Rapid Estimation of the Intermolecular Electronic Couplings and Charge-Carrier Mobilities of Crystalline Molecular Organic Semiconductors through a Machine Learning Pipeline

Vinayak Bhat[1], Baskar Ganapathysubramanian[2] and Chad Risko*[1]*

[1]*Department of Chemistry &*
Center for Applied Energy Research
University of Kentucky
Lexington, Kentucky 40506, USA

[2]*Department of Mechanical Engineering &*
Translational AI Center
Iowa State University
Ames, Iowa, 50010, USA

Corresponding Authors

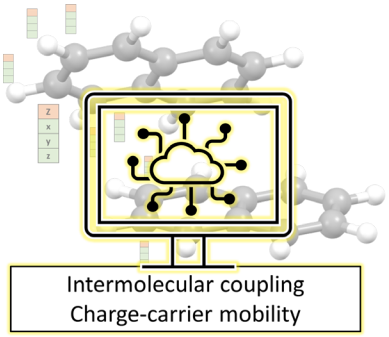
Chad Risko; *E-mail:* chad.risko@uky.edu

Baskar Ganapathysubramanian; *E-mail:* baskarg@iastate.edu

ABSTRACT

Organic semiconductors (OSC) offer tremendous potential across a wide range of (opto)electronic applications. OSC development, however, is often limited by trial-and-error design, with computational modeling approaches deployed to evaluate and screen candidates through a suite of molecular and materials descriptors that generally require hours to days of computational time to accumulate. Such bottlenecks slow the pace and limit the exploration of the vast chemical space comprising OSC. When considering charge-carrier transport in OSC, a key parameter of interest is the intermolecular electronic coupling. Here, we introduce a machine learning (ML) model to predict intermolecular electronic couplings in organic crystalline materials from their three-dimensional (3D) molecular geometries. The ML predictions take only a few seconds of computing time compared to hours by density functional theory (DFT) methods. To demonstrate the utility of the ML predictions, we deploy the ML model in conjunction with mathematical formulations to rapidly screen the charge-carrier mobility anisotropy for more than 60,000 molecular crystal structures and compare the ML predictions to DFT benchmarks.

TOC GRAPHIC



Intermolecular electronic couplings (or transfer integrals) in organic semiconductors (OSC) are critical parameters governing charge-carrier transport.[1-5] The intermolecular electronic couplings depend both on the geometric overlap of neighboring molecules (and, hence, intermolecular vibrational or phonon modes) and the molecular orbital (MO) overlap of these adjacent molecules – e.g., between the highest-occupied molecular orbitals (HOMO) of the two molecules (HOMO-HOMO coupling) for hole transport, or the lowest-unoccupied molecular orbitals (LUMO) of the two molecules (LUMO-LUMO coupling) for electron transport.[3, 6] The phases of the intermolecular electronic couplings are determined by the MO overlap symmetries.

Several approaches have been implemented to determine intermolecular electronic couplings.[4, 7-11] In the energy-splitting-in dimer method,[8, 12] the intermolecular electronic coupling is estimated to be one-half the energy difference between the HOMO and HOMO-1 of a (noncovalent) dimer formed by two adjacent molecules (see Figure 1 for representation of a molecular dimer geometry). While this method is effective for symmetrically arranged molecules, the method fails for systems where molecular asymmetry leads to polarization. This shortcoming is overcome in the fragment molecular orbital (FMO) approach, wherein an orthonormal basis is used to preserve the local character of the monomer orbitals.[10, 13, 14] Via the FMO approach, the effective intermolecular electronic coupling (V_{12}) between the adjacent molecules (denoted by the numbers 1 and 2) in a molecular dimer is given by

$$V_{12} = \frac{H_{12} - 0.5 \times S_{12} \times (H_1 + H_2)}{1 - {S_{12}}^2} \quad (1)$$

where H_{12} is the interaction energy or electronic coupling matrix, S_{12} is the overlap integral, and H_1 and H_2 are the monomer site energies. While there are extensions of the FMO approach beyond molecular dimers to include many-body interactions,[15] band dispersion interactions in crystalline organic systems are typically weak (0.1-0.5 eV), and restricting evaluations to molecular dimers yields reasonable

estimates of the intermolecular electronic couplings.[16] Application of the FMO approach typically makes use of wavefunction or density functional theory (DFT) calculations, which can afford high accuracy, though they can be time-consuming; semi-empirical wavefunction methods do offer faster evaluations, often at the cost of accuracy.

With determinations of the intermolecular electronic couplings in hand, OSC charge-carrier transport characteristics can be evaluated by including these descriptors with kinetic Monte Carlo methods,[17-19] molecular dynamics (MD) simulations,[20, 21] or transient localization theory.[22, 23] However, each of these approaches requires that a large number of intermolecular electronic couplings be evaluated with high accuracy and, ideally, limited computational cost. Recent efforts have sought to develop fast yet reliable machine learning (ML) models to predict intermolecular electronic couplings.[24-33] The underlying idea of training an ML model is to acquire accurate, near-real-time predictions of desired properties (also called fast online performance) while amortizing the cost via an expensive offline dataset creation, curation, and model training campaign. For intermolecular electronic couplings, a key step is that molecular dimer geometries must be transformed to ML model input. One of the commonly used transformations is the coulomb matrix,[34] wherein the matrix element between two atoms i and j is defined by

$$C_{ij} = \begin{cases} 0.5\, Z_i^{2.4} & \forall\ i = j \\ \dfrac{Z_i Z_j}{\left\| R_i - R_j \right\|} & \forall\ i \neq j \end{cases} \qquad (2)$$

where Z_i and Z_j represent the nuclear charges of the atoms, R_i and R_j are the atomic Cartesian position vectors, and $\left\| R_i - R_j \right\|$ is the Euclidean distance between atom i and atom j. Here, the size of the matrix depends on the number of atoms in the molecule. Gagliardi and coworkers used the coulomb matrix representation for dimer geometries as input for a kernel ridge regression (KRR)-based ML model[28] that was trained on a dataset that contained dimer geometries extracted from an MD simulation of a pentacene crystal. This work demonstrated that an ML model could be trained to yield intermolecular

electronic couplings for a pentacene crystal with low prediction error. Wang et al. adopted a similar approach and used 250,000 naphthalene dimers extracted from MD simulations to train a KRR and artificial neural network (ANN) model with a coulomb matrix.[27] ML models trained by this approach, however, are generally not transferable, as coulomb matrix sizes vary for different molecules; hence, one needs to generate data and train models for each new system to be investigated.[27, 35] The redundancy of training one model for every molecule can be reduced by training a single model with a fixed size coulomb matrix representation for all data, as demonstrated by Reiser et al.;[26] here, the size of the coulomb matrix representation for molecular dimers is fixed according to the largest molecule in the dataset, and the representation padded with zeros for smaller molecules to maintain the fixed size. Though this representation yields a transferable ML model for systems with a similar number of atoms, the large variation in atom numbers leads to sparsity in the representation, resulting in limited model performance.[35]

