue can be solved with a machine learning (ML) technique. With the help of ML, a quantum chemical property can be obtained with a small fraction of the computational costs.[18−28] For instance, ML models have been widely applied to predict potential energy surfaces,[19,29−43] as well as electronic couplings.[44−57]

Using machine learning algorithms to predict electronic coupling of electron transfer has been reported.[44−59] Many studies have focused on structures derived from crystals under thermal fluctuation,[45,47,48,54,55,58] while far fewer investigations have examined amorphous materials.[51,52,55] With sampling from crystals, errors of trained ML models can be less than 1 meV due to the small range of structural variation. Resulting models, however, cannot be generalized for amorphous phases on which most organic electronic devices are based. ML built for amorphous states is applicable to a much broader class of

Received: July 5, 2023
Revised: December 3, 2023
Accepted: December 6, 2023
Published: December 29, 2023



problems since there is far greater structural diversity, but the magnitude of errors reported, 1−20 meV, needs further reduction since effective ET coupling is typically in the range of 10^1 to 10^2 meV.

One of the practical challenges in establishing the relationship between an atomistic structure and its corresponding electronic coupling[44−59] is that electronic coupling is *not* a slowly varying function of the atomistic structure. For example, with a large amount of data points (120,000), we previously reported a testing mean absolute error (MAE) of 6.5 meV for coupling of a naphthalene pair using artificial neural networks (ANNs).[52] One of the difficulties is the sensitivity of electronic coupling to the spatial orientation of a dimer. As one of the fragments in a naphthalene pair rotates, electronic coupling oscillates and crosses zero several times, reflecting the oscillatory nature of active molecular orbitals (MO) in the coupling. Therefore, electronic coupling is a much more sensitive quantity than potential energy surfaces and energy gradients.

Such sensitivity can be captured by the overlap integral of MOs, as a strong correlation between MO overlaps and coupling has been reported in the literature,[60−64] implying a chance to predict electronic couplings with overlap integral values. In order to obtain this overlap integral, active MOs involved in electron transfer are needed. Building a transferable method to predict MOs of a general set of molecules has been an active research area. The sophisticated deep learning framework, SchNOrb, offers a general prediction of MOs for a wide range of molecules.[23,65] Nonetheless, for the charge-transfer coupling study, the precision of the overlap integral and subsequently the requirement for MO coefficients need to be very precisely predicted. Thus, it would be ideal to divide the task into two steps, with the first step being prediction of the MO and the second being the final coupling with the MO or the overlap values provided. In this way, the first step can be replaced by any MO-predicting model when the transferable models of MO prediction are mature enough.

In the present work, we exploit the observed correlation between MO overlap and electronic coupling and propose a new algorithm to predict electronic couplings. As depicted in Figure 1, we propose a scheme with two consecutive ML models. The first predicts MOs for each of the two molecules in the structure. The second predicts the electronic coupling of the molecular pair using a new feature spanned by the overlap integral of MOs of individual molecules in the pair, as obtained from ML1, together with the Coulomb matrix (CM) of the pair. We show that accuracy can achieve an MAE of 0.27 meV for ethylene dimers and 1.99 meV for naphthalene dimers, a 6−16 fold reduction compared to our previous results.[51] For both test cases, our model exhibits good distance dependence and phase prediction. The present study provides a way to use ML to predict quantum chemical properties.

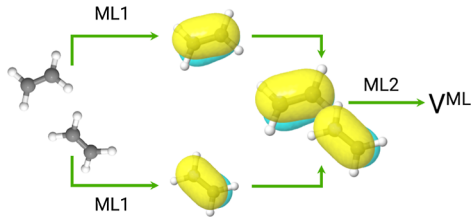


Figure 1. Schematic representation of machine learning workflows for electronic coupling prediction.

METHODOLOGY

Our proposed scheme to predict electronic coupling is depicted in Figure 1. This new two-step algorithm is implemented with two ML models. The first model (ML1 hereafter) predicts the MOs of individual fragments from their structures. The overlap between the predicted MOs is then calculated. In the second stage, the calculated overlap along with structural information is used to form the *feature input* and is then passed to the second model (ML2 hereafter) to predict electronic coupling.

Construction of ML1: MO Prediction. We first developed an ML model to predict the MO coefficients. Here, we propose that instead of learning MOs directly, we train a model to predict the variation of MO coefficients with the corresponding structural variation.

MOs under Structural Variation. Pople et al. showed that the gradient of MO coefficients with respect to the nuclear coordinate[66] can be calculated by

$$\sum_{\mu\nu} \frac{\partial C_{\mu a}}{\partial \mathbf{R}_I} S_{\mu\nu} C_{\nu a} = -\frac{1}{2} \sum_{\mu\nu} C_{\mu a} \frac{\partial S_{\mu\nu}}{\partial \mathbf{R}_I} C_{\nu a} \tag{1}$$

where $\mu(\nu)$ denotes the atomic orbitals, $S_{\mu\nu}$ is an element of an overlap matrix, $\mathbf{R}_I$ denotes the nuclear coordinate of the Ith atom, and $C_{\mu a}$ represents the coefficients of MO a on the μth atomic orbital. This expression implies that the gradient of MO coefficients is a functional of structure variation $\Delta\mathcal{G}$ and the gradient of the overlap matrix, which is also dependent on $\Delta\mathcal{G}$, which can be formally denoted as $\Delta\mathbf{C}[\Delta\mathcal{G}, \Delta\mathbf{S}[\Delta\mathcal{G}]]$.

Naturally, the *input feature* used to predict the variation of MO coefficients is formed by the difference between the Cartesian coordinates of composite atoms

$$\Delta\mathcal{G} = \{\Delta\mathbf{R}_1, \Delta\mathbf{R}_2, \ldots, \Delta\mathbf{R}_{N_{atom}}\} \tag{2}$$

$$\Delta\mathbf{R}_I = (x_I, y_I, z_I) - (x_I^{Ref}, y_I^{Ref}, z_I^{Ref}) \tag{3}$$

where I denotes the Ith atom and "ref" represents the reference molecule, which is a fully optimized structure in the ground state in the present work. All molecules used for ML1 training are rotated and translated into a standard orientation, and further details are included in Section 1.1 of the Supporting Information.

The target, variation of MO coefficients, as our *reference label*, can be calculated simply by

$$\begin{aligned} \Delta\psi &= \psi_{MD} - \psi_{Ref} \\ &= \sum_{\mu} C_{\mu}^{MD}\chi_{\mu} - \sum_{\mu} C_{\mu}^{Ref}\chi_{\mu} \\ &= \sum_{\mu} (C_{\mu}^{MD} - C_{\mu}^{Ref})\chi_{\mu} \end{aligned} \tag{4}$$

where ψ_{MD} and ψ_{ref} represent the MOs of molecules extracted from the MD trajectory and the reference molecule, respectively, with $\mathbf{C}^{MD}$ and $\mathbf{C}_{\mu}^{ref}$ being their MO coefficients, and χ_{μ} denotes an atomic orbital.

