REPRODUCIBLE PREDICTION OF ORGANIC SEMICONDUCTOR PROPERTIES THROUGH OPEN SOURCE SOFTWARE DEVELOPMENT

by

James Rushing

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in Materials Science and Engineering

Boise State University

August 2022

BOISE STATE UNIVERSITY GRADUATE COLLEGE

DEFENSE COMMITTEE AND FINAL READING APPROVALS

of the thesis submitted by

James Rushing

Thesis Title: Reproducible Prediction Of Organic Semiconductor Properties Through Open Source Software Development

Date of Final Oral Examination: 09 June 2022

The following individuals read and discussed the thesis submitted by student James Rushing, and they evaluated the student's presentation and response to questions during the final oral examination. They found that the student passed the final oral examination.

Eric Jankowski Ph.D.	Chair, Supervisory Committee
Jens Harlander Ph.D.	Member, Supervisory Committee
Mahmood Mamivand Ph.D.	Member, Supervisory Committee

The final reading approval of the thesis was granted by Eric Jankowski Ph.D., Chair of the Supervisory Committee. The thesis was approved by the Graduate College.

ACKNOWLEDGMENT

Thank you to Dr. Jankowski and the CMElab for making this thesis and my master's education possible. They made time to support and encourage me amid a once-in-a-century pandemic, suffocating wildfires, and myriad sociatal and personal calalmities. For that, I am eternally grateful.

ABSTRACT

Semiconducting materials made from carbon-based molecules are potential replacements for inorganic semiconductors, but with lower costs of processing. Devices made from organic semiconductors can be produced at scale by inkjet printing and roll-to-roll manufacturing of these molecules in solution or melt phases. The efficiency of these organic devices is dependent on the structure of the active layer, so controlling the morphology of organic molecules through self-assembly during manufacturing is a key challenge to realizing their utility. Molecular self-assembly depends on the chemical structure of the molecules, how key moieties interact with each other and with any solvent present, and the thermodynamic paths that are sampled during processing. Computer simulations of molecular self-assembly can predict the structure and properties of candidate systems, and can improve the amount of information gained from more expensive trials performed in a wet lab when used to guide and explain experiments. Here we focus on the prediction of charge mobility in organic semiconducting materials, which requires a sequence of modeling calculations spanning many orders of magnitude across both time and space. We describe an open-source 'pipeline' of calculations that serves as a virtual laboratory for the screening of organic semiconductors for their charge transport properties. We describe work on `Planckton`, a software package for managing molecular simulations of organic semiconductors, and `MorphCT`, a package for managing kinetic Monte Carlo simulations,

the modularization and testing of which improves their transparency, usability, reproducibility, and extensibility. We measure improvements to `Planckton` and `MorphCT` by using them to study two organic molecules of interest in the photovoltaics field. In the first case study, of semiconducting polymer Poly-(3-hexylthiophene) (P3HT), we validate qualitative trends of charge mobility against prior work from both simulation and experiment. In the second case we predict the morphology and charge transport of the semiconducting macromolecule 3,9-bis(2-methylene-(3-(1,1-dicyanomethylene)-indanone))-5,5,11,11-tetrakis(4-hexylphenyl)-dithieno[2,3-d:2',3'-d']-s-indaceno[1,2-b:5,6-b']dithiophene (ITIC). We find that our work modularizing `Planckton` improves the pace at which simulations can be iteratively tested. We validate the electronic structure predictions made by `pySCF` against those previously made by the more restrictively-licensed `orca` package. We measure specific features of local structure that contribute to large-scale mobility trends in P3HT and describe predictions of charge transport in ITIC. In sum we improve the software ecosystem for reproducibly predicting charge mobility in organic semiconductors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

BHJ Bulk Heterojunction

FREA Fused-ring Electron Acceptor

HOMO Highest Occupied Molecular Orbital

ITIC 3,9-bis(2-methylene-(3-(1,1-dicyanomethylene)-indanone))-5,5,11,11-tetrakis(4-hexylphenyl)-dithieno[2,3-d:2',3'-d']-s-indaceno[1,2-b:5,6-b']dithiophene

kMC Kinetic Monte Carlo

LUMO Lowest Unoccupied Molecular Orbital

MD Molecular Dynamics

MSD Mean Squared Displacement

OLED Organic Light Emitting Diode

OPV Organic Photovoltaic

OSC Organic Solar Cell

P3HT Poly-(3-hexylthiophene)

QCC Quantum Chemical Calculation

TRUE Transparent, Reproducable, Usable, Extensible

CHAPTER 1:

INTRODUCTION

The unique properties of organic semiconductors make them ideal candidates for many electronic applications. They are used today in ultra high resolution Organic Light Emitting Diode (OLED) TVs and cellphone displays. They enable foldable OLED screens and rollable TVs [1]. Organic semiconductors are also integral to the design of next generation medical devices owing to their self-healing properties and their biodegradability [2]. For more on bioelectronic materials, see Organic Electronic: Emerging Concepts and Technologies [3].

In this thesis, we focus on a subclass of organic semiconductors, Organic Photovoltaic (OPV) materials, and their use in the active layer of Organic Solar Cell (OSC) devices. A simple model for OSCs is that they they are OLEDs working in reverse, absorbing photons and generating current rather than converting current into light. The OSC schematic in Figure 1.1 summarizes the main processes involved in the generation of electrical current from photonic energy. Any organic semiconductor can exhibit a photovoltaic effect when photons with appropriate energy (equivalently, wavelength) are absorbed. Therefore tuning an OPVs absorption spectra represents one opportunity to optimize its properties to a particular application.

OSCs can improve on the flexibility, processability, and cost of manufacturing of traditional inorganic solar cells. They are also used in more revolutionary electronic

designs. For example, researchers have exploited the relatively narrow absorption spectrum in OPVs ($\sim 300nm$) to make windows that act as OSCs. They achieved this by tuning the active layer material to absorb radiation right above or right below the visible spectrum (into the NIR or UV spectrum respectively). Semi-transparent OSCs have already reached 11 % efficiency [4]. Non-transparent OSCs are approaching 20% efficiency [5].

OPVs absorb photons very differently than inorganic semiconductors due to their low permittivity. The coulombic attraction, V, between an excited electron and the hole it left in a molecular orbital is given by Coulombs law as follows:

$$V = k_e \cdot \frac{e^2}{\epsilon_r} \tag{1.1}$$

where k_e is Coulomb's constant and e is charge of an electron. Relative permittivity, ϵ_r, is a unitless quantity that describes a materials polarizability relative to that of free space. That is, relative permittivity describes the readiness of a material to polarize in response to an electric field. A low relative permittivity of ~ 3 in OPVs (for reference: $\epsilon_r \sim 12$ for silicon [6] and $\epsilon_r \sim 78$ for water [7]) means that OPVs are only 3 times more polarizable than free space in response to the electric field created between the electron and the hole. And, because the material occupying the space between electron and hole is not willing (or able?) to fight back against the electric field created between them, they stick together and behave as a quasipartcle. This bound electron-hole quasiparticle is referred to herein as an exciton.

This excitonic absorption introduces a unique design challenge. That is, to extract a charge from the device, the exciton must first be coerced apart. This coercion can take place at the interface between donor and acceptor molecules, where the slight

offset in energy levels creates a charge transfer state wherein it is more energetically favorable for the donor to undergo electron transfer with an adjacent acceptor than it is to radiatively decay to its ground state and photoemit. This means that, after photoabsorbtion, the exciton must diffuse to this interface for the charge to be extracted. Because an exciton can only diffuse so far ($\sim 10nm$ [8]) before it relaxes to its ground state, it is critical that absorption take place close to a donor/acceptor interface. Producing extremely thin active layers could conceivably achieve this. However, producing a layer this small is untenable from a manufacturing standpoint. Furthermore, extremely thin active layers restrict the amount of radiation that a device interacts with.

In 1986, Ching W. Tang showed that, processed under the right conditions, a blend of donor and acceptor molecules can self-assemble into a Bulk Heterojunction (BHJ) microstructure [9]. The interlocking phases of donor accepter molecules, ensure that an exciton will intersect with the boundary between accepter and donor domains, while also ensuring that there is a continuous escape route, albeit labyrinthine, for the free charges to travel on their way to their respective electrodes.

