Macromolecules

pubs.acs.org/Macromolecules Article

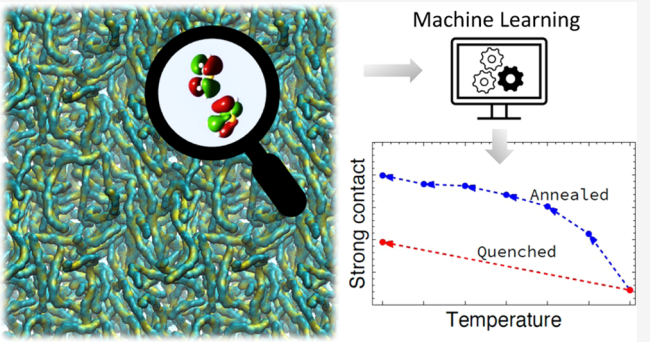
Machine Learning Identifies Strong Electronic Contacts in Semiconducting Polymer Melts

Puja Agarwala, Shane Donaher, Baskar Ganapathysubramanian, Enrique D. Gomez, and Scott T. Milner*

Cite This: *Macromolecules* 2023, 56, 5698−5707 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: In semiconducting polymers, charge transport depends on the electronic coupling between neighboring molecules and is governed by their local structures. Here, we identify the local arrangements that promote strong electronic contacts and assess how thermal annealing affects the probability of such contacts. We use molecular dynamics with virtual site coarse graining to simulate the bulk morphology of a semicrystalline donor polymer (P3HT) and then evaluate electronic coupling between monomers on neighboring chains. To avoid brute-force electronic calculations and tedious manual identification of strongly coupled pairs, we apply feature selection in machine learning to identify the most important configurations and develop a predictive model to predict electronic coupling from the most important geometric features of the local arrangement. We find that the key geometric features for strong contacts promoting hole transport closely relate to coherent overlap between HOMO wavefunctions on nearby moieties. Strong contacts in amorphous P3HT are rare but become more common with slow cooling, which leads to the formation of crystalline regions in which π-stacked configurations have more coherent overlap and thus stronger electronic couplings.



INTRODUCTION

The efficiency of semiconducting polymer-containing devices, such as field effect transistors,[1,2] light-emitting diodes,[3] and organic photovoltaics,[4] depends on charge transport through semiconducting polymeric domains.[5−10] Charge carriers travel most efficiently through crystalline domains, but transport through amorphous domains can also be crucial for device performance. For semicrystalline organic transistors to conduct, carriers must pass through disordered amorphous domains as they travel between crystalline domains.[7] In addition, charge dissociation in organic photovoltaic materials may occur in amorphous intermixed regions.[11]

Charges in the polymer may travel along the same chain (intrachain) or hop between different chains (interchain). Interchain charge transport rate is ten times slower compared to transport through all trans intrachain configuration and therefore is the rate-limiting step. Within crystalline domains, intermolecular charge hopping occurs between strongly coupled π-stacked chains.[12−15] In contrast, amorphous domains are conformationally disordered, which has been associated with decreased carrier mobilities.[16,17] However, some recently developed polymers are highly disordered, but nonetheless exhibit high charge mobilities in field effect transistors.[18] Efficient charge conduction in these polymers may result from enough strong contacts between amorphous chains such that well-connected pathways percolate through the material.

Recently, Pokuri et al. developed a graph-based technique to quantify charge carrier transport from molecular dynamics simulations.[19] It is an excellent method to estimate charge carrier mobility and depends on two key parameters: (1) molecular dynamics simulation generated morphology and (2) edge weight calculated from electronic coupling between monomers. In this paper, we address the computational challenges in obtaining these parameters.

Charge transport between pairs of nearby moieties is proportional to the square of electronic coupling,[20] which depends on the identity and relative placement of the moieties. To identify key local structures responsible for high electronic coupling, we must overcome three computational challenges:[21] (1) obtaining equilibrated local morphologies for slow-relaxing polymer systems; (2) computing electronic couplings for many thousands of local arrangements so that a heterogeneous

Received: May 19, 2023
Revised: June 30, 2023
Published: July 21, 2023

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707

system can be sampled with good statistics; and (3) identifying what aspects of geometry lead to high couplings among thousands of local arrangements. We meet these challenges by combining three computational advances: a new method for fast simulations of slowly relaxing stiff polymers, a simple and robust approach to computing coupling between pairs of moieties, and machine learning methods to identify key features of local arrangements that lead to high couplings.

Molecular dynamics (MD) simulations can represent local structures, but stiff semiconducting polymers are very slow-moving so that long simulations (many hundreds of ns) are necessary to equilibrate even oligomer melts.[22] This inconvenient fact limits previously reported atomistic simulations, where only short simulations (often, less than the time required for a chain to diffuse its own length) were performed for equilibration.[23−26] It is especially limiting to study properties, which extensively depend on morphology. For instance, Poelking et al. simulated poly(2,5-bis(3- tetradecylthiophen-2-yl)thieno[3,2-*b*]thiophene) (PBTTT) in a temperature scan from 300 to 500 K for only 1 ns, reporting no significant changes in morphology and charge transport with temperature.[27] Evidently, to study semiconducting polymers, it is crucial to accelerate MD simulations.[28]

Traditional coarse-graining techniques accelerate MD simulations by representing several atoms with a single particle, at the cost of detailed structural and electronic information.[29] For example, Do et al. represented an entire conjugated ring as a single spherical particle.[30] To preserve the planar ring structure of conjugated rings, Nguyen et al. replaced each ring with a disk-shaped particle.[31] Even in these simulations, the detailed positions of atoms are lost, which limits the usefulness of simulation configurations in computing electronic properties such as interchain couplings.[32] To predict the charge transport from a coarse-grained simulation, Van et al. "fine-grained" the system, by adding back atoms in plausible positions and relaxing the resulting configuration to obtain an atomistic structure.[33] However, this method is somewhat cumbersome, requiring multiple simulations to relax the chains into realistic local conformations.

In an effort to reduce the computational cost of back-mapping, machine learning methods have been applied.[34,35] In one such method, Jackson et al. performed all-atom simulations to prepare configurations for which electronic properties were computed.[34] The all-atom structure is then mapped onto various coarse-grained structures, to define a large set of candidate structural features that are needed to build machine learning models to predict electronic properties.

To circumvent the tradeoff between fast simulations and loss of atomic structure, we have developed a virtual-site coarse-graining method, which substantially accelerates simulations of stiff, slowly relaxing chains without giving up atomistic detail.[36] The method preserves atomistic structure by replacing real atoms with "virtual sites". Each stiff aromatic moiety is represented with only a few "real" atoms connected by fixed-length bonds; the rest of the atoms are represented by virtual sites, with positions determined by the "real" atoms. By this approach, the larger and stiffer the constituent aromatic moieties are, the larger the reduction in degrees of freedom can be achieved. For poly(3-hexylthiophene-2,5-diyl) (P3HT) polymer, virtual site simulations run ten times faster than conventional all-atom simulations.