An alternative to the coulomb representation is the graph representation, where atoms correspond to nodes of the graph and bonds are the edges.[36] For example, if one considers benzene to be represented as a graph, the carbon and hydrogen atoms are represented by nodes, and the bonds between the atoms are represented as edges; node features include atom type and hybridization state, while edge features include bond type and length. We demonstrated in previous work that the graph representation coupled with a message-passing neural network (MPNN) can be used to predict electronic, redox, and optical properties of organic π-conjugated molecules with DFT-level accuracy.[37] In MPNN, information from neighboring nodes is aggregated and processed at each node. This allows the ML model to learn how the local environment influences each atom. Unlike the coulomb matrix, the graph representation does not depend on the number of atoms in the system; hence, the graph representation offers a more transferable approach compared to the coulomb matrix. Notably, graph representations have been previously proposed to predict intermolecular electronic couplings.[35]

In this work, we use graph representations to predict intermolecular electronic couplings from molecular dimer geometry using SphereNet, a graph-based three-dimensional (3D) MPNN.[38] For a 3D MPNN, the input representation includes 3D coordinates of each atom written in the graph format described above, thereby capturing the molecular spatial arrangements, a crucial feature for predicting properties dependent on molecular shape / structure and intermolecular interactions. SphereNet has been used to predict molecular properties such as the dipole moment, polarizability, and free energy from a 3D molecular geometry.[38] The input for training SphereNet used here includes the atomic Cartesian coordinates (x, y, z) of molecules in a dimer and corresponding atomic numbers (Z), as shown in Figure 1. These coordinates are then transformed into a 3D graph representation using spherical coordinates (d, θ, ϕ). A brief discussion on SphereNet is included in the SI; however, for more in-depth information on SphereNet, readers are referred to the work by Liu et al.[38] We demonstrate that the SphereNet architecture, when trained with a diverse dataset of 438,000 DFT-derived intermolecular electronic couplings from over 25,000 molecular crystal structures in the OCELOT (Organic Crystals in Electronic and Light-Oriented Technologies) database,[39] provides a highly transferable ML model. Furthermore, we develop and deploy an open-access ML pipeline that uses the predicted intermolecular electronic couplings to estimate charge-carrier mobility anisotropies within the semi-classical Marcus theory approach proposed by Goddard and coworkers;[40] the reorganization energy parameters used to derive the Marcus theory hopping are also predicted via a pre-trained ML model.[37] Using this ML pipeline, we screen over 60,000 molecular organic crystals for their capacity to transport charge carriers.

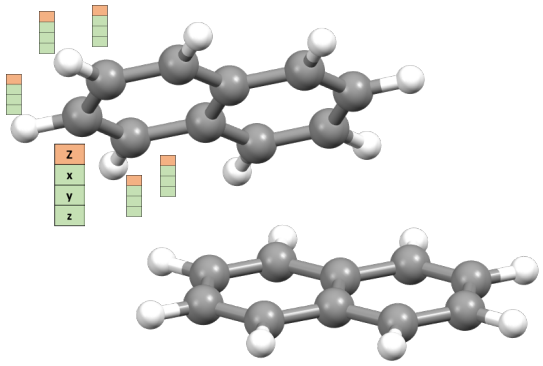


Figure 1. Representative molecular dimer geometry, which consists of two identical molecules at arbitrary relative separation and orientation. The atomic number (Z) and the Cartesian coordinates (x,y,z) features are labeled for the molecular dimer geometry and serve as input for the ML model.

