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

pubs.acs.org/JPCC
Article

Machine Learning Models for Predicting Electronic Coupling in TEMPO/TEMPO$^+$ Systems

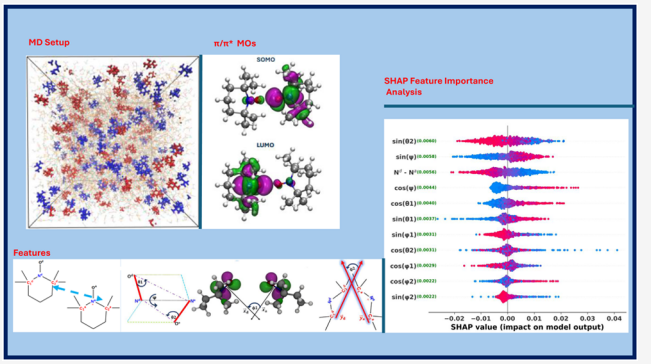
Souvik Mitra, Clara Zens, Stephan Kupfer, Andreas Heuer, and Diddo Diddens*

Cite This: J. Phys. Chem. C 2025, 129, 14667−14678

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Organic radical batteries (ORBs) based on the TEMPO (2,2,6,6-tetramethylpiperidin-1-yl oxyl) radical have drawn significant attention, owing to their unique redox properties. A key factor influencing ORB's redox properties, i.e., the kinetics of the electron transfer between the TEMPO−TEMPO$^+$ pairs, is the communication between the underlying redox-active states as given by the electronic coupling. However, due to the complex structure, predicting accurate electronic couplings for these pairs is computationally expensive and challenging. In this study, we introduce a machine learning (ML) workflow to predict the electronic coupling for TEMPO−TEMPO$^+$ pairs simply by their specific geometric orientations. For the ML models, a data set was generated through time-dependent density functional theory calculations coupled with the Generalized Mulliken Hush method to assess energies, (transition-)dipole moment, and couplings for specific TEMPO−TEMPO$^+$ configurations obtained from classical molecular dynamics simulations that mimic a realistic electrolyte environment. Our results demonstrate that, among the three ML models—linear regression, kernel ridge regression (KRR), and random forest—the KRR model, with its kernel-based approach, most effectively handles the correlated orientation-based descriptors. Moreover, our SHapley Additive exPlanations (SHAP)-based feature importance analysis indicates that multiple orientation factors jointly influence electronic coupling, rather than any single distance or angle dominating, with each parameter's impact strongly contingent on the values of the others which is in agreement with previous studies computational by the consortium.



INTRODUCTION

Organic radical batteries (ORBs) are a promising new generation of energy storage devices due to their unique electrochemical properties as well as environmental friendliness. Among the myriad of active materials being explored, TEMPO, or (2,2,6,6-tetramethylpiperidin-1-yl)oxyl-based polymers, has been one of the most studied. The radical stability of TEMPO and its reversible redox behavior make it a top-class candidate for ORBs with high energy conversion and storage efficiency.[1−4]

One of the factors influencing the overall ORB performance in these redox systems is the complexity of the electron transfer reactions. These electron transfer reactions are governed by the electronic coupling between the donor and acceptor states, implying that an in-depth understanding of the electronic communication between the respective redox-active states is required to adequately analyze these reactions.

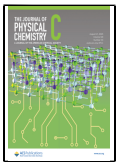
Thereby, the electronic coupling plays a crucial role in the kinetics of electron transfer processes, which can be quantitatively rationalized, e.g., by means of semiclassical Marcus theory.[5−7] Within the Marcus−Hush approach,[8,9] the electron transfer rate depends on electronic coupling ($H_{\delta\alpha}$) as

$$k_{\mathrm{hop}} = \frac{H_{\delta\alpha}^2}{\hbar}\left[\frac{\pi}{k_{\mathrm{B}}T\lambda}\right]^{1/2}\exp\left[-\frac{(\lambda + \Delta G^\circ)^2}{4\lambda k_{\mathrm{B}}T}\right] \quad (1)$$

Here, $\hbar$ is the reduced Planck constant, k_{B} is the Boltzmann constant, T is temperature, λ is the reorganization energy, and ΔG° is the free energy change. Finally, electronic coupling describes the interaction between the two diabatic (electronic) states. In energy-releasing reactions (exergonic reactions), a strong electronic coupling leads to increased electron transfer rates.[10]

In our recent study,[11] we examined the TEMPO/TEMPO$^+$ redox couple as a case example and rationalized the suitability of several computational methods to allow an accurate and cost-efficient evaluation of the electronic coupling. We demonstrated that an appropriate range-separated hybrid functional and basis set, when applied in time-dependent

Received: March 28, 2025
Revised: July 29, 2025
Accepted: July 30, 2025
Published: August 6, 2025



https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

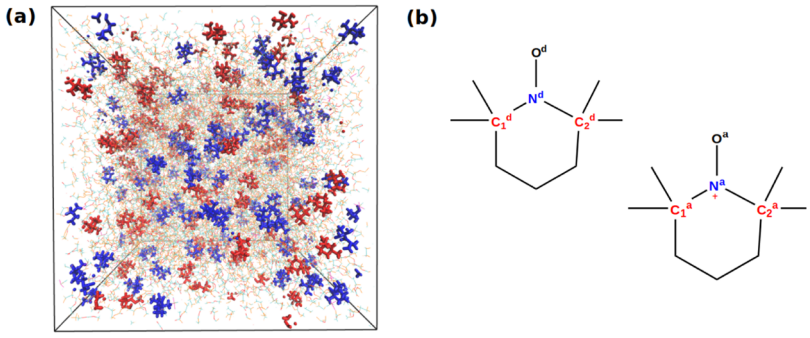


Figure 1. (a) MD snapshot, where TEMPO radicals and TEMPO cations are depicted in thick blue and thick red lines, respectively. The solvents EC and EMC, along with Li^+ and PF_6^- ions, are shown by using thin lines in red, yellow, green, and cyan, respectively. (b) Schematic representation of the donor molecule, TEMPO radical ("d") and acceptor molecule, TEMPO cation ("a").

density functional theory (TDDFT) alongside the Generalized Mulliken Hush (GMH) method,[12] provides such a fast and reliable approach. Here, the GMH method characterizes the electronic coupling as follows:

$$H_{\delta\alpha}^{\mathrm{GMH}} = \frac{|\vec{\mu}_{12}|(V_2 - V_1)}{\sqrt{(\vec{\mu}_2 - \vec{\mu}_1)^2 + 4(\vec{\mu}_{12})^2}} \quad (2)$$

with $\vec{\mu}_{12}$ representing the adiabatic transition dipole transition moment, while $\vec{\mu}_1$ and $\vec{\mu}_2$ denote the permanent dipoles of the respective adiabatic states. The term $(V_2 - V_1)$ corresponds to the vertical excitation energy. This study also highlighted the fact that the electronic coupling in the TEMPO−$TEMPO^+$ system is highly orientation-sensitive, and thus, the electronic coupling between them is challenging to predict with the help of simple statistical models.

In this context, machine learning (ML) offers a feasible solution to this problem. In the past, for simpler ethylene-type molecules, ML models demonstrated high accuracy to predict electronic couplings when trained on extensive data sets.[13] While computationally inexpensive density functional theory (DFT)-based Frontier molecular orbital (FMO) methods[14,15] were sufficient for the electronic coupling calculations for such system, our earlier research[11] indicates that these inexpensive methods are inadequate for the TEMPO−$TEMPO^+$ system, which limits the capability of obtaining comparably large training data sets.

In our current study, we focus on the prediction of electronic couplings for the TEMPO−$TEMPO^+$ system by incorporating certain orientations as descriptors within a moderately sized training data set. To construct a reliable data set, we employed classical molecular dynamics (MD) simulations and combined them with TDDFT/GMH-based electronic coupling calculations. This approach aims to address the challenges posed by the large degrees of freedom in the TEMPO−$TEMPO^+$ system for predicting electronic coupling and to establish an efficient descriptor selection approach for moderate data size.

COMPUTATIONAL METHODS

Molecular Dynamics. In this study, classical MD simulations[16−18] were performed to mimic the complex environment of TEMPO-based ORBs. The simulation setup contained 100 TEMPO radicals and 100 TEMPO cations to replicate a positively charged electrode. A mixture of 1000 ethylene carbonate (EC) molecules, 2000 ethyl methyl carbonate (EMC) molecules, and 70 lithium hexafluorophosphate ($LiPF_6$) molecules made up the electrolyte environment. This composition was enclosed within a periodic cubic box of $10 \times 10 \times 10$ nm^3, ensuring the continuity of interactions across boundaries. By making three of the six monomers on each polymer chain cationic and neutralizing each of the charged monomers with an additional PF_6^- anion, overall charge neutrality was accomplished. In order to avoid unphysically large repulsive interactions, PACKMOL[19] was used to construct the initial configurations for the simulation.