To investigate the effects of morphology and local arrangements on charge transport, we compute electronic couplings between nearby monomers on neighboring chains using density functional theory (DFT). The energy splitting dimer (ESD) method has been used to compute electronic coupling, but it is limited to symmetric arrangements of identical moieties.[5,37] In amorphous melts, local arrangements are rarely symmetric. For nonsymmetric arrangements or nonidentical moieties, both orbital energies and wavefunctions are required to evaluate electronic coupling.[38,39]

In previous work, we have developed an elementary "Frontier Orbital Numerical Projection" (FONP) method to compute electronic couplings for nonsymmetric systems.[40] The method relies on projecting frontier orbital wavefunctions of individual monomers onto the frontier orbital of the interacting pair. Orbitals are computed by DFT and represented as cube files, from which overlaps are computed by numerical integration. We use this approach to evaluate electronic couplings of local structures extracted from MD simulations.

MD simulations with virtual site coarse graining represent the local arrangements in a polymer melt, and FONP calculates the strength of electronic contacts between nearby monomers. Combining these methods, we can ask what local configurations facilitate charge transport and how often they occur, in a disordered melt or an ordered structure. But manually identifying geometric features that promote charge hopping from a plethora of local structures is arduous, so we apply machine learning algorithms to identify the key geometric features.[41−43]

In this paper, we study poly(3-hexylthiophene-2,5-diyl) (P3HT), as a well-studied representative of the class of semicrystalline semiconductor donor polymers. At high temperatures, the polymer is an amorphous melt, for which we calculate electronic couplings between neighboring monomers. We focus on interchain electronic couplings because they are small, of order tens of meV. Intrachain couplings are generally ten times larger in conjugated polymers in near-trans local configurations; thus, interchain couplings limit overall charge transport.[44] Disordered melt morphologies result in widely varying interchain electronic couplings, with only a few strong electronic contacts.

We identify critical geometric features that lead to strong electronic contacts by applying feature selection in machine learning.[45] Using feature selection, we find the physically reasonable and expected result that the center-to-center distance between the lobes of the highest occupied molecular orbital (HOMO) on neighboring moieties is the most important feature predictive of strong couplings that facilitate hole transport. Of course, machine learning models not only make qualitative predictions about which geometric features matter the most but also provide quantitative models that predict electronic couplings from the relative arrangement of nearby monomers, without the need for DFT calculations for every monomer pair. By using feature importance to select the most important features, we can develop an efficient, accurate, predictive model for coupling between thiophene monomers; using only six features and 5800 monomer pairs in the training set, our model achieves an R^2 of 0.94.

For amorphous P3HT, relatively few configurations with strong interchain couplings are expected to dominate charge transport. We expect the probability of such configurations to depend on morphology. To explore this dependence, we slowly cool P3HT, which leads to the formation of crystalline domains. We use our predictive model to explore how the

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707

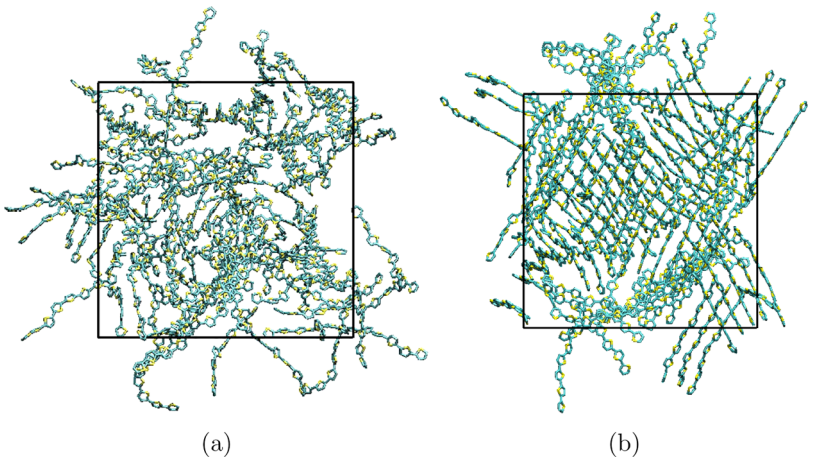


Figure 1. Simulation frame of (a) amorphous P3HT and (b) crystalline P3HT at 300 K (side groups not shown).

distribution of coupling values depends on morphology. As anticipated, crystal formation increases the probability of configurations that promote charge transport.[46−48]

MOLECULAR DYNAMICS SIMULATION OF P3HT

P3HT consists of a conjugated backbone that enables charge transport, and flexible hexyl alkane side groups that enable solution processing. The stiff backbone makes P3HT slow to move and hard to equilibrate in an all-atom molecular dynamics simulation. To overcome this challenge, we have developed a new method of coarse-graining using virtual sites.[36]

In the virtual site method of coarse-graining, each rigid thiophene ring is represented with three "real atoms" and two "virtual sites", as described in our previous work.[36] Unlike conventional coarse-graining all heavy atoms are explicitly present so detailed molecular structure is retained, and yet degrees of freedom are significantly reduced; the resulting simulations are ten times faster than an all-atom simulation. For 96 molecules of octamer P3HT, simulations run at 300 ns/day with 4 cores and 1 GPU, which is fast enough to enable thorough equilibration and subsequent annealing.

P3HT force field parameters were validated by comparing simulation results for density and persistence length to experimental values. The persistence length L_p is determined from the decay along the chain of the tangent-tangent correlation function.[49] This procedure is effective for oligomers as short as a persistence length (in fact, the length of our P3HT octamers is close to L_p), because the tangent decorrelates enough along such an oligomer to measure L_p reasonably well.

Even for oligomers, equilibrating chain conformations in the melt requires simulations of many tens of nanoseconds. For example, P3HT octomers have a mean square end-to-end distance of 2.63 nm; it takes about 20 ns for such a chain to diffuse that distance, which is comparable to the longest conformational relaxation time for an unentangled chain. If we doubled the oligomer length to 16 monomers, Rouse scaling implies the longest relaxation time would increase by a factor of four so that equilibrating even single-chain conformations in the melt would take at least 80 ns. Careful attention to expected and observed equilibration times is essential if we expect simulations of long, stiff, slowly relaxing molecules to be reliable guides of material properties.

We prepare an initial melt state by building an ordered array of 96 straight chains of 8 monomer P3HT at low density and simulate to disorder and densify the melt. To represent an amorphous morphology, we equilibrate the system at high temperatures and then quench to low temperatures, which traps the amorphous local structure. We first perform the high-temperature simulation for 500 ns at 600 K and 1 atm, using a velocity rescaling thermostat, Berendsen barostat, and 4 ps timestep. We then quench from 600 to 300 K at constant pressure and equilibrate for 200 ns.