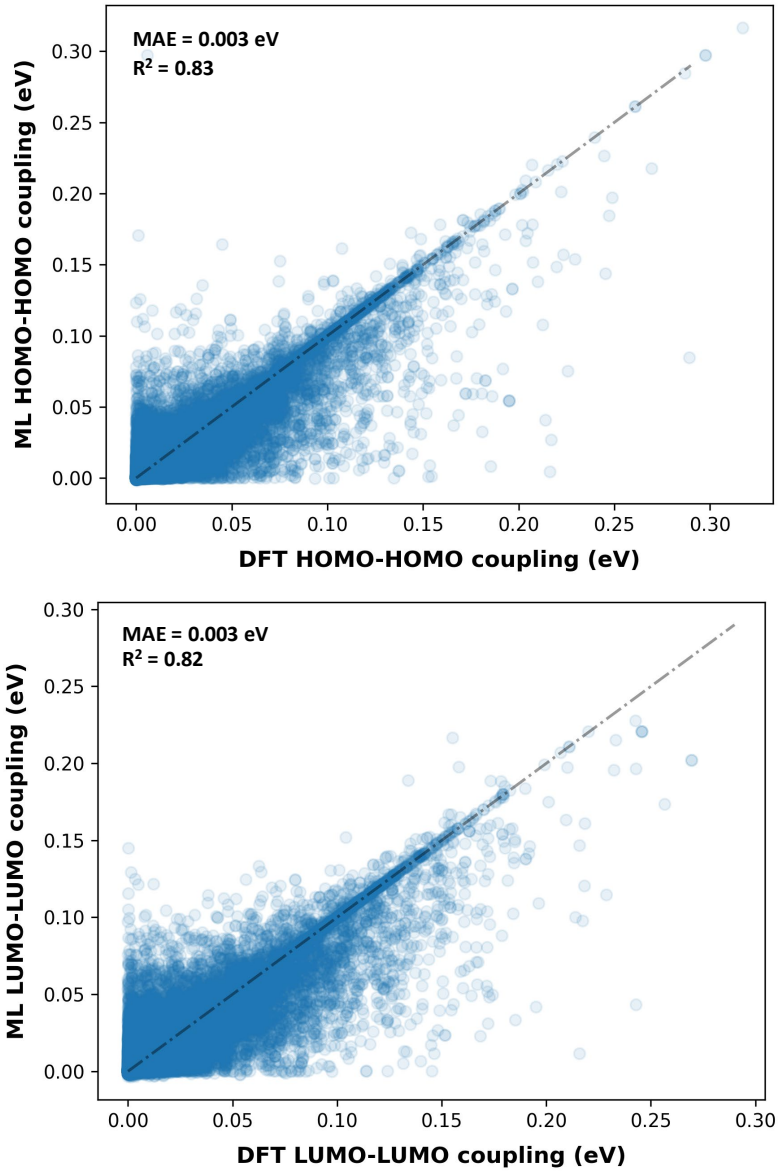


Figure 2. Scatter plot showing the correlation between the DFT estimated and ML predicted intermolecular HOMO-HOMO (top) and LUMO-LUMO (bottom) electronic couplings for the holdout test set consisting of 87,939 dimer configurations. The ML model was trained with the absolute values of intermolecular electronic couplings.

The OCELOT dimer v1 dataset, which contains more than 438,000 dimers extracted from more than 25,000 (experimental and DFT-minimized) crystal structures in the OCELOT database, was used to train the ML model. Compared to a dataset generated through MD simulations, the OCELOT dimer v1 dataset

may not capture all thermal molecular displacements. However, we hypothesize that the chemical diversity – the smallest molecular dimer in the dataset contains 20 atoms, while the largest has 392 atoms (see Figure S2) – represented by the crystal structures makes the ML model trained on the OCELOT dataset more generalizable than previous ML models trained on more limited chemical spaces.[28, 41] We note that while the signs of intermolecular electronic couplings are essential in determining the charge-carrier transport characteristics in molecular crystals through the transient localization theory model,[22, 42, 43] our initial efforts to train an ML model with the signs of the intermolecular electronic couplings yielded poor predictions. Hence, the model reported here was trained to predict absolute values of the intermolecular electronic couplings, which can be used as input in semi-classical evaluations of the electronic hopping rate constant in semi-classical Marcus-Hush theory. We used a 60:20:20 training:validation:test split of the dataset. Such a data split ensured that there were 125 unique crystal structures in the test set (see Table S1).

Figure 2 demonstrates that the ML model produces reliable predictions of the intermolecular electronic couplings derived from DFT: The intermolecular HOMO-HOMO and LUMO-LUMO electronic couplings have mean absolute errors (MAE) of 3 meV and Pearson correlations (R^2) of greater than 0.80. We implemented the natural logarithm of the absolute intermolecular electronic couplings for training to improve model performance, as demonstrated by Riderle et al.;[28] this training, however, did not significantly improve the performance but did lead to avoided saturation of values close to 0 meV (see Figure S4). To gain insights into possible ML prediction errors, we analyzed the average percent error for the test dataset. As shown in Figure S5, the average percent error is about 20%, suggesting that the ML model predictions are reliable over a large range of intermolecular electronic couplings. We note that, from the perspective of DFT evaluations of intermolecular electronic couplings, it is expected that the use of different DFT functionals and basis sets will lead to different coupling values;[44] hence, the ability to reproduce the trends of the intermolecular electronic couplings is more critical than reproducing the

absolute values when making comparisons amongst different systems and models. We further determined the Spearman's rank correlation between the DFT-derived and ML-predicted HOMO-HOMO intermolecular electronic couplings; the results largely suggest a positive correlation (see Figure S6), demonstrating that the ML model can predict well the trends in the DFT-derived intermolecular electronic couplings.

To further validate the observations, we analyzed the performance of the trained ML model to estimate the trends in intermolecular electronic couplings for pentacene. For the following discussion, we focus only on intermolecular HOMO-HOMO electronic couplings for a set of pentacene dimers with varied displacements – the dimer geometries were generated, using a Python code, by varying the interplanar, long-axis, and short-axis distances between the face-to-face packing of two molecules. Unlike previous ML models trained on over 10,000 molecular dimer geometries from MD snapshots for pentacene,[28] our dataset contains fewer than 400 pentacene dimer geometries from the 12 polymorphs and their DFT-relaxed geometries on which the ML model is trained. As shown in Figure 3 and Figures S7 and S8, the ML model correctly predicts the trends of DFT-derived intermolecular electronic couplings, especially for interplanar separations in the range of 3.5 – 5.0 Å. We highlight, though, the discrepancies for interplanar separations less than 3.5 Å and the underestimation of large intermolecular electronic couplings. These discrepancies arise from sparse sampling of these regions in the datasets, as evident from Figure S2, which is a consequence of the physics of the packing of π-conjugated organic molecules – there are very few crystal structures wherein the interplanar distance is less than 3.5 Å under standard experimental (temperature and pressure) conditions.

