

Deciphering conductivity in PEDOT guided by machine learning: From solvent baths to charge paths

Najmeh Zahabi^{1,*}, Ioannis Petsagkourakis^{2,*}, Nicolas Rolland³, Ali Beikmohammadi⁴, Xianjie Liu¹,
Mats Fahlman¹, Eleni Pavlopoulou⁵, and Igor Zozoulenko^{1,†}

¹Laboratory of Organic Electronics (LOE), Department of Science and Technology (ITN),
Campus Norrköping, [Linköping University](#), 60174 Norrköping, Sweden

²Bio and Organic Electronics Unit, Department of Smart Hardware, Division of Digital Systems,
[RISE Research Institutes of Sweden](#) AB, 60221 Norrköping, Sweden

³[Université Lille](#), CNRS, Centrale Lille, Université Artois, UMR 8181–UCCS–Unité de Catalyse et Chimie du Solide, F-59000 Lille, France

⁴Department of Computer and Systems Sciences, [Stockholm University](#), 16425 Stockholm, Sweden

⁵Institute of Electronic Structure and Laser, [Foundation for Research and Technology-Hellas](#), Heraklion, 71110 Crete, Greece



(Received 16 March 2025; revised 28 July 2025; accepted 24 September 2025; published 17 October 2025)

PEDOT:Tos is a promising conducting polymer for electronic and bioelectronic applications, yet its charge transport is affected by various factors and remains challenging to optimize. This study investigates the impact of solvent posttreatment on PEDOT:Tos thin films, exploring its influence on morphology and electrical conductivity. A combined experimental-theoretical approach is employed, integrating molecular dynamics, density functional theory, and transport calculations on one hand and conductivity, GIWAXS and XPS measurements on the other hand. Moreover, we developed a machine learning (ML) framework based on convolutional neural networks with the Coulomb matrix as the predictor, and transfer integrals for multiscale transport calculations as targets. Our results reveal that solvent-induced morphological changes strongly affect charge transport, with the ML model effectively reproducing observed conductivity. The developed ML model dramatically boosts the speed of mobility calculations, enabling the analysis of large-scale polymer films that were previously beyond computational reach.

DOI: [10.1103/5h7d-yvvd](https://doi.org/10.1103/5h7d-yvvd)

I. INTRODUCTION

Conducting polymers are appealing materials for a wide range of applications in electronic and bioelectronic devices due to their unique combination of electrical conductivity and flexibility [1–3]. One of the most studied conducting polymers is PEDOT (poly-(3,4-ethylenedioxythiophene) due to its excellent electronic (electrical conductivity up to 3000 S/cm), optical properties (transparency >80%) and electrochemical properties which allow the manufacturing of a variety of electrochemical devices [4–6]. PEDOT alone is insoluble in most solvents and for that reason it is often polymerized in the presence of other polyelectrolytes such as Poly(4-styrenesulfonate) (PSS) that act as the counterions of the conjugated polymers and enable the formation of stable water-based dispersions. Alternatively, PEDOT can be polymerized with *in-situ* wet chemical polymerization

or vapor phase polymerization, in which the counterions of the conjugated polymer originate from the oxidant used in the process [5,7]. These counterions range from simple ions like chloride (Cl^-) to organic molecules like tosylate (Tos) that compensate oxidized PEDOT chains that are positively charged [8,9]. Among these, PEDOT:Tos is particularly popular in cutting-edge applications because it combines high conductivity, processability, and stability [10–12].

One of the most important features influencing the efficiency of PEDOT:Tos-based applications is materials electrical conductivity σ . It is generally accepted that transport in PEDOT, as well as in most other conducting polymers and disordered organic semiconductors, is due to phonon-assisted hopping mechanisms [13–15]. Currently, there is a surge of interest in studies addressing the electrical conductivity of PEDOT:Tos [16], which can be modulated either by electron density n or mobility μ (Note the relation $\sigma = ne\mu$, where e is the electric charge). However, the charge carrier concentration of the polymer is typically linked to the oxidation level of the PEDOT:Tos films and the latter has been widely reported as $\sim 30\% - 33\%$ for as polymerized PEDOT:Tos in the literature [7,17], while any increase of the charge carrier concentration involves usually postprocessing with highly corrosive acids such as sulfuric acid [18]. On the other hand, the charge carrier mobility is heavily influenced by the morphological properties of the PEDOT film which can be modulated by wider range of more user-friendly additive

*These authors contributed equally to this work.

†Contact author: igor.zozoulenko@liu.se

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](#) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by [Bibsam](#).

manufacturing approaches [19–21]. Therefore, understanding and enhancing mobility is key to improving the performance of PEDOT:Tos in practical applications. Some of the most common approaches to modulate the charge carrier mobility and thus the electrical conductivity of PEDOT systems involve the use of solvents, either as additives in the solution during the polymer deposition or as a postprocessing treatment in a bath. The solvent treatment [e.g., using a common laboratory solvent such as dimethylsulfoxide (DMSO)], for instance, can enhance electrical conductivity of PEDOT:PSS systems by several orders of magnitude [22].

Despite considerable progress in improving mobility and electrical conductivity [19], the development of PEDOT:Tos and related polymers has been hindered by a limited understanding of the physical processes underlying the solvent treatment of this polymer, complicating the rational and systematic design of materials and processing methods. Our study aims to address this gap by providing new insights into the role of solvents in modulating the electrical properties of PEDOT:Tos through a combined theoretical and experimental approach.

Recently, theoretical approaches to calculating charge carrier mobility have focused on the relation between morphology and mobility, employing different levels of theory ([23–31]). These approaches often utilize molecular dynamics (MD) simulations to generate realistic representations of polymer thin films. A key cornerstone in connecting thin film morphology to carrier mobility is the calculation of the transfer integrals distribution. Specifically, charge carriers in conducting polymers, known as polarons [32], move between localized sites via thermally activated hopping [33]. The rate of this hopping process is governed by the transfer integrals, which quantify the electronic coupling between neighboring sites and are crucial in determining charge transport efficiency. For example, some of us, in Ref. [34], have developed a multiscale computational model that relies on analyzing the transfer integral distribution and applies percolation theory as a general framework for understanding transport mechanisms. However, calculating transfer integrals for a realistically large polymer film presents a significant bottleneck, as quantum mechanical simulations typically utilizing density functional theory (DFT) can take months to complete due to their high computational cost and memory demands, rendering them impractical. To overcome these limitations, machine learning (ML) techniques have emerged as a powerful and groundbreaking solution, enabling direct prediction of transfer integrals from morphology and significantly accelerating the calculations.

ML has been widely applied across various fields of research to address big data challenges for complex problems. ML identifies patterns and correlations between datasets to suggest new scientific insights without requiring specific knowledge of the physical principles. In the realm of materials science, ML has proven the ability to predict various properties, such as electrolyte stability, solvent selection, crystallinity, and the electronic properties of materials, just to mention a few [35–38]. Recently, ML has been successfully used to predict transfer integrals in different organic semiconductors [39–42]. Specifically, these studies have employed ML techniques such as kernel ridge regression, support

vector regression, and deep feed-forward neural networks, leveraging predictors (also called “features”) like system geometric characteristics, the full Coulomb matrix, off-diagonal Coulomb matrix elements, and the interaction part of the Coulomb matrix, see for details Sec. IV A 3, ‘Machine Learning Model.’ However, ML techniques have not yet been applied to predicting transfer integrals in conducting polymers. Due to the vastly different morphology of polymer films—often largely amorphous [2,4]—compared to crystalline organic semiconductors, it remains unclear whether these methods can be directly transferred to such materials. Furthermore, despite the progress made with existing ML models, the application of state-of-the-art techniques, such as convolutional neural networks (CNNs), remains underexplored.

Consequently, in this work, we address two key questions to advance our understanding of PEDOT:Tos and its electronic properties:

How does a solvent post treatment influence the charge carrier mobility and electrical conductivity of PEDOT:Tos thin films? To answer this, we have combined experimental and theoretical approaches for treatment with five commonly used laboratory solvents, DMSO, Ethylene Glycol, Ethanol, Isopropanol, and n-butanol (Butanol), to systematically investigate how changes in thin-film morphology due to solvent treatment influence electronic properties.

Can ML be used to accurately predict the transfer integrals and enhance our understanding of mobility in PEDOT:Tos? To address this, we implemented CNN techniques to predict transfer integrals in PEDOT:Tos thin films by leveraging the interactive part of the Coulomb matrix as the predictor of the ML model, with the transfer integral being the target.

Overall, the role of solvents in modulating the electrical properties of PEDOT:Tos is explored in a holistic approach, combining theoretical tools (DFT, MD simulation, and ML, as well as transport calculations) and experimental techniques (electrical, spectral, and structural characterizations). Here, we addressed a critical challenge in the computational modeling of charge transport in conducting polymers by developing a machine learning-based approach to predict transfer integrals directly from morphology without performing expensive quantum mechanical calculations. Next, the validity of the theoretically calculated conductivity is confirmed by experimental data, and the origin of solvent-induced conductivity variations is rationalized via an experimental study of the films structure. One of the most important outcomes of our work is that we demonstrated that ML can significantly accelerate mobility calculations, making it feasible to analyze large-scale polymer films that were previously computationally prohibitive. By extending ML techniques to conducting polymers, we provide a new framework for understanding charge transport in these systems, paving the way for more efficient material design and optimization, that has not been available before.

II. RESULTS AND DISCUSSIONS

This section presents the validation of the ML model for predicting transfer integrals between two PEDOT chains,

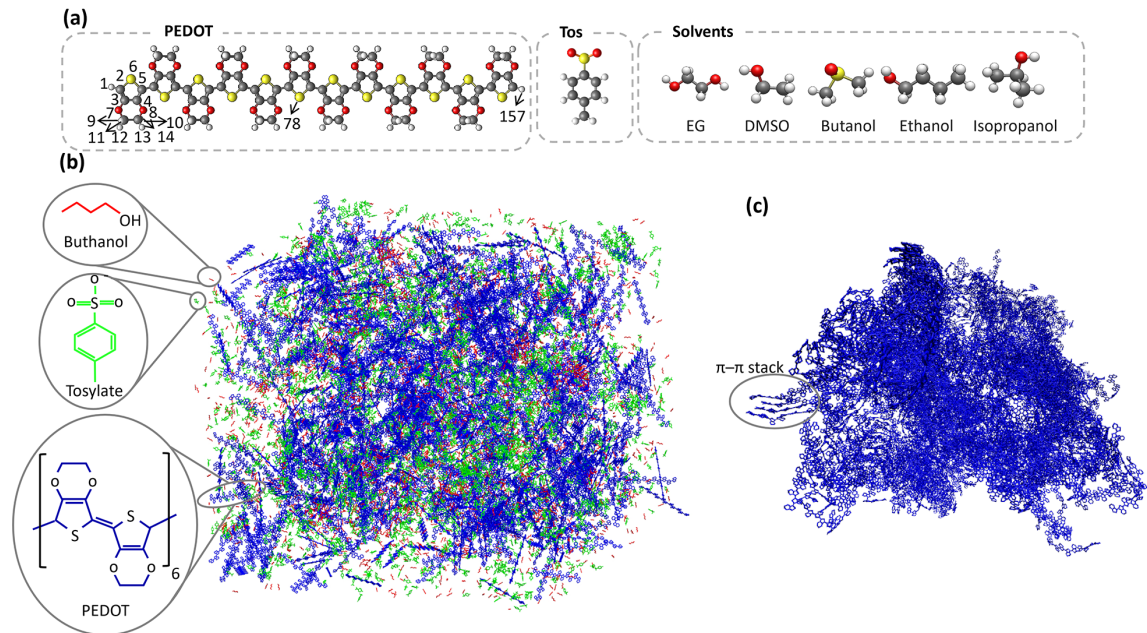


FIG. 1. (a) Optimized ball-and-stick model of the PEDOT chain containing $N = 12$ monomer units, Tosylate anions, Ethylene Glycol (EG), Ethanol, Dimethyl Sulfoxide (DMSO), Butanol, and Isopropanol. The numbers in this figure indicate the indices of the atoms used in the Coulomb matrix shown in Fig. 2. (b) Snapshot of the initial random system created using Packmol, consisting of 800 PEDOT chains, 3200 tosylate molecules, and 20% weight fraction of the solvent (Butanol). The chemical structures of Butanol (red), Tosylate (green), and PEDOT (blue) in $25 \times 25 \times 30 \text{ nm}^3$ computational box. (c) Snapshot of the final simulation stage showing a highlighted example of π - π stacked crystallite consisting of four PEDOT chains (encircled region); solvent is not shown in (c) for clarity.

followed by a comparison of mobility results obtained from ML and DFT. We also examine the effect of solvent on charge conductivity and sample morphologies, comparing experimental and simulation results. The studied system and its computational preparation are described in Sec. IV A, where Sec. IV A 1 describes computational morphology preparation, Sec. IV A 2 outlines electronic transport calculations, and Secs. IV A 3–IV A 5 present ML model and calculations. The experimental systems and methods are described in Sec. IV B.