To induce order, we cool slowly from 600 to 300 K in six steps. At each step, we lower the temperature by 50 K at a rate of 50 K/ns and then hold it constant for 100 ns. The virtual site method of coarse graining enables us to anneal P3HT long enough that the slowly relaxing polymer nonetheless forms ordered morphologies.

Rapidly quenching P3HT from 600 to 300 K results in an amorphous melt, as shown in Figure 1a. Quenching increases the density from 907 to 1075 kg/m^3, but the chains remain disordered. In contrast, slowly annealing P3HT to 300 K produces crystalline domains, in which polymer chains exhibit π-stacking,[50] as shown in Figure 1b. Neighboring chains along the lamellar stacking direction are randomly oriented, presumably because orienting chains parallel along the lamellar stacking direction is slow and thus difficult to achieve in MD simulations. However, charges do not hop between chains in the lamellar stacking direction, so the crystalline order produced in MD simulation is useful for calculating charge transport in crystalline P3HT.

COMPUTING ELECTRONIC COUPLINGS

In semiconducting polymers, charge moves between chains by hopping between nearby monomers. In Marcus theory, the charge transfer rate between neighboring monomers is proportional to the square of electronic coupling, given by

$$t_{AB} = \langle q_A | H | q_B \rangle \tag{1}$$

in which q_A and q_B are localized wavefunctions of sites A and B, respectively.

We extract local structures for electronic coupling calculations from simulation trajectories. We retain only the conjugated thiophene rings and ignore the electronically inactive side groups. Our simulations have no explicit

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707

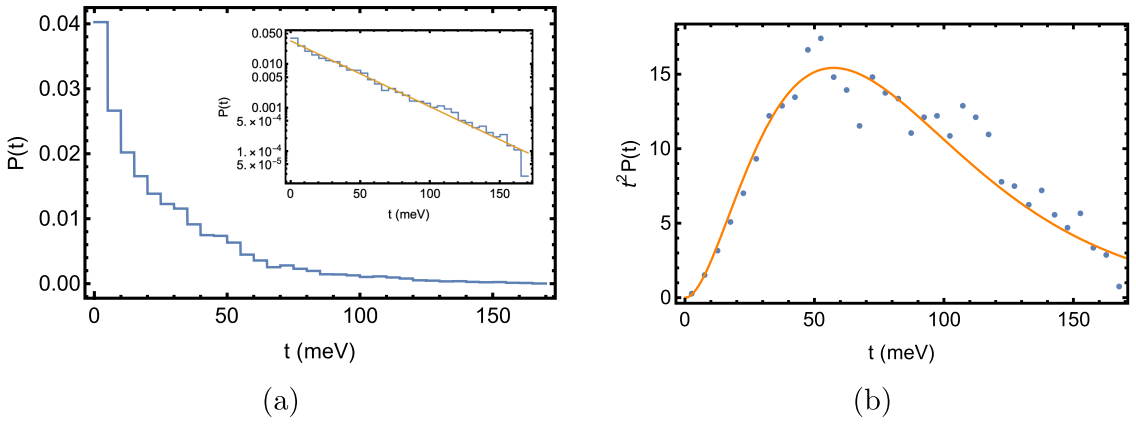


Figure 2. (a) Interchain electronic coupling distribution in amorphous P3HT and (b) product of electronic coupling square with corresponding probability.

hydrogen, so we add them at appropriate positions corresponding to the energy-minimized structure of thiophene.

We evaluate t_{AB} between monomer pairs using the frontier orbital numerical projection (FONP) method.[40] The method assumes that monomer frontier orbital hybridizes to form dimer frontier orbitals. The electronic coupling is then expressed by

$$t = \frac{\alpha}{1+\alpha^2}\Delta E \tag{2}$$

in which ΔE is the energy splitting, and α is the ratio of wavefunction amplitudes on the two moieties.

We automate the FONP method using a bash script to calculate electronic couplings for thousands of configurations. We focus on charge transfer between close contacts, with a separation distance of less than 0.6 nm, beyond which electronic coupling is negligibly small. We compute the electronic coupling for 7330 close pairs, extracted from amorphous P3HT at 600 K from 100 simulation frames over 100 ns.

We calculate the orbital wavefunctions and energies using the B3LYP functional and 6-31G basis set in `Gaussian`[51] for the thiophene dimer and constituent monomers. Previous studies suggest that electronic coupling results are sensitive to the choice of basis set.[38,52−56] We test the basis set sensitivity to electronic coupling by comparing results obtained with 6-31+G(d) and 6-31G for the entire set of 7330 close pairs. We find that coupling computed with 6-31+G(d) correlates well with those obtained using 6-31G, but are 5−10% higher. All results reported here use the computationally more efficient 6-31G basis set.

Electronic coupling for close pairs in amorphous P3HT exhibits a roughly exponential distribution $P(t) \approx e^{-t/\bar{t}}/\bar{t}$ with a characteristic energy $\bar{t}$ = 28.5 meV, as shown in Figure 2a. Assuming that the hopping matrix elements t are not strongly correlated to hopping barriers in an amorphous melt, we can separately average the prefactor t^2 in the Marcus expression for charge transport to arrive at a characteristic RMS t value given by $\sqrt{\langle t^2\rangle}$, where $\langle t^2\rangle$ is defined as

$$\langle t^2\rangle = \int_0^\infty \mathrm{d}t\; t^2 P(t) \tag{3}$$

Figure 2b displays the integrand $P(t)$, with the solid curve the approximate form $t^2 e^{-t/\bar{t}}/\bar{t}$ implied by the approximately exponential form of $P(t)$. The resulting RMS average t is 39.7 meV. Evidently, the main contributions to the RMS average come not from the many small values of t near zero, but those less-common but larger values of t in the vicinity of the peak in t^2Pt. In what follows, we focus on the configurations of pairs with such relatively larger electronic couplings, which we for definiteness we define as greater than 50 meV.

Identifying the structures that lead to high electronic coupling by poring over a plethora of configurations would be tedious. Instead, we use machine learning methods to identify the most important geometric features of high-coupling pairs. Each simulation frame contains thousands of close neighbor pairs, and DFT calculations for all those pairs would be expensive. To avoid this expense, we use machine learning to develop a predictive model to calculate electronic couplings directly from the geometry of monomer pairs.

MACHINE LEARNING

Machine learning algorithms help identify key features and develop predictive models for situations in which input and output variables have complex relations. The input parameters for electronic coupling between nearby monomers are the relative placement of the two monomers, defined by six degrees of freedom; the output parameter is electronic coupling, evaluated by complex DFT-based calculations.

Previous works have needed a large number of features to develop a good predictive model. For example, Rinderle et al.[34,57,58] used 1296 features for coupling between a pair of pentacene dimers to obtain a model with an R^2 value of 0.9. Such models, even if they successfully predict electronic coupling, suffer from a lack of interpretability. In the present work, we identify key features beforehand using feature importance methods, which enables us to develop predictive models using only a handful of the most important features.