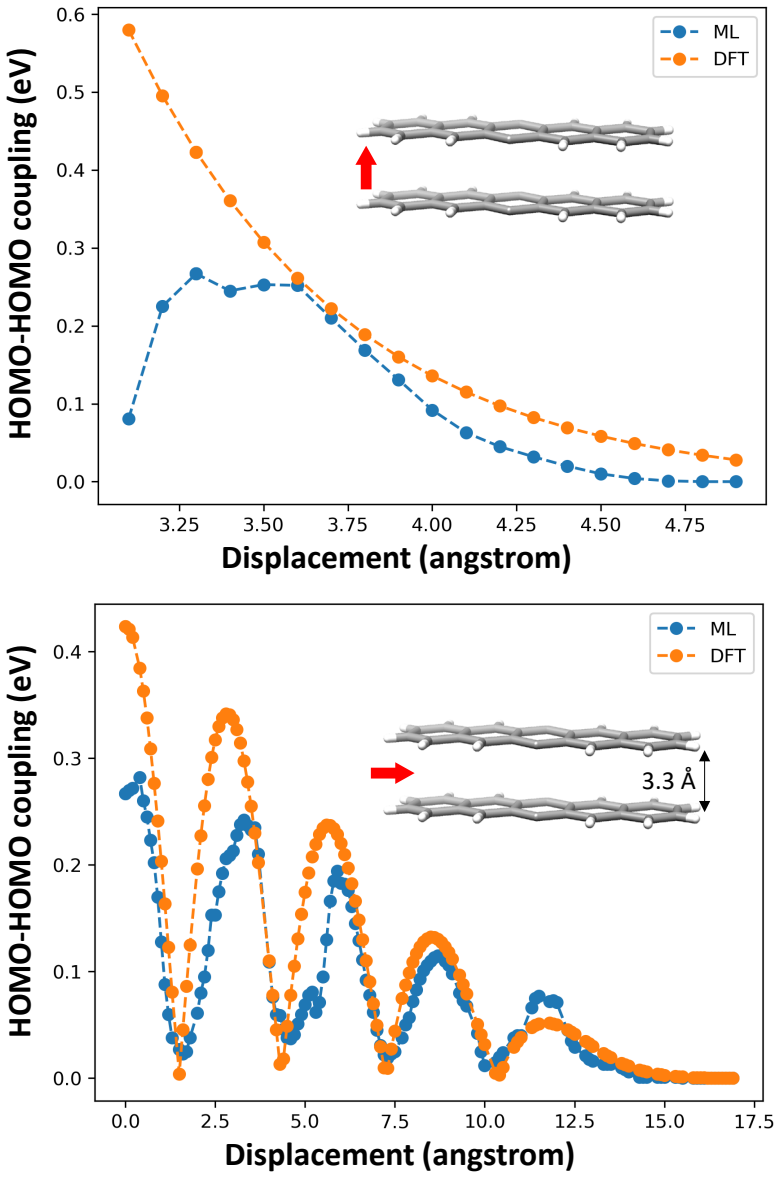


Figure 3. Variation of intermolecular HOMO-HOMO electronic coupling as a function of interplanar displacement (top) and displacement along the long-axis (bottom) for pentacene dimer. The direction of the red arrow between a dimer shows the displacement direction. DFT-derived data are in orange and ML predictions are in blue. The interplanar distance is set to 3.3 Å for the long-axis translation (bottom), and the long-axis displacement is set to 0 Å for the interplanar translation (top).

Though the ML model does generally underestimate the intermolecular electronic couplings with respect to the chosen DFT approach, predicting trends is sufficient for analyzing the relative values and trends among dimers in crystalline OSC and, thus, charge-carrier transport. With the ML-derived absolute

intermolecular electronic coupling values, we next evaluate the charge-carrier transport anisotropies of crystalline OSC via semi-classical Marcus theory through the method proposed by Goddard and coworkers, which is insensitive to the sign of intermolecular electronic coupling (see equation 3; Computational Methods).[40] We stress, of course, that this approach has significant limitations, as OSC charge-carrier mobilities can require descriptions from approaches that account for more delocalized charge-carrier wave functions, as described by, e.g., transient (de)localization or band transport mechanisms,[22, 45-47] and non-local electron-phonon couplings. To forge a full ML pipeline to evaluate charge-carrier mobilities, the ML intermolecular electronic couplings described here were coupled to ML-derived estimates of the intramolecular reorganization energies, as previously described.[37] The performance of the ML pipeline was evaluated with pentacene and rubrene crystals, which show different angular anisotropies for charge-carrier transport.[48, 49] That the ML predictions tend to underestimate the intermolecular electronic couplings when compared to the DFT approach, this feature is propagated to estimated charge-carrier mobilities. The ML-evaluated angular dependencies in pentacene and rubrene crystals agree reasonably with the experiment, as shown in Figure 4. While the angular dependence is accurately modeled for rubrene in Figure 4, there is a discrepancy in the direction of the highest charge-carrier mobility for pentacene. We note that this discrepancy is inherent in the Marcus theory approach, as observed in the original article by Goddard and coworkers, wherein DFT calculations were used to evaluate these systems.[40]