As illustrated in 1.1, a BHJ active layer harvests photons through the following steps: (1) photoabsorbtions, (2) exciton diffusion, (3) charge transfer, and (4) free charge diffusion [10]. When a photon is incident on an OSC active layer, if it intersects with a molecular segment in the donor, whose Highest Occupied Molecular Orbital (HOMO) has comparable energy levels to the photon, the energy can be absorbed via the promotion of an electron to the next available energy level; the LUMO. This forms an exciton as described above, which can then diffuse until it intersects the boundary. The excited electron would like to relax back into is lower energy level,

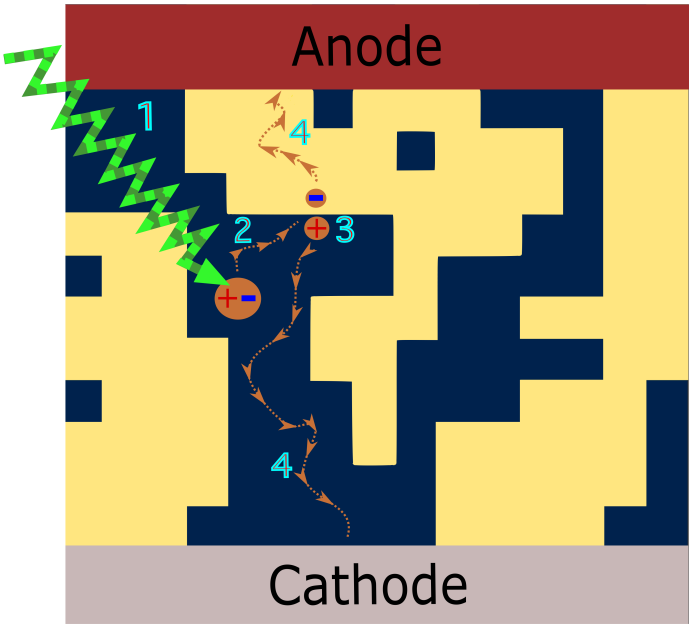


Figure 1.1: A cartoon representation of a BHJ device. All four stages involved in harvesting photonic energy in a BHJ device are represented. (1) The photon (green arrow) interacts with the material, exciting an electron and creating an quasiparticle referred to as an exciton. (2) The exciton diffuses about until it intersects the interface between donor and acceptor material domains. (3) The exciton is coerced apart by the energy offset between donor and acceptor molecules. (4) The, now unbound, hole and the electron are free to diffuse about until they reach their respective electrodes where they can be extracted.

but at the interface with the acceptor, it finds a better option. The acceptor's LUMO is engineered to sit between the energy of the donor's excited electron's energy level and the available lower energy level. Because of this, the electron cascades down to the acceptors LUMO through a charge transfer reaction. Finally the free charges can diffuse until they interact the electrode where it can be extracted. This is, of course, an ideal description, as there are loss mechanisms at all four stages.

The BHJ design pushed the efficiency beyond the 1% milestone. However, early BHJ devices utilized fullerene derivatives as acceptors which have a theoretical efficiency limit of $\sim$ 13% [11] and their spherical structure leads to difficult chemical purification, weak absorption in the visible-NIR spectrum, and rapid device degradation. In 2015, a group of researchers conceptualized the use of Fused-ring Electron Acceptor (FREA)s. FREAs consist of a fused-ring core and end groups that can be engineered to achieve specific electronic characteristics and side chains that can be engineered to achieve desired morphological features. The modularity of FREAs laid the stage for headspinning progress in the following years from 4% efficiency to 18% efficiency [12]. This designed modularity is a benefit to researchers, as they can alter one moiety and test the effect on resulting charge characteristics. For example, in a paper published in 2019 [13], researchers, knowing that flourine is an electronegative atom, swapped 4 hydrogen atoms for flourine atoms on the end groups the FREA molecule ITIC. Using *ab initio* DFT, they found that this substitution could lower the exciton binding energy (discussed above using Equation 1.1), thus facilitating exciton disassociation (stage 3 of BHJ). They also found that this flourination tends to reduce the reorganization energy, which we see in subsection 2.4.2, drastically increases charge mobility. This trend is borne out through our simulations in Figure 2.8.

Here we focus on (4), charge transport in BHJ active layers. We simulate the self-assembly and charge mobility of pure acceptor (or pure donor) domains post deposition. How the materials will self-assemble in the environment in which they are deposited determines the morphology which intern governs charge mobility [14][15]. We believe that building and refining a high-throughput simulation pipeline can aid researchers in screening these materials across a vast parameter space.

With that, we outline the data pipeline in Figure 1.2 through which we take atomistic representations like those shown in Figure 1.3 and, through a series of computations and simulations, arrive at data that is predictive of material properties. This pipeline is laid out in further detail by Jones et al. [16]. The pipeline begins with an atomistic description of a given molecule's atom types and bonding structure. On the basis of this description, forces between all atoms are defined. From the classical forces acting on particles in the system, equilibrium Molecular Dynamics (MD) simulations can predict the morphological structure. The structural data obtained from these simulations can then be fed as an input into kMC simulations that characterize charge mobility (conductivity) in these chemistries based on the energetics of electronically active molecular segments ('chromophores') within the morphology. We present `Planckton`, our package for navigation MD simulations of OPVs in subsection 2.4.1. The focus of this thesis, however, is the development of `MorphCT` for running kMC simulations, which we introduce more thoroughly in subsection 2.4.2.

We use the term "chromophore" liberally in this thesis. We take a brief aside to be more explicit about its meaning. The term chromophore arose in a biochemical context and is generally defined as a light-absorbing group or molecule [17]. In this context, a chromophore is so named because these molecules are associated with

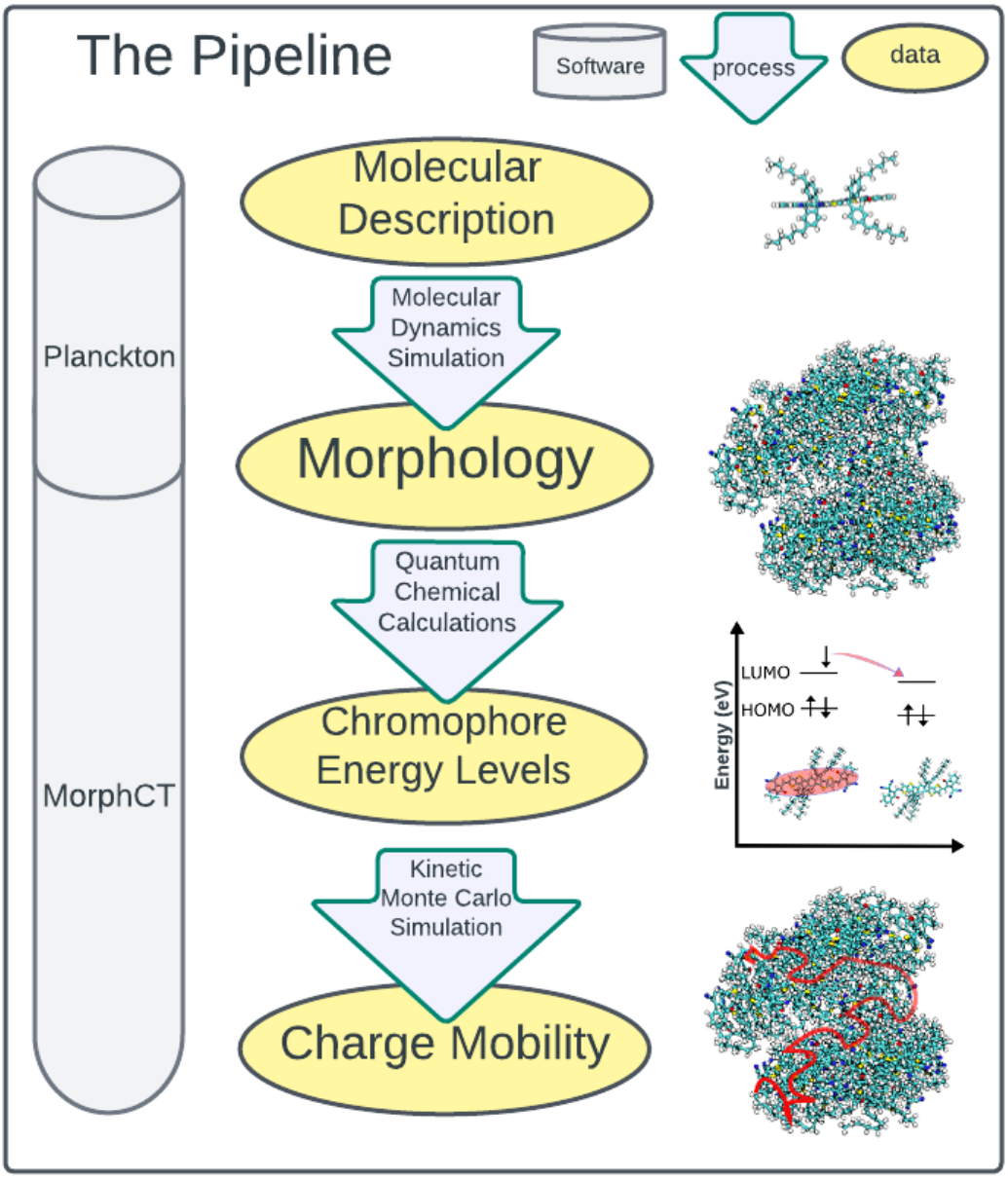


Figure 1.2: The pipeline for progressively graduating atomistic data to charge characteristics. On the left we see our packages developed for navigating the pipeline.