The GROMACS 2019 package was used to run the simulations.[20] At a time step of 0.5 fs, initial relaxation was performed for 2 ns. The reference temperature and pressure were maintained at 298.15 K and 100 bar, respectively, during this phase using the Berendsen thermostat and barostat,[21] with relaxation time constants of 1.0 ps. The aim of this step was to eliminate high-energy configurations from initialization and obtain a stable starting point for subsequent simulations. A particle-mesh Ewald (PME) method was used to tackle the long-range Coulomb interaction efficiently with a cutoff distance of 1.2 nm.

A 50 ns equilibration run was performed with a time step of 2 fs after relaxation. During equilibration, the system's pressure and temperature were regulated using the Parrinello−Rahman barostat and the Nosé−Hoover thermostat, respectively. Using the same time constants as in the relaxation run, the temperature and pressure were maintained at 298.15 K and 1 bar, respectively. This step allowed the system to achieve a thermodynamically stable state suitable for subsequent production runs.

Under the same conditions as those during the equilibration phase, the production run was performed for 100 ns. The simulation trajectories obtained from this phase were used to extract structures necessary for the electronic coupling calculations.

The OPLS all-atom force field[22] was used in all simulations. For the calculation of partial charges of each atom on the TEMPO radical and $TEMPO^+$ cation, electrostatic potential (ESP) fits (Figure S1) were used at the MP2 level of theory with the pVDZ basis set[23] as implemented in Gaussian16.[24]

Finally, around 1300 TEMPO/$TEMPO^+$ redox pairs were selected from 300 time frames extracted from the production trajectories, with the N^d−N^a distance ranging from 4.0 to 4.5 Å. The intervals for the time frame selections were chosen to be 250 ps, which is significantly higher than any correlation time scale for the TEMPO molecule (e.g., the vertical energy correlation time was found to be 4−5 ps for both TEMPO

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

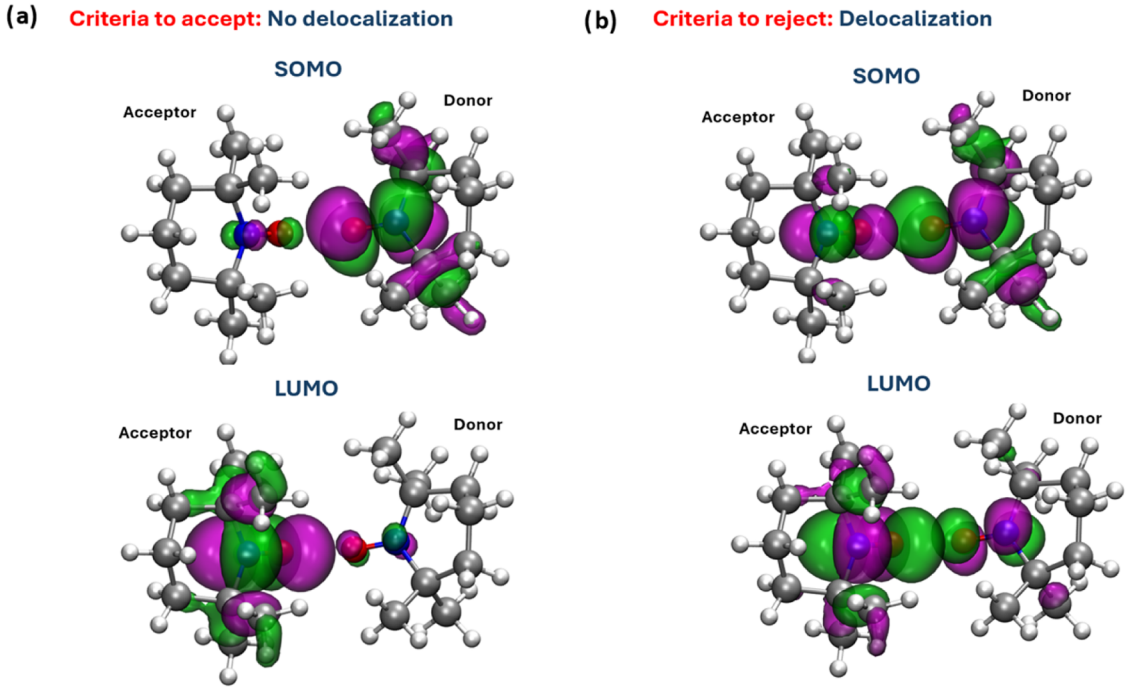


Figure 2. Data screening criteria: (a) If, in a TEMPO/TEMPO$^+$ pair, delocalization did not occur in the SOMO and LUMO, the pair was accepted. (b) Pairs with high delocalization were eliminated from the data set.

radical and TEMPO$^+$ cation[18]). These pairs were used for the calculation of the electronic couplings (Figure 1b).

(TD)DFT−GMH Setup. To calculate the electronic coupling in the TEMPO/TEMPO$^+$ redox pair, the GMH approach was employed in conjunction with (TD)DFT. The GMH approach evaluates electronic coupling based on ground and excited-state properties, such as ground-state energies, permanent dipole moments, excited-state energies, and transition dipole moments (see eq 2). These parameters were obtained through a combination of DFT and TDDFT calculations.

The ground-state electronic energy and ground-state permanent dipole moments were computed using DFT. However, the excited-state electronic energy, permanent dipole moment of the respective excited state, and transition dipole moment were obtained by using TDDFT.

As our recent work[11] suggests that the ωB97XD3 functional,[25] a range-separated hybrid functional incorporating empirical dispersion corrections, effectively describes long-range electronic interactions and accurately captures electron transfer excitations, we utilized this functional for our (TD)DFT calculations. For the basis set, ma-def2-TZVP,[26] a Karlsruhe-type basis set[27] was used. This basis set offers a balance between computational efficiency and accuracy.[11] These DFT/TDDFT calculations were performed using the ORCA v. 5.0.3 program package.[28,29]

Data Curation. Our recent work[11] suggests that several orientational factors have a significant impact on the electronic coupling for short N^d−N^a distances. This study also revealed that delocalization effects in (TD)DFT calculations may result in an overestimation of electronic coupling values in certain instances. The errors are especially large when the singly occupied molecular orbital (SOMO) and the lowest-unoccupied molecular orbital (LUMO) exhibit delocalization across the two TEMPO molecules (see Figure 2).

Taking motivation from this study, in our current work, we focused on analyzing the orientational dependence within the TEMPO/TEMPO$^+$ system, specifically considering short intermolecular N^d−N^a distances ranging from 4.0 to 4.5 Å, which yield around 1300 data points. To assess the delocalization effects across the two TEMPO molecules in our data set, a thorough visualization of the molecular orbitals using the Multiwfn package[30] was performed. This revealed that for TEMPO/TEMPO$^+$ pairs in this N^d−N^a distance range, any electronic coupling above 0.14 eV corresponds to delocalized molecular orbitals. To maintain the accuracy of the training data for our machine learning models, these configurations were removed from the final data set.

Additionally, analyzing the partial charges generated in the output from these electronic calculations also showed that in 28 TEMPO/TEMPO$^+$ configurations, donor−acceptor charges were swapped during TDDFT calculations as compared to the charges from the force field utilized to generate the structures. These configurations were removed from the data set as well in order to maintain consistency.