Training ML1 for MO. The workflow for training a KRR model to predict MO coefficients is schematically shown in Figure 2. After extracting the structures of molecules from an MD trajectory, we first translate and rotate them into standard

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271−280

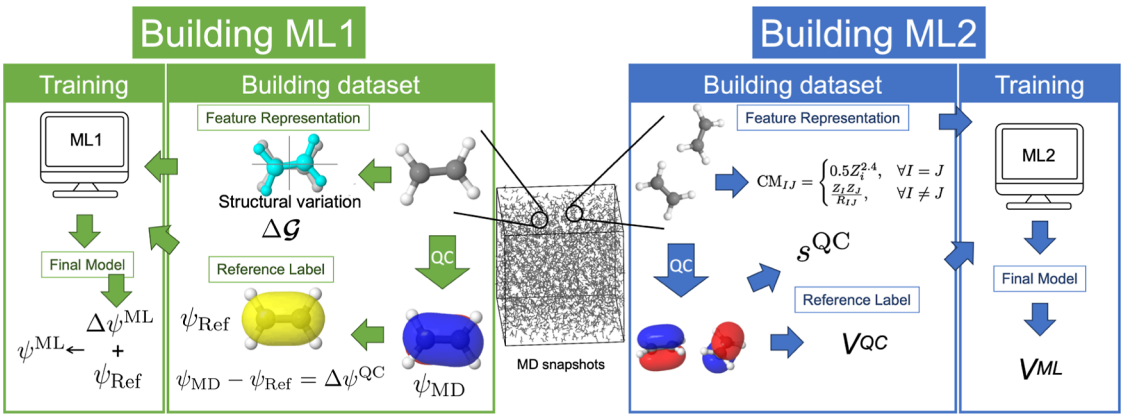


Figure 2. Workflow for building ML1, a KRR model that predicts the MO from the structural variation, and ML2, which predicts electronic coupling from the CM and the overlap. In training the ML2 stage, overlap *s* was obtained from quantum chemical calculations.

nuclear coordinates. Subsequently, MOs are calculated using DFT calculations. Phases of wave functions are arbitrarily determined by the linear algebra subroutines in the quantum chemical program. Hence, phases of MOs should be aligned, as illustrated in Figure S1, before computing the variation of MO coefficients.

For building ML1, the *feature vector* calculated with eq 2 and the *label vector* computed with eq 4 are employed to train a Gaussian KRR model to predict variation of the MO coefficients. Full MOs are then obtained by adding the MO of the reference molecule back to the predicted variation. Since we seek to describe hole transfer, we focus on highest occupied molecular orbital (HOMOs) in the present work. Our approach can be generalized to the lowest unoccupied molecular orbitals for electron-transfer coupling, or to other MOs.

ML2: Electronic Coupling Prediction. We developed a second-stage ML model to predict electronic coupling. The workflow is also illustrated in Figure 2. Geometric features are excellent descriptors to predict electronic coupling.[44−52,58] Moreover, previous studies have also demonstrated that overlap of MOs involved in an electron transfer is highly correlated with electronic coupling.[60−64] The same trend applies to this work (Figure S2). The overlap *s* and the coupling are linearly dependent even though the range of overlap is rather small. Nevertheless, V^{FMO} starts to deviate from the linear dependence in the range of large overlap ($|s| > 0.03$), indicating a nonlinear contribution in the dependence of *s*, as implied in eq 5. To capture the linear dependence and deviation simultaneously, we use both overlap *s* and structural information to form the *feature vector* of ML2.

The sign of electronic coupling may not be relevant to describing a charge-transfer process between two fragments. However, it is crucial to be able to track the sign with the reference geometry because randomly determined phases make the data unlearnable.[67] Moreover, to describe charge transfer with many molecules, it is important to maintain the phase of the basis at each site in determining Hamiltonian matrix elements. We employed a phase correction scheme to relate the structure and the phase.[51,52] Subsequently, *feature vectors* are spanned by the CM and the dimer overlap. Once *feature vectors* are constructed, they and the *reference label* are passed to the machine for training the model.

Prediction Algorithm. By combining ML1 and ML2, we built a scheme that predicts the electronic coupling of a molecular dimer. The workflow outlined in Figure 3 is also listed as follows:

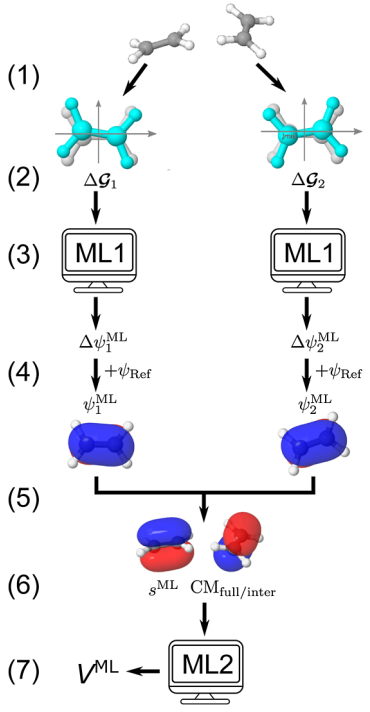


Figure 3. Workflow was used to predict electronic coupling from a molecular pair.

(1) Rotate and translate both fragments of a dimer into the standard orientation independently.
(2) Compute the structure deviation of two fragments ($\Delta\mathcal{G}_1$ and $\Delta\mathcal{G}_2$) with respect to the reference molecule.
(3) Predict variations of the MO coefficients of fragments ($\Delta\psi_1^{ML}$ and $\Delta\psi_2^{ML}$) with the ML1 model.
(4) Obtain full HOMOs of both fragments by adding back the HOMO of the reference molecule.
(5) Rotate and translate fragments and ML HOMOs back to their original positions.
(6) Calculate the ML overlap (s^{ML}) of the two HOMOs and then form the *input feature* together with the CM of the dimer.

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271−280

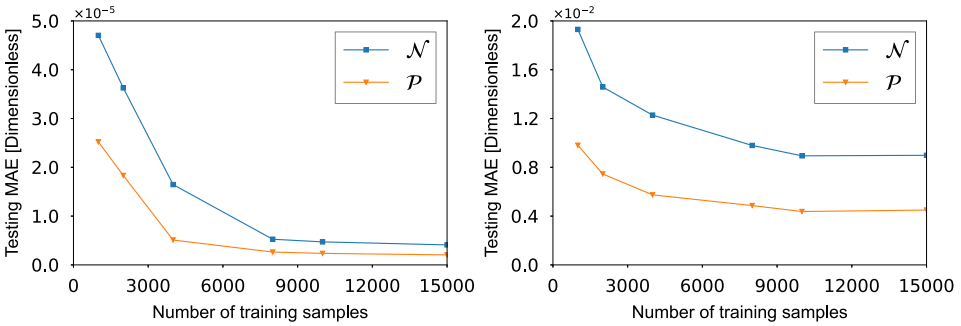


Figure 4. Testing MAEs in normalization and similarity of the ML HOMOs of (A) ethylenes and (B) naphthalenes with different numbers of training samples.

(7) Pass the *input feature* to the ML2 model for electronic coupling prediction.

Other Details. This section includes conditions and methods or further details to obtain the data points.

Sampling from MD Simulation. The first step in implementing machine learning is to acquire a large data set. For the sake of convenience, we reused the database that was generated for our previous investigations.[51,52] For both ethylene and naphthalene, 20,000 molecules were used for MD simulations in their liquid phase (130 and 400 K, respectively). The pressure for both systems was 1 bar, with optimized potentials for liquid simulation force fields employed.[68]

Since ML1 needs only monomolecular data, 20,000 molecules were extracted from the MD snapshot for training and testing. To build the ethylene data sets for ML2, structures of pairs of molecules were extracted from the MD trajectory. Center-of-mass (COM) distances of two molecules ranging from 3−4, 4−5, 5−6, 6−7, and 7−8 Å were collected separately, with each set containing 20,000 pairs of molecules for training and testing of ML2. In developing both ML1 and ML2, 90% of the data points were used for training and 10% were reserved for testing. An additional 1000 dimers were extracted from each of the five COM separation ranges for testing the final combined model. Therefore, 20,000 and 100,000 data points were used for training and testing ML1 and ML2, respectively, and for testing the overall two-step performance, 5000 data points were used. The process of building the data set for naphthalene was the same except that the range of the COM distances of pairs collected from the MD trajectory was 3−10 Å, with seven sets in each Å increment. Hence, numbers of data points for training (90%) and testing (10%) ML1, ML2, and the overall prediction algorithm are 20,000, 140,000, and 7000, respectively.