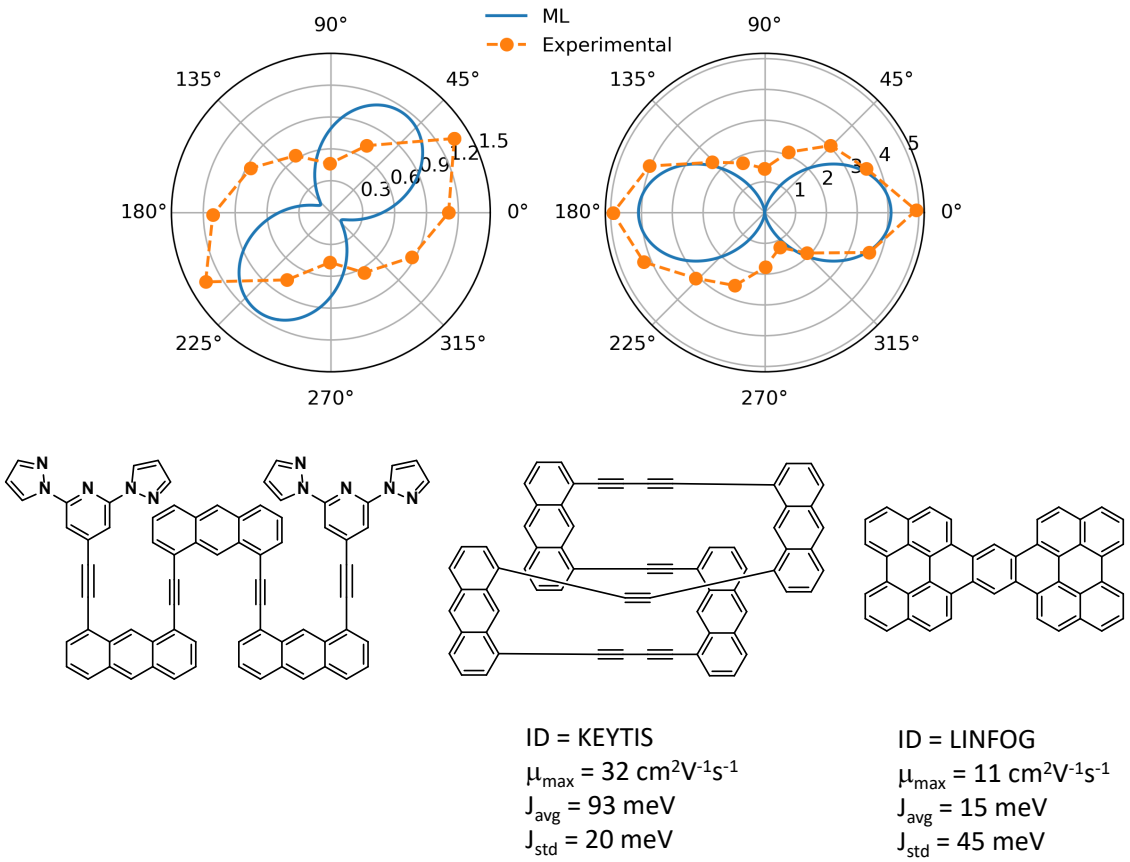


Figure 4. (Top) The angular dependence of mobility (in $cm^2V^{-1}s^{-1}$) in the ab plane for pentacene (left) and rubrene (right). Experimental estimates are in orange, and ML predictions are in blue. Experimental data for pentacene is from Ref.[48] with re-evaluations from Ref.[40] The experimental data for rubrene is from Ref.[49] (Bottom) The molecular structure of the top three crystals with high charge-carrier mobility as predicted by the ML pipeline. The ID corresponds to the Cambridge Structural Database (CSD) identifer.[50]

The ML pipeline can estimate the charge-carrier mobility of a single crystal structure in less than one minute on a standard desktop (1-core Intel Xenon E3-1241, 4GB RAM). We screened more than 110,000 crystal structures from the Cambridge Structural Database (2020.0.1 CSD release),[50] each of which consisted of one or more π-conjugated aromatic rings. We successfully screened approximately 60,000 of these structures for their charge-carrier mobilities (see SI for data); those structures that failed resulted from errors with parsing the crystal structure, missing hydrogen atoms, and the presence of metal atoms. The predictions indicate that highly π-conjugated molecules yield larger mobilities, as shown in Figure 4,

and only 372 structures presented an estimated maximum mobility (μ_{max}) greater than 1 $cm^2V^{-1}s^{-1}$. Most systems show large intermolecular electronic coupling anisotropies, as evident from the average (J_{avg}) and standard deviation (J_{std}) intermolecular electronic couplings extracted (unique) dimers in the crystal structures. To focus on systems with low intermolecular electronic coupling anisotropies, we further filtered the down-selected 372 structures by identifying those with a ratio of J_{avg} / J_{std} greater than 1. Demonstrating the successful proof-of-concept of the ML-based screening approach, the resulting 40 down-selected structures (see Table S2) contain derivatives of well-known OSC, namely polyacenes. Some of these materials have reported charge-carrier mobilities (example: OKANUK; predicted=2.2 $cmV^{-1}s^{-1}$; experimental=1.12 $cmV^{-1}s^{-1}$), while others (example: ATOWUD, SECPUO, BISYAG, and KAJTEX) remain unexplored for charge-carrier transport applications. Of the 40 structures, the crystal structure of AQOSIJ is reported to have a phase transition above 125 K and hence would not be suitable for room temperature applications. Only six crystal structures (IPODEX, JEBNAG, SECPUO, NEVGUT, EDIHUV, KOYMES, DOGCEI) were measured at temperatures of 290 K or greater, while others were obtained in the range of 90K to 200K. Hence, structural changes due to elevated temperatures should still be evaluated. We note, of course, that the charge-carrier transport analyses presented here are solely based on a hopping-based transport mechanism, and more thorough computational and experimental investigations are warranted to understand charge-carrier transport in these potential materials.

In summary, we report a ML model to predict intermolecular electronic couplings from 3D geometries of crystals comprised of π-conjugated organic molecules and an associated ML pipeline to evaluate charge-carrier mobility anisotropies in crystalline molecular OSC. Trained on the diverse 438,000 OCELOT dimers v1 dataset, we demonstrate that the MPNN–based model is reliable in predicting the trends in intermolecular electronic couplings. The ML model is transferable to molecules of different sizes, and we anticipate that this strategy can be deployed over a vast chemical space. Adopting the semi-classical Marcus theory formulation, an ML-based pipeline was created and deployed to evaluate charge-carrier

mobilities in organic crystals. More than 60,000 crystalline molecular materials were screened for their charge-carrier transport properties via the ML pipeline, and 40 potential candidates with charge-carrier mobilities greater than 1 $cm^2V^{-1}s^{-1}$ and low anisotropies in the intermolecular HOMO-HOMO electronic couplings were identified. Importantly, several of the down-selected molecular crystals are well-known and experimentally verified systems in terms of their charge-carrier transport properties, demonstrating the validity of the proof-of-concept screening approach. With rapid predictions of intermolecular electronic couplings and charge-carrier mobilities, these ML models can be used in fast analyses of charge-carrier transport that are coupled with MD simulation or kinetic Monte Carlo approaches and eventually in a complete materials discovery suite that involves searches of molecular space, crystal structure prediction, and material property prediction.

COMPUTATIONAL METHODS

Dataset generation. Molecular (noncovalent) dimers from more than 25,000 crystal structures, both as solved via x-ray crystallography and minimized via DFT, were collected from the curated OCELOT database.[39] The *screen_dimers* function from the *Hop* module of the OCELOT API[39] was used to extract dimer geometries from the crystal structures. The extraction process involved identifying all the unique molecules in the unit cell of the crystal structure and searching for neighboring molecules within 5 Å of any atom in the unique molecule. The duplicates were removed by analyzing relative interplanar, long-axis, and short-axis displacements. The approach yielded various dimers for each structure depending on the number of molecules in the unit cell. Including DFT relaxed crystal structure for some entries doubles the number of dimer geometries. For instance, pentacene crystal (csd_PENCEN) with two unique molecules in the cell yielded 12 dimers for X-ray crystal structure and 12 dimers for DFT relaxed crystal structure, thus providing 24 dimer geometries for csd_PENCEN. The maximum total dimer geometry for

a crystal structure in the dataset is 184. DFT single-point energy calculations were performed on the dimer geometries without further geometry optimization in Gaussian 16 A.03[51] at the PBE/6-31G(d,p) level of theory.[52] The intermolecular electronic couplings were evaluated with the fragment molecular orbital (FMO) approach implemented in the OCELOT API.[10, 13, 14, 39] As noted, the FMO method used here accounts for polarization effects that arise from weak van der Waals intermolecular interactions. No additional corrections to the DFT functional were made to account for van der Waals interactions. The curated dataset contains 438,709 dimer geometries and corresponding intermolecular HOMO-HOMO and LUMO-LUMO electronic coupling values. This dataset, called OCELOT dimers v1, can be downloaded programmatically and from the OCELOT web user interface. Dataset statistics are presented in the Supporting Information (SI).