A. Validation of the ML model for transfer integrals and mobility predictions

In this section, we aim to evaluate the prediction performance of the proposed ML model using the Coulomb matrix as the predictor for estimating transfer integrals (targets) between two PEDOT chains (hereafter referred to as dimers). Specifically, we investigate two key questions: (1) Can the ML model accurately predict transfer integral values for dimers derived from morphologies (see Fig. 1) influenced by solvents that were included in the training data (a procedure commonly referred to as internal validation in ML)? and (2) To what extent can the model predict transfer integral values for dimers from morphologies influenced by solvents that were not part of the training data (external validation in ML)?

We compare the transfer integral values directly obtained from the DFT calculations (hereafter referred to as DFT-based transfer integrals) with those predicted by the ML model (hereafter referred to as ML-based transfer integrals). Figure 2(c) presents the results of internal validation, which

addresses the first question regarding the model's ability to predict transfer integral values for dimers derived from morphologies influenced by solvents that the model has seen during training. The training and testing data for this validation are described in detail in Sec. IV A. The model achieves a high R^2 value of 0.99 and the mean absolute error [MAE, Eq. (13)] of 0.89 meV, which is significantly lower than the range of transfer integrals between -200 and 200 meV . This demonstrates that the ML model can accurately predict transfer integral values for such dimers. This confirms that the model performs well for unseen dimers provided by the full test dataset, which contains dimers from all morphologies prepared with different solvents. Note that, we also trained our ML model by utilizing the logarithm of the transfer integral as the target, following previous studies [42] that reported higher accuracy with this approach. The prediction results are depicted in Fig. S1 in Supplemental Material [43]. However, in our case, using the actual transfer integral as the target resulted in better performance with higher accuracy.

To address the second question posed above, we evaluate the model's generalization ability to predict transfer integrals for dimers derived from morphologies influenced by solvents that were not part of the training data. This ability ensures the model's practical applicability across diverse chemical environments for broader use. Figure 2(d) illustrates the results of this external validation. The dataset for this purpose is also described in Sec. IV A. In this case, the model achieves a high R^2 of 0.97 and an MAE of 1.02 meV. These results demonstrate that the ML model can accurately predict transfer integrals even for dimers from morphologies influenced by unseen solvents. This shows that, at the transfer integral

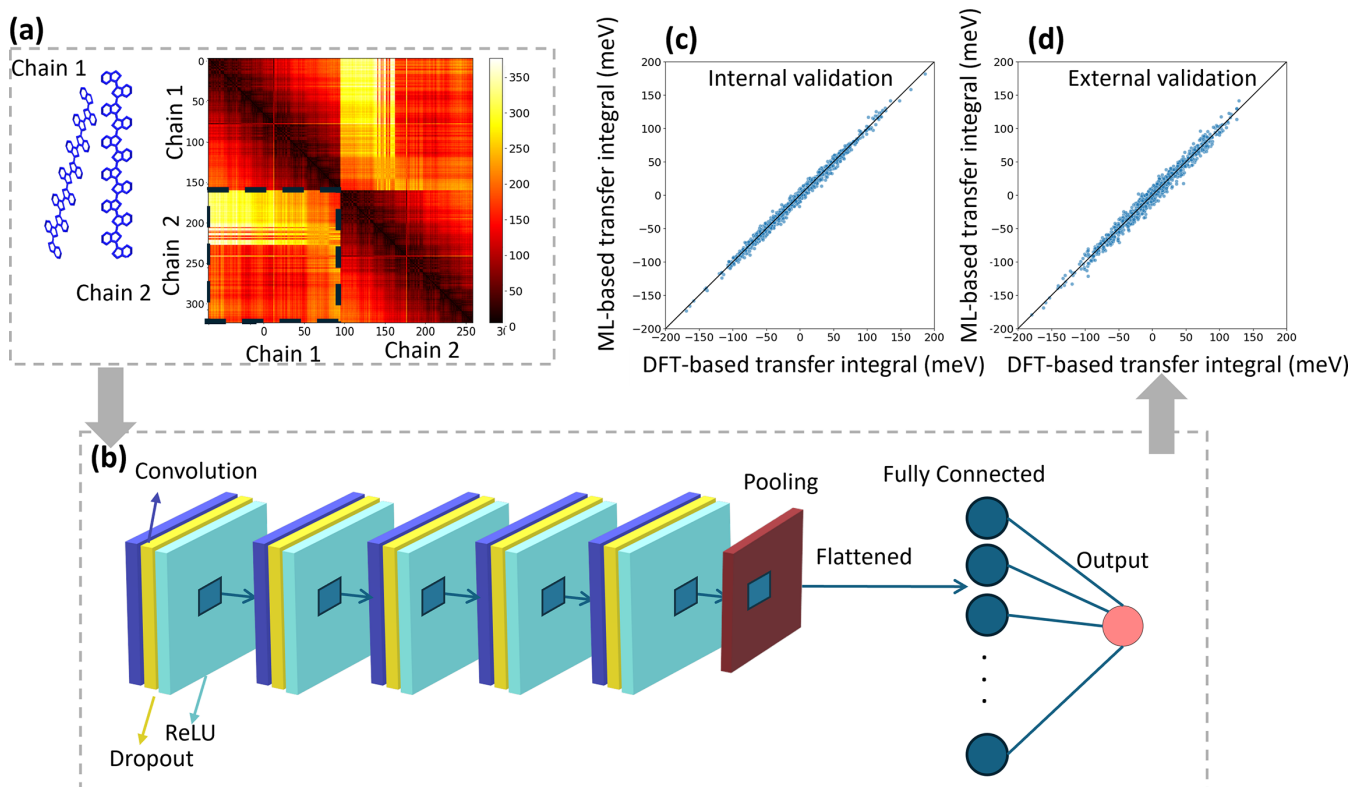


FIG. 2. (a) Heat map of the Coulomb matrix, with color intensity representing varying magnitudes. The dashed line highlights the interactive Coulomb matrix between two chains. For chain 1, atom indices are shown in Fig. 1, while for chain 2, indices continue sequentially from 158 to 314. (b) Schematic of a Convolutional Neural Network (CNN) architecture. The network consists of convolutional layers followed by activation functions (ReLU) and dropout layers for regularization. Pooling layers reduce spatial dimensions, and the output from the final pooling layer is flattened and passed through fully connected layers, leading to the final output. (c) and (d) Comparison of DFT-based and ML-based transfer integral values. (c) Internal validation results, highlighting the model's ability to predict transfer integral values for dimers derived from morphologies influenced by solvents present in the training dataset. (d) External validation results, showcasing the model's generalization ability to predict transfer integrals for dimers derived from morphologies influenced by solvents excluded from training.

level, the coupling between two chains in the dimer in a particular geometry is independent of solvent effects (i.e., it is not dependent on how this particular geometry has been achieved), enabling the model to generalize well across different chemical environments. The model's ability to predict transfer integrals for unseen data can be attributed to the fact that distribution of the transfer integrals is similar for all solvents, see Fig. S2 in the Supplemental Material [43].

Although the prediction accuracy of transfer integrals using our ML model has been validated, the implications of this accuracy for charge carrier mobility remain uncertain and therefore need to be ascertained. To address this, we calculate and compare the hole mobilities using DFT-based transfer integrals and ML-based transfer integrals.

DFT-based transfer integrals mobility calculations. Using the transfer integrals derived from DFT calculations, we calculated the hole mobility for systems prepared in solvents. The mean-field approach described in Sec. IV A was employed for these calculations. Note that to evaluate the isotropicity of mobility within these systems, we computed the mobility along the x , y , and z directions of the simulation box by applying an electric field in each respective direction. Mobility values across different directions show

negligible variation (see Table S1 in Supplemental Material [43]). Moreover, as outlined in Sec. IV A, three independent configurations were prepared for each solvent. Mobility values were calculated for all samples to assess reproducibility and to account for variations arising from sample preparation.

ML-based transfer integrals mobility calculations. After obtaining transfer integrals predicted by the ML model, we repeated the entire mobility calculation process explained above. This included the same mean-field approach, consideration of isotropic, and analysis of multiple configurations prepared with solvents, ensuring consistency in methodology between the DFT- and ML-based calculations.

Figure 3(a) presents a direct comparison between the DFT-based and ML-based mobility values. Note that the average mobility, as presented in Table S1 [43], was used for this comparison, and the variations between different configurations for each solvent are indicated by error bars. The results demonstrate excellent agreement between the two approaches, confirming the reliability of the ML model in predicting mobility values, with an MAE of $0.024 \text{ cm}^2/\text{V}\cdot\text{s}$. By confirming a full agreement between ML- and DFT-based results, this study establishes the ML model with the Coulomb matrix as the predictor as a viable and computationally efficient

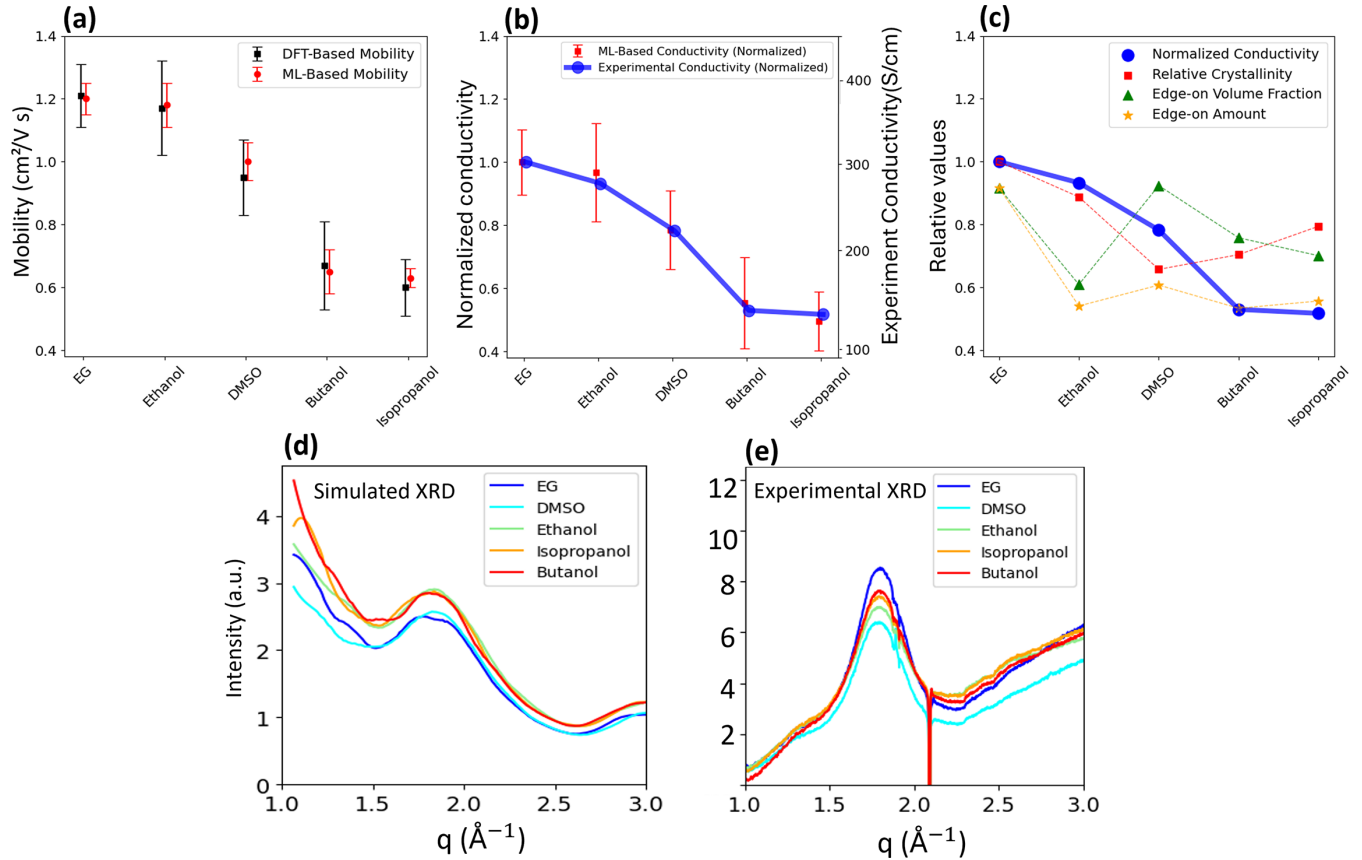


FIG. 3. (a) Comparison of DFT-based and ML-based mobility values. Error bars indicate variations across different configurations. (b) ML-based conductivity predictions compared with experimentally measured values, with the left axis showing after normalization with the conductivity of the most conducting sample, to allow comparison between the two approaches and the right axis showing actual conductivity values. (c) Relative crystallinity of the various PEDOT:Tos films, normalized to the crystallinity of the most crystalline film (the one immersed in EG). The volume fraction of edge-on oriented crystallites and the edge-on amount of PEDOT chain content, both derived from polar plot analysis ($I(\chi) \times \sin(\chi)$ vs χ) are also included in the same plot. (d) Simulated x-ray diffraction (XRD) patterns. (e) Experimental GIWAXS patterns recorded for the various samples. Only the q -range above 1 ($\text{\AA}^{-1}$) is presented herein, to allow direct comparison with the simulated XRD patterns.

way for mobility calculations. This is absolutely essential for large-scale or complex systems where traditional DFT calculations are computationally impractical.