Our approach differs from previous work in another respect; in previous work, machine learning models were designed to predict not the electronic coupling, but its logarithm. Presumably, this choice was made because the dynamic range of couplings is large, with many very small coupling values and a few large ones. But fitting the log results in a relatively uniform error in the log, i.e., errors of a constant factor, not a constant magnitude, across small and large couplings. In previous work, the typical error factor could be as large as 10, which makes for unreliable predictions of the couplings themselves.[34] Because large couplings are what matter most for transport, as shown in Figure 2b, we fit the

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707

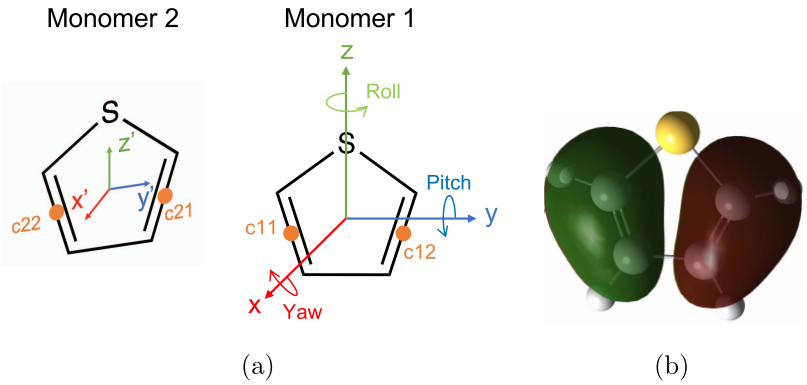


Figure 3. Representation of (a) thiophene pair local structure to define features and (b) highest occupied molecular orbital of a thiophene monomer.

coupling itself. Fitting done this way leads to a relatively uniform error across small and large couplings; weak couplings are predicted less accurately than when the log of the coupling is used, but this is of little consequence when strong couplings dominate transport.[27,34,57,58]

Using this general approach of feature identification and fitting the couplings rather than their logs, we compare the performance of random forest,[59] support vector,[60] and *k*-nearest neighbor[61] methods for developing a predictive model.

Feature Identification. Feature identification selects the most important features from a large set of candidate features, based on how well models that include these features predict the property of interest. Electronic coupling between two monomers depends only on their relative placement, which is completely specified by six degrees of freedom; these may be taken as displacement in the *x*, *y*, and *z* directions, and roll, pitch, and yaw angles of monomer 2 relative to monomer 1, as shown in Figure 3a. However, there may be better choices of geometric parameters, which relate more directly to high coupling values. To explore this possibility, we consider a modestly expanded set of plausibly important geometric features, as shown in Table 1.

In Table 1 and Figure 3a, *c*11 and *c*12 are the centers of the first and second C═C bonds in thiophene monomer 1, and *c*21 and *c*22 are the centers of the C═C bonds in monomer 2, with *c*21 closest to monomer 1.

The random forest regressor method calculates the relative importance of features listed in Table 1. The method is based on "Gini importance" or "mean decrease impurity" implemented by `scikit-learn`[62] package. The feature importance values are shown in Figure 4.

Table 1. Input Parameters for Machine Learning

no.	parameter	explanation
1	*c*11−*c*21	distance between *c*11 and *c*21
2	*c*12−*c*22	distance between other two sides, *c*12 and *c*22
3	*c*11−*c*22	distance between *c*11 and *c*22
4	*c*12−*c*21	distance between *c*12 and *c*21
5	*a*1	angle between C═C bonds at *c*11 and *c*21
6	*a*2	angle between C═C bonds at *c*12 and *c*22
7	*i*1	azimuthal angle between C═C bonds at *c*11 and *c*21
8	*i*2	azimuthal angle between C═C bonds at *c*12 and *c*22
9	*x*	displacement in the *x* direction
10	*y*	displacement in the *y* direction
11	*z*	displacement in the *z* direction
12	*d*	distance between ring centers
13	sd	distance between sulfur atoms
14	roll	roll angle
15	pitch	pitch angle
16	yaw	yaw angle
17	ra	angle between monomer planes

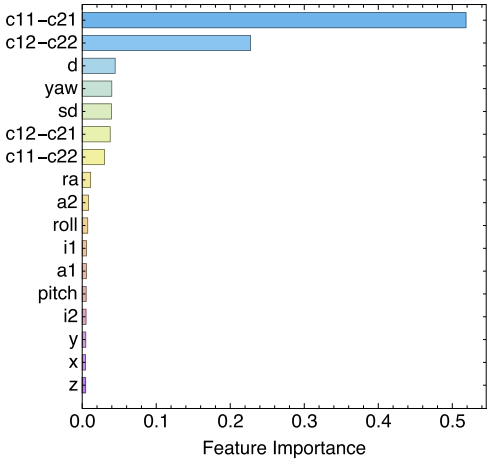
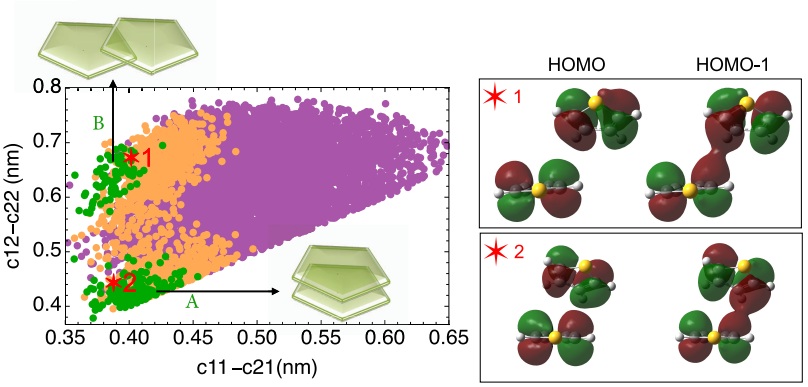


Figure 4. Feature importance of parameters in Table 1 with random forest regressor.

Figure 4 displays the features in the order of importance; the most important features are *c*11−*c*21 and *c*21−*c*22. Strong dependence on the distance between monomers is not unexpected; several prior studies have found that electronic coupling varies exponentially with the distance between the center of mass of monomers.[5,15,40,58] But our feature importance results show that the specific distances *c*11−*c*21 and *c*21−*c*22 (features 1 and 2) are much more important than the generic center of mass distance *d* (feature 12). This is physically reasonable: the thiophene HOMO wavefunctions are concentrated on the C═C bonds at *c*11, *c*12, *c*21, and *c*22 (see Figure 3b), so these distances relate directly to the overlap between the HOMO lobes on the two monomers.