Quantum Chemical Calculations. As mentioned above, ML1 and ML2 are built to predict MO and electronic coupling, respectively. To obtain ground truths for the learning process, DFT calculations are performed with the DZ* basis set and the long-range corrected functional LC-BLYP.[69,70] Here, we set the range−separation parameter μ to 0.41 bohr^{-1} for ethylene and 0.26^{-1} for naphthalene in order to properly describe the hole transfer.[71]

Electronic couplings of pairs are calculated with the frontier molecular orbital (FMO) approach, defined as

$$V^{\mathrm{FMO}} = \frac{f_{\mathrm{DA}} - s/2(f_{\mathrm{DD}} - f_{\mathrm{AA}})}{1 - s^2} \tag{5}$$

where $f_{\mathrm{DA}} = \langle \psi^{\mathrm{D}}_{\mathrm{FMO}} | \hat{f} | \psi^{\mathrm{A}}_{\mathrm{FMO}} \rangle$ with $\hat{\mathrm{f}}$ being the Kohn−Sham operator of the donor−acceptor system. $\psi^{\mathrm{D(A)}}_{\mathrm{FMO}}$ denotes frontier Kohn−Sham orbitals for the donor (acceptor), and throughout the present work, since we are building models for hole transfer, they are the HOMOs. s is the overlap between MOs of the donor and the acceptor, calculated as

$$\begin{aligned} s &= \langle \psi^{\mathrm{D}}_{\mathrm{FMO}} | \psi^{\mathrm{A}}_{\mathrm{FMO}} \rangle \\ &= \sum_{\mu \in \mathrm{D}} \sum_{\nu \in \mathrm{A}} C^{\mathrm{D}}_{\mu \mathrm{FMO}} C^{\mathrm{A}}_{\nu \mathrm{FMO}} \langle \chi_\mu | \chi_\nu \rangle \\ &= (\mathbf{C}^{\mathrm{D}}_{\mathrm{FMO}})^T \mathbf{S}_{\mathrm{DA}} \mathbf{C}^{\mathrm{A}}_{\mathrm{FMO}} \end{aligned} \tag{6}$$

where $\mathbf{S}_{\mathrm{DA}}$ is the overlap matrix in the atomic basis of the donor and acceptor. It is the off-diagonal block of the atomic orbital overlap matrix. $\mathbf{C}^{\mathrm{D/A}}_{\mathrm{FMO}}$ is the coefficient of FMO for the donor/acceptor as a vector, and $\mathbf{C}^T$ is its transpose. In the present work, we calculate the electronic coupling for the hole-transfer case. Hence, the HOMOs of the donor and acceptor are considered. We first calculate the HOMOs of individual fragments in their neutral singlet state, and diagonal and off-diagonal Kohn−Sham matrix elements, $f_{\mathrm{DD(AA)}}$ and f_{DA}, and the overlap s are computed. The final coupling value is then obtained from eq 5. All QC calculations are performed with a development version of Q-Chem.[72]

RESULTS AND DISCUSSION

In this section, we first assess the performance of the ML1 model for HOMO prediction, which is followed by the prediction accuracy of ML2, including the effect of CM representation and learning curves on the size of training samples. Finally, we report the performance of the full two-step algorithm, including learning curves and test results on separation distance dependence as well as phase prediction.

In this section, we first assess the performance of the ML1 model for HOMO prediction, followed by the prediction accuracy of ML2, including the effect of CM representation and learning curves with the size of training samples. We also report the performance of the full two-step algorithm, including learning curves and test results on separation distance dependence as well as phase prediction, at the end of this section.

ML1 Evaluation: Quality of ML MOs. The performance of the first ML model is assessed by the quality of the ML MOs. In the present work, the quality of an ML HOMO is quantified by normalization of predicted HOMOs and similarity to that of QC HOMOs. The normalization error of a molecule is calculated by

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271−280

$$\mathcal{N}_{\mathrm{err}} \equiv \langle \psi^{\mathrm{ML}}_{\mathrm{HOMO}} | \psi^{\mathrm{ML}}_{\mathrm{HOMO}} \rangle - 1 = \mathcal{N} - 1 \tag{7}$$

and the similarity is evaluated by projecting an ML HOMO onto a QC HOMO

$$\mathcal{P}_{\mathrm{err}} \equiv \langle \psi^{\mathrm{ML}}_{\mathrm{HOMO}} | \psi^{\mathrm{QC}}_{\mathrm{HOMO}} \rangle - 1 = \mathcal{P} - 1 \tag{8}$$

The MAEs of $\mathcal{P}$ and $\mathcal{N}$ for testing samples are summarized in Figure 4. Since both $\mathcal{P}$ and $\mathcal{N}$ are dimensionless quantities, so are their MAEs.

As expected, MAEs of $\mathcal{N}$ and $\mathcal{P}$ decrease with the size of training samples for both molecules. The prediction accuracy of an ML model tends to increase with a larger database. For ethylene, with 8000 training samples, the MAE of $\mathcal{N}$ and that of $\mathcal{P}$ converge to 0.53×10^{-5} and 0.26×10^{-5}, respectively. With only 1000 samples, MAEs in $\mathcal{N}$ and $\mathcal{P}$ are already lower than 6×10^{-5}, a rather small number that should be good enough to predict MOs.

The corresponding performance for naphthalene is lower. With 10,000 training samples, the MAEs of $\mathcal{N}$ and $\mathcal{P}$ converge to 9.0×10^{-3} and 4.4×10^{-3}, respectively. The better performance for ethylene is likely due to the structural simplicity of ethylene compared to naphthalene. The additional intramolecular degrees of freedom of naphthalene make the fitting problem harder to solve. In addition, in preparation of data, a higher temperature in sampling was used for naphthalene for a liquid state in MD; thus, more variable structures are obtained.

While both errors are less than 1%, we note that the normalization error can be simply corrected by rescaling after ML1. However, deviation from the ground-truth MO, as indicated by $\mathcal{P}$, is harder to fix. In the present work, we chose not to rescale the MO obtained by ML1, and the predictive power is demonstrated in section "Assessment of the Two-Step Scheme". We note that our current approach works well for learning just *one* MO under structural variation. A similar strategy can be applied to the full set of MOs, or the full wave function prediction, but additional consideration would be required, such as orthogonality between any two MOs.

ML2 Evaluation. *Effect of CM Representation.* Next, we examine the performance of ML2. Details of the training strategy follow our previous work.[51] In our earlier work, the Gaussian KRR model for ethylene dimers reached an MAE of 3.5 meV, while for naphthalene pairs, it was ∼11 meV.[51] To improve predictive power, the overlap integral value of HOMOs is included in the *feature vector*. Here, we take a look at the effect of the CM representation.

