Machine Learning & multiscale simulations: toward fast screening of organic semiconductor materials

Michael Rinderle and Alessio Gagliardi
Department of Electrical and Computer Engineering,
Technical University of Munich,
Karlstr. 45, 80333 Munich, Germany
Email: alessio.gagliardi@tum.de

Abstract—Organic semiconductor devices promise cost-efficient processability at low temperatures, but the usually amorphous materials suffer from low charge carrier mobility. The search for high mobility organic semiconductor materials has thrived data science and Machine Learning approaches to screen the vast amount of possible organic materials. We present a multiscale simulation model based on machine learned transfer integrals to compute the charge carrier mobility in organic thin films.

I. INTRODUCTION

Organic semiconductor materials are used in a variety of (opto-)electronic devices such as organic field effect transistors (OFETs) or organic solar cells (OPVs) [1], [2]. The device performance directly depends on the underlying semiconductor morphology [3]. Therefore, the study of structure—property relations in organic semiconductor materials is an important factor for the development of such devices. In particular, the relation between morphology, electronic structure, and charge transport is of interest [4], [5]. Modeling charge transport in organic semiconductor devices requires a multiscale simulation approach, where classical Molecular Dynamics (MD) is used to create molecular structures, kinetic Monte Carlo (kMC) methods are used to capture charge carrier trajectories, and Quantum Chemistry (QC) methods are used to compute the electronic structure of the organic molecules [6], [7]. Because of the high computational cost of calculating the electronic structure of amorphous organic materials, Machine Learning (ML) approaches have been proposed to efficiently predict QC properties [8], [9], [10].

One important step towards a successful ML model is to find a suitable representation of the molecular structure. In our previous work we developed a ML model based on Kernel Ridge Regression (KRR) to predict transfer integrals between small organic molecules and study the impact of different molecular representations on the models prediction accuracy [11], [12]. We showed that using a few geometric features tailored to the particular molecule under investigation are enough to reach good prediction accuracy. But with such a simple geometric representation, the ML model is not capable of distinguishing different molecules and the model therefore is not generalizable for a larger class of molecules. By using Coulomb Matrix based features, we were able to further improve the prediction accuracy and generalize the model for

multiple small molecules. In a different study, we utilized a graph neural network (GNN) based ML model to predict orbital energies of 62 000 molecules [13]. Using a graph, where the nodes correspond to atoms and the edges to bonds, is a natural choice to represent a molecule.

In this project, we extend our multiscale simulation framework for charge transport in organic materials [12]. In order to generalize the ML model for a larger class of molecules, we utilize a GNN based model.

II. MULTISCALE SIMULATION MODEL

In this section we summarize our multiscale simulation model. A schematic overview of the model is depicted in fig. 1.

The first step is to create realistic organic morphologies using classical Molecular Dynamics (MD). We deposit organic molecules on a frozen slab of crystalline SiO_2. While annealing the structure at $300\,K$ we take snapshots of the molecular structure. We extract neighboring molecule pairs from the morphology using the Voronoi tesselation. For these dimers, we compute the transfer integral using the dimer projection method (DIPRO) [14]. The needed molecular orbital expansion coefficients, overlap matrix, and orbital energies are computed with the Density Functional Theory (DFT) tool CP2K [15]. With this method we generate a training set of about 2000 data points per organic molecule.

In our previous work we utilized KRR models to predict transfer integrals between organic molecules [12]. One problem with this ML algorithm is the need for a fixed-size feature vector. A generalization for arbitrary sized molecules is possible by setting the feature vector size according to the largest molecule and using zero-padding for smaller molecules. The drawback of this method is a huge waste of computation time and storage. A graph can naturally represent molecules of any size and is therefore advantageous for a generalized ML model [13].

Graphs provide node and edge features. For the node features, we choose the atom number and for the edge features the Coulomb Matrix elements we previously used in our KRR models. A GNN is used to extract molecular features of arbitrarily sized molecules. The output layer of the GNN is used as an input to a Feed-Forward Neural Network that predicts the transfer integral. This framework is chosen to have
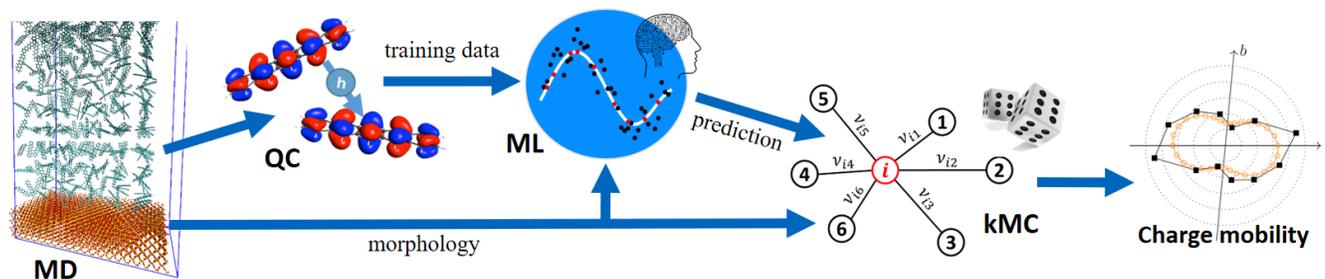
Fig. 1. Schematic of a multiscale simulation approach to compute charge transport in organic thin films. Molecular Dynamics (MD) is used to create realistic morphologies. Molecule pairs are extracted from the obtained structures and transfer integrals are computed using Quantum Chemistry (QC) calculations. The obtained data are used to train a Machine Learning (ML) model using a Graph Neural Network (GNN). The charge carrier mobility is obtained from kinetic Monte Carlo (kMC) simulations using the MD morphologies and predicted transfer integrals as its input values.

the possibility to add other molecule representations to the Neural Network as shown in [13].