Figure 5a presents a 2D scatter plot of electronic coupling versus *c*11−*c*21 and *c*12−*c*22, which illustrates how well these two features predict the coupling. Two distinct regions (green) emerge with high electronic coupling (t > 100 meV). Region *A*, in which *c*11−*c*21 and *c*12−*c*22 are both smaller than 0.45 nm, corresponds to roughly parallel configurations with both

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707



(a) (b)

Figure 5. (a) Electronic coupling as a function of most important features, less than 50 meV (purple), between 50 and 100 meV (orange), between 100 and 150 meV (green). (b) Frontier orbital of monomer pairs with high electronic coupling (t > 50 meV) corresponding to points 1 and 2 in (a).

HOMO lobes overlapping. Region B, in which $c11-c21$ is less than 0.45 nm, but $c12-c22$ is greater than 0.55 nm, corresponds to rings with only one HOMO lobe overlapping, shown in Figure 5a. Nearby configurations with slightly larger values of $c11-c21$ and $c12-c22$ give reasonably strong coupling (t > 50 meV) as shown in Figure 5a (orange). Figure 5b shows a few examples of configurations from these regions. Overall, these results confirm that the key feature for strong coupling is coherent overlap between frontier orbital lobes, which we expect to hold for other monomers as well.

Predictive Modeling. To construct a model that predicts electronic coupling from local arrangements, we compare the random forest (RF), k-nearest neighbor (KNN), and support vector regression (SVR) algorithms. We first use feature selection to select the most important input parameters. Then, we apply each algorithm to train a model, which we validate with a test data set. We also test for data sufficiency. Because the amorphous melt explores a broad range of configurations, we extract 7330 configurations from the 600 K simulation to train and test the models.

We use the forward sequential feature selector method to add features in order of importance from the list in Table 1. In forward feature selection, one feature is added at a time, and the feature that gives the largest increase in R^2 is selected at each step. We use the random forest and k-nearest neighbor estimators to obtain the feature sequence as shown in Table 2. Because there are only six degrees of freedom, we limit the model to seven parameters to avoid egregious overfitting.

Figure 6 plots R^2 for each method versus the number of features added in the order of Table 2. We follow the same sequence of feature addition for SVR and k-neighbors as obtained using random forest feature selection. The model R^2 increases with feature addition for all three methods. Even with only two most important features, all three models have R^2 greater than 0.6. But for better prediction, we include six features for SVR and random forest, and five features for k-nearest neighbor.

Table 2. Order of Features

rank	random forest	k-neighbors
1	$c11-c21$	$c11-c21$
2	$c12-c22$	$c12-c22$
3	d	d
4	yaw	$c12-c21$
5	$c12-c21$	$c11-c22$
6	$c11-c22$	roll
7	ra	yaw

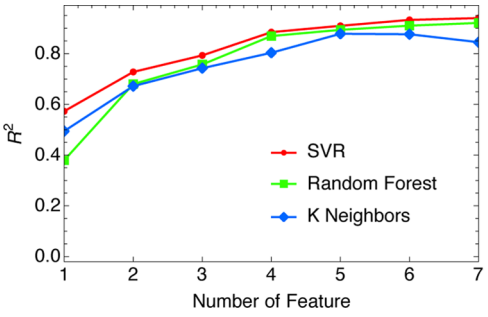


Figure 6. R^2 versus number of features added in a sequence shown in Table 2 for random forest, k-nearest neighbor, and support vector regression models.

For each method, we train the models on 80% of the data (5864 configurations of nearby monomer pairs), optimize using hyperparameter tuning, and test the resulting models on the rest of the data (1466 configuration). The random forest method uses 1400 estimators and gives an R^2 of 0.91. The data for SVR is normalized, and the optimized model uses the radial basis function (RBF) kernel with a gamma value of 0.1. The SVR model R^2 on the test data is 0.94. The optimized KNN model uses six near neighbors and gives an R^2 of 0.88.

To check for data sufficiency, we systematically increase the size of the training data set used to develop the model and observe how R^2 increases for the test data set, as shown in Figure 8. Evidently, SVR achieves a high R^2 value with a smaller training set than either random forest or k-nearest neighbor. The SVR model (Figure 7c) is particularly accurate at high electronic couplings and has a mean absolute error (MAE) of 4.32 meV for 1466 data. We use the SVR model for further predictions. The model and the data are available in GitHub.

Our approach gives good results with a much smaller training data set than was required by other studies.[34,57,58] For example, Wang et al. needed a training data set of size 10^4 to model the coupling between two ethylene molecules, a very similar problem conceptually.[58] Their need for such a large

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707

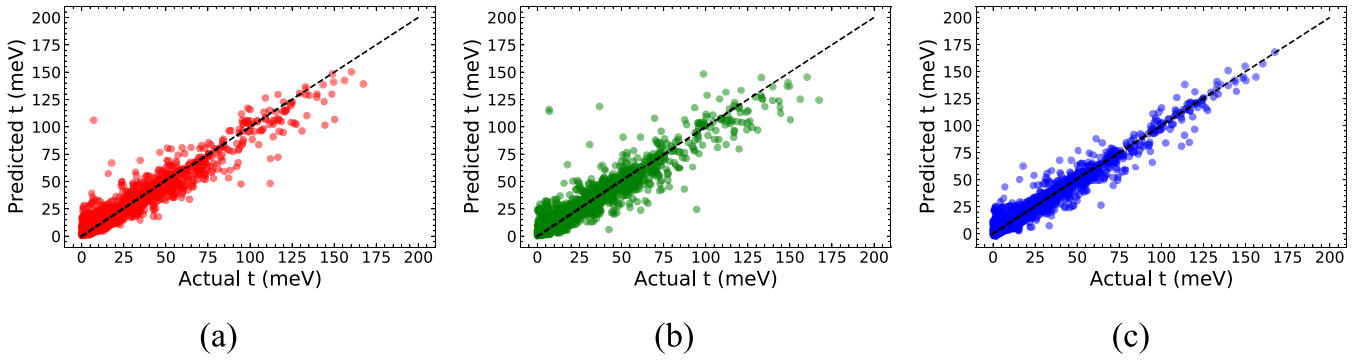


Figure 7. Actual versus predicted coupling from (a) random forest, (b) *k*-nearest neighbor, and (c) support vector regression algorithms.

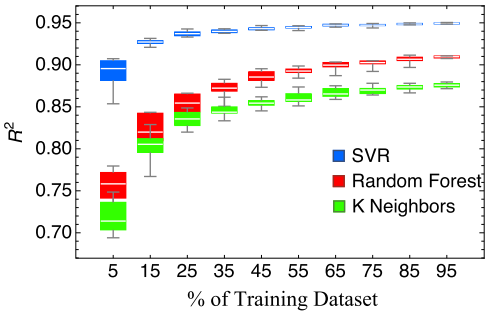


Figure 8. Data sufficiency test with machine learning algorithms.

training set is a direct result of building a model with a large number of indiscriminately chosen features. Because we selected the most important features beforehand, we obtain a good predictive model with only six features and a tenfold smaller training set.