the color of a material. This is because, mechanistically, a photon collides with a chromophore, the absorbed energy excites an electron from the HOMO to the LUMO. When the chromophore relaxes to its ground state, it releases a photon with wavelength (or color) $\lambda = \frac{\hbar c}{E}$, where $\hbar$ is Plank's constant, c is the speed of light, and E represents the energy difference between the HOMO and LUMO. In plants, this amounts to a rejection of the overabundant amount of light in the green region which can damage the plant's DNA. We continue with the use of the term chromophore for the sake of brevity. In what is to follow, chromophore is taken to be defined as a molecular segment over which the wave function of a free electron (hole) is assumed to be localized. It is under this assumption that we execute a hopping model of charge transfer between chromophores based on the Marcus rates of electrochemical reactions. We will also use chromophore to refer to the object in computer memory that stores all the information that we know about a chromophore within morphology (e.g., atom locations, neighbors, energy levels).

Engineering the charge mobility of pure donor and acceptor domains is critical to overall device performance [18]. An understanding of why this is the case can be found from the inspection of Figure 1.1. If the electrons in stage 4 reach the anode much faster than the holes reach the cathode, a traffic jam can occur at the anode which can result in recombination with surrounding holes and ultimately a loss of the charge via annihilation. An imbalanced charge carrier mobility can also lead to space-charge build up in the low mobility material that can screen the built in field [19]. This phenomena has been shown by space charge limited current experiments [20].

To improve the efficacy of the pipeline we focus on developing simulation tools that are Transparent, Reproducable, Usable, Extensible (TRUE) ([21]). These virtues

provide a lens through which to evaluate the adherence to basic scientific principles and to the best practices of distributed software development simultaneously. As enumerated by Jankowski et al., some emerging best practices for scientific software development include the following: (1) address cognitive load, (2) use version control, (3) automate repetitive tasks, (4) develop open-source code, (5) write code in the highest language possible [22]. "Addressing cognitive load" amounts to an admission that a learner has a finite capacity for new ideas [23]. With that, if a new user is spending substantial time navigating aspects of the pipeline that are not clearly germane to the science, we have an opportunity to improve our software.

Here we improve the pipeline outlined in Figure 1.2. To do this, we: (1) integrate an open source quantum chemical package into `MorphCT`, (2) perform verification and sensitivity analysis on benchmark P3HT morphologies and (3) deploy the pipeline from start to finish on ITIC. Informing this work is a new developer's perspective on cognitive load. Two specific challenges to this work were (1) understanding which facets of quantum theory were being applied, and (2) navigating the application programming interface of the `MorphCT`. Tutorials, documentation, github issues, and searchable collaborative workspace forums (Slack) provided a foundation for using and learning `MorphCT`. We aim to expand this foundation and address challenges (1) and (2) through the publication of interactive Jupyter notebook tutorials.

We perform case studies using the charge transport prediction pipeline of P3HT, a donor molecule, and ITIC, an acceptor molecule. With respect to organic electronic devices, 'acceptors' are the organic analogue of n-type inorganic semiconductors and 'donors' the analogue of the p-type. We make this note because in chapter 2 we describe a model of a charge hopping from one molecule to another. And as we

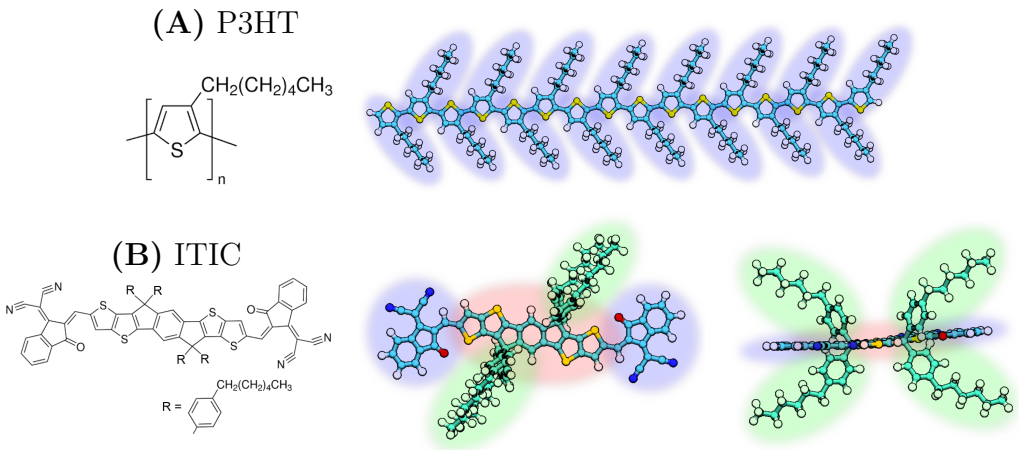


Figure 1.3: (A) P3HT monomer structure and an oligomer composed of 15 identical monomers imbued in blue. (B) ITIC structure and a view of the molecule viewed from above the plane of the backbone and in the plane of the backbone. ITIC molecule is imbued with blue, red, and green to call attention to the end groups, fused-ring core, and side chains respectively.

describe, our investigation involves hops within a pure donor domain or within a pure acceptor domain (not to and from donor and acceptor as would be the case at the interface). Any molecule can accept or donate a charge carrier. They merely receive a donor/acceptor designation as a result of how they are primarily utilized in electronic devices.

P3HT is a polymeric material that can self-assemble into a wide spectrum of crystallinities depending on how it is processed. Polythiophenes as a whole have been well investigated since 1883 when they were first characterized [24]. Seen in Figure 2.1, the molecular structure of P3HT is such that these molecules can pi-stack into lamellar structures that facilitate fast charge transport along the backbones. P3HT was first synthesized by Rick Mcullough in 1992. The first devices to utilize P3HT showed a low mobility due to a regio-randomness inhibiting the lamellar packing exhibited

by the now widely used regio-regular (head-to-tail) P3HT shown in Figure 1.3 [25]. ITIC is not polymeric. Rather, it a material consisting of small molecules with no long range bonding. ITIC is a FREA that was first synthesized in 2015 [26]. The structure of ITIC is shown in Figure 1.3 wherein we have highlighted the moieties.

In chapter 2, we provide an exposition of the theories and software packages used to model and simulate self-assembly and charge mobility in OPVs. The methods used in this thesis are centrally motivated by, and justified for, the application of our workflow to materials used in OSC design. However, the methods are not exclusively applicable to these materials and could be applied similarly to supplement the engineering of any organic electronic devices. In subsection 2.4.2, we verify that our current `MorphCT` workflow produces results that are consistent with previous implementations of `MorphCT`. We do this by recreating charge mobility predictions on three benchmark MD simulated morphologies of P3HT. Following that, again using P3HT, we test the sensitivity of our charge mobility prediction to various input parameters. We conclude subsection 2.4.2 by applying our pipeline, start to finish, to ITIC. In chapter 3, we outline the ramification of subsection 2.4.2 and detail the areas of future development of the workflow.

CHAPTER 2:

METHODS

Predicting morphology and charge transport properties in OPVs is accomplished through a combination of techniques, each modeling specific physical phenomena. In this section we outline the models and theory employed throughout the pipeline to describe the atomic structure of organic molecules, how they arrange, and how charges move through them. For each of these approaches, we then describe the open-source software tools that we create, modify, and use to implement these methods.

In section 2.1, we introduce the MD techniques used to predict self-assembly of OPVs. In section 2.2, we describe Marcus model of charge hopping between chromophores, the techniques we use to identify individual chromophores in simulated volumes, and how we use QCC to estimate the Marcus hopping rate between neighboring chromophores. In section 2.3, we describe the basics of stochastically modeling kinetic processes and our specific kinetic Monte Carlo approach to modeling charge transport in OPVs. Finally, in section 2.4, we enumerate the software used in this thesis and outline the means by which it is developed. Particular focus is given to two packages that are principally developed by members of the CMELab: `Planckton` and `MorphCT`.