Direct comparisons of coupling values from QC calculations V^{QC} to those from ML results V^{ML} are illustrated in Figure 5. Models trained with either $\mathrm{CM_{full}}$ or $\mathrm{CM_{inter}}$ are compared. Results for ethylene and naphthalene are shown in the top and bottom panels of Figure 5, respectively. For the sake of simplicity, only results for the testing set are shown here. The error in the training set is included in the Supporting Information accompanying this work. Generally speaking, ML models that were trained with $\mathrm{CM_{full}}$ (Figure 5A,C) yield slightly less accurate results, compared to those with $\mathrm{CM_{inter}}$ (Figure 5B,D). Nonetheless, the KRR model with two types of CM representations reproduces the testing data well. For ethylene, MAEs for $\mathrm{CM_{full}}$ and $\mathrm{CM_{inter}}$ are 0.66 and 0.38 meV, respectively. Errors for testing samples are roughly 3 times larger than those for the training data set (Figure S6 in the Supporting Information). All determinations of coefficients

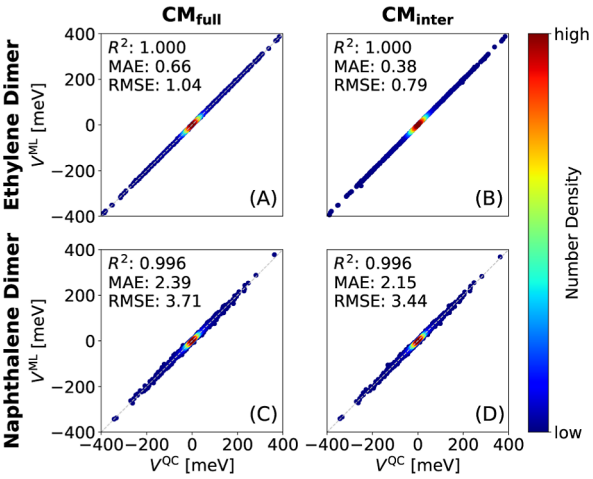


Figure 5. Comparison of electronic couplings obtained by quantum chemical calculations (V^{QC}) and by the ML2 (V^{ML}) for ethylene pairs (panels A and B) and naphthalene dimers (panels C and D) trained with $\mathrm{CM_{full}}$ (A,C) or $\mathrm{CM_{inter}}$ (B,D) belonging to the testing sets. All models were trained with N_{T} = 20,000 and tested with another 2000 samples. Color indicates the density of the data points. Both MAE and RMSE are expressed in meV.

(R^2) are reported as 1.000, which means that the predictive ability is powerful enough.

Electronic coupling between two naphthalene molecules is harder to predict precisely, as can be seen in Figure 5C,D. The MAE with $\mathrm{CM_{full}}$ is 2.39 meV, whereas using $\mathrm{CM_{inter}}$, the MAE is 2.15 meV, much higher than that for ethylene. In order to maintain a similar level of disorder for a liquid phase, naphthalenes were extracted from MD simulations at a higher temperature, which leads to greater structural diversity. Moreover, as a larger molecule, naphthalene has many more degrees of freedom, including several low-frequency ring wiggling or distortion modes that are easily populated during sampling. The HOMO of naphthalene has two nodal planes, leading to a more complicated orientation dependence. Both factors contribute to the observed difficulties in constructing an accurate ML model.[51,52]

With inclusion of the overlap, the improvement of prediction accuracy can be seen clearly in the scatter plots (Figure 5). The MAE for ethylene is now reduced from 4.25[51] to 0.38 meV in the case of $\mathrm{CM_{inter}}$ representation. For naphthalene, the MAE decreases to 2.15 meV, which is approximately one-fifth of previous errors.[51]

Learning Curves. Machine learning is a data-driven approach. More training data lead to a more accurate model. On the other hand, computational cost and time grow tremendously. Therefore, it is important to find a balance between accuracy and cost. In order to achieve this goal, we trained the Gaussian KRR model with different numbers of training samples for both molecules(Figure 6). For each size, the model was trained with randomly selected N_{T} samples and tested with another $N_{\mathrm{T}}/10$ samples. After this process was repeated 20 times, the mean of MAEs for the coupling value is shown in the plots, and their standard deviations (SDs) of the 20 tests are shown as error bars.

All learning curves for both ethylene and naphthalene gradually decreased with an increasing number of training samples. For ethylene dimers, the testing MAE in couplings of $\mathrm{CM_{full}}$ and $\mathrm{CM_{inter}}$ start from 1.5 ± 0.17 and 1.03 ± 0.08 meV and then slowly approach 0.47 ± 0.01 and 0.27 ± 0.01 meV.

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271−280

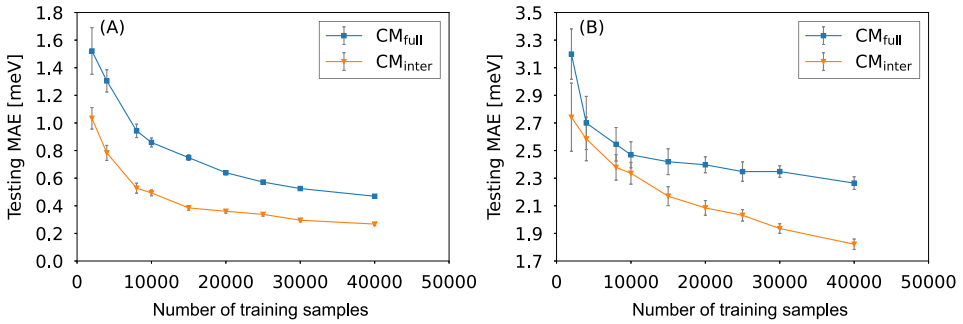


Figure 6. Learning curves for the ML models with different CM representations of (A) ethylene and (B) naphthalene dimers, which consist of MAEs of coupling values in the testing set as a function of training data size. The best set of hyperparameters for each KRR model is found by 5-fold cross-validation (further details are given in the Supporting Information). At each point, results are obtained by repeating 20 runs, with their SD indicated as error bars.

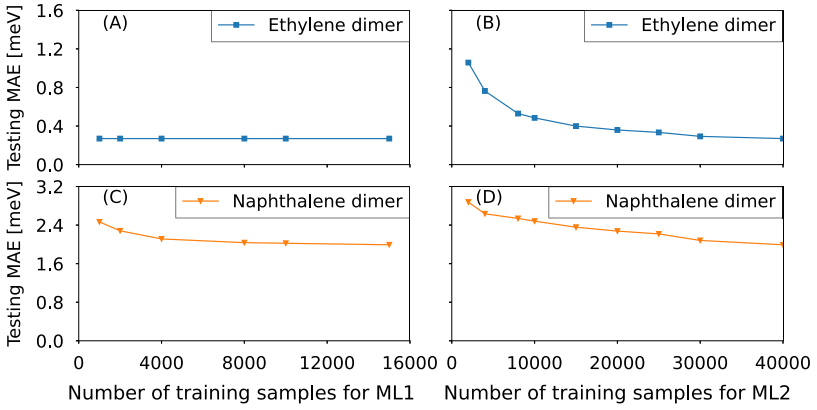


Figure 7. Learning curves for the full two-step algorithm. The testing errors are plotted as a function of training sample sizes for ML1 (A,C) and ML2 (B,D) and for ethylene (A,B) and naphthalene (C,D) dimers. In (A,C), ML2 was trained with 40,000 samples, while in (B,C), ML1 was trained with 15,000 samples.

The predictive power of the KRR model with CM_{inter} reaches convergence at N_T = 40,000. In addition, including CM_{full} in the feature results in a slower learning rate than those of CM_{inter}. The dimension of CM_{full} is much larger than CM_{inter}. The larger *feature vector* dimension results in larger input space and greater spareness in the data. A larger training data set is needed to reach the same accuracy level. Furthermore, the SDs, variations of performance in different samplings, of CM_{full} and CM_{inter} are 0.17 and 0.08 meV with 2000 training samples and then drop to ∼0.01 meV in larger data sets.

MAEs for coupling values from ML2 of naphthalene dimers (Figure 6B) also decreased with increasing numbers of training samples. The testing MAEs of CM_{full} and CM_{inter} start from 3.20 ± 0.18 and 2.74 ± 0.25 meV and then slowly approach 2.26 ± 0.05 and 1.82 ± 0.04 meV. Unlike the behavior of learning curves for ethylene dimers, the prediction accuracy for the CM_{inter} of naphthalene dimers still can be clearly improved even with 40,000 samples. We stop here because memory usage and computational time required exceeded our usual computational settings.