B. Solvent effects on conductivity: Experimental and simulation insights

The effect of different solvents on the electrical conductivity of PEDOT:Tos thin films is explored herein as a case study to validate the mobility obtained using machine-learned transfer integrals through comparison with experimental data. Such a comparison provides a robust and comprehensive validation of both the ML model's applicability and the entire multiscale approach employed in this study.

The electrical conductivity was calculated using the Drude equation [Eq. (6)], as detailed in Sec. IV A. To apply the Drude equation, the charge carrier concentration and mobility of the system are required. Since PEDOT:Tos is responsible for charge transport in the studied systems, the charge carrier concentration remains the same across all calculated systems (as well as in all experimental systems, see below) and is set to be $n = 2.16 \times 10^{21} \text{ cm}^{-3}$ (Sec. IV A). Using this charge car-

rier concentration along with the ML-predicted mobility for each solvent, we derived the relative ML-based conductivity, as presented in Fig. 3(b). Note that the arbitrary parameter ω_0 in the Miller-Abrahams formula [Eq. (3)] is used as a fitting parameter, which is customary for this approach.

In the same Fig. 3(b) we compare the ML predicted conductivity values with the experimentally measured results. The comparison demonstrates excellent agreement between the two. Specifically, the conductivity of PEDOT:Tos after immersion in ethylene glycol and ethanol is the highest, followed by DMSO, and finally, butanol and isopropanol show the lowest conductivity. In all cases, experimental conductivity lies within the error bar of the ML-predicted value. This agreement validates the ML model's ability to estimate conductivity in complex systems like PEDOT.

The trend observed in the solvent-dependent electrical conductivity can either be rationalized by changes in the oxidation levels of the materials or by changing of the thin film morphology. As seen in the x-ray photoelectron spectroscopy (XPS) data presented in the Supplemental Material (Fig. S4) [43], the oxidation level of the PEDOT systems is nearly the same after the different post-processing conditions. Consequently, the

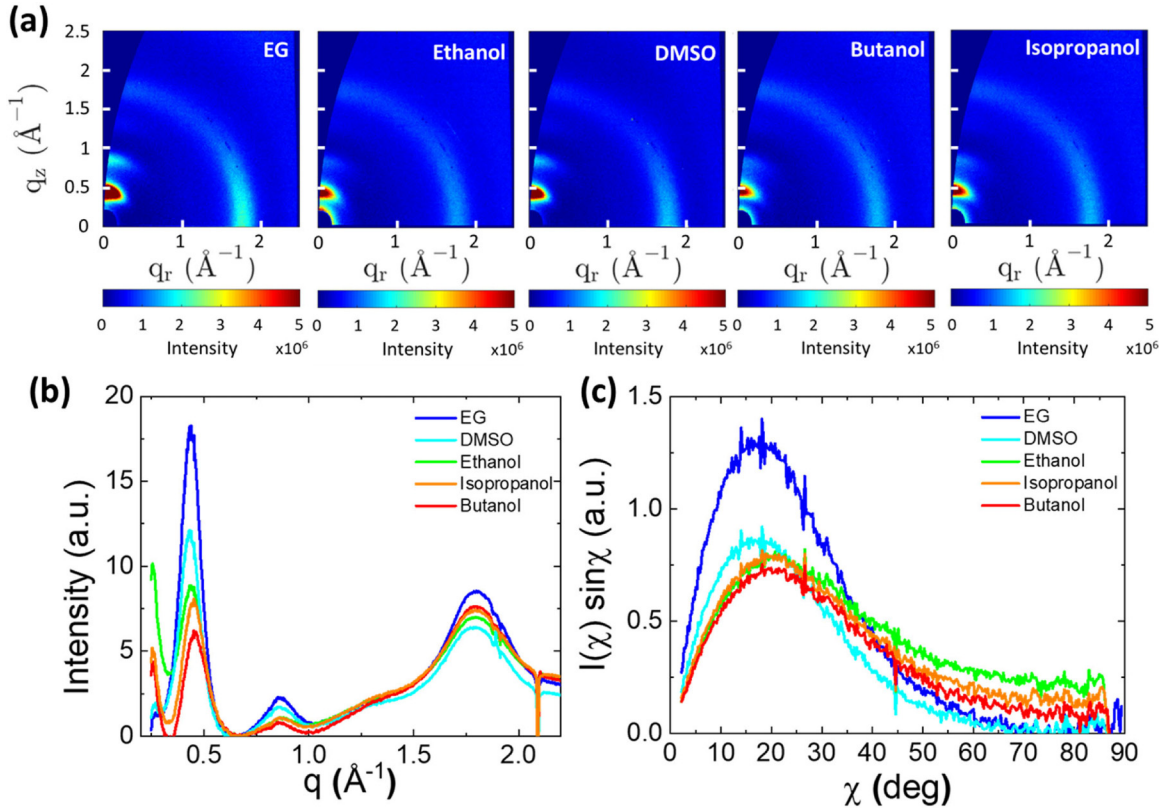


FIG. 4. (a) Wedge-corrected 2D GIWAXS images of the PEDOT:Tos films, after immersion in the various solvents. (b) The radially averaged scattered intensity with respect to the scattering vector q . (c) The polar plots ($I(\chi) \times \sin(\chi)$ vs χ) for the (100) peak of the various PEDOT:Tos thin films. In all cases the scattering intensity was normalized by film thickness and the substrate scattering was subtracted.

observed electrical behavior should be attributed to variations in the thin film mobility, caused by the difference in morphology. In order to study the structure of the PEDOT:Tos films we performed grazing incidence wide angle x-ray scattering (GIWAXS). We analyze three different structural parameters trying to correlate them with the observed behavior of conductivity. This includes (a) degree of crystallinity, (b) the edge-on volume fraction, and (c) the amount of edge-on crystallites.

(a) *Degree of crystallinity.* The wedge-corrected (q_r , q_z) 2D scattering patterns recorded for the PEDOT:Tos films after immersion in the various solvents, as well as the radially averaged 1D scattering patterns are presented in Figs. 4(a) and 4(b). Note that scattered intensity is presented after all necessary corrections and normalizations, to allow the quantitative comparison between samples. The strong peak at $0.44 \text{ \AA}^{-1}$ and the following peak at $0.86 \text{ \AA}^{-1}$ are assigned to the (100) and (200) reflections, respectively, and correspond to a lamellar distance of $14.28 \text{ \AA}$ [44]. The peak at $1.8 \text{ \AA}^{-1}$ is assigned to the (020) reflection and corresponds to the π - π stacking direction.

The variation of scattered intensity among the different samples can be used to derive a relative degree of crystallinity. We consider the PEDOT:Tos films as “in-plane powders” (i.e., isotropic in-plane orientation of the crystallites) and we calculate the $I(\chi)$ vs χ polar plots for the (100) peak, with (Fig. 4), χ being the polar angle defined with respect to the normal direction to the film plane [45]. The area of the

$I(\chi) \times \sin(\chi)$ vs χ plots is a measure of relative crystallinity [20] and will be used next to correlate electrical conductivity to thin film morphology.

(b) *The edge-on volume fraction* is a second structural parameter that can help elucidate charge transport. Polar plots shown in Fig. 3(a) are also employed to derive quantitative information on the orientation of PEDOT crystallites with respect to the substrate. Already, the anisotropy of scattering features along the in-plane (q_r) and near out-of-plane (q_z) directions in the 2D images demonstrates a preferential orientation of the PEDOT crystallites at an “edge-on” configuration with respect to the substrate, i.e., the polymer chains are aligned with the π - π stacking direction parallel to the substrate. (Note that formation of the “edge-on” configuration was explained in [46] by the hydrophilicity of the substrate surface.) Polar plots can be used to calculate the volume fraction of these edge-on oriented crystallites. Following the procedure reported elsewhere [20] the volume fraction of edge-on oriented crystallites can be estimated from the ratio between the $I(\chi) \times \sin(\chi)$ vs χ integral from 0° to 45° and the one from 0° to 90° , excluding the fraction of isotropically oriented crystallites [47].

(c) *The amount of edge-on crystallites.* Finally, the combination of relative crystallinity and edge-on fraction constitutes a third structural parameter that affects in-plane charge transport. In a previous study of PEDOT:Tos thin films treated with solvent additives, we have documented that in-plane charge transport is affected by the combination of film

crystallinity and crystallite edge-on orientation [20]. This combination was quantified as the product of the relative crystallinity and the edge-on fraction, which represents the amount of edge-on crystallites in the film. Since relative crystallinity is employed, this amount is a relative quantity that allows the qualitative comparison between the edge-on amounts of a series of samples.

In order to rationalize our experimental and theoretical results on conductivity, we attempt to draw herein correlations with the three aforementioned structural parameters. Crystallinity has a well-known positive effect in PEDOT:Tos conductivity [48]. Figure 3(c) presents the relative crystallinity (plotted after normalization to the crystallinity of the most crystalline film—the film immersed in EG) compared to the experimental conductivity. For the three more conducting films (EG, ethanol, and DMSO), the conductivity follows nicely the trend of crystallinity. Notably, the crystallinity of the films immersed in butanol and isopropanol is significant, intermediate to that of DMSO and ethanol, yet conductivity is much lower. This suggests that relative crystallinity is not the sole structural parameter affecting conductivity. The effect of crystallite orientation should be also evaluated, given that edge-on orientation is necessary to achieve high in-plane mobility because it promotes interchain and intrachain hopping transport along the backbone [49,50].

The volume fraction of edge-on oriented crystallites for the samples under study is presented in Fig. 3(c). A high volume fraction of 0.6 to 0.9 is observed for all samples, with no obvious trend and no clear correlation to conductivity. A closer look at the data reveals that bathing in butanol and isopropanol results in lower edge-on fraction than in EG or DMSO. This could justify the lower conductivity of these two samples and indicate that edge-on orientation is indeed an important parameter for charge transport in PEDOT:Tos, with higher edge-on fractions promoting higher conductivity. In that view, the sample immersed in ethanol should be considered as an outlier.

Finally, we turn to the amount of edge-on oriented crystallites [Fig. 3(c)]. Besides the film immersed in ethylene glycol which exhibits the highest edge-on amount, all other films have a similar edge-on amount of around 0.5–0.6. This observation implies that crystallinity and edge-on fraction are not enough to explain charge transport. Among the two parameters our results show that relative crystallinity matches better both the measured and calculated conductivities, suggesting that it has a stronger effect on charge transport. At the same time, we acknowledge that the absence of a perfect correlation between the studied structural parameters and the observed conductivities is not unexpected. Conductivity in these materials is influenced by multiple factors, making them inherently complex and not easily reducible to a single parameter. As we have demonstrated in a previous work [34], an efficient network of π – π stacked chains extending throughout the entire sample is essential to achieve high mobility. However, analysis of GIWAXS images does not provide sufficient information to support this argument and dedicated experiments should be performed. Yet, the present study represents a first step and hopefully will inspire further experimental and theoretical studies to fully enlighten the structure-charge transport relationships in conducting polymers.