ML model. The Dive into Graphs implementation of SphereNet was used here.[53] Default hyperparameters were used, as tuning with Optuna version 2.10[54] did not yield better performance. A 60:20:20 training: validation: test split of the dataset was used with mean square error (MSE) loss for training the ML model. The ML models were trained for 120 epochs, with a batch size of 32, Adam optimizer[55] with a learning rate of 0.0005, and a decay factor of 0.5 for 15 steps. Two ML models were trained – one for intermolecular HOMO-HOMO electronic coupling (for hole transport) and another for intermolecular LUMO-LUMO electronic coupling (for electron transport). ML model training was performed in PyTorch version 1.10 and used Cuda 11.4 for GPU acceleration on a single NVIDIA Tesla V100 GPU.[56, 57] Each training epoch took 25 minutes.

Charge-carrier mobility. We implemented the formalism proposed by Goddard and coworkers to estimate charge-carrier mobility anisotropies.[40] The hopping rate W is evaluated using the semi-classical Marcus-Hush equation[58]

$$W = \frac{V^2}{\hbar}\sqrt{\frac{\pi}{\lambda k_B T}}exp\left(-\frac{\lambda}{4k_B T}\right) \quad (3)$$

where V is the intermolecular electronic coupling, λ is the reorganization energy, T is the temperature and k_B is the Boltzmann constant. The total reorganization constitutes the inner-sphere reorganization (intramolecular vibrational relaxations) and outer-sphere reorganization (polarization of the surrounding medium). Here, we assume the outer sphere reorganization energy contribution is of similar magnitude across the systems in the investigation; hence, we only consider changes in the inner-sphere reorganization energy when estimating the total reorganization energy. The values of V are obtained from the trained SphereNet model, and λ is predicted by a previously published ML model.[37] The angular dependence of mobility is computed using the following equation.

$$\mu_\phi = \frac{e}{2k_B T}\sum_i W_i r_i^2 P_i cos^2\gamma_i cos^2(\theta_i - \phi) \quad (4)$$

$$P_i = \frac{W_i}{\sum_i W_i} \quad (5)$$

where i represents a specific hopping path with a hopping distance of r_i, hopping rate W_i, hopping probability P_i. $(\theta_i - \phi)$ is the angle between the conducting channel and the hopping path, ϕ is the orientation of the conducting channel relative to the reference axis and γ_i is the angle between the hopping paths and the reference plane.

ML pipeline. The input to the pipeline is a crystallographic information file (CIF) from which the dimers and the largest, contiguous π-conjugated fragment of the molecule are extracted with OCELOT API. The reorganization energy is estimated for 2D SMILES representation of the largest, contiguous π-conjugated fragment using the fourth-generation pre-trained ML models from Bhat et al.[37] The intermolecular electronic coupling predictions obtained from the SphereNet model are then used to compute the anisotropic charge-carrier mobility along the various crystallographic planes. The temperature is set to

298 K. The ML pipeline is open-access and is deployed on the OCELOT ML infrastructure, as shown in Figure S9 (see Data Availability section for link).

ASSOCIATED CONTENT

Supporting Information

- Supplementary figures and table with 40 potential candidates (PDF)
- Mobility data on 60k structures (CSV)

Data availability

The code used for training and testing is available on GitHub at https://github.com/caer200/ocelotml_coupling. The *OCELOT dimers v1* dataset is available on the OCELOT website at https://oscar.as.uky.edu/datasets. The ML pipeline can be accessed on the OCELOT ML infrastructure at https://oscar.as.uky.edu/ocelotml_coupling. The mobility data on 60k structures is available in the CSV file of SI.

AUTHOR INFORMATION

Corresponding authors

Chad Risko, chad.risko@uky.edu

Baskar Ganapathysubramanian, baskarg@iastate.edu

Author contributions

V.B.: conceptualization, data curation, investigation, methodology, writing–original draft, writing–review & editing. B.G.: supervision, writing–review & editing. C.R: supervision, funding acquisition, writing–review & editing.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work at the University of Kentucky was sponsored by the National Science Foundation through the Designing Materials to Revolutionize and Engineer our Future (NSF DMREF) program under award numbers 1627428 and 2323422. The work at Iowa State University was supported by the Office of Naval Research through award number N00014-19-12453. We acknowledge the University of Kentucky Center for Computational Sciences and Information Technology Services Research Computing for their fantastic support and collaboration and use of the Lipscomb Compute Cluster and associated research computing resources. Computational resources were also provided through the NSF Extreme Science and Engineering Discovery Environment (XSEDE) program on Stampede2 through allocation award TG-CHE200119.

REFERENCES

(1) Thorley, K. J.; Risko, C. Mapping the configuration dependence of electronic coupling in organic semiconductors. *J. Mater. Chem. C Mater* **2016**, *4* (17), 3825-3832. DOI: 10.1039/c5tc03765d

(2) Zhang, Z.; Jiang, L.; Cheng, C.; Zhen, Y.; Zhao, G.; Geng, H.; Yi, Y.; Li, L.; Dong, H.; Shuai, Z.; et al. The Impact of Interlayer Electronic Coupling on Charge Transport in Organic Semiconductors: A Case Study on Titanylphthalocyanine Single Crystals. *Angew. Chem. Int. Ed.* **2016**, *55* (17), 5206-5209. DOI: 10.1002/anie.201601065

(3) Brédas, J. L.; Calbert, J. P.; Da Silva Filho, D. A.; Cornil, J. Organic semiconductors: A theoretical characterization of the basic parameters governing charge transport. *PNAS* **2002**, *99* (9), 5804-5809. DOI: 10.1073/pnas.092143399

(4) Coropceanu, V.; Cornil, J.; Da Silva Filho, D. A.; Olivier, Y.; Silbey, R.; Brédas, J.-L. Charge Transport in Organic Semiconductors. *Chem. Rev.* **2007**, *107* (4), 926-952. DOI: 10.1021/cr050140x

(5) Bhat, V.; Callaway, C. P.; Risko, C. Computational Approaches for Organic Semiconductors: From Chemical and Physical Understanding to Predicting New Materials. *Chem. Rev.* **2023**, *123* (12), 7498-7547. DOI: 10.1021/acs.chemrev.2c00704