Feature Engineering. Specific geometric orientations between the two NOCC segments of TEMPO−TEMPO$^+$ pairs (Figure 1b) are critical to define their electronic coupling, especially for short N^d−N^a distances, as we found in our previous study.[11] For our current work, we chose a more geometry-specific method to feature selection rather than depending on a generic descriptors like smooth-overlap-of-atomic-positions (SOAP)[31] or Coulomb matrices (CM).[13,32] These widely used descriptors frequently generate a large feature space in order to capture every potential orientation, particularly in complex systems such as TEMPO−TEMPO$^+$ pairs. This can be problematic because it creates the risk of overfitting for moderate data sets and makes it difficult to identify the orientations that influence electronic coupling. To solve this, we crafted our descriptor set to directly capture any

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

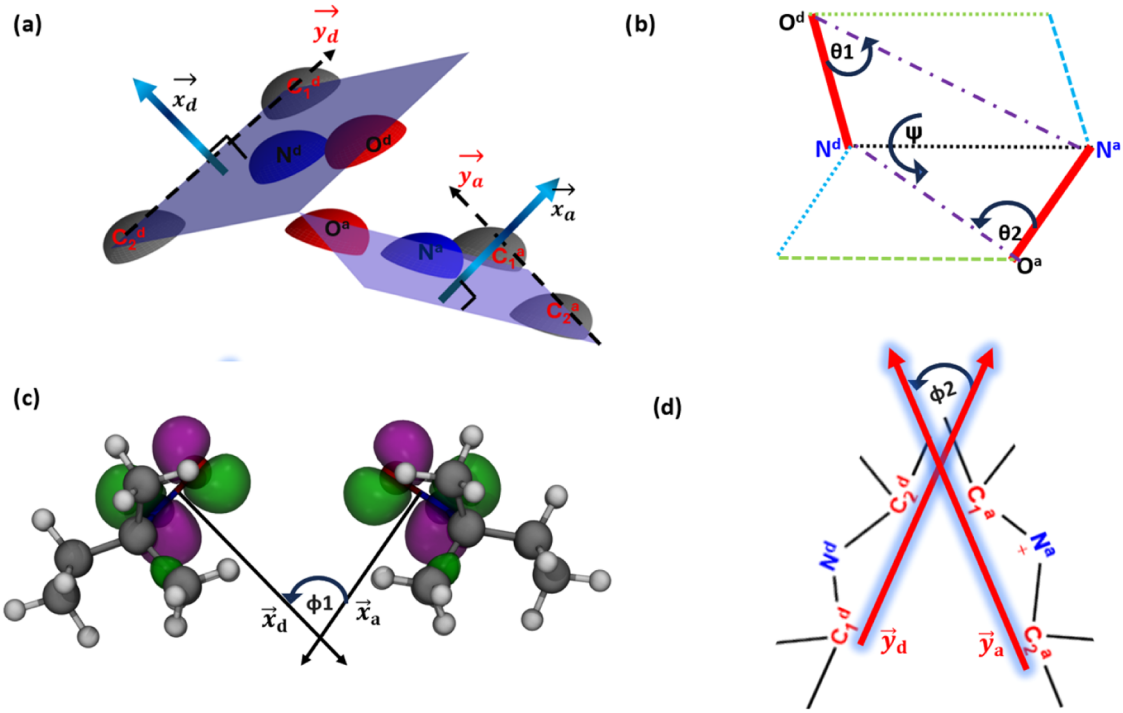


Figure 3. Geometric descriptors used in this study: (a) General representation of the two planes spanned by the respective nitrogen atom and the two adjacent carbon atoms (same atom labels as in Figure 1b), (b) angles $\theta 1$, $\theta 2$, and ψ, describing the relative orientation of the NO−NO bond in the TEMPO/TEMPO$^+$ system; (c) angle $\phi 1$, representing the relative orientation between the normal vectors of the two NCC planes; and (d) angle $\phi 2$, describing the relative orientation between the two C−C bonds.

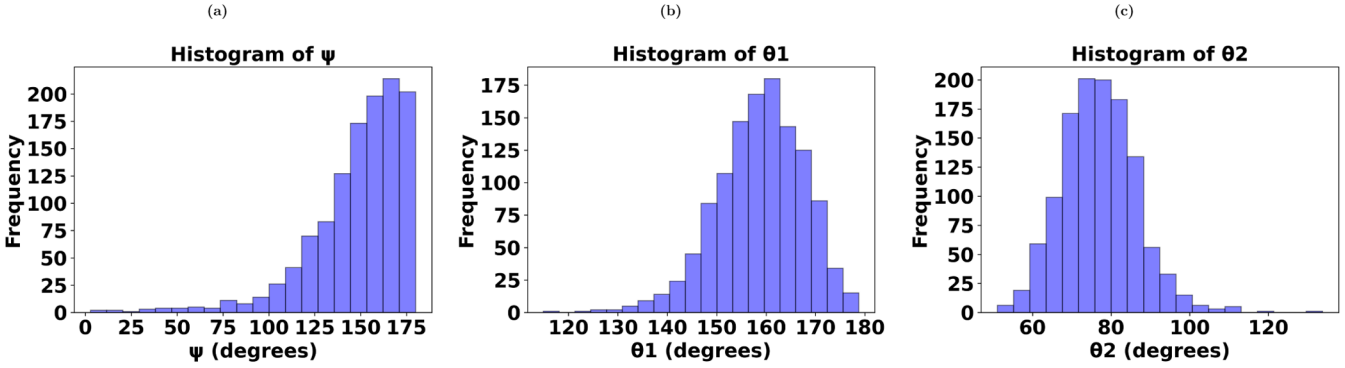


Figure 4. Histograms of angles ψ, $\theta 1$, and $\theta 2$.

orientation that could exist between the NOCC segments, influencing the electronic coupling values.

In particular, the relative orientations of the two NOCC segments are characterized by treating them as rigid bodies with one distance and five angular parameters. Our descriptor set includes angles $\theta 1$, $\theta 2$, and ψ (Figure 3b) to describe the orientation of the NO−NO bond in the TEMPO/TEMPO$^+$ system; angle $\phi 1$ (Figure 3c) for the relative orientation between the normal vectors of the two NCC planes; angle $\phi 2$ (Figure 3d) for the relative orientation of the two C−C bonds; and the intermolecular N^d−N^a distance to measure the separation between the donor and acceptor nitrogen atoms.

To maintain uniformity, carbon atoms on the donor and acceptor NOCC segments were labeled such that C_2^d−C_1^a always represents the shortest distance. Considering the symmetry between the two NOCC fragments, we can define such a structure that either $\phi 1$ or $\phi 2$ always lies in the first quadrant. Hence, we defined our angles $\theta 1$, $\theta 2$, ψ, and $\phi 1$, constrained in the range 0−180°, while $\phi 2$ is defined over 0 to 90°.

It is worth noting that similar geometry-based descriptors have been employed in similar studies. Bag et al.[33] used distance-based descriptors capturing the relative orientation of Guanine bases. Miller et al.[34] considered a descriptor with nine specific features that include distances, angles, as well as one energy difference term to characterize P3HT dimer configurations. Lederer et al.[35] also used geometric features (six specific angles and distances) as descriptors to predict the electronic coupling between pentacene molecules.

Histograms of these angles (Figures 4 and S2) showcase the variability and range of intermolecular orientations sampled from equilibrated MD trajectories for N^d−N^a distances between 4.0 and 4.5 Å. For $\theta 1$, the distribution shows a strong preference for values between 120 and 180°, while $\theta 2$ predominantly lies between 60 and 120°. The angle ψ broadly favors values in the 100 to 175° range. Together, these trends

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

suggest that these angular orientations of the NO−NO bond minimize steric hindrance and Coulombic repulsion to stabilize TEMPO−TEMPO$^+$ pairs in this short N^d−N^a distance range. For the similar reason, angles $\phi 1$ and $\phi 2$ (see Figure S2) also show some preference (60 to 120° range for $\phi 1$ and 50 to 90° range for $\phi 2$) to stabilize the steric hindrance most likely to arise due to the methyl groups attached with the C atoms in each of the NOCC segments.

The wide range of orientation angles in Figures 4 and S2 gives rise to a large variation of the transition dipole moment and, consequently, the electronic coupling over several orders of magnitude (Figure S3), in line with our previous work.[11]

To address the periodicity of angular variables, our final descriptor set includes both sine and cosine functions of all five angles to ensure that the model can learn continuous relations between angular orientations and electronic coupling without discontinuities at the boundaries of the angular domain.

Machine Learning Model. In this study, we utilized three different ML regression models: Linear Regression (LR), Kernel Ridge Regression (KRR), and Random Forest (RF), for a performance comparison with respect to our geometry-based feature space.

The LR model implemented in this work is L^2 Ridge Regression.[36] This assumes a linear relationship between the target variable and the input features, while incorporating regularization to prevent overfitting. Here, the target variable y is predicted with a weighted sum of the input features, $\mathbf{x}$, along with a bias term b, as

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x} + b$$

$\mathbf{w}$ and b are obtained by minimizing a regularized cost function

$$\min_{\mathbf{w},b} \frac{1}{2}\sum_{i=1}^{N}(y_i - \mathbf{w}^{\mathrm{T}}\mathbf{x}_i - b)^2 + \frac{\lambda}{2}\left\|\mathbf{w}\right\|_2^2$$

with λ as the regularization parameter controlling the penalty applied to the L^2 norm of the weights. Solving this optimization problem for the matrix of input features, $\mathbf{X}$, and target vector, $\mathbf{y}$, gives

$$\mathbf{w} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$$

This solution stabilizes the model by avoiding overfitting and ensuring numerical stability through regularization.