Given the demonstrated excellent agreement between experimental and ML-derived conductivity, and the insights into the structure-charge transport relationship we gained from GIWAXS, we moved further and calculated the XRD patterns for the various samples, in order to compare them with the experimental GIWAXS data. At this point, it should be stressed that XRD refers to diffraction while GIWAXS to scattering, as the two experiments are performed in different sample-detector geometries. However, the recorded patterns are similar and the information they contain is practically the same. Figures 3(c) and 3(d) present the simulated and experimental data at the scattering vector range $1\text{--}3\text{ \AA}^{-1}$, focusing on the π – π peak. In both approaches, the (020) π – π peak position is practically constant. From simulations, the corresponding π – π stacking distance is around $3.44\text{ \AA}$, while experimental data show a slightly larger stacking distance of $3.48\text{ \AA}$. Despite this minor difference in the π – π stacking distance, the overall agreement between the two suggests that immersion of PEDOT:Tos in different solvents does not significantly influence the π – π stacking interaction.

However, discrepancies between simulation and experiment arise when comparing the range of structural disorder of the π – π stacks. The simulation shows a broad π – π peak, indicating significant disorder. In contrast, the experimental results reveal a sharper π – π peak, suggesting less disorder and larger crystallite sizes. We attribute the discrepancy between the simulated XRD pattern and the experiment GIWAXS to the absence of a substrate in the simulation, a limitation of the computational approach. For further details on the impact of the substrate, refer to our recent study [46]. Employing the Scherrer equation [51] to the simulated data, the crystallite size along the π – π direction is calculated to be $7.4\text{--}8\text{ \AA}$, suggesting that three polymer chains are packed along the π – π stacking per crystal. On the other hand, experimental data show that the crystallite size varies slightly with solvent, ranging from $13.5\text{ \AA}$ for ethanol to $15.2\text{ \AA}$ for EG, implying that five polymer chains are involved in π – π stacking per crystal. These results suggest that, in practice, crystallite size along this direction does not affect charge transport.

III. CONCLUSION

This study provides an exploration of how solvent post-treatments influence the charge carrier mobility and electrical conductivity of PEDOT:Tos thin films through a combined experimental and theoretical approach guided by the ML model.

The first key finding is that solvent post-treatment significantly affects the morphology of PEDOT:Tos thin films, ultimately influencing charge carrier mobility and electrical conductivity. Specifically, the conductivity of PEDOT thin films follows this order based on the solvent used: Ethylene Glycol, Ethanol, DMSO, Butanol, and Isopropanol. Using the XPS spectroscopy we found that the oxidation level of all systems was nearly the same; consequently, the difference in conductivity was attributed to variations in the thin film mobility, caused by the difference in morphology. In order to study the morphological structure of the PEDOT:Tos films we performed grazing incidence wide angle x-ray scattering (GIWAXS). We analyze three different structural parameters including (a) degree of crystallinity, (b) the edge-on volume

fraction, and (c) the amount of edge-on crystallites, trying to correlate them with the observed behavior of conductivity. Among all these parameters the relative crystallinity matches better the measured and calculated conductivities, suggesting that it has a stronger effect on charge transport. However, better understanding of the relationship between solvent-affected structure and charge transport in conducting polymers is desirable. Therefore, this study marks an initial step and aims to inspire further experimental and theoretical investigations to gain a better understanding of the relationship between solvent-affected structure and charge transport in conducting polymers.

To quantify the link between morphology and conductivity and to overcome the computational limitations of traditional multiscale transport calculations, we developed an ML model as an integral part of the study. This model was specifically designed to predict transfer integrals—key parameters for charge transport—from morphological data, paving the way for understanding charge transport in the system of realistic dimensionality, that has not been possible before. We employed CNN techniques to predict transfer integrals, which are the bottleneck of multiscale calculations, using the interactive part of the Coulomb matrix as the predictor and the transfer integral as the target. Our results show that the ML model achieved an R^2 value of 0.99 for internal validation and an R^2 of 0.97 for external validation, confirming the model's high accuracy and robustness.

The ML-guided mobility simulations closely reproduce the experimental results, establishing a computationally efficient tool for large-scale analysis of complex systems by directly computing transfer integrals from the generated morphology without the need for costly quantum mechanical calculations. In this way, the ML framework was not developed as a standalone tool but as a critical link that connects the experimental observations to charge transport mechanisms, enabling a predictive understanding of structure–property relationships. We believe this integrated approach provides a valuable methodology for future investigations of conducting polymers. The ML files are freely available at our GitHub repository [52], allowing interested users to adapt this model for various polymeric systems.

While this study marks a significant breakthrough in ML-guided conductivity calculations for polymer films, it represents only an initial step, with many aspects open for further exploration. For instance, our current approach utilized all-atom calculations, which significantly constrained the system size. Implementing a coarse-grained approach with consecutive backmapping [53], combined with ML techniques, would enable the study of conductivity in realistically large systems. Additionally, accounting for the influence of the substrate can be important, as it is known to affect thin-film morphology in real-world applications [46]. Further systematic investigations into the role of solvents would also be valuable, including exploring various factors such as temperature and different polymerization setups. Finally, we note that the multiscale calculations and the developed ML model are specifically designed for PEDOT, a polymer with short chains where intrachain motion can be neglected and interchain hopping dominates charge transport. For polymers with longer chains, both intra- and interchain transport mechanisms may play

significant roles. Addressing such systems would require extending the current ML model or developing new models capable of capturing these additional complexities.

IV. METHODS

A. Simulation methods

This section describes the method used to calculate the conductivity step-by-step. A workflow summarizing the process and the software used in each step is shown in Fig. 5.

1. Morphology preparation

DFT. Geometry optimization was performed using DFT for a PEDOT chain consisting of $N = 12$ monomer units, which aligns with the experimentally estimated range of $N = 5–20$ monomer units [54,55]. Since the relative concentration of charge carriers (holes) per monomer in PEDOT is 33%, [5,7,17,56], the total charge for a 12-monomer chain is set to be $Q = +4e$. Geometry optimizations were also carried out for Tos anions, and solvent molecules, including EG, Ethanol, DMSO, Butanol, and Isopropanol individually. All chemical structures are shown in Fig. 1(a). Following the geometry optimizations, partial atomic charges for all components in the system were determined using DFT calculations. These charges were obtained by fitting to the results of Mulliken [57] population analysis. All DFT calculations were performed using Gaussian software [58] with the B3LYP functional [59] and the 6–31+G(d) basis set [60].

Generating initial configuration. To study the effect of solvents on PEDOT conductivity, we generated different solutions, each consisting of charged PEDOT oligomers and Tos anions immersed in one of the mentioned solvents (i.e., EG, Ethanol, DMSO, Butanol, and Isopropanol). PEDOT chains, Tos anions, and chosen solvent molecules with the optimized geometries of each molecule were placed randomly in a $25 \times 25 \times 30$ nm computational box using Packmol [61,62]. To ensure the system is electrically neutral, the number of Tos anions was four times the number of PEDOT chains (specifically 3200 Tos anions and 800 PEDOT chains) to compensate for the positive charges of the PEDOT oligomers. The initial solvent concentration was set to 20% w/w, but the number of solvent molecules varies depending on the type of solvent used. To be more specific, the number of molecules for EG, Ethanol, DMSO, Butanol, and Isopropanol solvents was 7430, 10010, 6220, 6220, and 7675, respectively.

Molecular dynamics simulation. Having generated the initial morphologies from Packmol, we performed MD simulations within the LAMMPS software [63]. For this purpose, Moltemplate [64] is used to assign force field parameters, based on the general AMBER force field (GAFF) [65], and to generate the input files for LAMMPS. The system was initially minimized and equilibrated for 2 ns in the canonical NVT ensemble [i.e., Constant Number of particles (N), Volume (V), and Temperature (T)] at 300 K using the Nose–Hoover thermostat. Time integration was performed using the Verlet algorithm [66] with a time step of 1 fs. Following NVT equilibration, NPT [i.e., Constant Number of particles (N), Pressure (P), and Temperature (T)] equilibration was conducted using the isothermal-isobaric NPT

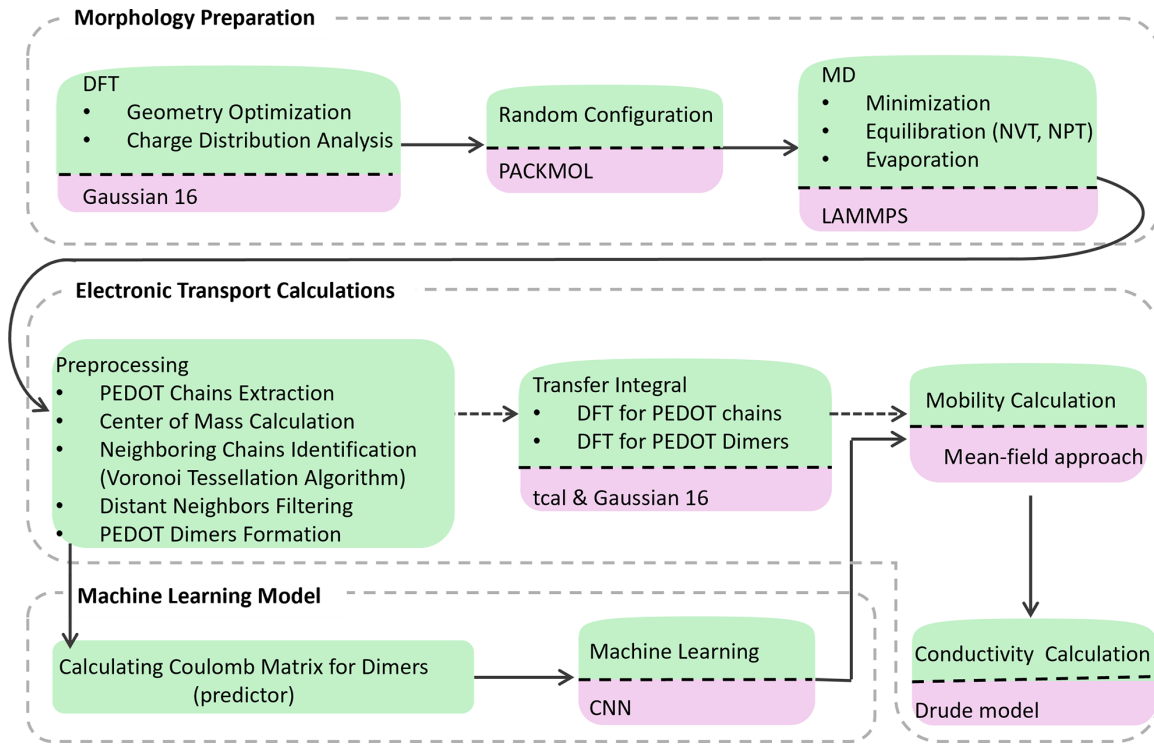


FIG. 5. Flowchart illustrating the simulation steps for the conductivity calculations.

ensemble at 1 atm and 300 K for 2 ns, with both the barostat and thermostat controlled by the Nose–Hoover method. After equilibration, the solvent concentration was reduced from 20% w/w, representing the initially hydrated phase, to 10% w/w, simulating a partially dry phase. This step simulates the solvent evaporation process in the experimental setup. Finally, NPT equilibration was conducted again for 2 ns after the evaporation step to allow the system to adjust to the new box size. The boxes for the initial and final configurations are shown in Figs. 1(b) and 2(c). The periodic boundary conditions were applied in all directions to simulate bulk behavior.

In Note 1 of Supplemental Material [43], we compared the current morphology with the morphology obtained using techniques from previous studies [54] where a high solvent content (about 80% w/w) was initially used, followed by a gradual and iterative solvent evaporation process. This comparison highlights that the fabrication technique used in this study yields similar morphology. However, our approach significantly reduces both the time and complexity of the process by speeding up solvent evaporation process.

X-ray diffraction pattern. X-ray diffraction patterns were simulated based on the approach outlined by Coleman *et al.* [67] and implemented using the LAMMPS software [63]. The relationship between the Bragg reflection angle (2θ) and the scattering vector Q is given by $Q = \frac{4\pi}{\lambda} \sin(2\theta)$ where λ is the $Cu_{k\alpha 1}$ wavelength ($\lambda_{k\alpha 1} = 1.540593 \text{ \AA}$), commonly applied in GIWAXS studies [68–70].