MORPHOLOGY EFFECT ON COUPLING

The P3HT melt exhibits a wide distribution of electronic coupling, with few values greater than 50 meV (see Figure 2a). We expect crystalline domains to have more high-coupling configurations. To explore the effect of morphology on coupling, we obtain the coupling distribution for quenched and annealed morphologies as a function of temperature.

To induce crystalline order, we annealed amorphous P3HT from 600 to 300 K, as described in the previous section. Along the annealing path, at intervals of 50 K, we extract close contact configurations from 100 frames at 1 ns intervals and

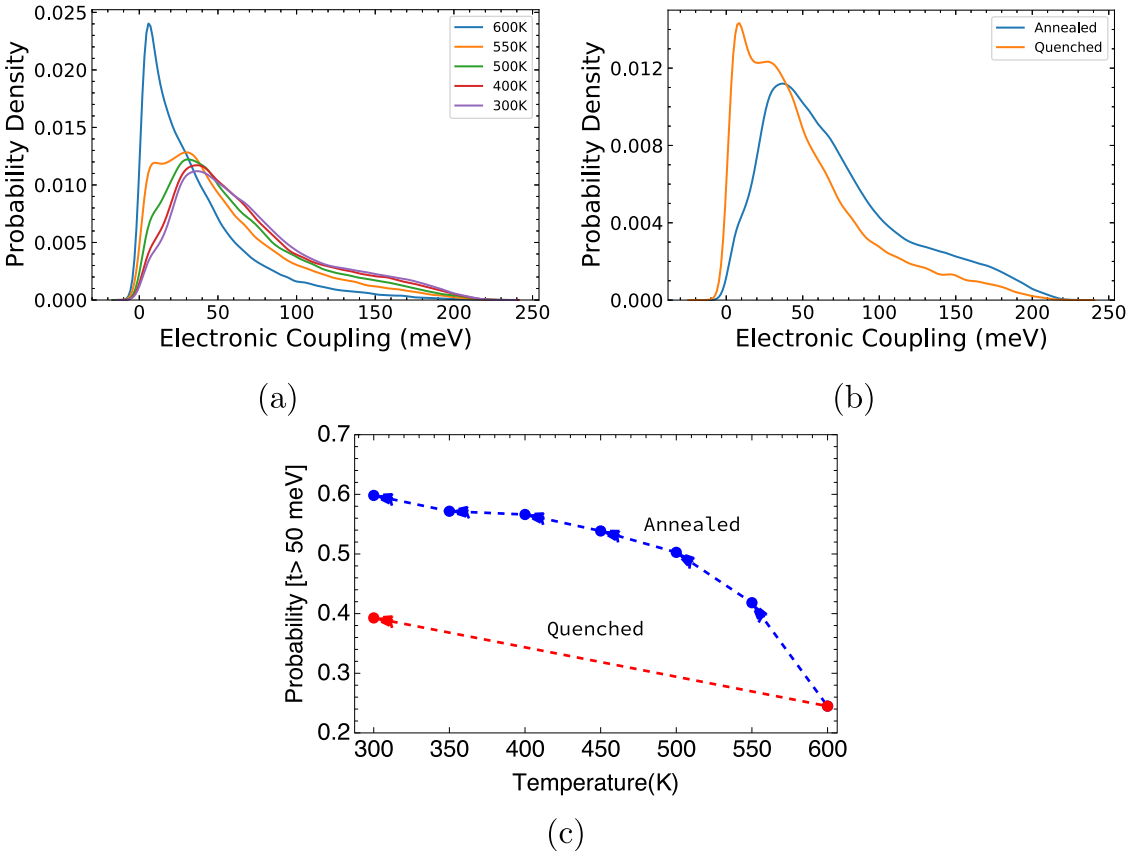


Figure 9. (a) Annealing effect on coupling probability distribution and (b) comparison with the quenched system at 300 K. (c) Fraction of neighbors with electronic coupling greater than 50 meV on annealing (blue) and quenching (red).

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707

compute the corresponding coupling distributions using the SVR predictive model. For comparison, we also quenched P3HT to 300 K, which represents amorphous regions of the polymer at room temperature, and extracted close contact configurations to compute the coupling distribution.

All morphologies exhibit a wide distribution of electronic couplings, as shown in Figure 9a. For the 600 K amorphous melt, the coupling distribution peaks near zero. Annealing shifts the peak to nearly 50 meV, with a corresponding increase in the high-coupling tail. In contrast, quenching results in a distribution peaked near zero, with a much weaker high-coupling tail (see Figure 9b).

We are particularly interested in configurations that contribute strongly to charge transfer. Figure 9c shows the probability of configurations with high electronic coupling (defined here as $t > 50$ meV) at each temperature. Evidently, annealing is more effective than quenching in producing high-coupling pairs. Annealing produces crystalline domains, which consist of π-stacked chains with high coupling. In contrast, quenching produces high-coupling pairs mainly by increasing the density, which brings monomers closer together and encourages a few parallel arrangements with higher couplings.

In this paper, we have simulated P3HT oligomers rather than long polymer chains to investigate the effect of polymer morphology on interchain electronic coupling. We studied oligomers because with shorter chains we can observe effects of annealing within the timescale (of order 1 μs) our simulations can access. We emphasize that interchain electronic couplings are quite local, depending only on the arrangement of pairs of close-approaching monomers on different chains. These geometries are very well represented by our oligomer simulations. Finally, we reiterate that our work provides effective methods for (a) generating well-equilibrated configurations of semiconducting chains, and (b) computing hopping matrix elements between nearby monomers, both essential ingredients of a mesoscale kinetic model of carrier transport in organic semiconductors.

CONCLUSIONS

This work investigates charge transport in amorphous and crystalline semiconducting polymers. We study amorphous and crystalline morphologies in P3HT using virtual site coarse-graining molecular dynamics simulation. We use machine learning feature identification methods to identify key geometric features that give rise to enhanced charge transport. The most important features are the center-to-center distances between C=C bonds of nearby monomers. The lobes of the HOMO reside on C=C bonds; thus, the key features directly relate to the coherent overlap of HOMO wavefunctions, which facilitate hole transport.

We identify two families of configurations that give high coupling: (a) face-to-face π-stacked thiophenes with both HOMO lobes overlapping and (b) parallel-displaced π-stacked thiophenes with only one HOMO lobe overlapping. Our result suggests that efficient coupling for charge transfer generally depends on the coherent overlap between frontier orbitals, but may not require a highly ordered π-stacking geometry.

Using feature selection methods, we identify the minimum set of features required to build a predictive model for electronic coupling. We developed a machine learning model using SVR to predict coupling for configurations of neighboring monomers, with an R^2 value of 0.94. We used this predictive model to investigate the effect of annealing and quenching on high-coupling pairs. We find that annealing increases the probability of strong contacts as a result of π-stacking, and quenching weakly increases the probability as a result of higher density and local packing.