Error bars in Figure 6 indicate variation in the performance of different models with different sampled training data. As in the case of ethylene, error bars become smaller with increasing numbers of training samples, as they converge to approximately 0.035 meV with 30,000 samples. Data for naphthalene dimers have larger error bars, indicating that the variation in the training sample data affects model performance, implying that increasing the training data improves the trained model. This is consistent with the slopes of the learning curves at the end of our tests.

To further improve the accuracy for naphthalene dimers or larger molecules, an efficient training strategy is required. For instance, an incremental learning or online learning technique allows one to retrain a trained model with incoming data with less memory.[73] In the future, we hope to develop and integrate such techniques into our algorithms for efficient learning.

Assessment of the Two-Step Scheme. *Overall Performance.* We further combined ML1 and ML2 for the two-step scheme. In this algorithm, the overlap s in the *input feature* of ML2 is computed with the HOMOs given by ML1. We first assess the overall performance of the two-step scheme with learning curves in which one of the two steps is trained with different sample sizes. The test for sample size in ML1 models was combined with an ML2 trained with 40,000 samples. As shown in Figure 7A, the MAE of ethylene dimers had already reached the accuracy limit with 1000 training samples. On the other hand, as shown in Figure 7C, the predictive power of our algorithm for naphthalene pairs is improved by increasing the number of training samples. The MAE gradually decreases with the number of training samples and reaches convergence at 15,000.

The effect of training sample size for ML2 was also tested, together with the fixed ML1 trained with 15,000 samples. As shown in Figure 7B,D, the number of training sets for the models has a significant effect on the predictive power. The MAE gradually approaches 0.26 meV for ethylene dimers. As

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271−280

for naphthalene pairs, the MAE also slowly decreases with the number of training samples and reaches 1.99 meV, which is roughly equivalent to the testing MAE of ML2 by using s^{QC}. As shown in Table 1, using one more ML model does not introduce additional errors into this algorithm. We also conclude that the prediction accuracy of this algorithm is essentially determined by the accuracy of ML2.

Table 1. Testing Performance (in MAE) of ML2, ML1 + ML2, and CM with the 40,000 Training Samples

molecules	ML2	ML1 + ML2	CM_{inter} [a]
ethylene	0.27[b]	0.27	3.5
naphthalene	1.82	1.99	~13

[a]From ref 51, where the feature is CM_{inter} without overlap. [b]In the units of meV.

Phase Prediction and Distance Dependence. It is important to test for physical characteristics of charge-transfer coupling, such as distance dependence and phase.[51,52] Obtaining the correct (relative) phase is essential for the simulation of charge or polaron dynamics of molecular solids. In the previous work, KRR models for ethylene with CM_{inter} representation could reach 98.5%,[51] and with ANN models for naphthalene, it was over 90%.[52] Those studies offer reliable phase prediction in principle, and errors only occur in the region of small couplings where signs of coupling are changing. Here, we quantitatively examine the phase prediction of our two-step scheme.

Prediction errors of naphthalene and ethylene dimers with different ranges of $|V|$ are summarized in Table 2. In the case of ethylene dimers, the accuracy of the phase prediction is very high. As $|V|$ is below 1 meV, cases of incorrect phase prediction are <2%. Beyond that range, we do not encounter any phase prediction problems. For naphthalene pairs, accuracy is lower. As $|V|$ is below 1 meV, the prediction error is as high as 13.44%. When $|V|$ is in the range between 1 and 10 meV, the error sharply decreases to 1.12%. For both ethylene and naphthalene pairs, one can conclude that incorrect phase prediction occurs mainly in small coupling regions.

Table 2. Percentage of Errors in Phase Prediction for Ethylene and Naphthalene Dimers

range in $\|V\|$	ethylene dimer	naphthalene dimer
0–1[a]	1.98[b]	13.44
1–10	0.00	1.12
10–20	0.00	0.04

[a]The range of absolute coupling value employed in the statistics, in meV. [b]Percentage of cases with an error in the phase.

Finally, we tested the distance dependence of the electronic coupling. Exponential decay with respect to the distance between the two fragments is expected, which can be described as

$$V = V_0 \exp(-\beta d) \tag{9}$$

where d is the distance between individual fragments of a pair and β is the decay rate. Practically, a MO is essentially a one-electron wave function, which is expanded by atomic-centered Gaussian basis sets; thus, the decay would have a Gaussian component at a large separation.[74]

To test the distance dependence, we arbitrarily picked three pairs for ethylene and naphthalene dimers from a subset of 3–4 Å separation. The two molecules are moved along the vector connecting their COM. We first compare the results for three ethylene pairs (Figure 8) obtained from quantum chemical calculations (QC) and ML2 models trained with 20,000 (20k) and 40,000 (40k) samples. ML models for all pairs perfectly capture the dependence of decay with increasing distance above 1 meV. The deviation can be clearly observed when the coupling is smaller than 0.1 meV. The discrepancy is not improved even if the model is trained with a larger training data set. No significant improvement is observed between a 20k model and one with 40k. We also note that the phase cannot be correctly predicted by the model. We take pair 3, for example. Starting from a distance of >7 Å, the oscillatory behavior can be clearly seen. In fact, the signs (phase) are not consistent in this region either, as the 20k model reports an opposite sign at 7.48 Å (details are given in Supporting Information Table S8).

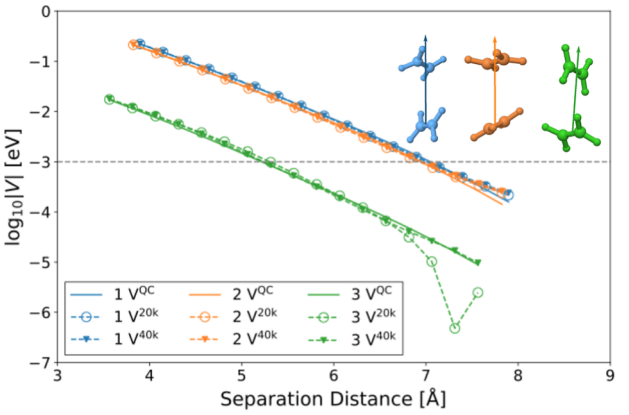


Figure 8. Distance dependence of electronic coupling for three randomly selected ethylene dimers. Solid lines are for data from quantum chemical calculations, and dashed lines with open circles and filled triangles are for ML prediction with 20k and 40k training samples, respectively. Structures of the three pairs tested are shown. Arrows represent vectors connecting the COMs of the two molecules along which the two molecules are separated.