2. Electronic transport calculations

To calculate polaron mobilities in PEDOT films we first start with the calculated polymer morphology and follow a procedure for a multiscale calculations as described in

[34,53]. As in [34,53], the entire PEDOT chain is treated as a transport site that localizes one charge carrier. The intra-chain transport in PEDOT is much faster than interchain one; therefore the primary transport mechanism in this system is the hopping of charge carriers between neighboring PEDOT chains (dimer) [71]. Note that the hopping of charge carriers between non-neighboring PEDOT chains was disregarded, as they rarely contribute to the overall charge transport [34].

To further clarify the hopping process, the energy diagram of PEDOT chains in a dimer and corresponding polaron state are depicted in Fig. 6, illustrating a representative case of polaron hopping between two chains. The polaron represents an unoccupied state in the gap between the valence and conduction bands. (For details of the polaron energetics in the conducting polymers, see [32,72]). The hopping process involves three key steps as follows. In the initial state, the polaron is localized on chain 1. Then, in the intermediate step, the polaron moves from chain 1 to chain 2, which means that an unoccupied orbital in chain 1 is filled by an electron that jumps from chain 2, as illustrated in Figs. 6(a) and 6(b). In the final state, the polaron becomes localized on chain 2, completing the hopping process, see Fig. 6(c). The hopping process is quantified by the transfer integral, which describes the coupling between two transport sites (PEDOT chains). Because the polaron hopping corresponds to the transfer of electrons between the valence bands of two chains, the hopping integrals are calculated between the electronic states in the valence band of these chains. Once the hopping rates are determined using a modified Miller–Abrahams formalism, the carrier mobility is calculated by solving the Master Equation [Eq. (5)] for charge transport. This approach bridges the gap between the system’s morphology and the material’s overall conductivity using the transfer integrals and the hopping rates.

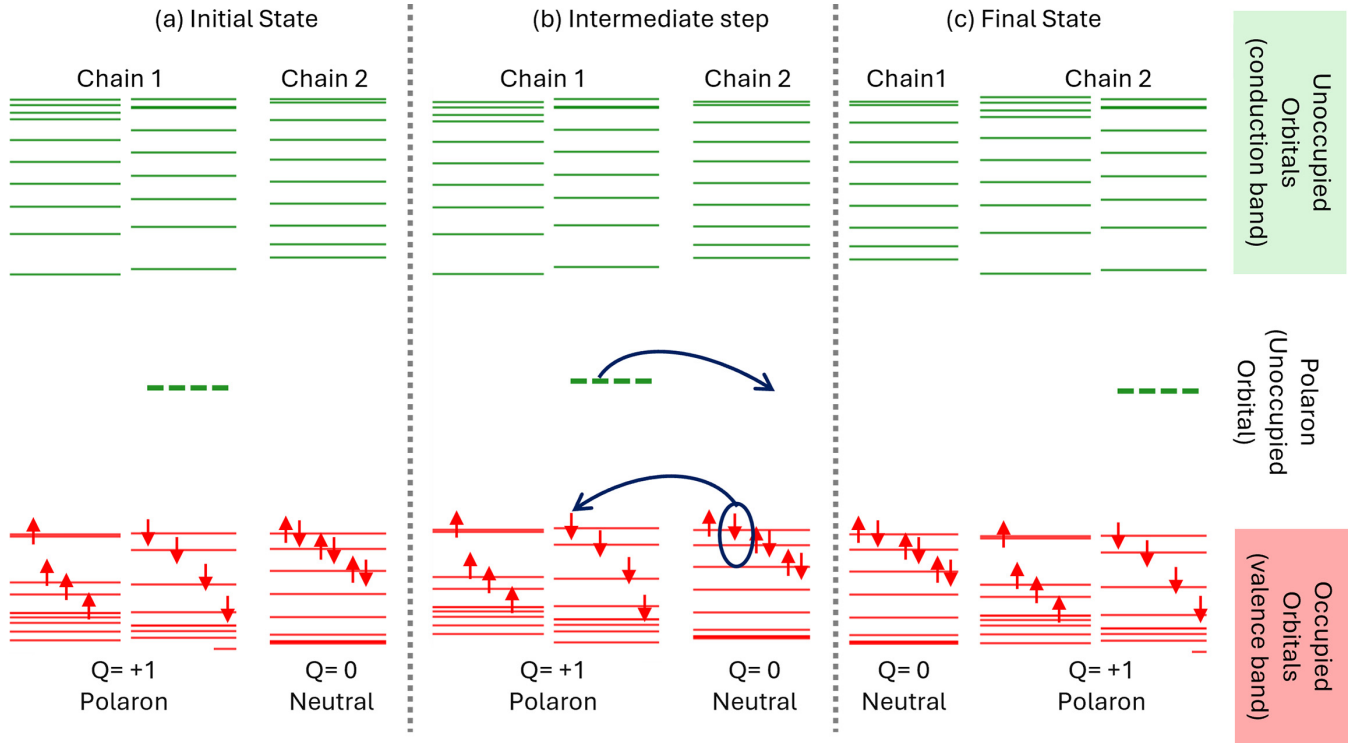


FIG. 6. Electronic structure of two PEDOT chains, highlighting the hopping of a polaron between them corresponding to the transfer of electrons between the valence bands of two chains.

In what follows, we focus on calculating the conductivity for each morphology generated using the simulation technique outlined in Sec. IV A. Here, we will provide a detailed explanation of each step involved in determining the conductivity of different morphologies.

Preprocessing. To calculate the transfer integral, we first calculate morphologies of polymer films with different solvents as described in Sec. IV A, see Fig. 1(c) for a representative snapshot. Then we identify dimers that contribute to charge transport. This process begins by extracting the PEDOT chains' atomic positions from the final configuration in all morphologies and calculating chains' centers of the mass. Neighboring molecules were determined using the Voronoi tessellation algorithm [54], and dimers with a center-of-mass distance greater than 15 Å were excluded, as their contribution to charge transport was considered negligible.

Transfer integral. For each dimer, the transfer integral was calculated using the Python library “tcal” [73], which extracts the electronic structure of both individual chains and their corresponding dimers. Specifically, tcal generates Gaussian input files for each chain comprising the dimer and subsequently calls Gaussian to perform DFT calculations for both individual chains and the dimer as a whole. These calculations provide the coefficients of the basis functions, the overlap matrix $\langle\psi_i|\psi_j\rangle$, and the Fock matrix, which are then used to compute the transfer integrals.

Particularly, in order to calculate the transfer integral between sites i and j , tcal obtains the highest occupied molecular orbitals (HOMO) $|\psi_i\rangle$ and $|\psi_j\rangle$ from each individual site calculation, and derives the Fock matrix, F , which is given by

$$F = T + V + G, \quad (1)$$

where T is the kinetic energy term, V is the electron-nuclear attraction term, and G is the two-electron repulsion term [74], as implemented in Gaussian software [58].

Finally, the intermolecular transfer integral, H_{ij} , was determined using the following equation [75]:

$$H_{ij} = \frac{\langle\psi_i|F|\psi_j\rangle - \frac{1}{2}(\epsilon_i + \epsilon_j)\langle\psi_i|\psi_j\rangle}{1 - \langle\psi_i|\psi_j\rangle^2}, \quad (2)$$

where $\epsilon_i = \langle\psi_i|F|\psi_i\rangle$ and $\epsilon_j = \langle\psi_j|F|\psi_j\rangle$.

Mobility calculation. With the transfer integral for each dimer available, the calculation of mobility relies on the hopping rate as a key intermediary. In this paper, we determined hopping rates, w_{ij} , using a modified Miller–Abrahams formalism [14,76,77]

$$w_{ij} = \begin{cases} w_0 |H_{ij}|^2 & \text{if } E_i \geq E_j \\ w_0 |H_{ij}|^2 \exp\left(\frac{-\Delta E_{ij}}{k_B T}\right) & \text{if } E_i < E_j \end{cases}, \quad (3)$$

where w_0 is a prefactor, set to $8 \times 10^3 \text{ J}^{-2} \text{ s}^{-1}$, which serves as a phenomenological parameter chosen arbitrarily to ensure that the calculated conductivity matches the experimental value, and $\Delta E_{i,j} = E_j - E_i - eFr_{ij,F}$ in which E_i and E_j are the HOMO energy of chain i and j . They are computed using Gaussian software [58] by single point DFT energy calculations based on the geometry obtained from the MD simulations. Note that an alternative way to obtain E_i and E_j can be to draw them from the Gaussian distribution that approximates the density of states (DOS) as it was done in our previous work [34]. $r_{ij,F}$ is the distance between the sites i and j in the direction of the applied electric field shown by F . In this paper, we chose F to be $10^4 \frac{\text{V}}{\text{cm}}$ so that it can force charge to hop from site i to another site and T is the

temperature ($T = 300$ K). It is worth noting that the Miller–Abrahams formalism used in Eq. (3) can be viewed as an empirical model for phonon-assisted tunneling, where the influence of the environment is included implicitly through the activation energy. A more rigorous and physically grounded description is provided by Marcus theory, which treats electron transfer as quantum nuclear tunneling between coupled electronic and vibrational states [14,76]. In Marcus theory, the environment’s effect is accounted for via the reorganization energy, offering a more accurate representation of delocalized polarons. However, it has been shown that both models yield very similar results, [78–80] largely due to the presence of a tunable phenomenological parameter in the Miller–Abrahams expression. Therefore, to maintain simplicity without sacrificing accuracy, we employ the Miller–Abrahams model in our work, as given by Eq. (3).

Then, the mobility μ was calculated using a mean-field approach, as described by Yu *et al.* [81], according to the following equation:

$$\mu = \frac{\sum_{i,j,i \neq j} w_{ij} p_i (1 - p_i) r_{ij,F}}{PNF}. \quad (4)$$

In this equation, w_{ij} are the hopping rates between sites i and j , p_i is the occupation probabilities for each site. The term P refers to the average probability, and N is the total number of sites in the system. Following our previous work [34], P is set to $P = 4 \times 10^{-2}$. Note that in this study the

experimental oxidation level (and therefore the charge carrier concentration) is fixed and not varied. Therefore, we do not consider here the dependence of the mobility on the charge density. The occupation probabilities, p_i , for each hopping site are determined by solving the master equation as follows:

$$\sum_{j \neq i} [w_{ij} p_j (1 - p_i) - w_{ji} p_i (1 - p_j)] = 0. \quad (5)$$

The solution is obtained iteratively until the probabilities converge. Once the values are known, the charge carrier mobility can be readily calculated.

Conductivity calculation. The conductivity σ is calculated using the Drude model equation

$$\sigma = ne\mu, \quad (6)$$

where n is the charge carrier concentration, e is the elementary charge, and μ is the charge carrier mobility obtained from the previous calculation. To determine the charge carrier concentration (n), we use the PEDOT:Tos density and oxidation level of PEDOT. To be more specific, the density of PEDOT:Tos was taken as 1.49 g/cm^3 [44,82]. Each PEDOT chain, consisting of 12 monomers, carries four charges due to a 33% oxidation level. Using the molecular weight of a PEDOT chain (1636.188 g/mol) and Avogadro’s number, we calculated the charge carrier concentration as follows:

$$n = \left(\frac{\text{Density of PEDOT}}{\text{Molecular weight of a PEDOT chain}} \right) \times \text{Avogadro's Number} \times \text{Charges per Chain}. \quad (7)$$

This calculation yields a charge carrier concentration of $n = 2.16 \times 10^{21} \text{ cm}^{-3}$. Note that this value is typical for PEDOT, which is highly charged in its pristine (i.e., as-polymerized) form, and is consistent with previous studies that report similar or even higher values [8,20,83–85].

3. Machine learning model

In this subsection, we explore the application of the ML model to calculate transfer integrals for polymer thin films. Recent studies have demonstrated the potential of ML in predicting transfer integrals and charge mobilities in organic crystals with acceptable accuracy by leveraging system-specific features and innovative molecular representations. For instance, Lederer *et al.* [39] focused on thermally disordered pentacene crystals and employed kernel ridge regression as their ML model. By leveraging system-specific geometric features (predictors), their approach accurately predicted transfer integrals and successfully captured mobility anisotropies consistent with experimental observations. Çaylak *et al.* [86] turned their attention to Alq3 molecules and applied a deep feed-forward neural network to predict transfer integrals. Their model utilized the full Coulomb matrix as the predictor, producing transfer integrals (targets) that were subsequently used to compute mobilities. Wang *et al.* [40,41] explored naphthalene as their material system and used kernel ridge regression with Gaussian and Laplacian

kernels to predict transfer integrals. By employing the Coulomb matrix feature, they achieved moderate accuracy. When the same ML framework was applied to naphthalene, however, prediction errors increased significantly. Recently, Rinderle *et al.* [42] focused on pentacene thin films and evaluated both kernel ridge regression and support vector regression models. Through a systematic comparison of different molecular representations, including geometric and Coulomb matrices, they identified the interaction part of the Coulomb matrix as the most effective predictor. Remarkably, the interaction part of the Coulomb matrix also demonstrated strong transferability to other materials, including tetracene and DNTT, making it a versatile and powerful representation as the target in ML models in this field. Despite this progress, the application of state-of-the-art ML techniques, such as CNNs, remains underexplored.