The optimum ridge parameter was λ = 10.0 after standardization (see below). With a total of 11 input features, the LR (Ridge) model contains 12 trainable coefficients (11 weights plus one bias). Because we restrict ourselves to linear terms in the regression, it cannot capture nonlinearity.

The KRR model[37] extends this L^2 Ridge Regression by introducing the concept of kernel matrix. Unlike the LR approach, instead of modeling the relationship between input features and target values directly, KRR maps input data $\mathbf{x}$ into a higher-dimensional feature space via a mapping function, $\Phi(\mathbf{x})$.

In this feature space, the weighted sum, $\mathbf{w}$, becomes

$$\mathbf{w} = \sum_{i=1}^{n}\alpha_i\Phi(\mathbf{x}_i),\ \boldsymbol{\alpha} = (\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y}$$

Here, the kernel matrix, $\mathbf{K}$, enters inside the coefficient matrix, α, with entries $K(x_i, x_j) = \Phi(\boldsymbol{x}_i)^{\mathrm{T}}\Phi(\mathbf{x}_j)$, and λ as before, is the regularization parameter. This kernel trick ensures that $\mathbf{w}$ depends only on the training data, even if the dimensionality of the feature space exceeds the number of data points.

In reality, since the model operates directly on kernel matrix $\mathbf{K}$, explicit computation of $\Phi(\mathbf{x})$ is avoided. This gives the prediction for a new data point $\mathbf{x}$ as

$$\hat{y}(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\Phi(\mathbf{x}) = \sum_{i=1}^{n}\alpha_i K(\mathbf{x}, \mathbf{x}_i) = \mathbf{k}(\mathbf{x})^{\mathrm{T}}(\mathbf{K} + \lambda\mathbf{I})^{-1}\mathbf{y}$$

where $\mathbf{k}(\mathbf{x})$ is a vector with ith component $K(\mathbf{x}, \mathbf{x}_i)$. Now, if one chooses a Gaussian kernel

$$K(x_i, x_j) = \exp\left(-\frac{\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2}{2\sigma^2}\right)$$

the model effectively handles high-dimensional feature spaces. Here, σ is the Gaussian kernel width.

The optimum hyperparameters were λ = 0.1, σ = 0.1. In KRR, the coefficient α_i has one entry per training sample, so the model stores the same number of effective parameters as the total number of training data (on the order of 1000).

The RF model,[38] an ensemble-based technique, constructs multiple decision trees during training. For an input $\mathbf{x}$, each tree, $T_m(\mathbf{x})$, is grown from a random bootstrap sample of the data and uses a randomly selected subset of features at each split, as

$$T_m(\mathbf{x}) = f(\mathbf{x}|\Theta_m)$$

where $f(\cdot)$ is the tree-based decision function, and Θ_m captures the random decisions made while building the mth tree. Then, the overall prediction for $\mathbf{x}$ is obtained by averaging the outputs of all trees

$$y = \frac{1}{M}\sum_{m=1}^{M}T_m(\mathbf{x})$$

with M as the total number of trees.

Being a tree-based approach, RF estimates how much each feature influences the overall model performance. In RF, this feature importance evaluation is done either by aggregating contributions to impurity reduction or by using permutation-based methods.[39] However, it is important here to note that interpreting feature importance when dealing with correlated features requires caution to avoid misleading conclusions.[39]

In our current work, all the ML calculations were performed using scikit-learn,[40] a Python-based package. For each of these three models, to optimize model performance, hyperparameters were tuned by using grid-based cross-validation. For the LR model, the Ridge regularization parameter λ was tested over values {0.1, 0.5, 1.0, 10.0}. For KRR, both λ and the Gaussian kernel width σ were varied across {0.1, 0.5, 1.0, and 10.0}. In the case of RF, the maximum tree depth was varied across ({2, 4, 8}) and the number of trees was varied across ({10, 20, 50, 80, 100}).

The optimum parameters were 80 trees and a maximum tree depth of eight. With these settings, each tree can hold up to $2^8 - 1 = 255$ decision nodes, yielding at most $80 \times 255 \approx 2.0 \times 10^4$ split parameters. This large number of parameters gives the model flexibility but can also pose a risk of overfitting. Additionally, when working with correlated features, RF can be biased toward particular features, affecting its performance.[39]

To ensure stable convergence, feature standardization[41] was implemented for all ML models, imposing that each feature

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

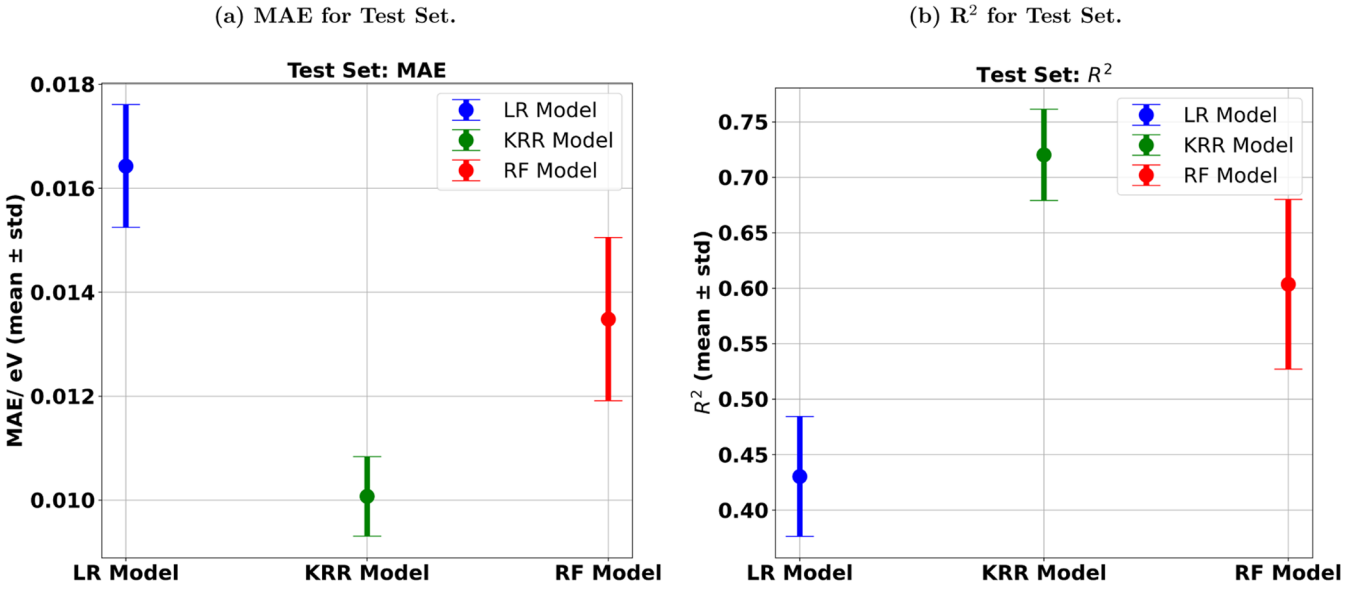


Figure 5. (a) MAE and (b) R^2 values for the three ML models evaluated on the test set. Maximum absolute errors for LR, KRR, and RF are 0.1, 0.1, and 0.08 eV, respectively (see full distributions in Figure S5).

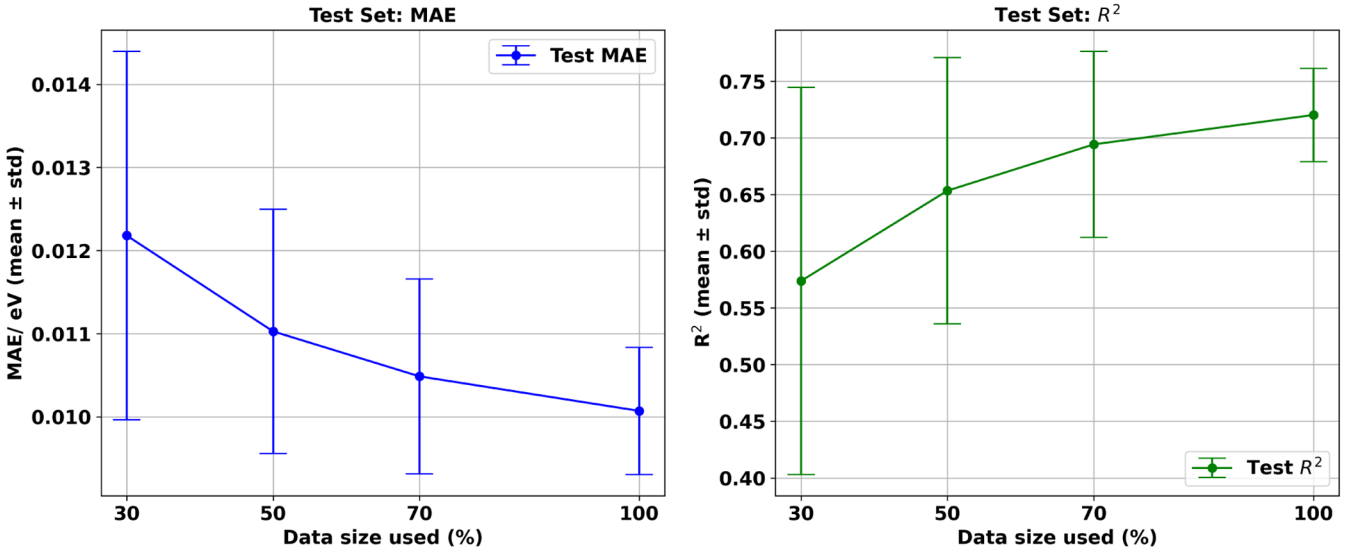


Figure 6. Impact of training data set size on the performance of the KRR model.

had zero mean and unit variance. For the purpose of splitting the overall data into training and test sets, the data set was divided in a 4:1 ratio, and a 10-fold cross-validation was performed on the training set, while the testing set was reserved for final evaluation.