The distance dependence of naphthalene dimers is shown in Figure 9. Naphthalene models are less accurate, with good agreement seen in couplings above 10^{-3} eV. In the case of pair 1, the 20,000 model offers an accurate prediction until 9 Å. On the other hand, the 40k model starts to significantly underestimate the strength of couplings from d = 6.73 Å and gives the wrong phase from d = 7.53 Å. For pair 2, the 20k model underestimates the couplings until d = 6.97 Å where a sudden drop and a change of phase occur, and the two-step ML scheme starts to yield the wrong phase in coupling. In this case, a model trained with a larger database improves the prediction. The prediction starts to deviate from the QC results below 1 meV. Last, for pair 3, the 20k model also gives reasonable couplings until d = 6.77 Å. The 40k model does not accurately predict couplings, either. The prediction for coupling in naphthalene is much more complex than that for ethylene. There exist two additional nodal planes in the HOMO, leading to higher oscillatory behavior, and the learned ML model is more complicated.[52] A larger set of training samples is required to capture its behavior. Nonetheless, the 40k model provides reasonable results in the range of 4–6 Å in

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271–280

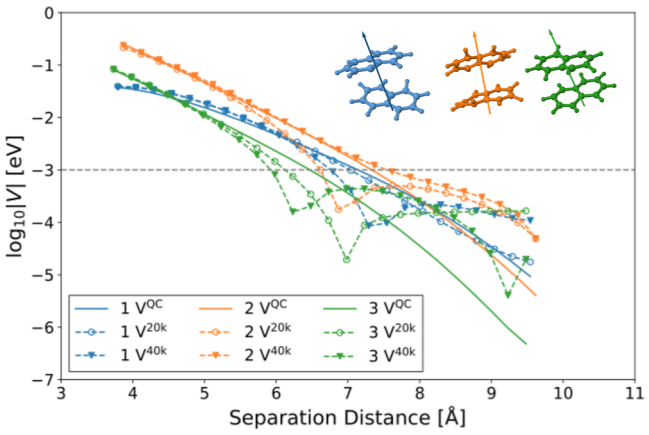


Figure 9. Distance dependence of electronic coupling for the three randomly selected naphthalene dimers. Solid lines are for data from quantum chemical calculations, and dashed lines with open circles and filled triangles are for ML prediction with 20k and 40k training samples, respectively. Structures of the three pairs tested are shown. Arrows representing vectors connecting the COM of the two molecules, along which the two molecules are separated.

COM separation, typical intermolecular separation in a condensed phase, so our model should be sufficient for such situations.

Gaussian KRR is a widely used regression tool for interpolating between data points scattered in a high-dimensional space of *input features*. In principle, the more training data one uses, the more accurate the model one obtains. The trade-off is the computation time of quantum chemical calculations in acquiring the ground truth for training and the training time and memory usage to obtain and use the models. In our present and previous works, molecular pairs for training a model are randomly selected from thermal sampling in MD simulations. We may repeatedly pick data that are near the equilibrium of the molecules. Hence, it will be useful to devise a selection strategy for meaningfully extracting data points based on the feature "difference" in future development.

Computational Efficiency. To estimate the benefits of the ML technique, we compare the computational time of QC calculations and that of the ML prediction. Using a server equipped with Intel Xeon CPU E5-2620 v4 8 cores at 2.10 GHz, NVIDIA Tesla P100, and Memory 256 GB, we built a charge-transfer Hamiltonian for 576 ethylene molecules, containing 165,600 pairs. The QC calculation took roughly 8932 min (0.863 min/pair × 165,600 pairs/16 threads), while the ML model took only 1.41 min with GPU.

CONCLUSIONS

In this work, we developed a new algorithm to predict electronic couplings of ethylene and naphthalene dimers. The algorithm contained two consecutive KRR models. The first ML model is used to predict the HOMO of a molecule with a structure variation. With the HOMO, we introduced the overlap of the two MOs in the *input vectors* for the second ML model. This strategy allows us to increase the ML prediction accuracy of the coupling, as MAEs for ethylene and naphthalene are reduced by 10-fold and 3-fold, respectively, reaching MAEs of ∼0.26 meV for ethylene and ∼1.99 meV for naphthalene.

Instead of increasing the number of training samples, in the present work, we increase the quality of the prediction with a strategy that first predicts individual MOs and their overlap. We show that such a step successfully increases the precision of electronic coupling prediction. The high sensitivity of electronic coupling with respect to intermolecular structures is embedded, at least partially, in the MOs and their overlap factor *s*. The prediction accuracy of hole-transfer coupling, essentially determined by the overlap of the HOMOs of fragments, is thereby increased. This work improves the ML prediction accuracy for electronic couplings.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.3c04524.

Methodology and performances of ML1 and ML2 and raw data for Figures 8 and 9 are included. Filename: The trained model for hole-transfer coupling for ethylene and naphthalene dimers are included online. Please refer to https://github.com/cherrihsu/A-two-step-machine-learning-approach-for-charge-transfer-coupling (PDF)

AUTHOR INFORMATION

Corresponding Author

Chao-Ping Hsu − *Institute of Chemistry, Academia Sinica, Taipei 115, Taiwan; Division of Physics, National Center for Theoretical Sciences, Taipei 106, Taiwan;* orcid.org/0000-0002-7187-427X; Phone: +886 2 5572 8659; Email: cherri@sinica.edu.tw; Fax: +886 2 2783 1237

Authors

Hung-Hsuan Lin − *Institute of Chemistry, Academia Sinica, Taipei 115, Taiwan; Molecular Science and Digital Innovation Center, Genetics Generation Advancement Corp, Taipei 114, Taiwan*

Chun-I Wang − *Institute of Chemistry, Academia Sinica, Taipei 115, Taiwan; Department of Chemistry, University of Illinois at Urbana−Champaign, Urbana, Illinois 61801, United States*

Chou-Hsun Yang − *Institute of Chemistry, Academia Sinica, Taipei 115, Taiwan*

Muhammad Khari Secario − *Institute of Chemistry, Academia Sinica, Taipei 115, Taiwan; Taiwan International Graduate Program on Sustainable Chemical Science & Technology, Academia Sinica Institute of Chemistry, Taipei 115, Taiwan; Department of Applied Chemistry, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan;* orcid.org/0000-0003-1271-3588

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.3c04524

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We gratefully acknowledge the support from Academia Sinica and the National Science and Technology Council of Taiwan through project 112-2123-M-001-002 and 111-2123-M-001-003.

REFERENCES

(1) Wang, C.; Dong, H.; Jiang, L.; Hu, W. Organic semiconductor crystals. *Chem. Soc. Rev.* **2018**, *47*, 422−500.