To compute the charge carrier mobility in an organic thin film we use the kinetic Monte Carlo method. The morphology is taken from the MD simulation and the molecules are reduced to their center of mass position. The charge hopping rates between these positions is calculated using Marcus theory [16]. The transfer integrals needed as an input for the Marcus rate equation are predicted by the ML model. By tracking the charge carrier movement we obtain the charge carrier mobility.

III. Outlook

We will investigate the prediction accuracy of GNN based ML models for the prediction of transfer integrals depending on network size, number of molecules in the training data and distribution of the training data. As a base case, we will use the Coulomb Matrix elements as edge features and compare the GNN performance to our previous KRR study. In a further step, we will investigate different molecular representations and their impact on the prediction accuracy.

References

[1] H. Klauk, "Organic thin-film transistors," *Chem. Soc. Rev.*, vol. 39, no. 7, pp. 2643–2666, 2010.

[2] A. J. Heeger, "25th anniversary article: bulk heterojunction solar cells: understanding the mechanism of operation," *Adv. Mater.*, vol. 26, no. 1, pp. 10–28, 2014.

[3] J. Locklin, M. E. Roberts, S. C. Mannsfeld, and Z. Bao, "Optimizing the thin film morphology of organic field-effect transistors: The influence of molecular structure and vacuum deposition parameters on device performance," *J. Macromolecular Sci. C: Polymer Rev.*, vol. 46, no. 1, pp. 79–101, 2006.

[4] S. D. Baranovskii, "Theoretical description of charge transport in disordered organic semiconductors," *Phys. Status Solidi B*, vol. 251, no. 3, pp. 487–525, 2014.

[5] H. Oberhofer, K. Reuter, and J. Blumberger, "Charge transport in molecular materials: An assessment of computational methods," *Chem. Rev.*, vol. 117, no. 15, pp. 10319–10357, 2017.

[6] 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.*, vol. 98, no. 22, p. 227402, 2007.

[7] P. Kordt, S. Stodtmann, A. Badinski, M. Al Helwi, C. Lennartz, and D. Andrienko, "Parameter-free continuous drift-diffusion models of amorphous organic semiconductors," *Phys. Chem. Chem. Phys.*, vol. 17, no. 35, pp. 22778–22783, 2015.

[8] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller, "Assessment and validation of machine learning methods for predicting molecular atomization energies," *J. Chem. Theory Comput.*, vol. 9, no. 8, pp. 3404–3419, 2013.

[9] F. A. Faber, L. Hutchison, B. Huang, J. Gilmer, S. S. Schoenholz, G. E. Dahl, O. Vinyals, S. Kearnes, P. F. Riley, and O. A. von Lilienfeld, "Prediction errors of molecular machine learning models lower than hybrid dft error," *J. Chem. Theory Comput.*, vol. 13, no. 11, pp. 5255–5264, 2017.

[10] O. Caylak, A. Yaman, and B. Baumeier, "An evolutionary approach to constructing a deep feedforward neural network for prediction of electronic coupling elements in molecular materials," *J. Chem. Theory Comput.*, 2019.

[11] J. Lederer, W. Kaiser, A. Mattoni, and A. Gagliardi, "Machine learning–based charge transport computation for pentacene," *Adv. Theory Simul.*, vol. 2, no. 2, p. 1800136, 2019.

[12] 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*, vol. 124, no. 32, pp. 17733–17743, 2020.

[13] O. Rahaman and A. Gagliardi, "Deep learning total energies and orbital energies of large organic molecules using hybridization of molecular fingerprints," *J. Chem. Inf. Model.*, vol. 60, no. 12, pp. 5971–5983, 2020.

[14] B. Baumeier, J. Kirkpatrick, and D. Andrienko, "Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies," *Phys. Chem. Chem. Phys.*, vol. 12, no. 36, pp. 11103–11113, 2010.

[15] T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl, M. Watkins, J. VandeVondele, M. Krack, and J. Hutter, "Cp2k: An electronic structure and molecular dynamics software package - quickstep: Efficient and accurate electronic structure calculations," *The Journal of Chemical Physics*, vol. 152, no. 19, p. 194103, 2020.

[16] R. A. Marcus, "Electron transfer reactions in chemistry. theory and experiment," *Rev. Mod. Phys.*, vol. 65, no. 3, p. 599, 1993.