AUTHOR INFORMATION

Corresponding Author

Scott T. Milner − *Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, United States; Department of Materials Science and Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, United States;* orcid.org/0000-0002-9774-3307; Email: stm9@psu.edu

Authors

Puja Agarwala − *Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, United States;* orcid.org/0000-0003-3561-6633

Shane Donaher − *Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, United States*

Baskar Ganapathysubramanian − *Department of Mechanical Engineering, Iowa State University, Ames, Iowa 50011, United States*

Enrique D. Gomez − *Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, United States; Department of Materials Science and Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, United States; Materials Research Institute, The Pennsylvania State University, University Park, Pennsylvania 16802, United States;* orcid.org/0000-0001-8942-4480

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.macromol.3c00987

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support from the National Science Foundation (DMREF-1921854) and Office of Naval Research (N00014-19-1-2453) are acknowledged.

REFERENCES

(1) Li, H.; Brédas, J.-L. Developing molecular-level models for organic field-effect transistors. *Natl. Sci. Rev.* **2021**, *8*, No. nwaa167.

(2) Liu, K.; Ouyang, B.; Guo, X.; Guo, Y.; Liu, Y. Advances in flexible organic field-effect transistors and their applications for flexible electronics. *npj Flexible Electron.* **2022**, *6*, No. 1.

(3) Kuik, M.; Wetzelaer, G.-J. A.; Nicolai, H. T.; Craciun, N. I.; De Leeuw, D. M.; Blom, P. W. 25th anniversary article: charge transport and recombination in polymer light-emitting diodes. *Adv. Mater.* **2014**, *26*, 512−531.

(4) Maturová, K.; Van Bavel, S. S.; Wienk, M. M.; Janssen, R. A.; Kemerink, M. Description of the morphology dependent charge transport and performance of polymer: fullerene bulk heterojunction solar cells. *Adv. Funct. Mater.* **2011**, *21*, 261−269.

(5) 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*, 926−952.

(6) Troisi, A. Charge transport in high mobility molecular semiconductors: classical models and new theories. *Chem. Soc. Rev.* **2011**, *40*, 2347−2358.

(7) Noriega, R.; Rivnay, J.; Vandewal, K.; Koch, F. P.; Stingelin, N.; Smith, P.; Toney, M. F.; Salleo, A. A general relationship between disorder, aggregation and charge transport in conjugated polymers. *Nat. Mater.* **2013**, *12*, 1038−1044.
(8) Ebenhoch, B.; Thomson, S. A.; Genevičius, K.; Juška, G.; Samuel, I. D. Charge carrier mobility of the organic photovoltaic materials PTB7 and PC71BM and its influence on device performance. *Org. Electron.* **2015**, *22*, 62−68.
(9) Jones, M. L.; Huang, D.; Chakrabarti, B.; Groves, C. Relating molecular morphology to charge mobility in semicrystalline conjugated polymers. *J. Phys. Chem. C* **2016**, *120*, 4240−4250.
(10) Kline, R. J.; McGehee, M. Morphology and charge transport in conjugated polymers. *J. Macromol. Sci., Polym. Rev.* **2006**, *46*, 27−45.
(11) Vandewal, K.; Himmelberger, S.; Salleo, A. Structural factors that affect the performance of organic bulk heterojunction solar cells. *Macromolecules* **2013**, *46*, 6379−6387.
(12) Bombile, J. H.; Shetty, S.; Janik, M. J.; Milner, S. T. Polaron hopping barriers and rates in semiconducting polymers. *Phys. Chem. Chem. Phys.* **2020**, *22*, 4032−4042.
(13) Bombile, J. H.; Janik, M. J.; Milner, S. T. Tight binding model of conformational disorder effects on the optical absorption spectrum of polythiophenes. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12521−12533.
(14) Bajpai, M.; Srivastava, R.; Dhar, R.; Tiwari, R. Role of reduced pi-pi stacking in the charge transport in polyfluorene. *Mater. Sci. Eng. B* **2016**, *212*, 62−70.
(15) Brédas, J. L.; Calbert, J. P.; da Silva Filho, D.; Cornil, J. Organic semiconductors: A theoretical characterization of the basic parameters governing charge transport. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 5804−5809.
(16) Miller, E. D.; Jones, M. L.; Jankowski, E. Tying together multiscale calculations for charge transport in P3HT: Structural descriptors, morphology, and tie-chains. *Polymers* **2018**, *10*, 1358.
(17) Himmelberger, S.; Vandewal, K.; Fei, Z.; Heeney, M.; Salleo, A. Role of molecular weight distribution on charge transport in semiconducting polymers. *Macromolecules* **2014**, *47*, 7151−7157.
(18) Zhang, W.; Smith, J.; Watkins, S. E.; Gysel, R.; McGehee, M.; Salleo, A.; Kirkpatrick, J.; Ashraf, S.; Anthopoulos, T.; Heeney, M.; McCulloch, I. Indacenodithiophene semiconducting polymers for high-performance, air-stable transistors. *J. Am. Chem. Soc.* **2010**, *132*, 11437−11439.
(19) Pokuri, B. S. S.; Ryno, S. M.; Noruzi, R.; Risko, C.; Ganapathysubramanian, B. Computational characterization of charge transport resiliency in molecular solids. *Mol. Syst. Des. Eng.* **2022**, *7*, 651−660.
(20) Marcus, R. A. Electron transfer reactions in chemistry. Theory and experiment. *Rev. Mod. Phys.* **1993**, *65*, 1111−1121.
(21) Mosquera, M. A.; Fu, B.; Kohlstedt, K. L.; Schatz, G. C.; Ratner, M. A. Wave functions, density functionals, and artificial intelligence for materials and energy research: Future prospects and challenges. *ACS Energy Lett.* **2018**, *3*, 155−162.
(22) Gartner, T. E., III; Jayaraman, A. Modeling and simulations of polymers: a roadmap. *Macromolecules* **2019**, *52*, 755−786.
(23) Dilmurat, R.; Lemaur, V.; Olivier, Y.; Gali, S. M.; Beljonne, D. Tuning Short Contacts between Polymer Chains To Enhance Charge Transport in Amorphous Donor−Acceptor Polymers. *J. Phys. Chem. C* **2022**, *126*, 3118−3126.
(24) Kupgan, G.; Chen, X.-K.; Brédas, J.-L. Molecular packing in the active layers of organic solar cells based on non-fullerene acceptors: Impact of isomerization on charge transport, exciton dissociation, and nonradiative recombination. *ACS Appl. Energy Mater.* **2021**, *4*, 4002−4011.
(25) Xiang, C.; Zhao, Q.; Liu, W.; Cao, J.; Zou, Y.; Zhou, H. Theoretical exploration of molecular packing and the charge transfer mechanism of organic solar cells based on PM6: Y6. *J. Mater. Chem. A* **2022**, *10*, 25611−25619.
(26) Goldey, M. B.; Reid, D.; de Pablo, J.; Galli, G. Planarity and multiple components promote organic photovoltaic efficiency by improving electronic transport. *Phys. Chem. Chem. Phys.* **2016**, *18*, 31388−31399.
(27) Poelking, C.; Cho, E.; Malafeev, A.; Ivanov, V.; Kremer, K.; Risko, C.; Brédas, J.-L.; Andrienko, D. Characterization of charge-carrier transport in semicrystalline polymers: electronic couplings, site energies, and charge-carrier dynamics in poly (bithiophene-alt-thienothiophene)[PBTTT]. *J. Phys. Chem. C* **2013**, *117*, 1633−1640.
(28) Gemünden, P.; Poelking, C.; Kremer, K.; Andrienko, D.; Daoulas, K. C. Nematic ordering, conjugation, and density of states of soluble polymeric semiconductors. *Macromolecules* **2013**, *46*, 5762−5774.
(29) Dhamankar, S.; Webb, M. A. Chemically specific coarse-graining of polymers: Methods and prospects. *J. Polym. Sci.* **2021**, *59*, 2613−2643.
(30) Do, K.; Huang, D. M.; Faller, R.; Moulé, A. J. A comparative MD study of the local structure of polymer semiconductors P3HT and PBTTT. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14735−14739.
(31) Nguyen, H. T. L.; Huang, D. M. Systematic bottom-up molecular coarse-graining via force and torque matching using anisotropic particles. *J. Chem. Phys.* **2022**, *156*, No. 184118.
(32) Heeger, A. J. Semiconducting polymers: the third generation. *Chem. Soc. Rev.* **2010**, *39*, 2354−2371.
(33) Van, E.; Jones, M.; Jankowski, E.; Wodo, O. Using graphs to quantify energetic and structural order in semicrystalline oligothiophene thin films. *Mol. Syst. Des. Eng.* **2018**, *3*, 853−867.
(34) Jackson, N. E.; Bowen, A. S.; Antony, L. W.; Webb, M. A.; Vishwanath, V.; de Pablo, J. J. Electronic structure at coarse-grained resolutions from supervised machine learning. *Sci. Adv.* **2019**, *5*, No. eaav1190.
(35) Simine, L.; Allen, T. C.; Rossky, P. J. Predicting optical spectra for optoelectronic polymers using coarse-grained models and recurrent neural networks. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 13945−13948.
(36) Agarwala, P.; Gomez, E. D.; Milner, S. T. Fast, Faithful Simulations of Donor-Acceptor Interface Morphology. *J. Chem. Theory Comput.* **2022**, *18*, 6932−6939.
(37) Newton, M. D. Quantum chemical probes of electron-transfer kinetics: the nature of donor-acceptor interactions. *Chem. Rev.* **1991**, *91*, 767−792.
(38) 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.
(39) Oberhofer, H.; Reuter, K.; Blumberger, J. Charge transport in molecular materials: An assessment of computational methods. *Chem. Rev.* **2017**, *117*, 10319−10357.
(40) Donaher, S.; Agarwala, P.; Milner, S. T. A simple approach to hopping matrix elements between nearby molecules, 2022, https://arxiv.org/abs/2212.11831.
(41) Wang, C.-I.; Joanito, I.; Lan, C.-F.; Hsu, C.-P. Artificial neural networks for predicting charge transfer coupling. *J. Chem. Phys.* **2020**, *153*, No. 214113.
(42) 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.
(43) Roch, L. M.; Saikin, S. K.; Hase, F.; Friederich, P.; Goldsmith, R. H.; León, S.; Aspuru-Guzik, A. From absorption spectra to charge transfer in nanoaggregates of oligomers with machine learning. *ACS Nano* **2020**, *14*, 6589−6598.
(44) MacKenzie, R. C. I.; Frost, J. M.; Nelson, J. A numerical study of mobility in thin films of fullerene derivatives. *J. Chem. Phys.* **2010**, *132*, No. 064904.
(45) Liu, H.; Yucel, B.; Wheeler, D.; Ganapathysubramanian, B.; Kalidindi, S. R.; Wodo, O. How important is microstructural feature selection for data-driven structure-property mapping? *MRS Commun.* **2022**, *12*, 95−103.
(46) Vakhshouri, K.; Gomez, E. D. Effect of crystallization kinetics on microstructure and charge transport of polythiophenes. *Macromol. Rapid Commun.* **2012**, *33*, 2133−2137.