(2) Giannini, S.; Blumberger, J. Charge Transport in Organic Semiconductors: The Perspective from Nonadiabatic Molecular Dynamics. *Acc. Chem. Res.* **2022**, *55*, 819−830.
(3) Cheung, D. L.; Troisi, A. Theoretical study of the organic photovoltaic electron acceptor PCBM: Morphology, electronic structure, and charge localization. *J. Phys. Chem. C* **2010**, *114*, 20479−20488.
(4) Fratini, S.; Ciuchi, S.; Mayou, D.; de Laissardière, G. T.; Troisi, A. A map of high-mobility molecular semiconductors. *Nat. Mater.* **2017**, *16*, 998−1002.
(5) Troisi, A. Charge transport in high mobility molecular semiconductors: classical models and new theories. *Chem. Soc. Rev.* **2011**, *40*, 2347.
(6) Giannini, S.; Ziogos, O. G.; Carof, A.; Ellis, M.; Blumberger, J. Flickering polarons extending over ten nanometres mediate charge transport in high-mobility organic crystals. *Adv. Theory Simul.* **2020**, *3*, 2000093.
(7) Marcus, R. A. On the theory of oxidation-reduction reactions involving electron transfer. I. *J. Chem. Phys.* **1956**, *24*, 966−978.
(8) Kenkre, V. M.; Andersen, J. D.; Dunlap, D. H.; Duke, C. B. Unified theory of the mobilities of photoinjected electrons in naphthalene. *Phys. Rev. Lett.* **1989**, *62*, 1165−1168.
(9) Ehrenfest, P. Bemerkung über die angenäherte Gültigkeit der klassischen Mechanik innerhalb der Quantenmechanik. *Z. Phys.* **1927**, *45*, 455−457.
(10) Tully, J. C. Molecular dynamics with electronic transitions. *J. Chem. Phys.* **1990**, *93*, 1061−1071.
(11) Kubař, T.; Elstner, M. Efficient algorithms for the simulation of non-adiabatic electron transfer in complex molecular systems: application to DNA. *Phys. Chem. Chem. Phys.* **2013**, *15*, 5794−5813.
(12) Kubař, T.; Elstner, M. A hybrid approach to simulation of electron transfer in complex molecular systems. *J. R. Soc., Interface* **2013**, *10*, 20130415.
(13) Heck, A.; Kranz, J. J.; Kubař, T.; Elstner, M. Multi-scale approach to non-adiabatic charge transport in high-mobility organic semiconductors. *J. Chem. Theory Comput.* **2015**, *11*, 5068−5082.
(14) Spencer, J.; Gajdos, F.; Blumberger, J. FOB-SH: Fragment orbital-based surface hopping for charge carrier transport in organic and biological molecules and materials. *J. Chem. Phys.* **2016**, *145*, 64102.
(15) Wang, L.; Prezhdo, O. V.; Beljonne, D. Mixed quantum-classical dynamics for charge transport in organics. *Phys. Chem. Chem. Phys.* **2015**, *17*, 12395−12406.
(16) Xie, W.; Holub, D.; Kubař, T.; Elstner, M. Performance of mixed quantum-classical approaches on modeling the crossover from hopping to bandlike charge transport in organic semiconductors. *J. Chem. Theory Comput.* **2020**, *16*, 2071−2084.
(17) Carof, A.; Giannini, S.; Blumberger, J. Detailed balance, internal consistency, and energy conservation in fragment orbital-based surface hopping. *J. Chem. Phys.* **2017**, *147*, 214113.
(18) Behler, J. Perspective: Machine learning potentials for atomistic simulations. *J. Chem. Phys.* **2016**, *145*, 170901.
(19) Behler, J. First principles neural network potentials for reactive simulations of large molecular and condensed systems. *Angew. Chem., Int. Ed.* **2017**, *56*, 12828−12840.
(20) Brockherde, F.; Vogt, L.; Li, L.; Tuckerman, M. E.; Burke, K.; Müller, K. R. Bypassing the Kohn-Sham equations with machine learning. *Nat. Commun.* **2017**, *8*, 872.
(21) Hu, D.; Xie, Y.; Li, X.; Li, L.; Lan, Z. Inclusion of machine learning kernel ridge regression potential energy surfaces in on-the-fly nonadiabatic molecular dynamics simulation. *J. Phys. Chem. Lett.* **2018**, *9*, 2725−2732.
(22) Grisafi, A.; Fabrizio, A.; Meyer, B.; Wilkins, D. M.; Corminboeuf, C.; Ceriotti, M. Transferable machine-learning model of the electron density. *ACS Cent. Sci.* **2019**, *5*, 57−64.
(23) Schütt, K. T.; Gastegger, M.; Tkatchenko, A.; Müller, K. R.; Maurer, R. J. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nat. Commun.* **2019**, *10*, 5024.
(24) Dral, P. O. Quantum chemistry in the age of machine learning. *J. Phys. Chem. Lett.* **2020**, *11*, 2336−2347.
(25) *Machine Learning Meets Quantum Physics*; Schütt, K. T., Chmiela, S., von Lilienfeld, O. A., Tkatchenko, A., Tsuda, K., Müller, K.-R., Eds.; Springer International Publishing, 2020.
(26) Westermayr, J.; Marquetand, P. Machine learning for electronically excited states of molecules. *Chem. Rev.* **2021**, *121*, 9873−9926.
(27) Kalita, B.; Li, L.; McCarty, R. J.; Burke, K. Learning to approximate density functionals. *Acc. Chem. Res.* **2021**, *54*, 818−826.
(28) Chen, M. S.; Zuehlsdorff, T. J.; Morawietz, T.; Isborn, C. M.; Markland, T. E. Exploiting Machine Learning to Efficiently Predict Multidimensional Optical Spectra in Complex Environments. *J. Phys. Chem. Lett.* **2020**, *11*, 7559−7568.
(29) Brorsen, K. R. Reproducing global potential energy surfaces with continuous-filter convolutional neural networks. *J. Chem. Phys.* **2019**, *150*, 204104.
(30) Kamath, A.; Vargas-Hernández, R. A.; Krems, R. V.; Carrington, T.; Manzhos, S. Neural networks vs Gaussian process regression for representing potential energy surfaces: a comparative study of fit quality and vibrational spectrum accuracy. *J. Chem. Phys.* **2018**, *148*, 241702.
(31) Manzhos, S.; Dawes, R.; Carrington, T. Neural network-based approaches for building high dimensional and quantum dynamics-friendly potential energy surfaces. *Int. J. Quantum Chem.* **2015**, *115*, 1012−1020.
(32) Kanamori, K.; Toyoura, K.; Honda, J.; Hattori, K.; Seko, A.; Karasuyama, M.; Shitara, K.; Shiga, M.; Kuwabara, A.; Takeuchi, I. Exploring a potential energy surface by machine learning for characterizing atomic transport. *Phys. Rev. B* **2018**, *97*, 125124.
(33) Xu, S.; Li, Y.; Wang, D.; Fang, C.; Luo, C.; Deng, J.; Hu, L. H.; Li, H.; Li, H. Efficient prediction for high precision CO-N₂ potential energy surface by stacking ensemble DNN. *J. Comput. Chem.* **2022**, *43*, 244−254.
(34) Kong, F.-C.; Li, Y.-F.; Shang, C.; Liu, Z.-P. Stability and phase transition of cobalt oxide phases by machine learning global potential energy surface. *J. Phys. Chem. C* **2019**, *123*, 17539−17547.
(35) Zuo, Y.; Chen, C.; Li, X.; Deng, Z.; Chen, Y.; Behler, J.; Csányi, G.; Shapeev, A. V.; Thompson, A. P.; Wood, M. A.; et al. Performance and cost assessment of machine learning interatomic potentials. *J. Phys. Chem. A* **2020**, *124*, 731−745.
(36) Tong, Q.; Luo, X.; Adeleke, A. A.; Gao, P.; Xie, Y.; Liu, H.; Li, Q.; Wang, Y.; Lv, J.; Yao, Y.; et al. Machine learning metadynamics simulation of reconstructive phase transition. *Phys. Rev. B* **2021**, *103*, 054107.
(37) Scherbela, M.; Hörmann, L.; Jeindl, A.; Obersteiner, V.; Hofmann, O. T. Charting the energy landscape of metal/organic interfaces via machine learning. *Phys. Rev. Mater.* **2018**, *2*, 43803.
(38) Nandi, A.; Qu, C.; Houston, P. L.; Conte, R.; Bowman, J. M. Δ -machine learning for potential energy surfaces: a PIP approach to bring a DFT-based PES to CCSD(T) level of theory. *J. Chem. Phys.* **2021**, *154*, 51102.
(39) Liu, Q.; Zhou, X.; Zhou, L.; Zhang, Y.; Luo, X.; Guo, H.; Jiang, B. Constructing high-dimensional neural network potential energy surfaces for gas-surface scattering and reactions. *J. Phys. Chem. C* **2018**, *122*, 1761−1769.
(40) Kolb, B.; Luo, X.; Zhou, X.; Jiang, B.; Guo, H. High-dimensional atomistic neural network potentials for molecule−surface interactions: HCl scattering from Au(111). *J. Phys. Chem. Lett.* **2017**, *8*, 666−672.
(41) Zhang, L.; Wang, H.; Muniz, M. C.; Panagiotopoulos, A. Z.; Car, R.; Weinan, E. A deep potential model with long-range electrostatic interactions. *J. Chem. Phys.* **2022**, *156*, 124107.
(42) Rossi, M.; Ceriotti, M.; Manolopoulos, D. E. Nuclear quantum effects in H^+ and OH^{--} diffusion along confined water wires. *J. Phys. Chem. Lett.* **2016**, *7*, 3001−3007.
(43) Unke, O. T.; Meuwly, M. PhysNet: A neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678−3693.