(6) Sutton, C.; Marshall, M. S.; Sherrill, C. D.; Risko, C.; Brédas, J.-L. Rubrene: The Interplay between Intramolecular and Intermolecular Interactions Determines the Planarization of Its Tetracene Core in the Solid State. *J. Am. Chem. Soc.* **2015**, *137* (27), 8775-8782. DOI: 10.1021/jacs.5b04066

(7) Kubas, A.; Gajdos, 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. II. *Phys. Chem. Chem. Phys.* **2015**, *17* (22), 14342-14354. DOI: 10.1039/c4cp04749d

(8) Lemaur, V.; Da Silva Filho, D. A.; Coropceanu, V.; Lehmann, M.; Geerts, Y.; Piris, J.; Debije, M. G.; Van De Craats, A. M.; Senthilkumar, K.; Siebbeles, L. D. A.; et al. Charge Transport Properties in Discotic Liquid Crystals: A Quantum-Chemical Insight into Structure−Property Relationships. *J. Am. Chem. Soc.* **2004**, *126* (10), 3271-3279. DOI: 10.1021/ja0390956

(9) Holstein, T. Studies of polaron motion: Part II. The "small" polaron. *Ann. Phys. (NY)* **1959**, *8* (3), 343-389. DOI: 10.1016/0003-4916(59)90003-X

(10) 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* (30), 9882-9886. DOI: 10.1021/ja061827h

(11) Cave, R. J.; Newton, M. D. Calculation of electronic coupling matrix elements for ground and excited state electron transfer reactions: Comparison of the generalized Mulliken–Hush and block diagonalization methods. *J. Chem. Phys.* **1997**, *106* (22), 9213-9226. DOI: 10.1063/1.474023

(12) Hutchison, G. R.; Ratner, M. A.; Marks, T. J. Intermolecular Charge Transfer between Heterocyclic Oligomers. Effects of Heteroatom and Molecular Packing on Hopping Transport in Organic Semiconductors. *J. Am. Chem. Soc.* **2005**, *127* (48), 16866-16881. DOI: 10.1021/ja0533996

(13) Newman, C. R.; Frisbie, C. D.; Da Silva Filho, D. A.; Brédas, J.-L.; Ewbank, P. C.; Mann, K. R. Introduction to Organic Thin Film Transistors and Design of n-Channel Organic Semiconductors. *Chem. Mater.* **2004**, *16* (23), 4436-4451. DOI: 10.1021/cm049391x

(14) Senthilkumar, K.; Grozema, F. C.; Guerra, C. F.; Bickelhaupt, F. M.; Lewis, F. D.; Berlin, Y. A.; Ratner, M. A.; Siebbeles, L. D. A. Absolute Rates of Hole Transfer in DNA. *J. Am. Chem. Soc.* **2005**, *127* (42), 14894-14903. DOI: 10.1021/ja054257e

(15) Kobori, T.; Sodeyama, K.; Otsuka, T.; Tateyama, Y.; Tsuneyuki, S. Trimer effects in fragment molecular orbital-linear combination of molecular orbitals calculation of one-electron orbitals for biomolecules. *J. Chem. Phys.* **2013**, *139* (9). DOI: 10.1063/1.4818599

(16) Giannini, S.; Blumberger, J. Charge Transport in Organic Semiconductors: The Perspective from Nonadiabatic Molecular Dynamics. *Acc. Chem. Res.* **2022**. DOI: 10.1021/acs.accounts.1c00675
(17) Troisi, A.; Cheung, D. L.; Andrienko, D. Charge Transport in Semiconductors with Multiscale Conformational Dynamics. *Phys. Rev. Lett.* **2009**, *102* (11). DOI: 10.1103/physrevlett.102.116602
(18) Brédas, J.-L.; Beljonne, D.; Coropceanu, V.; Cornil, J. Charge-Transfer and Energy-Transfer Processes in π-Conjugated Oligomers and Polymers: A Molecular Picture. *Chem. Rev.* **2004**, *104* (11), 4971-5004. DOI: 10.1021/cr040084k
(19) Cornil, J.; Beljonne, D.; Calbert, J.-P.; Brédas, J.-L. Interchain Interactions in Organic π-Conjugated Materials: Impact on Electronic Structure, Optical Response, and Charge Transport. *Adv. Mater.* **2001**, *13* (14), 1053-1067. DOI: 10.1002/1521-4095(200107)13:14<1053::AID-ADMA1053>3.0.CO;2-7
(20) 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* (16), 5794. DOI: 10.1039/c3cp44619k
(21) Kubař, T.; Elstner, M. A hybrid approach to simulation of electron transfer in complex molecular systems. *Journal of The Royal Society Interface* **2013**, *10* (87), 20130415. DOI: 10.1098/rsif.2013.0415
(22) Fratini, S.; Mayou, D.; Ciuchi, S. The Transient Localization Scenario for Charge Transport in Crystalline Organic Materials. *Adv. Funct. Mater.* **2016**, *26* (14), 2292-2315. DOI: 10.1002/adfm.201502386
(23) Ciuchi, S.; Fratini, S.; Mayou, D. Transient localization in crystalline organic semiconductors. *Phys. Rev. B.* **2011**, *83* (8). DOI: 10.1103/physrevb.83.081202
(24) Lederer, J.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine Learning–Based Charge Transport Computation for Pentacene. *Adv. Theory Simul.* **2019**, *2* (2), 1800136. DOI: 10.1002/adts.201800136
(25) 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* (42), 10457-10464. DOI: 10.1021/acs.jpclett.1c03053
(26) 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* (6), 3750-3759. DOI: 10.1021/acs.jctc.1c00191
(27) Wang, C.-I.; Joanito, I.; Lan, C.-F.; Hsu, C.-P. Artificial neural networks for predicting charge transfer coupling. *J. Chem. Phys.* **2020**, *153* (21), 214113. DOI: 10.1063/5.0023697
(28) Rinderle, M.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine-Learned Charge Transfer Integrals for Multiscale Simulations in Organic Thin Films. *Journal of Physical Chemistry C* **2020**, *124* (32), 17733-17743. DOI: 10.1021/acs.jpcc.0c04355
(29) 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* (38), 7658-7664. DOI: 10.1021/acs.jpca.0c04368
(30) Tan, T.; Wang, D. Machine learning based charge mobility prediction for organic semiconductors. *J. Chem. Phys.* **2023**, *158* (9), 094102. DOI: 10.1063/5.0134379
(31) Cignoni, E.; Cupellini, L.; Mennucci, B. Machine Learning Exciton Hamiltonians in Light-Harvesting Complexes. *J. Chem. Theory. Comput.* **2023**, *19* (3), 965-977. DOI: 10.1021/acs.jctc.2c01044
(32) 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* (7), 074111. DOI: 10.1063/5.0016009
(33) Schäfer, C.; Fojt, J.; Lindgren, E.; Erhart, P. Machine Learning for Polaritonic Chemistry: Accessing Chemical Kinetics. *J. Am. Chem. Soc.* **2024**. DOI: 10.1021/jacs.3c12829