In the present study ML is used to estimate transfer integrals in a polymer film by learning the relationship between molecular morphology data from MD simulations and the corresponding transfer integrals computed via quantum mechanical methods. Molecular morphology is represented by the Coulomb matrix for each dimer, which encodes important electrostatic and geometric information. [For definition of the Coulomb matrix see below, Eq. (8)]. The Coulomb matrix serves as the predictor for the ML model, and the model learns to predict the transfer integrals (targets) by identifying patterns within the Coulomb matrix. (Note that in ML models the

“predictors” can also be called as “features”, “variables”, or “attributes”, whereas “targets” are called as “labels”, “output”, or “outcomes”). Once trained, the ML model can quickly estimate the transfer integral for new dimers by simply inputting their Coulomb matrices, thus eliminating the need for expensive quantum mechanical calculations. (The calculation of the Coulomb matrices is by far faster than transfer integrals). This approach enables rapid and accurate predictions of transfer integrals across large molecular systems, making it a powerful tool for studying charge transport. In the following, we provide a detailed explanation of the dataset, target engineering, and CNN model used in this study.

Data collection. To leverage a data-driven model for predicting transfer integrals, we first need to collect a dataset to feed into the ML model. As datasets, we use our morphologies detailed in Sec. IV A. Specifically, we generated bulk systems of PEDOT immersed in five different solvents (i.e., EG, Ethanol, DMSO, Butanol, and Isopropanol) individually. We repeated this process three times to create a larger and more diverse dataset. Overall, considering that each system contains 800 PEDOT chains, initiated with three different distributions ($800 \times 3 = 2400$), and each treated in five different solvents, we ended up with a total of $2400 \times 5 = 12\,000$ PEDOT chains. After preprocessing the data, as outlined in Sec. IV A 2, we ended up with 19 078 dimers, which form the dataset for our ML model. Then, we used the calculated transfer integrals as the targets for this dataset.

To evaluate the accuracy and generalization of our model, we perform both internal and external validation. For internal validation, we divide the full dataset, which contains dimers from all morphologies prepared with different solvents, into training and test subsets. Specifically, 80% of the data is used for training, while the remaining 20% is used for testing. Note that the way the data is divided between the training and test sets is consistent, ensuring that the test set fairly represents the entire dataset without any bias. The train/test distribution across our dataset is presented in Fig. S3(a) in Supplemental Material [43].

For external validation, we aim to assess the model’s generalization ability by predicting the transfer integral for PEDOT dimers immersed in previously unseen solvents. In this case, the model is trained on dimers from all solvents except those prepared in Butanol. The test set consists solely of dimers from the PEDOT morphology prepared in Butanol. Figure S3(b) in Supplemental Material [43] illustrates the train/test distribution across our dataset in this case. This approach ensures that the model’s performance is evaluated on a completely independent dataset, confirming its extendibility and applicability to new untrained morphologies.

Feature engineering. Feature engineering is the process of transforming raw data, such as molecular structures in different morphologies, into meaningful features (predictors). In our case, these features can include geometric aspects like the center of mass distance between two molecules, the distance vector between molecule centers, or the angles that describe the relative orientation of the molecule backbones, as seen in previous studies [39,42]. Recently, the Coulomb matrix has also been used as a feature (predictor) [40,42,86]. Feature engineering plays a critical role in enhancing model performance in traditional ML methods like kernel ridge regression,

support vector regression, or linear regression. However, deep learning models offer the advantage of automatic feature extraction, often yielding better results with less effort.

Therefore, in this paper, instead of focusing on improving features, we use deep learning to allow the algorithm to extract the most relevant features automatically. In this regard, we utilize the Coulomb matrix as a representation of the molecular electronic structure. The matrix form of the Coulomb matrix is compatible well with the strengths of deep learning architectures, as its grid-like structure allows for efficient processing and feature extraction.

The Coulomb matrix, C_{ij} , is defined as

$$C_{ij} = \begin{cases} 0.5Z_i^{2.4} & \text{if } i = j \\ \frac{Z_i Z_j}{\|R_i - R_j\|} & \text{if } i \neq j \end{cases} \quad (8)$$

where Z_i is the atomic number of atom i , and R_i is its position. The Coulomb matrix captures the pairwise interactions between atoms in a molecule, and by converting it into a 2D array, we can leverage the CNNs model that is optimized for processing such data structures.

For our system, we follow prior work [42] by focusing only on the interactive part of the Coulomb matrix. Specifically, only the intermolecular elements C_{ij} between atoms i of chain 1 and atom j of chain 2 are considered. A visual representation of the full Coulomb matrix and its interactive part is provided in Fig. 2(a). All intramolecular elements, where both atoms i and j belong to the same chain, are excluded from the matrix. This approach is based on the understanding that the transfer integral is fundamentally an intermolecular property; therefore, using the interaction part of the Coulomb matrix as a feature is expected to provide a more accurate prediction. As a result, the input matrix for each PEDOT dimer is reduced from dimensions 314×314 to 157×157 , representing the interactions between the two chains. This reduction not only simplifies the feature set but also leads to a significant decrease in computational costs, making the input data four times smaller. This efficiency gain has a huge impact on both the training time and scalability of the model.

Convolutional neural networks (CNN) architecture. To predict transfer integrals for any morphology containing PEDOT chains, we used CNN, a deep learning method well-suited for processing image and grid-like data. As said earlier, CNNs automatically extract features and identify patterns, reducing the need for manual feature engineering. This approach offers a clear advantage over traditional methods like kernel ridge regression and support vector regression, which have been explored in previous studies [39,40,42].

Our CNN model architecture, shown in Fig. 2(b), consists of several layers that work together to extract and learn features from the input data. These layers include five convolutional layers, each followed by Rectified Linear Unit (ReLU) activation functions and dropout layers. A convolutional layer is a key component of the CNN and serves as its main building block. It contains a set of filters each of size $M \times N$, which are small matrices of numbers. These filters are applied to the input data using a process known as the convolution operation. Essentially, a filter slides over the data grid and performs a convolution operation, combining the filter values with the underlying data values in the input. The

convolution operation can be described as [87]

$$O_k(i, j) = \sum_{m=1}^M \sum_{n=1}^N I(i+m, j+n) \times K_k(m, n), \quad (9)$$

where $O_k(i, j)$ is the value of the output feature map at position (i, j) generated by k -th filter, $I(i+m, j+n)$ represents the input data at position $(i+m, j+n)$, $K_k(m, n)$ is the k -th filter value at position (m, n) .

The result of this operation helps extract important features from the data. By performing this convolution operation across the entire data grid, the layer creates a feature map, highlighting the presence of specific features in different parts of the data grid. In our CNN model, each convolutional layer has a specific number of filters, which allows the model to learn different aspects of the data at each layer. These layers are structured with increasing numbers of filters, which are 32, 64, 128, 256, and 128, respectively. Using multiple filters at each layer allows the model to capture various features, from simple ones to more complex structures in the deeper layers.

The size of all the filters used in our model is 3×3 , which means each filter is a 3×3 grid of pixels. This filter size determines how much of the input data the filter will “look at” at once. The filter starts scanning data from the top-left corner of the input data and scans it from left to right. After completing each row, it then moves down to the next row and continues scanning from left to right. This allows the network to capture detailed features with high resolution. To determine how much the filter moves each time it scans over the data, we use a parameter called stride. The stride controls the step size of the filter as it moves across the input. In our model, the stride is set to 1, meaning the filter moves one pixel at a time. To ensure that the filters can fully process the input data, we apply padding. Padding is essential for preserving the spatial dimensions of the input as it moves through the convolutional layers. Without padding, information from the borders of the input might be lost. In our architecture, padding is set to 1, meaning we add one pixel around all sides of the input data. This allows the filters to process the entire input, including the borders, ensuring that important information is not discarded. After each convolutional layer, we use a ReLU activation function. This function helps the model learn nonlinear patterns by keeping positive values and setting negative values to zero, which is defined as follows:

$$\text{ReLU}(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}, \quad (10)$$

where x represents the input value.

To prevent overfitting, dropout layers are incorporated after each convolutional layer. These layers randomly set a fraction of input units to zero during each forward pass. The mathematical representation of dropout is as follows:

$$\text{Dropout}(x, p) = \begin{cases} 0 & \text{with probability } p \\ \frac{x}{1-p} & \text{otherwise} \end{cases}, \quad (11)$$

where x represents the input value, and p is the dropout probability.

Our dropout rates are 0.5, 0.3, 0.3, 0.2, and 0.1 for each corresponding convolutional layer.

An average pooling layer is then applied to reduce computational complexity and the number of parameters. Specifically, this layer calculates the average value of each vector along a specific axis, which helps create a fixed-length summary of the input data. The idea behind this layer is that it captures the overall, important information from the input while ignoring small, unnecessary details or noise, resulting in a compact and meaningful representation of the data.

After the convolutional and pooling operations, the feature maps are flattened to convert the 2D data into a 1D vector, by sequentially stacking all values row by row, making it suitable for input into the fully connected layer. This step is necessary because fully connected layers require a 1D input, rather than a 2D matrix, to process the learned features. The fully connected layer, consisting of 128 neurons, takes this flattened vector as input, where each neuron receives a weighted sum of all the features. These neurons integrate the features learned by the convolutional layers, allowing the model to make decisions based on all the features it has learned. In this study, the task is a regression problem, which is formulated using the standard linear regression equation [88]

$$y = W \times x + b, \quad (12)$$

where y is the output, W is the weight vector, x is the input vector (the flattened feature map), and b is the bias term. Finally, the output layer consists of a single neuron with a linear activation function, which outputs a continuous value representing the predicted transfer integral.

Training hyperparameters. The CNN model was implemented and trained using PyTorch [89], a popular deep learning framework. To help the model learn from the data and perform as well as possible, we used a set of parameters called “hyperparameters.” These are arbitrary values that we choose before starting the training process, and they control various aspects of how the model is trained. One of the key hyperparameters we set is the number of epochs, which refers to how many times the model sees the entire dataset during training. In our case, we chose 1000 epochs, meaning the model would process the data 1000 times, allowing sufficient training for the model to converge. To determine the number of samples to work through during each training step, we used the batch size hyperparameter. A batch size of 64 was chosen to strike a balance between computational efficiency and stable gradient estimates. The gradient estimate tells the model how model’s parameters should be adjusted to minimize errors.

The model’s performance during training was evaluated using mean MAE, which is described in Sec. IV A [see below, Eq. (13)], and also called as a loss function. The loss function serves as a guide for the model by telling it how far off its predictions are. By minimizing this error, the model was guided toward making more accurate predictions. We used the Adam optimizer [90] with a learning rate of 0.001 to update the model’s parameters during training. Adam optimizer is widely regarded for its ability to dynamically adapt the learning rate, often resulting in faster convergence and improved performance compared to other optimization algorithms.

To avoid unnecessary computation, we implemented early stopping with a patience value of 10 epochs. This meant that if the model’s performance failed to improve for 10 consecutive

epochs, the training process was stopped early. Finally, the training data was shuffled at the start of each epoch. This step ensured that the model did not learn unintended patterns in the data order, promoting better generalization when applied to unseen data.

4. Performance metrics

To assess the performance of our model, we use two widely used performance metrics: MAE and R^2 [91]. The MAE calculates the average absolute difference between the predicted and actual values. It is a simple measure that captures the magnitude of the prediction errors without considering their direction. Mathematically, it is defined as

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|, \quad (13)$$

where y_i is the actual value, $\hat{y}_i$ represents the predicted value, and n is the number of values.

A perfect model would have an MAE of 0, which represents no error in predictions. However, in practice, the MAE is generally greater than 0. To determine if an MAE value is acceptable, it should be considered relative to the scale of the target variable. The R^2 value, or coefficient of determination, measures how well the model accounts for the variance in the dependent variable. It can be calculated as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (14)$$

where $\bar{y}$ is the mean of the actual values.

An R^2 value of 1 shows perfect predictions, where the model accounts for all the variability in the data. Conversely, an R^2 value of 0 implies that the model fails to explain any of the variance in the dependent variable.