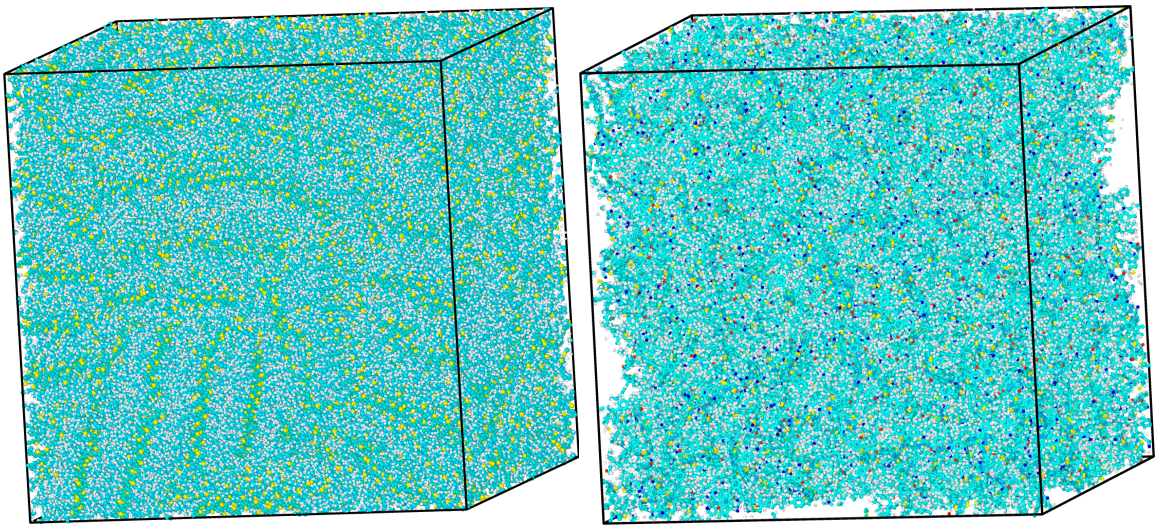


Figure 2.1: Left: 1000 oligomer atomistic morphology of P3HT. Right: 1000 molecule atomistic morphology of ITIC.

2.1 Molecular Dynamics

Modelling charge transport in OPVs demands a methodology that is accurate across orders of magnitude of resolution. Electronic wave functions operate at the atomic scale while the morphological features (grain boundaries, crystallinity etcetera) that govern charge transport occur across many nanometers. MD simulations are suited to this task because they enable the combination of coarse-grained models with atomistic representations to predict experimentally relevant length scales [27].

Molecular dynamics is a method of computer simulation for predicting the equilibrium geometries of molecular systems. MD simulations proceed iteratively by solving Newtons laws of motion in accordance with a predefined interatomic interaction potentials. The non-bonded interaction potentials are modeled with a classic Leonard-Jones(LJ) potential [28]. Parameters can also be defined for equilibrium bond lengths, angles between 3 bonded atoms, and angles between 4 bonded atoms (dihedral an-

gles). At each iteration, a numerical integration over these potentials provide an update to the velocities and positions of particles in the system.

Using the canonical ensemble (NVT) for example, conserving the number of particles, the volume, and the temperature allows for the exploration of the potential energy surface of the system and sample the microstates from equilibrated region the simulation for statistical analysis. Simulation ensembles are regulated via the Nosé-Hoover thermostat [29] to maintain temperature using the MTK equations [30][31]. The system can be considered equilibrated when the sum of all interatomic potentials no longer decreases with time. Determining the equilibrated region of the simulation can be fleshed out statistically from the progression of the systems potential energy. By binning the microstates into distinct regions, working backwards in time, a bin can be added to the equilibrated region if its standard deviation in system potential energy is less than twice that of the previous bin [32].

MD simulations can predict the self-assembly of OPV materials. To connect the chemistry to the conductivity of the material we use Marcus theory coupled with kMC.

2.2 Marcus Model

The movement of a free charge through a morphology can be modeled as a series of Marcus nonadiabatic electron transfer reactions, or 'hops' between relatively weakly interacting chromophores in the system. Each hop then is modeled as a thermally activated process, the rate of which can be solved for analytically from the intersection of two parabolic potential energy surfaces. Each parabola in Figure 2.2 represents the potential energy well of a charge that is localized on one of two chromophores in a dimer complex. With electron transfer much much faster than the movement

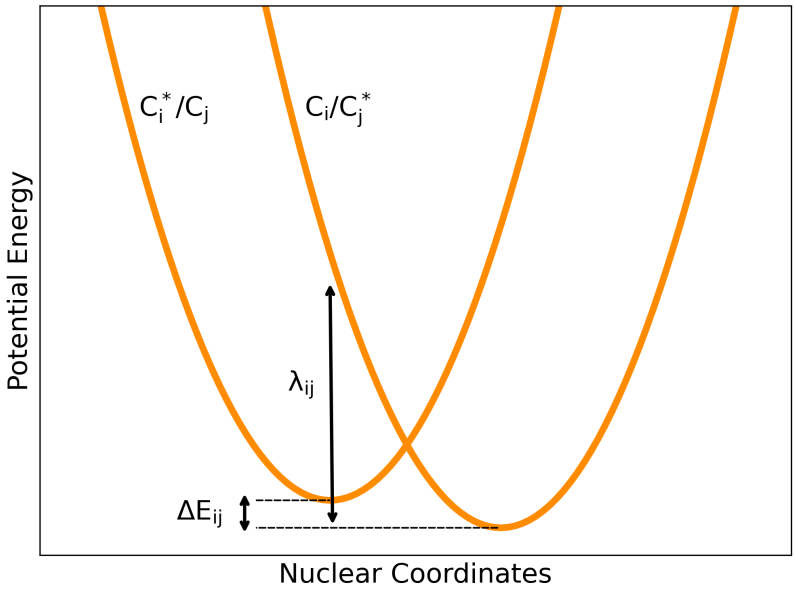


Figure 2.2: Two intersecting dimer potential energy surfaces annotated with λ_{ij}, the reorganization energy, ΔE_{ij}, the free energy difference between dimers. C_i^*/C_j and C_i/C_j^* represent the dimer with charge on chromophore i and on chromophore j respectively. C^* represents an excited chromophore, with the superscript representing an electron that has been promoted to the LUMO (in the case of acceptor transport).

of nuclear coordinates (Franck-Condon principle), the intersection of the parabolas corresponds to the unique nuclear geometry, and distinct vibrational mode, at which charge transfer is assumed to take place.

Within this framework, the rate at which a charge will hop from chromophore i to chromophore j, k_{ij}, is given by the following equation:

$$k_{ij} = |T_{ij}|^2 \frac{2\pi}{\hbar\sqrt{4\pi\lambda_{ij}k_BT}} \exp\left[-\frac{(\Delta E_{ij} - \lambda_{ij})^2}{4\pi\lambda_{ij}k_BT}\right] \tag{2.1}$$

with Boltzmann's constant, k_B, and Planck's constant, $\hbar$. The parameters T_{ij}, λ_{ij}, ΔE_{ij}, T represent the electronic overlap, the reorganization energy, the free energy

difference between chromophores, and temperature. These are discussed individually in the results section, wherein we test the sensitivity of the kMC results to the choice or computation of these parameters individually. The accuracy of the Marcus rate is thus dependent on the accuracy with which the inputs can be estimated. In our work, λ_{ij} and T are set as constants. In the following section we outline our quantum chemical treatment of both T_{ij} and ΔE_{ij} for all potential hops throughout the morphology.

Simulating hops through a morphology requires the identification of individual chromophores and the calculation of hop rates to neighboring chromophores. Each local molecular environment requires its own justification for where a chromophore may be considered to harbor a localized free charge. In disordered organic systems, unlike in metals or single crystal organic materials, electronic states are localized to the frontier molecular orbitals of tightly bound packets of atoms within the morphology. The spacial extent of a chromophore correspond roughly to the boundary between packets of strongly interacting molecular regions.

For ITIC, the candidate location of chromophores naturally arises from its composition of distinct macromolecules. The charges are taken to be localized on individual molecules of ITIC. The frontier molecular orbitals have negligible electron density along the side chains. Therefore, the simplest model for chromophores then is taken to be the backbone of individual molecules. Significant computational resources can be conserved by leaving the side chain atoms out of the of the QCCs. To test this on ITIC, we delineate the backbone and the whole molecule and compare carrier mobility in section 2.8.