Finally, the performance of all models was assessed based on mean absolute error (MAE) and the coefficient of determination (R^2)

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i|$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

where N is the total number of samples, y_i are true values, $\hat{y}_i$ are predicted values, and $\bar{y}$ is the mean of the true values.

In addition, the maximum absolute error has been calculated.

RESULTS AND DISCUSSION

Performance Comparison of ML Models. To evaluate the performance of the three ML models (LR, KRR, and RF), MAE and R^2 scores on the same test set (Figure 5) were compared. Each data point with an error bar represents the mean and standard deviation from 10-fold cross-validation.

KRR achieved the best performance among these three models, with an MAE of approximately 10 meV (as a comparison, it is worth mentioning here that the mean electronic coupling value of our data set is 36 meV) and an R^2 score of around 0.73 on the test set. On the training set (Figure S4), KRR exhibited an MAE of about 7 meV and an R^2 around 0.9. The difference in these metrics between the training and test sets is relatively modest, suggesting that overfitting is not severe for KRR. Despite its strong performance, further

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

improvement may be limited by some additional complexity of the system (e.g., internal degrees of freedom of the molecules) and inherent uncertainties in the data imposed by our electronic coupling scheme. This is also reflected by the comparatively large maximum absolute error of 0.1 eV, suggesting that for some configurations substantial outliers are present. In this context, however, one has to keep in mind that also range of potential coupling values varies over several orders of magnitude.

The RF model performs worse than the KRR on the test set, with an MAE of approximately 13 meV and an R^2 of around 0.6 (shown as red bars in Figure 5). For the training set (Figure S4), RF achieved an MAE of about 8 meV and an R^2 around 0.9, indicating that the RF model partially overfits. The weaker performance of RF on the test set can be attributed to its sensitivity to correlated features.[39] Highly correlated features (as shown in the correlation matrix in Figure S9) can cause RF to overemphasize or underemphasize certain features, reducing its predictive accuracy. In contrast, the KRR model, through a kernel mapping into a high-dimensional feature space, is better able to accommodate such correlations by distributing the weight more evenly across correlated features.

The LR model showed the poorest performance among the three, with the largest MAE (around 16 meV) and the lowest R^2 (around 0.4) on both training and test sets. Since LR showed significantly poorer performance compared to KRR, this clearly implies the importance of capturing nonlinear dependencies in the ML models.

To assess how the accuracy of models is affected by increasing the training data volume, we subsampled the data set at 30, 50, and 70% of its total size, in addition to using the full data set (100%). For each case, the data were split into training and test sets in a 4:1 ratio, and 10-fold cross-validation was applied to the training set. To ensure variability along the data set, each time, the subsampling process was repeated 10 times with different random selections of data, and the average performance was computed.

As shown in Figure 6, the test-set MAE of the KRR model steadily improved from about 12 meV (for 30% data) to 10 meV (for 100% data), and the R^2 score increased from below 0.6 to approximately 0.73. Furthermore, the linear fit of the plot of inverse of R^2 versus the inverse of the data size used (see Figure S8a) reveals that as the training data set size tends to infinity, the R^2 score converges to 0.82 (see the SI for details). This implies that the KRR model cannot achieve perfect predictions ($R^2 \rightarrow 1$). This limitation is likely due to three factors: First, although KRR performs best among the three models, it may still be insufficient to capture situations such as abrupt changes in the electronic coupling values that occur with only slight variations in a feature value. Second, the feature space may still be incomplete. Under our rigid-body assumption, the selected features are all intermolecular. To verify whether intramolecular orientations are also important, a set of intramolecular distances (e.g., $N^{d(a)} - O^{d(a)}$, $C_1^{d(a)} - C_2^{d(a)}$, etc.) and intramolecular angles (e.g., $O^{d(a)} - N^{d(a)} - C_1^{d(a)}$, $O^{d(a)} - N^{d(a)} - C_2^{d(a)}$, $C_1^{d(a)} - N^{d(a)} - C_2^{d(a)}$, etc.) were thoroughly examined, but no significant improvement in ML performance was observed. Nonetheless, due to the large degrees of freedom present in our TEMPO−TEMPO$^+$ system, some important orientations may still be missing. Third, the error may also arise directly from the inherent noise in our TDDFT/GMH coupling data set, even though the final data set was obtained via multiple screening steps.

Figure S6 demonstrates less pronounced but still noticeable improvements for RF if more data is added. Figure S8b shows that as the data set size tends to infinity, the R^2 score converges to 0.69. As discussed before, in this correlated feature space, RF cannot achieve performance as good as KRR. Figures S7 and S8c show that unlike KRR or RF, LR exhibits minimal performance improvement with an increased training data set (R^2_∞(LR) ≈ 0.46). This difference with KRR (R^2_∞(KRR) ≈ 0.82), again highlights the importance of incorporating nonlinearities in the ML models.

In a subsequent section, we employ the KRR model to explore the molecular orientations that most significantly influence electronic coupling.

Feature Importance. Since the KRR model demonstrated the best performance among all three models, it was used to identify which molecular orientations most strongly influence electronic coupling. An initial global assessment of feature importance was conducted by systematically removing individual features—as well as sine and cosine pairs of the same angles, from the feature space and then evaluating the resulting MAE of the model (see Figure S10). If the MAE increases upon removal of a certain feature (or feature pair), it shows that the feature has a strong "global" impact on the model's performance.

Figure S10 shows that the feature pair associated with $\theta 2$ is the most influential, as its removal causes the highest increase in the MAE (around 15 meV). The next most impactful pairs are related to ψ and $\theta 1$, whose removal results in MAE values around 13 meV. Interestingly, even removing the single feature N^d−N^a increases the MAE to nearly 12 meV. This highlights that even in this short N^d−N^a range (4−4.5 Å), electronic coupling is heavily distance-dependent. Among all these features, the sine-cosine pairs of $\phi 1$ and $\phi 2$ have the least impact on the overall performance.

However, it should be noted that this feature importance approach can become biased in the presence of correlated features, similar to the permutation-based feature importance methods used in Random Forest models.[39] In fact, the correlation matrix (Figure S9) in our feature space shows correlations not only between sine and cosine functions of the same angle (e.g., $\sin(\theta 1)$ and $\cos(\theta 1)$) but also between different angles (e.g., $\sin(\theta 1)$ and $\sin(\phi 1)$). This also implies that removing one feature (e.g., $\sin(\theta 1)$) or a pair (e.g., $\sin(\phi 1)$ and $\cos(\phi 1)$) may not significantly degrade the model's performance because the model can rely on their correlated counterparts.

Given these correlations in our feature space, a more robust technique is needed to quantify each feature's contribution in the presence of other correlated features. This situation is very similar to the Game Theory problem, where the contribution of one player is dependent on how other players are active. In Game theory, SHapley Additive exPlanations (SHAP)[42−44] provides an interesting framework to attribute contribution from each player. So, in our study, to assess feature importance efficiently for this correlated feature space, the SHAP approach is utilized.