5. Implementation and reproducibility of ML model

To ensure full reproducibility of our machine learning experiments, we provide all relevant codes, trained model parameters, and supporting files in a publicly accessible GitHub repository [52]. The repository is organized into the following components: Data loader, Model's architecture, Main loop, Hyperparameters, Info loading, Model Training, and Post-training Analysis.

B. Experiment method

1. Materials

The oxidant solution Clevios CB54 was supplied from Heraeus. 3,4-ethylenedioxythiophene (EDOT), n-butanol, ethanol, isopropanol, DMSO, and Ethylene Glycol were supplied by Sigma Aldrich.

2. Sample synthesis

Glass substrates were cleaned with Acetone and isopropanol for 15 min respectively and then dried with pressurized air. The PEDOT samples were synthesized following a recipe reported elsewhere [20]. Clevios CB54 was

diluted with n-butanol to form a solution of 40% w/v. Then, 33 μ l of EDOT were mixed with 100 μ l of the oxidant solution for 2 s and the precursor solution was spincoated at 1500 rpm for 30 s on the aforementioned cleaned glass substrates. The coated substrates are then placed on a hot plate at 100 C for 15 min. The samples are then washed with twice in n-butanol baths in order to remove excess oxidant and edot from the film, and then dried with pressurized nitrogen. For the GIWAXS samples, the fabrication procedure was the same, but the substrate was silicon substrate with a native oxide.

3. Exposure

The samples from Sec. III B 2 are immersed in baths of a solvent of choice (DMSO, Ethylene Glycol, Ethanol, n-butanol, Isopropanol) for 15 min. Subsequently, the samples are removed from the bath and the excess solvent is removed with pressurized nitrogen and the samples are placed in a desiccator with dynamic vacuum overnight. In order to secure that the solvent has been removed, the samples are finally annealed at 100 C for 15 min.

4. Electrical characterization

The electrical conductivity was extracted as the inverse product of the sheet resistance with the thickness of the film. The sheet resistance was measured with a four-point probe setup connected to a Keithley 2400, while the thickness was measured with a Dektak profilometer [92].

5. X-ray photoelectron spectroscopy

X-ray photoelectron spectroscopy (XPS) experiments were carried out under the base pressure of 2×10^{-10} mbar with ESCA 200 electron analyzer and a monochromatic Al ka x-ray source ($h\nu = 1486.6$ eV). All spectra were collected at normal emission and room temperature. The spectrometer was well calibrated by a sputter-cleaned Au film with the Fermi level at 0 eV and Au 4f 7/2 peak at 84.0 eV with its full width at half maximum being 0.65 eV.

6. GIWAXS

The crystalline structure of the PEDOT:Tos films have been studied by means of Grazing Incidence Wide Angle X-ray Scattering (GIWAXS) experiments. GIWAXS measurements have been performed on the Dutch-Belgian Beamline (DUBBLE CRG), station BM26B, at the European Synchrotron Radiation Facility (ESRF), Grenoble, France [93,94]. The energy of the x-rays was 12.42 keV, the sample-to-detector distance was 12.07 cm and the angle of incidence, α_i , was set at 0.17° . The diffracted intensity was recorded by a Frelon CCD camera and was normalized by the incident photon flux and the acquisition time. All necessary corrections (flat field, polarization, solid angle, efficiency) have been applied to the 2D GIWAXS images. The scattering vector q was defined with respect to the center of the incident beam and has a magnitude of $q = (4\pi/\lambda) \sin(\theta)$, where 2θ is the scattering angle and λ is the wavelength of x-ray beam. Wedge correction has been applied on the images, with q_r ($q_r^2 = q_x^2 + q_y^2$) and q_z being

the in-plane and near out-of-the-plane scattering vectors, respectively.

ACKNOWLEDGMENTS

The ESRF and the NWO are acknowledged for allocating beam time at the Dutch-Belgian beamline (DUBBLE, ESRF, Grenoble) for the GIWAXS experiments. I.Z. and N.Z. acknowledge the support from the European Commission through the Marie Skłodowska-Curie projects HORATES (Grant No. GA-955837). I.Z. acknowledges support from Swedish Government Strategic Research Area in Materials Science on Advanced Functional Materials at Linköping University (Faculty Grant SFO-Mat-LiU No. 2009-00971), and Swedish Research Council (2024-04449). The computations were performed on resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS) at NSC. I.P. acknowledges the support from the funding from

Sweden's Innovation Agency, Vinnova, DEvelopment of TRANSPARENT CONductors for transparent photovoltaic cells. "DETRACON", Diary No.: 2024-00598.

DATA AVAILABILITY

The data that support the findings of this article are not publicly available upon publication because it is not technically feasible and/or the cost of preparing, depositing, and hosting the data would be prohibitive within the terms of this research project. The data are available from the authors upon reasonable request. All relevant machine learning codes, trained model parameters, and supplementary files are publicly available in a GitHub repository [52]. The repository is structured into the following components: data loading, model architecture, main training loop, hyperparameter configuration, information preprocessing, model training, and post-training analysis.

-
- [1] N. J. Pataki *et al.*, A rolled organic thermoelectric generator with high thermocouple density, *Adv. Funct. Mater.* **34**, 2400982 (2024).
 - [2] S. T. Keene, V. Gueskine, M. Berggren, G. G. Malliaras, K. Tybrandt, and I. Zozoulenko, Exploiting mixed conducting polymers in organic and bioelectronic devices, *Phys. Chem. Chem. Phys.* **24**, 19144 (2022).
 - [3] C. Cea, G. D. Spyropoulos, P. Jastrzebska-Perfect, J. J. Ferrero, J. N. Gelinas, and D. Khodagholi, Enhancement-mode ion-based transistor as a comprehensive interface and real-time processing unit for in vivo electrophysiology, *Nat. Mater.* **19**, 679 (2020).
 - [4] I. Zozoulenko *et al.*, Electronic, optical, morphological, transport, and electrochemical properties of PEDOT: A theoretical perspective, *Macromolecules* **54**, 5915 (2021).
 - [5] A. Elschner, S. Kirchmeyer, W. Lovenich, U. Merker, and K. Reuter, *PEDOT: Principles and Applications of an Intrinsically Conductive Polymer* (CRC Press, Boca Raton, 2010).
 - [6] K. Sun *et al.*, Review on application of PEDOTs and PEDOT:PSS in energy conversion and storage devices, *J. Mater. Sci. Mater. Electron.* **26**, 4438 (2015).
 - [7] I. Petsagkourakis *et al.*, A nuanced understanding of the doping of poly(3,4-ethylenedioxythiophene) with tosylate, *Discov. Mater.* **3**, 10 (2023).
 - [8] M. N. Gueye, A. Carella, J. Faure-Vincent, R. Demadrille, and J.-P. Simonato, Progress in understanding structure and transport properties of PEDOT-based materials: A critical review, *Prog. Mater. Sci.* **108**, 100616 (2020).
 - [9] S. Chen *et al.*, Unraveling vertical inhomogeneity in vapour phase polymerized PEDOT:Tos films, *J. Mater. Chem. A* **8**, 18726 (2020).
 - [10] G. E. Fenoy, C. von Bilderling, W. Knoll, O. Azzaroni, and W. A. Marmisollé, PEDOT:Tosylate-polyamine-based organic electrochemical transistors for high-performance bioelectronics, *Adv. Elect. Mater.* **7**, 2100059 (2021).
 - [11] K. Wijeratne, M. Vagin, R. Brooke, and X. Crispin, Poly(3,4-ethylenedioxythiophene)-tosylate (PEDOT-Tos) electrodes in thermogalvanic cells, *J. Mater. Chem. A* **5**, 19619 (2017).
 - [12] D. R. Evans, Understanding PEDOT doped with tosylate, *Chem. Commun.* **58**, 4553 (2022).
 - [13] S. D. Baranovskii, Theoretical description of charge transport in disordered organic semiconductors, *Phys. Status Solidi (b)* **251**, 487 (2014).
 - [14] J. Lee, Physical modeling of charge transport in conjugated polymer field-effect transistors, *J. Phys. D Appl. Phys.* **54**, 143002 (2021).
 - [15] N. Rolland, J. F. Franco-Gonzalez, and I. Zozoulenko, Can mobility negative temperature coefficient be reconciled with the hopping character of transport in conducting polymers? *ACS Appl. Polym. Mater.* **1**, 2833 (2019).
 - [16] S. Perrot, F. Pawula, S. Pechev, G. Hadziioannou, and G. Fleury, PEDOT:Tos electronic and thermoelectric properties: Lessons from two polymerization processes, *J. Mater. Chem. C* **9**, 7417 (2021).
 - [17] D. Kim and I. Zozoulenko, Why is pristine PEDOT oxidized to 33%? A density functional theory study of oxidative polymerization mechanism, *J. Phys. Chem. B* **123**, 5160 (2019).
 - [18] J. Rehmen *et al.*, Structural control of charge storage capacity to achieve 100% doping in vapor phase-polymerized PEDOT/Tosylate, *ACS Omega* **4**, 21818 (2019).
 - [19] I. Petsagkourakis, N. Kim, K. Tybrandt, I. Zozoulenko, and X. Crispin, Poly(3,4-ethylenedioxythiophene): Chemical synthesis, transport properties, and thermoelectric devices, *Adv. Elect. Mater.* **5**, 1800918 (2019).
 - [20] I. Petsagkourakis *et al.*, Structurally-driven enhancement of thermoelectric properties within poly(3,4-ethylenedioxythiophene) thin films, *Sci. Rep.* **6**, 30501 (2016).
 - [21] L. V. Lingstedt *et al.*, Effect of DMSO solvent treatments on the performance of PEDOT:PSS based organic electrochemical transistors, *Adv. Elect. Mater.* **5**, 1800804 (2019).