In the case of conducting polymers like P3HT this boundary is affected by the how the chains are twisted or bent. A material that spans the gambit of disorder and crys-

tallinity can be difficult to model, because charge transport can occur in a hopping way, as described here, and also occur in an adiabatic metal-like transport. Therefore, while experimental studies have suggested that charges localize along roughly 7 monomer chains in P3HT, our application of our model to three morphologies of P3HT with vastly different structure mean that this may not be broadly applicable. To disambiguate the results in this thesis we chose to take the simplest model. Individual monomers in the system to be chromophores. The single monomer chromophore model has been shown to produce good results [16]. Deciding where a chromophore should be expected is one step in the workflow that requires nuanced scientific justifications. However, after that decision has been made, the significant procedural hurdle of computationally ascribing the atoms to their respective chromophores remains. We describe our approach in subsection 2.4.2.

With the morphology chopped into chromophores and stored in memory, the energetics between neighboring chromophores can be estimated with quantum chemical methods.

2.2.1 Quantum Chemical Methods

Calculating the rate at which a charge hops from one chromophore to the next using Marcus theory requires an understanding of the energy changes associated with the hop, which requires a calculation of chromophore's electronic orbital structure. Quantum chemistry allows for the estimation of the energy levels of electrons (holes) whose molecular orbitals are implied by the chromophore's current atomic configuration. Quantum chemical calculations comprise a set of methods, including *ab initio* calculations implemented in Density Functional Theory packages that work from first principles, and semi-empirical methods that use experimental data to make modeling

approximations [33].

We assume that the electrons occupying the frontier molecular orbitals are the sole participants in the hopping that is going on between chromophores. That is, if an electron hops from i to j, it will hop into the LUMO of j, and out of the HOMO of i. The driving force for a one electron charge transfer reaction, with a rate described by Equation 2.1, is the difference between the energy that our electron currently possesses on chromophore i, and the energy that it could enjoy over on chromophore j. This is writtin as follows:

$$\Delta E_{ij} = E_{homo,i} - E_{homo,j}. \tag{2.2}$$

Quantum chemically, the values $E_{homo,i}$ and $E_{homo,j}$ represent the eigenvalues of the time-independent Shrodinger equation corresponding to the HOMOs of chromophore i and j respectively. In our work, these values are approximated with the MINDO/3 method, a variation of the intermediate neglect of differential overlap (INDO) method. This method seeks recreate the *ab initio* Hartree-Fock results, where Hatree-Fock theory allows for an iteratively convergent numerical solution to the Shrodinger equation [34].

The value T_{ij} in Equation 2.1 is a measure of the electronic orbital overlap between chromophores. This value can be obtained using the the dimer splitting method [35]. This method compares the HOMO energies of chromophores i and j in isolation to the energies of the frontier molecular orbitals of a dimer consisting of the two chromophores. This difference is written as $(E_{homo,dimer} - E_{homo-1,dimer})$ where $E_{homo,dimer}$ and $E_{homo-1,dimer}$ are the two highest energy occupied molecular orbitals of the dimer. MINDO/3 is used again to approximate the eigenvalues of the frontier molecular or-

bitals, but this time of the dimer Hamiltonian.

The dimer method can be imagined as a dimer being pulled apart. At some distance the two highest occupied energy levels of the dimer will be the HOMOs of their respective chromophores. If the chromophores are not interacting, then the two highest energy molecular orbitals of the dimer will be the HOMO of chromophore i and the HOMO of chromophore j. If there is electronic overlap, a comparison between the two highest occupied molecular orbitals of the dimer and the HOMOs of the chromophores calculated in isolation can quantify the degree of electronic overlap. Indeed, T_{ij} is written as follows:

$$T_{ij} = \frac{1}{2}\sqrt{(E_{homo,dimer} - E_{homo-1,dimer})^2 - (\Delta E_{ij})^2}. \tag{2.3}$$

Solving Shrodinger's equation, with any level of accuracy, across an entire molecular arrangement is a prodigious computational lift. Other studies have implemented *ab initio* DFT methods at this stage of predicting mobility from molecular arrangement to good effect [36]. These more rigorous methods are untenable on the scale of the morphologies studied in this thesis. While INDO methods are less precise, the results of using this method have shown good agreement with experimental and *ab initio* DFT methods [37]. Computational quantum chemistry is a nascent and evolving field of its own, with quickly increasing efficiencies and accuracies. A particular choice of method comes down to how well we can organize a workflow and integrate the QCC portion of the workflow modularly to facilitate upgrading the QCC as more efficient methods and/or software emerges.

In our implementation, a quantum chemical calculation must be performed for every chromophore and every chromophore pair. To understand the scale of this lift,

the reader is reminded that 'n choose k' notation gives the number of ways to choose 'k' objects from a set of 'n' objects as follows

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{2.4}$$

Therefore, the upper bound of possible chromophore pairs (pairs setting $k = 2$) is given by the $\binom{n}{2}$ where n is the total number of chromophores in the simulated volume. With this, the effort of exhaustively calculating all chromophore pairs scales as $\frac{n^2-n}{2}$, or using big Oh notation, $O(n^2)$. This quadratic scaling of computational effort before performing kMC simulations can represent a bottleneck, so we investigate and apply approaches for identifying and calculating only the chromophore pairs that are spatially proximal.

In the following section we introduce our methods for determining which chromophore pairs to consider using Voronoi analysis.

2.2.2 Voronoi Analysis

To minimize the number of dimer calculations, we use Vornoi analysis to locate the spatially nearest neighboring chromophores. This analysis is performed on the Cartesian coordinates of the geometric center of the chromophore. A polyhedron cell is constructed around the geometric center. The polyhedron cell consists of every point in space that is closer to that chromophore center than any other chromophore center. Chromophores are considered neighbors if their Voronoi cells abut one another.

For simplicity, we construct and visualize a Voronoi diagram of the xy components of the chromophore centers of the crystalline P3HT system described in section 2.1. To carry out this analysis, `MorphCT` incorporates the Voronoi class provided by `freud`;

a python package for analysing and visualizing simulation data[38]. This class is compatible with 2D or 3D simulation data.

Shown in Figure 2.3, 15,000 thousand dots represent the chromophore's geometric centers projected in the xy-plane. In this 2D analogue, cell edges are drawn in a Euclidean way, with lines between polygons representing the set of points equidistant from that point and its geometrically closest neighbor across the line. In the 3D case, this analysis reduced the pairwise calculation from $\binom{15000}{2} = 112,492,500$ to $113,315$.

Euclidean space searching algorithms of this sort are an efficient way to parse space. They are known to scale with $O(n \log n)$ in the worst case and as low as $O(n)$ in the average case [39].

However, an artifact of constructing neighbor lists in this way is that some neighbors are too far apart to interact electronically, but are nevertheless closer to each other than they are to any other chromophores and are therefore counted as neighbors. Inspection of Figure 2.4 reveals how this phenomena can arise from this type of construction. Because charges will not hop between these pairs, including them in the pair list will result in superfluous QCCs.

In light of this, we introduce a parameter by which we further pare down the neighbor list. This parameter is a naive cutoff distance, referred to as 'd_{cut}' in this thesis. We visualize various d_{cut} values as black circles in Figure 2.4. It is clear from this image that the choice of d_{cut} could drastically effect the neighbor list and the resulting charge mobility calculation. Note that the z-direction has been collapsed, and the distances do not necessarily correlate to the distance between chromophores in the system.

A proper choice of d_{cut} will depend on the material under investigation, as the size