This approach starts with the concept of a baseline prediction, $E[f_{full}(\boldsymbol{x})]$, representing the average model output over the entire data set $\mathbf{x} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_m\}$

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

$$E[f_{\text{full}}(\mathbf{x})] = \frac{1}{m}\sum_{k=1}^{m} f_{\text{full}}(\mathbf{x}_k)$$

where $f_{\text{full}}(\boldsymbol{x}_k)$ denotes the model's prediction for data point $\boldsymbol{x}_k$ using all the features in the feature space. Then, $f_{\text{full}}(\boldsymbol{x}_k)$ decomposes into

$$f_{\text{full}}(\mathbf{x}_k) = E[f_{\text{full}}(\mathbf{x})] + \sum_{i=1}^{n} C_i$$

Here, the SHAP value of the ith feature, C_i, is introduced, which can intuitively be thought of as the influence of feature i on the data point, $\boldsymbol{x}_k$, to decide how much the model's prediction will be changed compared to the average model output over the entire data set.

To compute C_i, the set of all feature indices, N, is divided into subsets $S \subseteq N \setminus \{i\}$, as

$$C_i = \sum_{S\subseteq N\setminus\{i\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[f_{S\cup\{i\}}(\mathbf{x}_k) - f_S(\mathbf{x}_k)]$$

Here, $f_S(\mathbf{x}_k)$ denotes the model's prediction using only the features in subset, S, and $[f_{S\cup\{i\}}(\mathbf{x}_k) - f_S(\mathbf{x}_k)]$ measures the marginal change of feature i when added to S. So, computing the SHAP value for a specific feature requires examining this marginal change over all possible subsets that exclude this specific feature. Although this approach can be computationally expensive for large feature sets, it remains feasible for our study due to the small number of features.

Intuitively, one can imagine a vertical reference line at mean, $E[f_{\text{full}}(\mathbf{x})]$, which represents the SHAP value zero (baseline). For each data point, positive SHAP value for each feature pushes the prediction to the right (above the mean), whereas negative SHAP value pushes it to the left (below the mean) of the vertical reference. In this way, it can be visualized how each feature influences each data point "locally".

For this analysis, a python-based SHAP package[45] was used. With this visualization approach, using the KRR model, SHAP values for all features in our data set, for each data point, are plotted in Figure 7. In this plot, each row corresponds to a feature and the dots represent SHAP values for individual data points. In a feature row, the dot colors represent how large or small this particular feature is on a relative scale for each data point (blue refers to low feature values, and red refers to high feature values). Globally, the features are ranked by their mean absolute SHAP value (shown in green next to each feature name). Notably, sin(θ2), sin(ψ), and N^d–N^a all show similar mean absolute SHAP values (~0.0056–0.006), suggesting they dominate the model's performance at a global level.

In particular, N^d–N^a shows a behavior that aligns well with physical intuition: longer distances (red dots) are associated with negative SHAP values, reducing the predicted electronic coupling below the average, while shorter distances (blue dots) yield positive SHAP values, enhancing the electronic coupling predictions above the average model performance. This is consistent with the idea that larger separations diminish electronic coupling, whereas shorter distances enhance it.

SHAP plot also highlights the locally extreme cases (which might be outliers), either far to the right or to the left of the baseline. Table 1 highlights three such instances along with all of the orientation parameters related to the respective data points. In the first case, sin(θ1) has a highly negative SHAP value, implying that this feature lowers the prediction for the

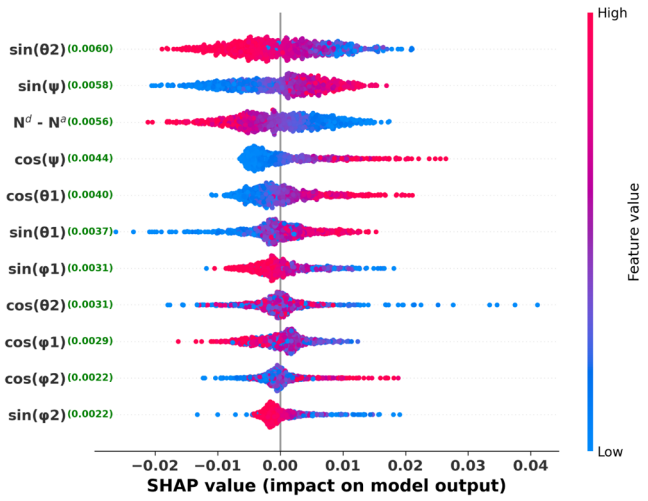


Figure 7. SHAP summary plot showing the contribution of individual features to the model predictions. Each row corresponds to a specific feature, and each dotted line represents the SHAP value of that feature for a single data point. The vertical line, corresponding to the average of the model prediction over all of the data points, is set at zero (baseline). The SHAP value indicates the feature's impact on the model output, with positive values pushing predictions to the right and negative values pushing them to the left of the baseline. The features are ranked in descending order based on their mean absolute SHAP value, shown in green next to each feature name.

respective data point compared to the average model prediction (baseline). Angular values for this data point show that $\theta1 \approx 180°$. Histogram of the angle, $\theta1$ (see Figure 4b), suggests that in our data set, there are few data points corresponding to this value. Also, with $\theta1 \approx 180°$, one can imagine a certain NO–NO orientation, which makes the orbital overlap very low. So, this can imply that this orientation largely lowers the prediction compared to the average value over all the data points present in our data set. This can be further verified with the electronic coupling vs $\theta1$ plot in Figure S11b, where both the running average and linear fit imply a trend of minimum electronic coupling at this $\theta1$ value. Hence, there is a large negative SHAP contribution. A detailed list of data points corresponding to the maximum and minimum SHAP values for each feature is included in Table S2.

By contrast, two data points with especially high positive SHAP values occur for cos(θ2) and cos(ψ). The data point with the highest SHAP value for the cos(θ2) feature corresponds to a $\theta2$ value of approximately 106°, while the data point with the highest SHAP value for the cos(ψ) feature corresponds to a ψ value of approximately 80°, both of which are less frequent in our data set (see Figure S2c,a, respectively). For these two data points, positive SHAP values can be attributed to the electronic coupling vs $\theta2$ (Figure S11c) and electronic coupling vs ψ trends (Figure S11a), where in both cases, the running average suggests a trend of increasing electronic coupling at these $\theta2$ and ψ values, respectively. Even though the linear fit, particularly for the electronic coupling vs $\theta2$ plot, shows a slightly decreasing trend at $\approx 106°$, this may be a result of the very low linearity ($R^2 \approx 0.01$) as well as the limited availability of data points in this region.

As the SHAP analysis indicates, $\theta2$ is one of the crucial factors in determining the electronic coupling, in both a global and local sense. To further examine its role, we explored how

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

Table 1. Local SHAP Value Analysis for Selected Features

feature	SHAP value	coupling (meV)	$\theta1$	$\theta2$	ψ	$\phi1$	$\phi2$	N^d-N^a (Å)
sin($\theta1$)	min: −0.0263	3.5	178.15	72.10	75.08	86.94	82.08	4.1956
cos($\theta2$)	max: 0.0411	119.7	158.28	106.28	143.27	76.49	54.94	4.3557
cos(ψ)	max: 0.0264	82.1	167.78	69.87	79.61	94.63	33.05	4.3408