- [22] M. Modarresi and I. Zozoulenko, Why does solvent treatment increase the conductivity of PEDOT: PSS? Insight from molecular dynamics simulations, *Phys. Chem. Chem. Phys.* **24**, 22073 (2022).
- [23] R. Volpi, S. Kottravil, M. S. Nørby, S. Stafström, and M. Linares, Effect of polarization on the mobility of C₆₀: A kinetic Monte Carlo study, *J. Chem. Theory Comput.* **12**, 812 (2016).
- [24] Z. Shuai, H. Geng, W. Xu, Y. Liao, and J.-M. André, From charge transport parameters to charge mobility in organic semiconductors through multiscale simulation, *Chem. Soc. Rev.* **43**, 2662 (2014).
- [25] M. Jakobsson, M. Linares, and S. Stafström, Monte Carlo simulations of charge transport in organic systems with true off-diagonal disorder, *J. Chem. Phys.* **137**, 114901 (2012).
- [26] V. Rühle *et al.*, Microscopic simulations of charge transport in disordered organic semiconductors, *J. Chem. Theory Comput.* **7**, 3335 (2011).
- [27] L. Wang, Q. Li, Z. Shuai, L. Chen, and Q. Shi, Multiscale study of charge mobility of organic semiconductor with dynamic disorders, *Phys. Chem. Chem. Phys.* **12**, 3309 (2010).
- [28] A. Troisi, Prediction of the absolute charge mobility of molecular semiconductors: The case of rubrene, *Adv. Mater.* **19**, 2000 (2007).
- [29] J. Kirkpatrick, V. Marcon, J. Nelson, K. Kremer, and D. Andrienko, Charge mobility of discotic mesophases: A multiscale quantum and classical study, *Phys. Rev. Lett.* **98**, 227402 (2007).
- [30] M. Unge, H. Aspöcker, F. Nilsson, M. Pierre, and M. S. Hedenqvist, Coarse-grained model for prediction of hole mobility in polyethylene, *J. Chem. Theory Comput.* **19**, 7882 (2023).
- [31] E. D. Miller, M. L. Jones, and E. Jankowski, Tying together multiscale calculations for charge transport in P3HT: Structural descriptors, morphology, and tie-chains, *Polymers* **10**, 1358 (2018).
- [32] I. Zozoulenko, A. Singh, S. K. Singh, V. Gueskine, X. Crispin, and M. Berggren, Polarons, bipolarons, and absorption spectroscopy of PEDOT, *ACS Appl. Polym. Mater.* **1**, 83 (2019).
- [33] N. Zahabi and I. Zozoulenko, Band versus hopping transport in conducting polymers by *ab initio* molecular dynamics: Exploring the effect of electric field, trapping and temperature, *Adv. Elect. Mater.* **11**, 2400239 (2024).
- [34] N. Rolland, J. F. Franco-Gonzalez, R. Volpi, M. Linares, and I. V. Zozoulenko, Understanding morphology-mobility dependence in PEDOT:Tos, *Phys. Rev. Mater.* **2**, 045605 (2018).
- [35] S.-P. Peng and Y. Zhao, Convolutional neural networks for the design and analysis of non-fullerene acceptors, *J. Chem. Inf. Model.* **59**, 4993 (2019).
- [36] P. Friederich, S. León, J. D. Perea, L. M. Roch, and A. Aspuru-Guzik, The influence of sorbitol doping on aggregation and electronic properties of PEDOT:PSS: A theoretical study, *Mach. Learn.: Sci. Technol.* **2**, 01LT01 (2021).
- [37] G. Montavon *et al.*, Machine learning of molecular electronic properties in chemical compound space, *New J. Phys.* **15**, 095003 (2013).
- [38] J. Li *et al.*, Machine learning-assisted property prediction of solid-state electrolyte, *Adv. Energy Mater.* **14**, 2304480 (2024).
- [39] J. Lederer, W. Kaiser, A. Mattoni, and A. Gagliardi, Machine learning-based charge transport computation for pentacene, *Adv. Theory and Sims.* **2**, 1800136 (2019).
- [40] C.-I. Wang, M. K. E. Braza, G. C. Claudio, R. B. Nellas, and C.-P. Hsu, Machine learning for predicting electron transfer coupling, *J. Phys. Chem. A* **123**, 7792 (2019).
- [41] C.-I. Wang, I. Joanito, C.-F. Lan, and C.-P. Hsu, Artificial neural networks for predicting charge transfer coupling, *J. Chem. Phys.* **153**, 214113 (2020).
- [42] M. Rinderle, W. Kaiser, A. Mattoni, and A. Gagliardi, Machine-learned charge transfer integrals for multiscale simulations in organic thin films, *J. Phys. Chem. C* **124**, 17733 (2020).
- [43] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/5h7d-yvvd> for comparison of morphology preparation methods; Comparison of DFT-based and ML-based logarithmic values of transfer integral; Histogram of DFT-based transfer integrals (meV) for different solvents; Distribution of the train and test datasets in order to perform internal and external validations; Calculated hole mobility values for systems prepared in different solvents; XPS measurements (the S2p signal for the PEDOT:Tos thin films treated with various solvents).
- [44] K. E. Aasmundtveit, E. J. Samuelsen, L. A. A. Pettersson, O. Inganäs, T. Johansson, and R. Feidenhans'l, Structure of thin films of poly(3,4-ethylenedioxythiophene), *Synth. Met.* **101**, 561 (1999).
- [45] J. Rivnay, S. C. B. Mannsfeld, C. E. Miller, A. Salleo, and M. F. Toney, Quantitative determination of organic semiconductor microstructure from the molecular to device scale, *Chem. Rev.* **112**, 5488 (2012).
- [46] J. F. Franco-Gonzalez, N. Rolland, and I. V. Zozoulenko, Substrate-dependent morphology and its effect on electrical mobility of doped poly(3,4-ethylenedioxythiophene) (PEDOT) thin films, *ACS Appl. Mater. Interfaces* **10**, 29115 (2018).
- [47] K. A. Page, A. Kusoglu, C. M. Stafford, S. Kim, R. J. Kline, and A. Z. Weber, Confinement-driven increase in ionomer thin-film modulus, *Nano Lett.*, **14**, 2299 (2014).
- [48] A. B. Kaiser and V. Skákalová, Electronic conduction in polymers, carbon nanotubes and graphene, *Chem. Soc. Rev.* **40**, 3786 (2011).
- [49] H. Sirringhaus *et al.*, Two-dimensional charge transport in self-organized, high-mobility conjugated polymers, *Nature (London)* **401**, 685 (1999).
- [50] E. J. W. Crossland *et al.*, Anisotropic charge transport in spherulitic poly(3-hexylthiophene) films, *Adv. Mater.* **24**, 839 (2012).
- [51] D.-M. Smilgies, Scherrer grain-size analysis adapted to grazing-incidence scattering with area detectors, *J. Appl. Crystallogr.* **42**, 1030 (2009).
- [52] N. Zahabi and A. Beikmohammadi, ML4TransferIntegral, <https://github.com/AliBeikmohammadi/ML4TransferIntegral>.
- [53] N. Rolland, M. Modarresi, J. F. Franco-Gonzalez, and I. Zozoulenko, Large scale mobility calculations in PEDOT (poly(3,4-ethylenedioxythiophene)): Backmapping the coarse-grained MARTINI morphology, *Comput. Mater. Sci.* **179**, 109678 (2020).
- [54] J. F. Franco-Gonzalez and I. V. Zozoulenko, Molecular dynamics study of morphology of doped PEDOT: From solution to dry phase, *J. Phys. Chem. B* **121**, 4299 (2017).
- [55] D. Kim, J. F. Franco-Gonzalez, and I. Zozoulenko, How long are polymer chains in poly(3,4-ethylenedioxythiophene):Tosylate films? an insight from molecular

- dynamics simulations, *J. Phys. Chem. B* **125**, 10324 (2021).
- [56] S. Kirchmeyer and K. Reuter, Scientific importance, properties and growing applications of poly(3,4-ethylenedioxythiophene), *J. Mater. Chem.* **15**, 2077 (2005).
- [57] R. S. Mulliken, Electronic population analysis on LCAO–MO molecular wave functions. I, *J. Chem. Phys.* **23**, 1833 (1955).
- [58] M. J. Frisch *et al.*, Gaussian~16 Revision C.01. (Gaussian, Inc., Wallingford CT, 2016).
- [59] A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange, *J. Chem. Phys.* **98**, 5648 (1993).
- [60] V. A. Rassolov, M. A. Ratner, J. A. Pople, P. C. Redfern, and L. A. Curtiss, 6-31G* basis set for third-row atoms, *J. Comput. Chem.* **22**, 976 (2001).
- [61] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, PACKMOL: A package for building initial configurations for molecular dynamics simulations, *J. Comput. Chem.* **30**, 2157 (2009).
- [62] J. M. Martínez and L. Martínez, Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking, *J. Comput. Chem.* **24**, 819 (2003).
- [63] A. P. Thompson *et al.*, LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales, *Comput. Phys. Commun.* **271**, 108171 (2022).
- [64] A. I. Jewett *et al.*, Moltemplate: A tool for coarse-grained modeling of complex biological matter and soft condensed matter physics, *J. Mol. Biol.* **433**, 166841 (2021).
- [65] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Development and testing of a general amber force field, *J. Comput. Chem.* **25**, 1157 (2004).
- [66] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten, Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions, *Mol. Simul.* **6**, 121 (1991).
- [67] S. P. Coleman, D. E. Spearot, and L. Capolungo, Virtual diffraction analysis of Ni [0 1 0] symmetric tilt grain boundaries, *Modelling Simul. Mater. Sci. Eng.* **21**, 055020 (2013).
- [68] T. Takano, H. Masunaga, A. Fujiwara, H. Okuzaki, and T. Sasaki, PEDOT nanocrystal in highly conductive PEDOT:PSS polymer films, *Macromolecules* **45**, 3859 (2012).
- [69] C. M. Palumbiny, J. Schlipf, A. Hexemer, C. Wang, and P. Müller-Buschbaum, The morphological power of soap: How surfactants lower the sheet resistance of PEDOT:PSS by strong impact on inner film structure and molecular interface orientation, *Adv. Elect. Mater.* **2**, 1500377 (2016).
- [70] M. N. Gueye *et al.*, Structure and dopant engineering in PEDOT thin films: Practical tools for a dramatic conductivity enhancement, *Chem. Mater.* **28**, 3462 (2016).
- [71] N. Zahabi, G. Baryshnikov, M. Linares, and I. Zozoulenko, Charge carrier dynamics in conducting polymer PEDOT using *ab initio* molecular dynamics simulations, *J. Chem. Phys.* **159**, 154801 (2023).
- [72] I. Sahalianov, J. Hynynen, S. Barlow, S. R. Marder, C. Müller, and I. Zozoulenko, UV-to-IR absorption of molecularly p-doped polythiophenes with alkyl and oligoether side chains: Experiment and interpretation based on density functional theory, *J. Phys. Chem. B* **124**, 11280 (2020).
- [73] K. Ozawa, T. Okada, and H. Matsui, Statistical analysis of interatomic transfer integrals for exploring high-mobility organic semiconductors, *Sci. Technol. Adv. Mater.* **25**, 2354652 (2024).
- [74] K. N. Kudin, G. E. Scuseria, and E. Cancès, A black-box self-consistent field convergence algorithm: One step closer, *J. Chem. Phys.* **116**, 8255 (2002).
- [75] V. Coropceanu, J. Cornil, D. A. Da Silva Filho, Y. Olivier, R. Silbey, and J.-L. Brédas, Charge transport in organic semiconductors, *Chem. Rev.* **107**, 926 (2007).
- [76] N. Tessler, Y. Preezant, N. Rappaport, and Y. Roichman, Charge transport in disordered organic materials and its relevance to thin-film devices: A tutorial review, *Adv. Mater.* **21**, 2741 (2009).
- [77] A. Miller and E. Abrahams, Impurity conduction at low concentrations, *Phys. Rev.* **120**, 745 (1960).
- [78] R. Saxena *et al.*, Role of the reorganization energy for charge transport in disordered organic semiconductors, *Phys. Rev. B* **103**, 165202 (2021).
- [79] I. Jurić, I. Batistić, and E. Tutiš, Recombination at heterojunctions in disordered organic media: Modeling and numerical simulations, *Phys. Rev. B* **77**, 165304 (2008).
- [80] S. V. Novikov, D. H. Dunlap, V. M. Kenkre, P. E. Parris, and A. V. Vannikov, Essential role of correlations in governing charge transport in disordered organic materials, *Phys. Rev. Lett.* **81**, 4472 (1998).
- [81] Z. G. Yu, D. L. Smith, A. Saxena, R. L. Martin, and A. R. Bishop, Molecular geometry fluctuations and field-dependent mobility in conjugated polymers, *Phys. Rev. B* **63**, 085202 (2001).
- [82] M. Modarresi, J. F. Franco-Gonzalez, and I. Zozoulenko, Morphology and ion diffusion in PEDOT:Tos. A coarse grained molecular dynamics simulation, *Phys. Chem. Chem. Phys.* **20**, 17188 (2018).
- [83] N. Massonnet, A. Carella, A. De Geyer, J. Faure-Vincent, and J.-P. Simonato, Metallic behaviour of acid doped highly conductive polymers, *Chem. Sci.* **6**, 412 (2015).
- [84] K. Zhang *et al.*, Effect of host-mobility dependent carrier scattering on thermoelectric power factors of polymer composites, *Nano Energy* **19**, 128 (2016).
- [85] Q. Wei, M. Mukaida, Y. Naitoh, and T. Ishida, Morphological change and mobility enhancement in PEDOT:PSS by adding co-solvents, *Adv. Mater.* **25**, 2831 (2013).
- [86] O. Çaylak, A. Yaman, and B. Baumeier, Evolutionary approach to constructing a deep feedforward neural network for prediction of electronic coupling elements in molecular materials, *J. Chem. Theory Comput.* **15**, 1777 (2019).
- [87] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning, in *Adaptive Computation and Machine Learning*. (The MIT Press, Cambridge, Massachusetts, 2016).
- [88] N. R. Draper and H. Smith, Applied regression analysis, *Wiley Series in Probability and Statistics*, 3rd ed. (Wiley, New York, 1998).
- [89] A. Paszke *et al.*, PyTorch: An imperative style, high-performance deep learning library, *Adv. Neural Inf. Process. Syst.* **32**, 8024 (2019).
- [90] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [91] A. Colin Cameron and F. A. G. Windmeijer, An R-squared measure of goodness of fit for some common nonlinear regression models, *J. Econometrics* **77**, 329 (1997).

- [92] I. Petsagkourakis *et al.*, Improved performance of organic thermoelectric generators through interfacial energetics, *Adv. Sci.* **10**, 2206954 (2023).
- [93] G. Portale *et al.*, Polymer crystallization studies under processing-relevant conditions at the SAXS/WAXS DUB-BLE beamline at the ESRF, *J. Appl. Crystallogr.* **46**, 1681 (2013).
- [94] W. Bras *et al.*, Recent experiments on a small-angle/wide-angle X-ray scattering beam line at the ESRF, *J. Appl. Crystallogr.* **36**, 791 (2003).