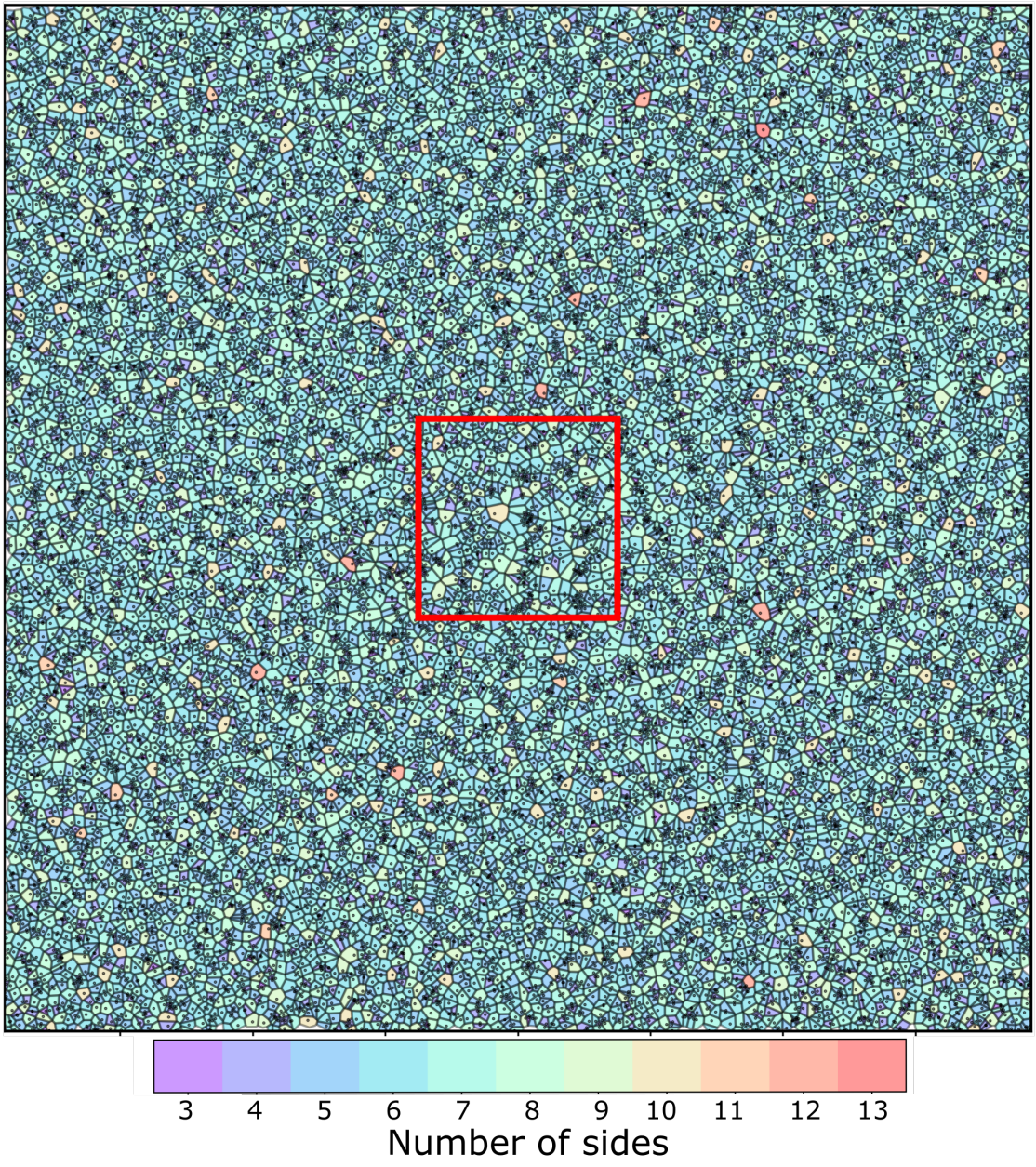


Figure 2.3: A 2D Voronoi diagram that was drawn from the xy components of a crystalline P3HT morphology. Dots represent chromopore centers. Lines represent points that are equidistant to the chromophore centers. Polygons represent all points that are closer to the chromophore center contained within than any other chromophore center. Even in 2d, the lamellar crystal structure is visible in the clustering of chromophores. RED SQUARE: Figure 2.4 shows the 15nm section of the sample zoomed in for detail.

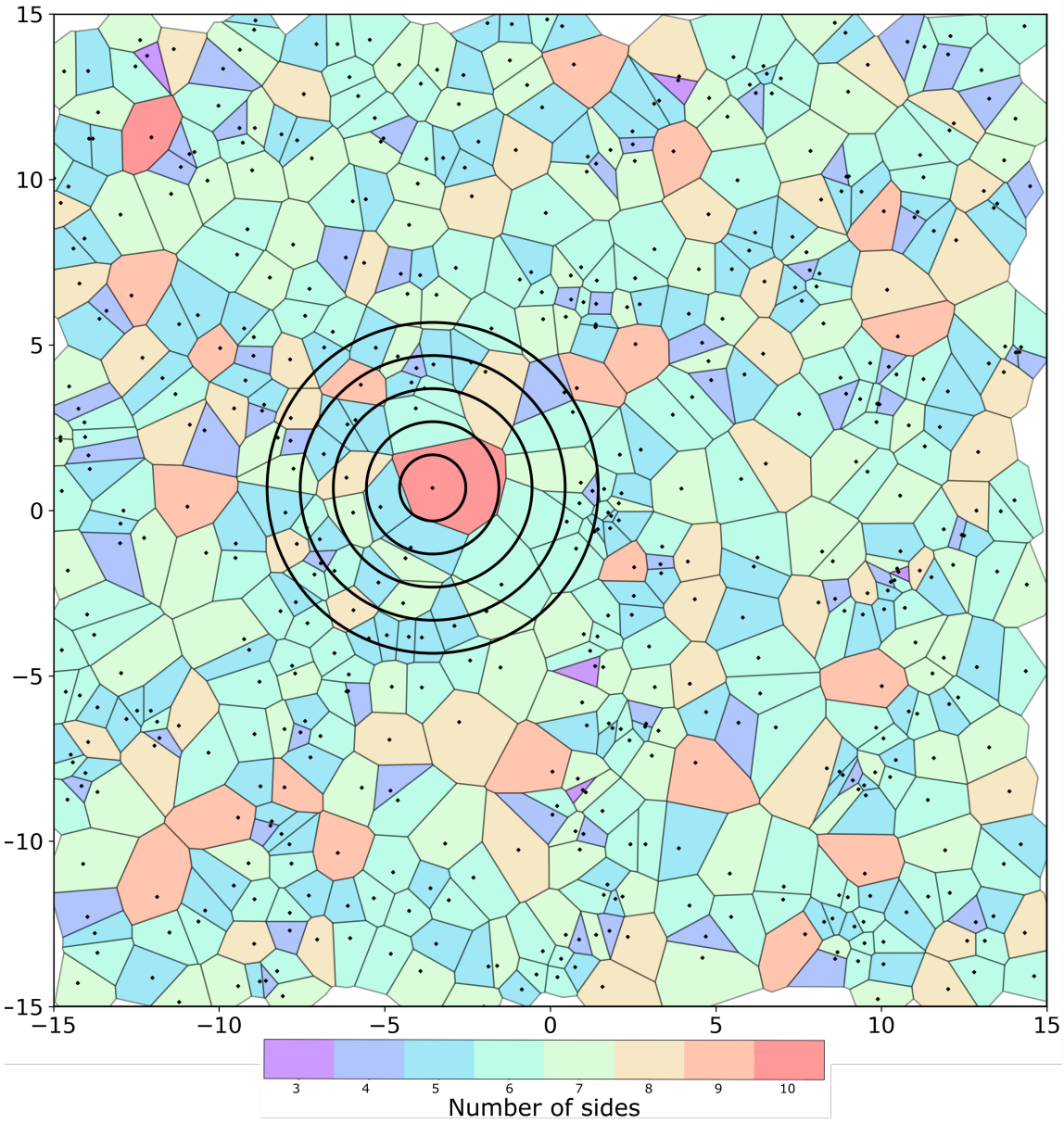


Figure 2.4: Zoomed section of Figure 2.3 wherein we see a cartoon version of a d_{cut} radius cutoff used to supplement Voronoi analysis. Circles represent the neighbor cutoff radius (d_{cut}) beyond which we truncate chromophores from the neighbor list. Cells are colored by number of neighbors. This chromophore has ten Voronoi neighbors. If we applied the d_{cut} radii implied in this figure, the chromophore would have 0,2,4,7 and 10 neighbors for the increasing d_{cut} radii.

of the individual chromophores will vary. In subsection 2.7.1, we test the sensitivity of our results to the value of d_{cut} for the crystalline P3HT data. From this testing, we consider if the computational juice is worth the radial squeeze.

2.3 Kinetic Monte Carlo

With the MD data generated, the data chopped into individual chromophores, the chromophore energetics quantified with QCC, and the Marcus hop rate calculated, a single charge's movement through the morphology can be simulated with the application of a kMC algorithm.

Monte Carlo algorithms use pseudorandom numbers to solve computational problems. Our implementation can be described as a first choice method kMC algorithm, where the kinetics involved is the rate of one electron charge transfer reactions and the first choice is that of the fastest available hop for a given charge.

Using `MorphCT`, a charge is implanted as quasi-particle into a random chromophore within the morphology. In this model, we assume that the only events that can take place in the system are hops between chromophores. With this, the rate of all possible events in the system are known and are given by Equation 2.1.

With the charge implanted, the hop rate, k_{ij}, to any neighboring chromophore is taken to be inversely proportional to the amount of time, τ, that the system will have to wait before that hop will take place. The τ of all available hops forms a queue of hops from shortest wait time (fastest hop), to longest wait time (slowest hop). From this queue, the shortest wait time (first choice) can be selected and the system can be moved forward in time by τ.