(47) Singh, C. R.; Gupta, G.; Lohwasser, R.; Engmann, S.; Balko, J.; Thelakkat, M.; Thurn-Albrecht, T.; Hoppe, H. Correlation of charge transport with structural order in highly ordered melt-crystallized poly (3-hexylthiophene) thin films. *J. Polym. Sci., Part B: Polym. Phys.* **2013**, *51*, 943−951.
(48) Bull, T. A.; Pingree, L. S.; Jenekhe, S. A.; Ginger, D. S.; Luscombe, C. K. The role of mesoscopic PCBM crystallites in solvent vapor annealed copolymer solar cells. *ACS Nano* **2009**, *3*, 627−636.
(49) Zhang, W.; Gomez, E. D.; Milner, S. T. Predicting chain dimensions of semiflexible polymers from dihedral potentials. *Macromolecules* **2014**, *47*, 6453−6461.
(50) Winokur, M.; Spiegel, D.; Kim, Y.; Hotta, S.; Heeger, A. Structural and absorption studies of the thermochromic transition in poly (3-hexylthiophene). *Synth. Met.* **1989**, *28*, 419−426.
(51) Frisch, M. J.et al. *Gaussian Inč16 Revision C.01*; Gaussian Inc.: Wallingford CT, 2016.
(52) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self—consistent molecular orbital methods. XII. Further extensions of Gaussian—type basis sets for use in molecular orbital studies of organic molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(53) Hariharan, P. C.; Pople, J. A. The influence of polarization functions on molecular orbital hydrogenation energies. *Theor. Chim. Acta* **1973**, *28*, 213−222.
(54) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(55) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6-31G*basis set for atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223−1229.
(56) Sutton, C.; Sears, J. S.; Coropceanu, V.; Bredas, J.-L. Understanding the density functional dependence of DFT-calculated electronic couplings in organic semiconductors. *J. Phys. Chem. Lett.* **2013**, *4*, 919−924.
(57) Rinderle, M.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine-learned charge transfer integrals for multiscale simulations in organic thin films. *J. Phys. Chem. C* **2020**, *124*, 17733−17743.
(58) 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.
(59) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5−32.
(60) Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273−297.
(61) Altman, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **1992**, *46*, 175−185.
(62) Pedregosa, F.; et al. Scikit-learn: Machine learning in Python. *J. Mach.Learn. Res.* **2011**, *12*, 2825−2830.

https://doi.org/10.1021/acs.macromol.3c00987
Macromolecules 2023, 56, 5698−5707