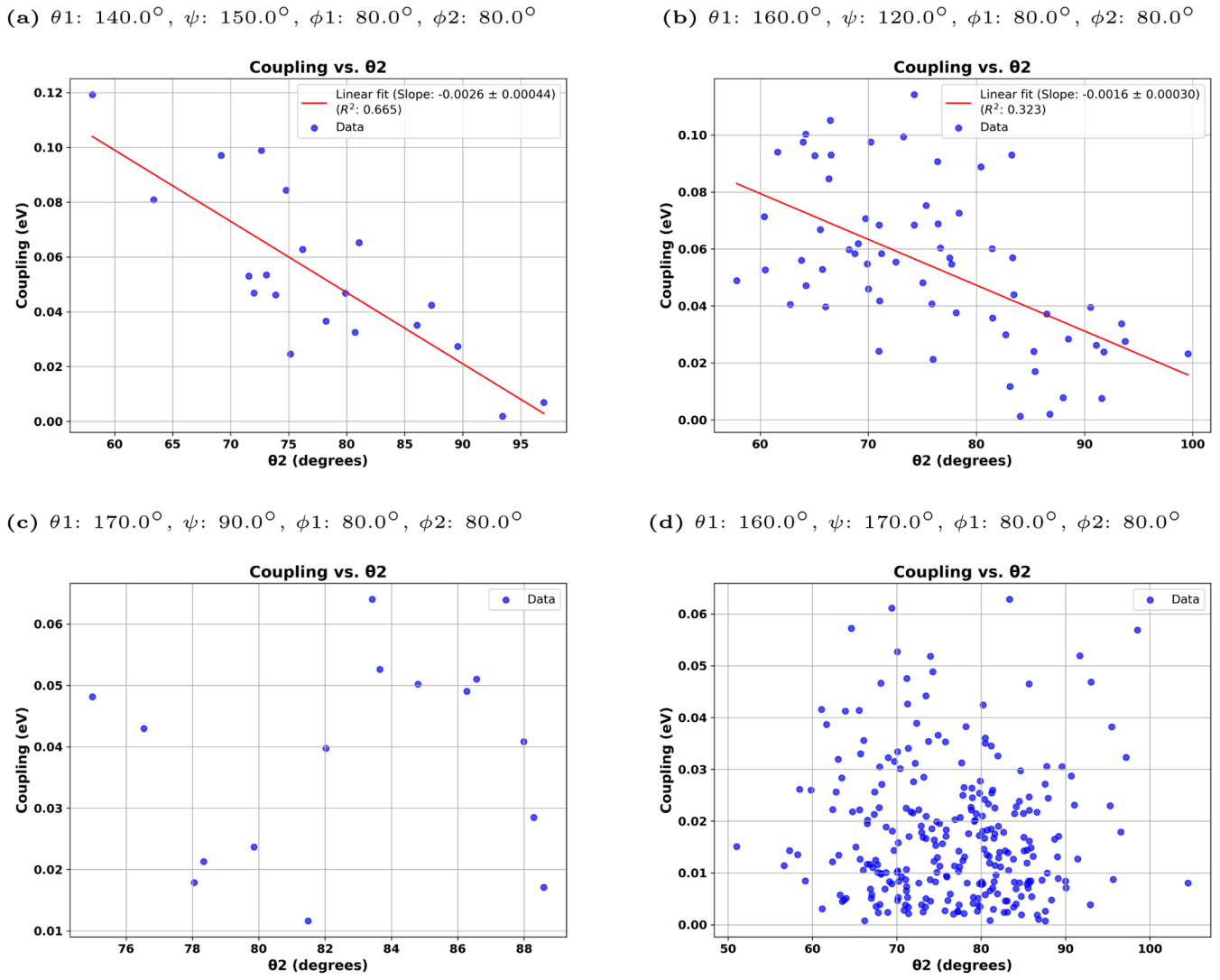


Figure 8. Dependency of the electronic coupling on $\theta2$, for four different sets of $\theta1$, and ψ.

the electronic coupling changes as a function of $\theta2$, while keeping other angles within a range of $\delta \pm 10°$ around specific reference values (Figure 8). These plots confirm that $\theta2$ can strongly influence electronic coupling, although the extent of this dependence varies heavily depending on the other angles, especially the angle of ψ.

For instance, in Figure 8a,8b, linear fits show negative slopes, implying that smaller $\theta2$ values promote higher electronic coupling. In Figure 8a, where $\theta1 = 140°$, $\psi = 150°$, and $\phi1 = \phi2 = 80°$, the correlation is stronger ($R^2 = 0.665$), whereas in Figure 8b ($\theta1 = 160°$, $\psi = 120°$), the slope is still negative but the correlation is weaker ($R^2 = 0.323$).

By contrast, certain orientations can show no noticeable dependence on $\theta2$, suggesting a negligible impact from $\theta2$ (Figure 8c,d).

For example, in Figure 8c, $\psi = 90°$, which implies that this perpendicular NO−NO orientation diminishes the relevance of $\theta2$. Similarly, in Figure 8d, $\psi = 180°$, which implies that an antiparallel NO−NO orientation also suppresses any clear $\theta2$ trend.

Overall, these findings reinforce the SHAP-based observation that in the short N^d−N^a data range, no single orientation alone can govern the electronic coupling between TEMPO−TEMPO$^+$ pairs. Rather, a combinatory effect, especially arising from the NO−NO orientations (angles: $\theta1$, $\theta2$, ψ; distance: N^d−N^a), plays a pivotal role in determining the electronic coupling. The influence of one parameter on the electronic coupling is heavily dependent on the range in which the other parameters lie. For example, in some geometries, small changes in $\theta2$ can increase orbital overlap, amplifying the electronic coupling, while in other alignments, $\theta2$ has far less effect. Thus, understanding the orientation dependence of electronic coupling is crucial for designing and tuning electron transfer reactions in TEMPO-based ORBs.

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678

CONCLUSIONS

In this ML-based study, we present a predictive workflow for the electronic coupling between TEMPO and TEMPO$^+$ pairs in TEMPO-based ORBs. For the sample structure generation, we opt for classical MD simulation by replicating a similar electrolyte environment around TEMPO-based electrode. For accurate electronic coupling calculations, we have chosen (TD)DFT/GMH-based calculations using the ωB97XD3 functional and ma-def2-TZVP basis set, as motivated from our previous work.[11] These computationally expensive calculations limited us to having a moderate data set (around 1300 data points) for our ML models. To get a reliable and consistent data set, a proper screening was performed that eliminated any pairs for which TDDFT calculations impose delocalization errors. For the feature engineering, we focused on specific geometric orientation-based descriptors for two reasons: first, to understand how exactly specific orientation influences the electronic coupling, and second, because this handpicked feature space would be much smaller compared to the generic features like SOAP or CM, which, due to their larger feature space, have a high risk of overfitting when dealing with moderate−sized data sets.

For the ML models, a comparison between LR, KRR, and RF was performed. Our results show that KRR, due to its kernel-based approach, works best in our correlated feature space, whereas both RF and LR are unable to efficiently deal with such features. A slight data set-size dependence for our KRR model suggests that increasing the data size can improve the model's performance by increasing data points in the less frequent region of our data set, which represents less stable configurations. A similar dependence on data size was also observed in the work of Wang et al. in their ML study on predicting the electronic coupling of ethylene molecules.[13]

Finally, to understand how important each orientation in this correlated feature space is in determining the electronic coupling, we used the SHAP analysis technique from game theory, which not only provides the feature importance "globally" but also "locally" for each data point. SHAP analysis showed that there is no single orientation alone that influences the electronic coupling; rather, a combinatory effect, especially arising from angles and distance related to the NO−NO orientations (angles: $\theta1$, $\theta2$, ψ; distance: N^d−N^a) is important to decide the model's prediction. One particular angle's (or distance's) influence on the electronic coupling can be varied heavily based on the values of other orientational factors. For example, when the dihedral angle between NO and NO is either 90° (perpendicular) or 180° (antiparallel), the angle $\theta2$ (which represents the tilt of the donor N on the acceptor NO) has no significant effect on electronic coupling, whereas in most other NO−NO alignments, an increase in $\theta2$ leads to a decrease in electronic coupling.

This approach can be easily adapted to similar systems in which the frontier orbitals are rather localized. Of course, for delocalized orbitals in, for example, pi-stacked monomers, the features might be extended or modified. As a general orientation description, one could use the principal components of the gyration tensor.

We believe that our current work highlights an efficient framework for predicting electronic coupling between TEMPO−TEMPO$^+$ pairs based on specific geometric orientations.

For example, in a reactive MD approach, in which classical MD simulations are augmented by physics-derived reaction rates to model chemical reactivity,[46−49] our scheme could be utilized for a computationally efficient on-the-fly estimation of the electronic coupling when modeling electron transfer in polymer materials.

In reality, the actual ORB cathodes, being multicomponent systems, the typical orientations can depend on the solvent composition employed as well as salt concentration.[18] Hence, tuning these factors will definitely impact electron transfer reactions in TEMPO-based ORBs by favoring certain geometric orientations.

ASSOCIATED CONTENT

Data Availability Statement

All the data sets and ML scripts used in this work are available at Zenodo (10.5281/zenodo.15003344).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcc.5c02094.

Atomic site charges of TEMPO; histograms for distance and angles; histograms for electronic coupling and transition dipole moment; ML models performance for training set; distribution of absolute errors for the ML models; data set-size dependence of RF and LR models; derivation of the dependence of R^2 on data size; correlation matrix for the features; feature importance based on removed features; electronic coupling vs angles; SHAP analysis for all the features; frontier orbitals for selected configurations with minimum and maximum SHAP value; and screening approach based on the ML models (PDF)

AUTHOR INFORMATION

Corresponding Author

Diddo Diddens − *Helmholtz-Institute Münster (IMD-4), Münster 48149, Germany;* orcid.org/0000-0002-2137-1332; Email: d.diddens@fz-juelich.de

Authors

Souvik Mitra − *Institute of Physical Chemistry, Universität Münster, Münster 48149, Germany;* orcid.org/0009-0005-9476-980X

Clara Zens − *Institute of Physical Chemistry, Friedrich Schiller University Jena, Jena 07743, Germany*

Stephan Kupfer − *Institute of Physical Chemistry, Friedrich Schiller University Jena, Jena 07743, Germany;* orcid.org/0000-0002-6428-7528

Andreas Heuer − *Helmholtz-Institute Münster (IMD-4), Münster 48149, Germany; Institute of Physical Chemistry, Universität Münster, Münster 48149, Germany;* orcid.org/0000-0003-2592-0287

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcc.5c02094

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

S.M., A.H., and D.D., as well as C.Z. and S.K., acknowledge funding by the Deutsche Forschungsgemeinschaft [Priority Program SPP 2248 "Polymer-based Batteries" (project

numbers 441255373 and 441265816)]. The authors are grateful to Dr. Josef Granwehr, Dr. Davis Thomas Daniel, and Dr. Moumita Maiti for their valuable insights. Sukanya Majumdar is also acknowledged for her guidance.