However, hopping processes at the angstrom level do not proceed deterministically. Our implementation, and others like it [40][41], have successfully captured the

stochasticity of these systems via a random shuffling of the hopping queue. The randomly shuffled wait time for every potential hop from occupied chromophore i onto a neighboring chromophore j is calculated as follows:

$$\tau = \frac{1}{k_{ij}} \cdot \ln \frac{1}{x} \tag{2.5}$$

where x is a random number between 0 and 1 and $\log(1/x)$ is a scaling factor. From the rationally shuffled queue, the shortest wait time is chosen and the charge is moved to its new chromophore host. The system is then considered to have moved forward in time by τ. This proceeds until the charge carrier stalls out or hops past a prespecified lifetime.

A core tenet of Monte Carlo sampling is ensuring a large sample size. With a large enough sampling, significant reshuffling of the queue will take place, allowing for a rare hop to jump the queue. For example, a charge carrier that hopped a million times, and had exactly 5 neighbors at each hop, would require forming a million wait time queues and 5 million individual wait time calculations. Shuffling each wait time by $\ln(1/x)$, we can expect to see $\sim$50 wait times that have been shortened by 5 orders of magnitude, $\sim$500 shortened by 4 orders of magnitude, and so on.

This is a lot to track in computer memory and this is carried out for hundreds or thousands of individual charge carriers. A benefit to Monte Carlo analysis of this type is that charge carriers can be simulated simultaneously. It is considered to be 'embarrassingly parallel' in that the subprocesses (charge carriers) require no communication. We discuss this from a software engineering point of view in subsection 2.4.2 How we analyze and aggregate data from thousands of single kMC simulations to ob-

tain macroscopic charge mobility is the subject of the following section.

2.3.1 KMC Analysis

Running a kMC simulation with `MorphCT` requires the choice of three parameters: the kMC temperature, the number of individual kMC simulations to perform, and the charge carrier lifetimes. Here we outline how we combine the data from the specified number of individual kMC runs.

The choice of carrier lifetimes effectively serve as checkpoints at which the displacement of charge carriers is recorded. For each specified lifetime, the prescribed number of individual kMC simulations is run as in section 2.3. When a given charge carrier hops past the specified lifetime, that is, the addition of the current iterations τ advances the simulation beyond the specified lifetime, its displacement from its starting location is stored in the carrier object. The Mean Squared Displacement (MSD) of a charge over the course of that lifetime is taken to be the average squared displacement across all the individual kMC runs. In other words, the MSD over a given time period is the standard deviation in position for a free charge walking randomly through this electronic environment.

It is known that the MSD of a diffusive particle increases linearly as time goes to infinity. The slope of the MSD (diffusion coefficient), D, can be estimated as a linear fit between the MSD values at the specified lifetimes. There is no objective best practice for determining the slope of the MSD as time goes to infinity from simulation data of this kind [42]. With that, we seek primarily to simplify the MSD analysis as much as possible. Doing so will make for more accurately reported results and easier apples to apples comparisons with future results. With that in mind, we choose only two lifetimes in this work.

Finally, the results of the MSD analysis can be related to zero-field mobility using the following Einstein-Smoluchowski relation:

$$\mu_0 = \frac{eD}{6k_BT}, \tag{2.6}$$

where e is the elemental charge of a charge carrier, D is the diffusion coefficient, k_B is Planck's constant and T is temperature.

The conductivity, σ, of a material is given by

$$\sigma = n \cdot e \cdot \mu, \tag{2.7}$$

where n is the number of charge carriers, e is the charge of an electron. μ is empirically defined as drift velocity, v, over the electric field, E as follows [43]:

$$\mu = \frac{v}{E} \tag{2.8}$$

With that, our kMC simulation most closely models a measure of conductivity in a bulk material in a controlled environment. That is, conditions wherein Equation 2.7 is measured with negligible n and E. This is the case for time-of-flight experiments carried out on very thin films under low charge density conditions [44].

2.4 Software Development

Implementing these methods requires the active development of scientific software. We develop two packages, `Planckton` and `MorphCT`, for performing and analyzing MD and kMC simulations.

We manage the development of `Planckton` and `MorphCT` in public repositories

hosted at github.com[45]. As with any open-source software project, these repositories serve as a central hub for developers to collaborate and integrate code into the code base. Repositories are version controlled with a snapshot of the code base saved at each iteration. This allows researchers to reproduce each others work with the exact version used to carry out the analysis. It also provides a documented and controlled way to merge together divergent code through pull requests. Unit testing and continuous integrating techniques provide passive protection against bugs in the software that might emerge from active development.

Modular, python based code allows for the curation and publication of Jupyter notebooks workflows and tutorials for performing reproducible analysis. Jupyter notebooks are a document format for publishing code that is executable and interactive [46].

Having had no prior experience with these materials and/or materials simulation prior to joining the CMElab, I was able to take an investigation of ITIC from molecular structure to a charge mobility; a macroscopic property. We hope that the combination of these two packages can make *in silico* screening of OPV materials realistically attainable by any aspiring researcher.

All the tools used to implement, analyze, and visualize this work are freely available. The packages and tools are enumerated in Table 2.1. We now describe `Planckton` and `MorphCT` in more detail.

2.4.1 Software for Molecular Dynamics Simulation

`Planckton` is a convenience package that integrates, `Mbuild`, `foyer`, and `HOOMD-Blue` for performing MD simulations of self-assembly in OPVs. `Planckton` provides clearly documented template scripts for inititializing and running MD simulations.

Table 2.1: Packages

Package/tool	Description
`foyer`	python package for applying atom-typing rules [47]
`freud`	python package for analyzing particle simulations [38]
`HOOMD-Blue`	general purpose toolkit for performing simulations. [48]
`mBuild`	python based molecule builder [47]
`MorphCT`	python package for simulating and analyzing charge transport from snapshots of MD simulations [16][45]
OVITO basic	tool for visualization simulation data [49]
`packmol`	python package for creating initial configurations of simulations [50]
`Planckton`	python based convenience package for running `HOOMD-Blue` simulations of OPVs [45]
`Planckton-flow`	python based package that supports the use of `Planckton` on high performance clusters[45]
`pySCF`	open-source collection of electronic structure modules [51]
`signac`	python based framework for managing large heterogeneous data spaces [52]
VMD	a molecular visualization program for displaying, analyzing, and animating large biomolecular systems [53]

`Planckton` uses `foyer` to interpret the forcefields used to generate the MD data. The MD simulations performed for this thesis employ the Generalized Amber Force Field (GAFF)[54]. The Amber forcefield was designed for use in modeling protein and nucleic acid systems. Serendipitously, the generalized Amber forcefield has parameters for organic molecules comprised of H,C,N,O, and P and can produces accurate simulations of organic molecules for use in OPVs. Also provided are files that define the atomic structure of many of the most commonly studied OPVs for OSC research in a format that are compatible with GAFF forcefields.

`Planckton` is built using `HOOMD-Blue` simulation toolkit [48]. The native file format of `HOOMD-Blue` is the `gsd` file. `gsd` files store simulation data in a binary format. As MD simulations proceed, the `gsd` files are populated with the microstate of the system at regular intervals. `MorphCT` is developed to operate on particular microstates stored within `gsd` files.

`Planckton-flow` is a dataspace manager that uses singularity [55] and `docker` [56] to 'contanerize' `Planckton` [45]. Containers are virtual machines that contain all the dependencies, configurations, code and data necessary to reproduce results [57]. `Docker` images are binary files that contain the entire software stack necessary to execute some code. This allows researchers to minimize dependency issues and increase reproducibility. However, `docker` has no native support for the use of GPUs and is not compatible with the more draconian permissions often present on HPCs. With that, `Planckton-flow` uses `singularity`, a software designed to overcome these shortcomings, to pull a `docker` image of `Planckton` to a container on the server.

`Planckton-flow` allows for a container of `Planckton` to be taken off the shelf and pulled to a high performance computing cluster without having to build the

software stack or write the simulation scripts from scratch. Screening OPV materials across many thermodynamic parameters can results in cumbersome data sets. `Planckton-flow` automates the management of large these large multidimensional data sets with `signac`.

2.4.2 Software for Kinetic Monte Carlo Simulation

`MorphCT` is a python package for running kMC simulations of charge transport in organics systems. `MorphCT` operates on `gsd` files containing the atom types and Cartesian coordinates of a system system. In this work we use `MorphCT` on equilibrium microstates obtained from MD simulations. However, `MorphCT` is indifferent to how the morphological data was obtained and does not measure physical accuracy of that data.