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271−280

(44) Krämer, M.; Dohmen, P. M.; Xie, W.; Holub, D.; Christensen, A. S.; Elstner, M. Charge and exciton transfer simulations using machine-learned hamiltonians. *J. Chem. Theory Comput.* **2020**, *16*, 4061−4070.
(45) Musil, F.; De, S.; Yang, J.; Campbell, J. E.; Day, G. M.; Ceriotti, M. Machine learning for the structure-energy-property landscapes of molecular crystals. *Chem. Sci.* **2018**, *9*, 1289−1300.
(46) Lederer, J.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine learning−based charge transport computation for pentacene. *Adv. Theory Simul.* **2019**, *2*, 1800136.
(47) Rinderle, M.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine-Learned Charge Transfer Integrals for Multiscale Simulations in Organic Thin Films. *J. Phys. Chem. C* **2020**, *124*, 17733−17743.
(48) Çaylak, O.; Yaman, A.; Baumeier, B. Evolutionary approach to constructing a deep feedforward neural network for prediction of electronic coupling elements in molecular materials. *J. Chem. Theory Comput.* **2019**, *15*, 1777−1784.
(49) Bag, S.; Aggarwal, A.; Maiti, P. K. Machine learning prediction of electronic coupling between the guanine bases of DNA. *J. Phys. Chem. A* **2020**, *124*, 7658−7664.
(50) Miller, E. D.; Jones, M. L.; Henry, M. M.; Stanfill, B.; Jankowski, E. Machine learning predictions of electronic couplings for charge transport calculations of P3HT. *AIChE J.* **2019**, *65*, No. e16760.
(51) Wang, C.-I.; Braza, M. K. E.; Claudio, G. C.; Nellas, R. B.; Hsu, C.-P. Machine learning for predicting electron transfer coupling. *J. Phys. Chem. A* **2019**, *123*, 7792−7802.
(52) Wang, C.-I.; Joanito, I.; Lan, C.-F.; Hsu, C.-P. Artificial neural networks for predicting charge transfer coupling. *J. Chem. Phys.* **2020**, *153*, 214113.
(53) Brian, D.; Sun, X. Charge-transfer landscape manifesting the structure-rate relationship in the condensed phase via machine learning. *J. Phys. Chem. B* **2021**, *125*, 13267−13278.
(54) Bai, X.; Guo, X.; Wang, L. Machine learning approach to calculate electronic couplings between quasi-diabatic molecular orbitals: the case of DNA. *J. Phys. Chem. Lett.* **2021**, *12*, 10457−10464.
(55) Farahvash, A.; Lee, C. K.; Sun, Q.; Shi, L.; Willard, A. P. Machine learning Frenkel Hamiltonian parameters to accelerate simulations of exciton dynamics. *J. Chem. Phys.* **2020**, *153*, 153.
(56) Aggarwal, A.; Vinayak, V.; Bag, S.; Bhattacharyya, C.; Waghmare, U. V.; Maiti, P. K. Predicting the DNA conductance using a deep feedforward neural network model. *J. Chem. Inf. Model.* **2021**, *61*, 106−114.
(57) Reiser, P.; Konrad, M.; Fediai, A.; Léon, S.; Wenzel, W.; Friederich, P. Analyzing dynamical disorder for charge transport in organic semiconductors via machine learning. *J. Chem. Theory Comput.* **2021**, *17*, 3750−3759.
(58) Tan, T.; Wang, D. Machine learning based charge mobility prediction for organic semiconductors. *J. Chem. Phys.* **2023**, *158*, 094102.
(59) Li, C.; Liang, C.; Rouzhahong, Y.; Wang, B.; Li, H. Transferable prediction of intermolecular coupling achieved by hierarchical material representation. *Sci. China Mater.* **2023**, *66*, 819−826.
(60) Zhang, L. Y.; Friesner, R. A.; Murphy, R. B. Ab initio quantum chemical calculation of electron transfer matrix elements for large molecules. *J. Chem. Phys.* **1997**, *107*, 450−459.
(61) Gajdos, F.; Valner, S.; Hoffmann, F.; Spencer, J.; Breuer, M.; Kubas, A.; Dupuis, M.; Blumberger, J. Ultrafast estimation of electronic couplings for electron transfer between π-conjugated organic molecules. *J. Chem. Theory Comput.* **2014**, *10*, 4653−4660.
(62) Troisi, A.; Orlandi, G. Hole migration in DNA: A theoretical analysis of the role of structural fluctuations. *J. Phys. Chem. B* **2002**, *106*, 2093−2101.
(63) Oberhofer, H.; Blumberger, J. Insight into the mechanism of the $Ru2^{+}$−$Ru3^{+}$ electron self-exchange reaction from quantitative rate calculations. *Angew. Chem., Int. Ed.* **2010**, *49*, 3631−3634.
(64) Kubas, A.; Hoffmann, F.; Heck, A.; Oberhofer, H.; Elstner, M.; Blumberger, J. Electronic couplings for molecular charge transfer: benchmarking CDFT, FODFT, and FODFTB against high-level ab initio calculations. *J. Chem. Phys.* **2014**, *140*, 104105.
(65) Gastegger, M.; McSloy, A.; Luya, M.; Schütt, K. T.; Maurer, R. J. A deep neural network for molecular wave functions in quasi-atomic minimal basis representation. *J. Chem. Phys.* **2020**, *153*, 153.
(66) Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Derivative studies in hartree-fock and møller-plesset theories. *Int. J. Quantum Chem.* **2009**, *16*, 225−241.
(67) Westermayr, J.; Gastegger, M.; Menger, M. F. S. J.; Mai, S.; González, L.; Marquetand, P. Machine learning enables long time scale molecular photodynamics simulations. *Chem. Sci.* **2019**, *10*, 8100−8107.
(68) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(69) Savin, A. *Theoretical and Computational Chemistry*; Elsevier, 1996; pp 327−357.
(70) Iikura, H.; Tsuneda, T.; Yanai, T.; Hirao, K. A long-range correction scheme for generalized-gradient-approximation exchange functionals. *J. Chem. Phys.* **2001**, *115*, 3540−3544.
(71) You, Z.-Q.; Hung, Y.-C.; Hsu, C.-P. Calculating electron-transfer coupling with density functional theory: the long-range-corrected density functionals. *J. Phys. Chem. B* **2015**, *119*, 7480−7490.
(72) Epifanovsky, E.; Gilbert, A. T.; Feng, X.; Lee, J.; Mao, Y.; Mardirossian, N.; Pokhilko, P.; White, A. F.; Coons, M. P.; Dempwolff, A. L.; et al. Software for the frontiers of quantum chemistry: an overview of developments in the Q-Chem 5 package. *J. Chem. Phys.* **2021**, *155*, 084801.
(73) Chen, B.-W.; Binti Abdullah, N. N.; Park, S.; Gu, Y. Efficient multiple incremental computation for Kernel Ridge Regression with Bayesian uncertainty modeling. *Future Gener. Comput. Syst.* **2018**, *82*, 679−688.
(74) You, Z. Q.; Shao, Y.; Hsu, C. P. Calculating electron transfer couplings by the spin-flip approach: Energy splitting and dynamical correlation effects. *Chem. Phys. Lett.* **2004**, *390*, 116−123.

https://doi.org/10.1021/acs.jpca.3c04524
J. Phys. Chem. A 2024, 128, 271−280