(34) 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* (5). DOI: 10.1103/physrevlett.108.058301

(35) Rinderle, M.; Gagliardi, A. Machine Learning & multiscale simulations: toward fast screening of organic semiconductor materials. In *2021 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*, 2021-09-13, 2021; IEEE. DOI: 10.1109/nusod52207.2021.9541414.

(36) Sanchez-Lengeling, B.; Reif, E.; Pearce, A.; Wiltschko, A. B. A gentle introduction to graph neural networks. *Distill* **2021**, *6* (9), e33. DOI: 10.23915/distill.00033

(37) Bhat, V.; Sornberger, P.; Pokuri, B. S. S.; Duke, R.; Ganapathysubramanian, B.; Risko, C. Electronic, redox, and optical property prediction of organic π-conjugated molecules through a hierarchy of machine learning approaches. *Chem. Sci.* **2023**, *14* (1), 203-213. DOI: 10.1039/d2sc04676h

(38) Liu, Y.; Wang, L.; Liu, M.; Zhang, X.; Oztekin, B.; Ji, S. Spherical Message Passing for 3D Graph Networks. *ArXiv* **2021**. DOI: 10.48550/arXiv.2102.05013

(39) Ai, Q.; Bhat, V.; Ryno, S. M.; Jarolimek, K.; Sornberger, P.; Smith, A.; Haley, M. M.; Anthony, J. E.; Risko, C. OCELOT: An infrastructure for data-driven research to discover and design crystalline organic semiconductors. *J. Chem. Phys.* **2021**, *154* (17), 174705. DOI: 10.1063/5.0048714

(40) Wen, S.-H.; Li, A.; Song, J.; Deng, W.-Q.; Han, K.-L.; Goddard, W. A. First-Principles Investigation of Anistropic Hole Mobilities in Organic Semiconductors. *J. Phys. Chem. B.* **2009**, *113* (26), 8813-8819. DOI: 10.1021/jp900512s

(41) 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* (36), 7792-7802. DOI: 10.1021/acs.jpca.9b04256

(42) Troisi, A. Charge transport in high mobility molecular semiconductors: classical models and new theories. *Chem. Soc. Rev.* **2011**, *40* (5), 2347. DOI: 10.1039/c0cs00198h

(43) Fratini, S.; Ciuchi, S.; Mayou, D.; De Laissardière, G. T.; Troisi, A. A map of high-mobility molecular semiconductors. *Nat. Mater.* **2017**, *16* (10), 998-1002. DOI: 10.1038/nmat4970

(44) Fonari, A.; Sutton, C.; Brédas, J.-L.; Coropceanu, V. Impact of exact exchange in the description of the electronic structure of organic charge-transfer molecular crystals. *Phys. Rev. B.* **2014**, *90* (16). DOI: 10.1103/physrevb.90.165205

(45) Giannini, S.; Peng, W.-T.; Cupellini, L.; Padula, D.; Carof, A.; Blumberger, J. Exciton transport in molecular organic semiconductors boosted by transient quantum delocalization. *Nat. Commun.* **2022**, *13* (1). DOI: 10.1038/s41467-022-30308-5

(46) Fratini, S.; Ciuchi, S. Dynamical localization corrections to band transport. *Physical Review Research* **2020**, *2* (1). DOI: 10.1103/physrevresearch.2.013001

(47) Leblanc, O. H. Band Structure and Transport of Holes and Electrons in Anthracene. *J. Chem. Phys.* **1961**, *35* (4), 1275-1280. DOI: 10.1063/1.1732038

(48) Lee, J. Y.; Roth, S.; Park, Y. W. Anisotropic field effect mobility in single crystal pentacene. *Appl. Phys. Lett.* **2006**, *88* (25), 252106. DOI: 10.1063/1.2216400

(49) Reese, C.; Bao, Z. High-Resolution Measurement of the Anisotropy of Charge Transport in Single Crystals. *Adv. Mater.* **2007**, *19* (24), 4535-4538. DOI: 10.1002/adma.200701139

(50) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. The Cambridge structural database. *Acta. Crystallogr. B. Struct. Sci. Cryst. Eng. Mater.* **2016**. DOI: 10.1107/S2052520616003954

(51) Frisch, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; et al. Gaussian 16, Rev. A.03. **2016**.

(52) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865-3868. DOI: 10.1103/PhysRevLett.77.3865

(53) Liu, M.; Luo, Y.; Wang, L.; Xie, Y.; Yuan, H.; Gui, S.; Xu, Z.; Yu, H.; Zhang, J.; Liu, Y.; et al. DIG: A Turnkey Library for Diving into Graph Deep Learning Research. *J. Mach. Learn. Res.* **2021**, *22*, 240:241-240:249.

(54) Takuya, A.; Shotaro, S.; Toshihiko, Y.; Takeru, O.; Masanori, K. Optuna: A Next-generation Hyperparameter Optimization Framework In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Association for Computing Machinery: Anchorage, AK, USA, 2019; pp 2623–2631.

(55) Kingma, D. P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**. DOI: 10.48550/arXiv.1412.6980

(56) Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*, 2019.

(57) Nickolls, J.; Buck, I.; Garland, M.; Skadron, K. Scalable Parallel Programming with CUDA. *Queue* **2008**, *6* (2), 40-53. DOI: 10.1145/1365490.1365500

(58) Marcus, R. A. Chemical and Electrochemical Electron-Transfer Theory. *Annu. Rev. Phys. Chem.* **1964**, *15* (1), 155-196. DOI: 10.1146/annurev.pc.15.100164.001103