REFERENCES

(1) Chang, Z.; Henkensmeier, D.; Chen, R. Shifting redox potential of nitroxyl radical by introducing an imidazolium substituent and its use in aqueous flow batteries. *J. Power Sources* **2019**, *418*, 11−16.
(2) Karlsson, C.; Suga, T.; Nishide, H. Quantifying TEMPO Redox Polymer Charge Transport toward the Organic Radical Battery. *ACS Appl. Mater. Interfaces* **2017**, *9*, 10692−10698.
(3) Kubala, D.; Regeta, K.; Janečková, R.; Fedor, J.; Grimme, S.; Hansen, A.; Nesvadba, P.; Allan, M. The electronic structure of TEMPO, its cation and anion. *Mol. Phys.* **2013**, *111*, 2033−2040.
(4) Kemper, T. W.; Larsen, R. E.; Gennett, T. Relationship between Molecular Structure and Electron Transfer in a Polymeric Nitroxyl-Radical Energy Storage Material. *J. Phys. Chem. C* **2014**, *118*, 17213−17220.
(5) Marcus, R. A. On the Theory of Electron-Transfer Reactions. VI. Unified Treatment for Homogeneous and Electrode Reactions. *J. Chem. Phys.* **1965**, *43*, 679−701.
(6) Marcus, R.; Sutin, N. Electron Transfers in Chemistry and Biology. *Biochim. Biophys. Acta, Rev. Bioenerg.* **1985**, *811*, 265−322.
(7) Marcus, R. A. Electron Transfer Reactions in Chemistry: Theory and Experiment (Nobel Lecture). *Angew. Chem., Int. Ed.* **1993**, *32*, 1111−1121.
(8) Hush, N. Distance Dependence of Electron Transfer Rates. *Coord. Chem. Rev.* **1985**, *64*, 135−157.
(9) Marcus, R.; Sutin, N. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta, Rev. Bioenerg.* **1985**, *811*, 265−322.
(10) Blumberger, J. Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions. *Chem. Rev.* **2015**, *115*, 11191−11238.
(11) Mitra, S.; Zens, C.; Kupfer, S.; Diddens, D. Toward robust electronic coupling predictions in redox-active TEMPO/TEMPO+ systems. *J. Chem. Phys.* **2024**, *161*, No. 214106.
(12) Blancafort, L.; Voityuk, A. A. CASSCF/CAS-PT2 Study of Hole Transfer in Stacked DNA Nucleobases. *J. Phys. Chem. A* **2006**, *110*, 6426−6432.
(13) 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.
(14) Baumeier, B.; Kirkpatrick, J.; Andrienko, D. Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11103−11113.
(15) Valeev, E. F.; Coropceanu, V.; da Silva Filho, D. A.; Salman, S.; Brédas, J.-L. Effect of Electronic Polarization on Charge-Transport Parameters in Molecular Organic Semiconductors. *J. Am. Chem. Soc.* **2006**, *128*, 9882−9886.
(16) Daniel, D. T.; Oevermann, S.; Mitra, S.; Rudolf, K.; Heuer, A.; Eichel, R.-A.; Winter, M.; Diddens, D.; Brunklaus, G.; Granwehr, J. Multimodal investigation of electronic transport in PTMA and its impact on organic radical battery performance. *Sci. Rep.* **2023**, *13*, No. 10934.
(17) Daniel, D. T.; Mitra, S.; Eichel, R.-A.; Diddens, D.; Granwehr, J. Machine Learning Isotropic g Values of Radical Polymers. *J. Chem. Theory Comput.* **2024**, *20*, 2592−2604.
(18) Mitra, S.; Heuer, A.; Diddens, D. Electron transfer reaction of TEMPO-based organic radical batteries in different solvent environments: comparing quantum and classical approaches. *Phys. Chem. Chem. Phys.* **2024**, *26*, 3020−3028.
(19) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157−2164.
(20) Abraham, M.; Alekseenko, A.; Bergh, C.; Blau, C.; Briand, E.; Doijade, M.; Fleischmann, S.; Gapsys, V.; Garg, G.; Gorelov, S.et al. GROMACS 2023 Manual. 2023 DOI: 10.5281/zenodo.7588711.
(21) Rapaport, D. *The Art of Molecular Dynamics Simulation*; Cambridge University Press, 2004.
(22) 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.
(23) Dunning, T. H.; Thom, H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(24) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.et al. *Gaussian16 Revision C.01*; Gaussian Inc.: Wallingford CT, 2016.
(25) Chai, J.-D.; Head-Gordon, M. Long-Range Corrected Hybrid Density Functionals with Damped Atom-Atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(26) Zheng, J.; Xu, X.; Truhlar, D. G. Minimally augmented Karlsruhe basis sets. *Theor. Chem. Acc.* **2011**, *128*, 295−305.
(27) Hellweg, A.; Rappoport, D. Development of new auxiliary basis functions of the Karlsruhe segmented contracted basis sets including diffuse basis functions (def2-SVPD, def2-TZVPPD, and def2-QVPPD) for RI-MP2 and RI-CC calculations. *Phys. Chem. Chem. Phys.* **2015**, *17*, 1010−1017.
(28) Neese, F. Software update: the ORCA program system, version 5.0. *WIRES Comput. Mol. Sci.* **2022**, *12*, No. e1606.
(29) Neese, F. The ORCA program system. *WIREs Comput. Mol. Sci.* **2012**, *2*, 73−78.
(30) Lu, T.; Chen, F. Multiwfn: A multifunctional wavefunction analyzer. *J. Comput. Chem.* **2012**, *33*, 580−592.
(31) Gugler, S.; Reiher, M. Quantum Chemical Roots of Machine-Learning Molecular Similarity Descriptors. *J. Chem. Theory Comput.* **2022**, *18*, 6670−6689.
(32) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, No. 058301.
(33) 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.
(34) 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.
(35) Lederer, J.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine Learning−Based Charge Transport Computation for Pentacene. *Adv. Theory Simul.* **2019**, *2*, No. 1800136.
(36) Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning*, 2nd ed.; Springer: New York, NY, USA, 2009.
(37) Schölkopf, B.; Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
(38) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(39) Gregorutti, B.; Michel, B.; Saint-Pierre, P. Correlation and variable importance in random forests. *Stat. Comput.* **2017**, *27*, 659−678.
(40) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825−2830.
(41) Bishop, C. M. *Pattern Recognition and Machine Learning*; Springer, 2006.
(42) Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. and Inf. Syst.* **2014**, *41*, 647−665.
(43) Lundberg, S.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions, 2017. arXiv:1705.07874. arXiv.org e-Print archive. https://arxiv.org/abs/1705.07874.

(44) Wood, T. R.; Kelly, C.; Roberts, M.; Walsh, B. An interpretable machine learning model of biological age. *F1000Research* **2019**, *8*, No. 17.
(45) Lundberg, S. M.; Lee, S.-I. SHAP: SHapley Additive exPlanations 2017 https://shap.readthedocs.io/en/latest/.
(46) Biedermann, M.; Diddens, D.; Heuer, A. rs@ md: Introducing reactive steps at the molecular dynamics simulation level. *J. Chem. Theory Comput.* **2021**, *17*, 1074−1085.
(47) Biedermann, M.; Diddens, D.; Heuer, A. Connecting the quantum and classical mechanics simulation world: Applications of reactive step molecular dynamics simulations. *J. Chem. Phys.* **2021**, *154*, No. 10.1063/5.0048618.
(48) Abbott, J. W.; Hanke, F. Kinetically corrected monte carlo−molecular dynamics simulations of solid electrolyte interphase growth. *J. Chem. Theory Comput.* **2022**, *18*, 925−934.
(49) Mabrouk, Y.; Safaei, N.; Hanke, F.; Carlsson, J.; Diddens, D.; Heuer, A. Reactive molecular dynamics simulations of lithium-ion battery electrolyte degradation. *Sci. Rep.* **2024**, *14*, No. 10281.

https://doi.org/10.1021/acs.jpcc.5c02094
J. Phys. Chem. C 2025, 129, 14667−14678