`MorphCT` is a collection of python scripts for defining, instantiating, reading and writing from the following three python objects: System, Chromophore, and Carrier. In general, the `MorphCT` workflow proceeds via interfacing with the System object. The System object can be instantiated from a `gsd` file.

Methods for adding chromophores, computing chromophore energies, and running kMC are coded into the System object. The method for adding chromophores creates a chromophore object based on a list of atom indices for every chromophore.

Every atom in the MD morphology has a unique index. All of the methods that follow hinge on assigning the prescribed atom indices to their respective chromophore. In this work, we manually index these chromophores. A tutorial for using VMDs graphical user interface to visually select chromophores is maintained for `MorphCT`. We also maintain workflows for automating the identification of chromophore positions using smarts matching.

The method for running kMC simulations `MorphCT` utilizes the python multiprocessing module to divide the prescribed number of charge carriers to be simulated across all available cores.

In the development of `MorphCT-flow`, for analogous reasons to those outlined for `Planckton`, we found that ORCA, the software used to perform QCC in `MorphCT` inhibited containerization because it required a licensing agreement. For this reason, ORCA was replaced with the fully open source `pySCF` (Python-based Simulations of Chemistry Framework) [51]. This framework was chosen in the interest of reproducability and extensibility. `pySCF` is implemented almost entirely in the Python language, which is becoming increasingly ubiquitous in the scientific computing community. The modularity of `pySCF` allowed for the entire QCC code to be implemented in five lines of code. These lines of code in asked asks `pySCF` to approximate the frontier energy levels given the chromophore's molecular arrangement.

With each step in the `MorphCT` corresponding to a method called on the system object, saving progress becomes trivially easy. With the python pickle module, complex objects can be saved as a binary file and subsequently brought back into computer memory. In its current form, pickling is critical to the `MorphCT` workflow. For example, a complete kMC simulation and analysis of a large P3HT of may take 5 days of computation, with a day to create all the chromophore objects, a day to perform the QCCs, and 3 days to run the kMC simulation. Pickling the system object after these steps allows us to return to these morphologies for further analysis without having to reconstruct the objects from scratch. Importantly, this means that the energetics of the chromophores need only be calculated once per morphology.

Workflow examples are maintained as jupyter notebook tutorials on github.

2.5 Summary

We present our results in three parts.

In section 2.6, we use `MorphCT` to test the performance of `pySCF` at the level of the chromophore. Two experiments are reported. In the first, we calculate the frontier molecular orbital energies for fused ring oligomers of increasing length. We compare the results of our experiment to the results of the same experiment done with more rigorous DFT methods. In the second, we test the performance of the `pySCF` dimer calculations outlined in chapter 2. We take two simple chromophores, two thiophene rings, in 5680 orientations and calculate the electronic coupling between them. We do so to test if our dimer calculation correlates sensibly to the angle and distance between chromophores. These experiments are broadly meant to confirm that we have integrated `pySCF` into `MorphCT` properly, and that the quantities produced comport with the physics of these systems. We then deploy `MorphCT` on three benchmark P3HT morphologies to obtain charge mobility and compare the results to previously reported values for these morphologies. This section provides verification of the current implementation of `MorphCT`. In section 2.7, we test the sensitivity of our `MorphCT` calculated mobilities to d_{cut}, chromophore reorganization energy, kMC temperature, and choice of charge carrier lifetimes for MSD analysis. Using a benchmark P3HT morphology, holding all other parameters constant and sweeping across relevant scales provides context for how to treat these parameters in future investigations.

In section 2.8, we extend our simulation pipeline to ITIC.

2.6 Implementation Verification

2.6.1 Quantum Chemical Calculation Verification

As outlined in subsection 2.2.1, QCC is used in two distinct ways in our workflow. First it is used to estimate free energy difference between individual chromophores as the difference in their HOMO (or LUMO) energy levels. Secondly, it is used to estimate the electronic overlap (T_{ij}) with the dimer splitting method which involves calculating the HOMO (or LUMO) of the dimer formed by the two chromophores. We present the results from two experiments meant to evaluate `pySCF`'s suitability for performing these duties: (1) we compare frontier molecular energies of single chromophores given by `pySCF` to those given by more rigorous *ab initio* DFT and (2) we evaluate the performance `pySCF`'s dimer calculation.

Experiment 1 Methods

At the level of a single chromophore, we calculate the HOMO-LUMO gap for fused-ring oligomers of increasing length. The difference between the HOMO and LUMO energy levels, the HOMO-LUMO gap, is an approximation of the amount of energy necessary to promote an electron to a higher energy level. Fused-ring geometries are of particular interest for accepter molecules, as discussed for FREAs in the introduction. The fused thiophenes in this experiment represent a generic FREA core, whose frontier molecular orbitals are the landing cites for a charge propagating through an acceptor material.

To recreate these experiments, using `mBuild` [58], oligomers composed of 4-8 planar fused thiophene rings were initialized and saved to a `gsd` file. The `gsd` files

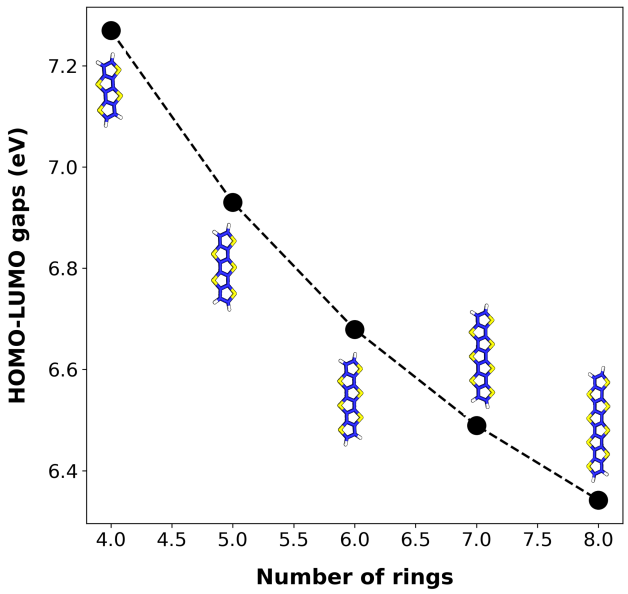


Figure 2.5: HOMO-LUMO gaps obtained using `MorphCT` for fused-ring molecules of oligmer length $4-8$.

were fed into `MorphCT` which uses `pySCF` to quantify the frontier orbital energy levels subsection 2.2.1.

Experiment 1 Results

Our values for the HOMO-LUMO gap are plotted by oligomer length in Figure 2.5. Our HOMO-LUMO gap ranges between $7.27eV$ and $6.34eV$. Consistent with our findings using a larger data set in section 2.8, the wall time for these individual calculations take place on the order of seconds.

It is known that there is a near linear relationship between HOMO-LUMO gap and oligomer length. We find that our use of `pySCF` (MINDO/3) replicates this trend and this is clear from Figure 2.5.

Our absolute values are in the range expected for INDO methods, which are known to overestimate DFT values by a factor of 2-3 [59]. Our HOMO-LUMO gap values compare well to those found using *ab initio* DFT (between $5.26eV$ and $4.92eV$)[60].

Experiment 2 Methods

At the level of the chromophore pair, we explore the effect of angle and distance on the electronic orbital overlap, T_{ij}, between two non-bonded thiophene rings. Two thiophene rings are positioned in electronic proximity using `mBuild` and saved to `gsd`.

A reference thiophene was placed at the origin in the xy-plane with the y-axis running through the sulfur atom as shown in Figure 2.6(a). For the second thiophene, two sets of 12 rotations about the x-axis and z-axis (with $-\frac{\pi}{2} < \theta < \frac{\pi}{2}$) and two sets of 12 translations between $0nm$ and $0.5nm$ along the x-axis and z-axis were defined. The Cartesian product of these sets define a space of $12^4 = 20736$ orientations for second thiophene.

Orientations resulting in atoms that were less than $0.3nm$ were removed from the data set as distances shorter than this are considered unphysical. The remaining 5680 atomic arrangements were saved to a `gsd` and the T_{ij} was quantified for each with `MorphCT`.

Experiment 2 Results

The data resulting from this experiment, 5680 orientations of electronically interacting non-bonded thiophenes and the corresponding T_{ij} between them, provide evidence that `MorphCT` is rationally capturing the orbital overlap between chromophores. The T_{ij} resulting from these 5680 orientations are shown in Figure 2.6. The figure shows

that, as expected, a decrease in center-to-center distance results in more orbital overlap and thus an increase in T_{ij}. Also observable in the figure is that a negative rotation about the x-axis orients the sulfur downward, resulting in a smaller sulfur-to-sulfur distance and a greater T_{ij}.

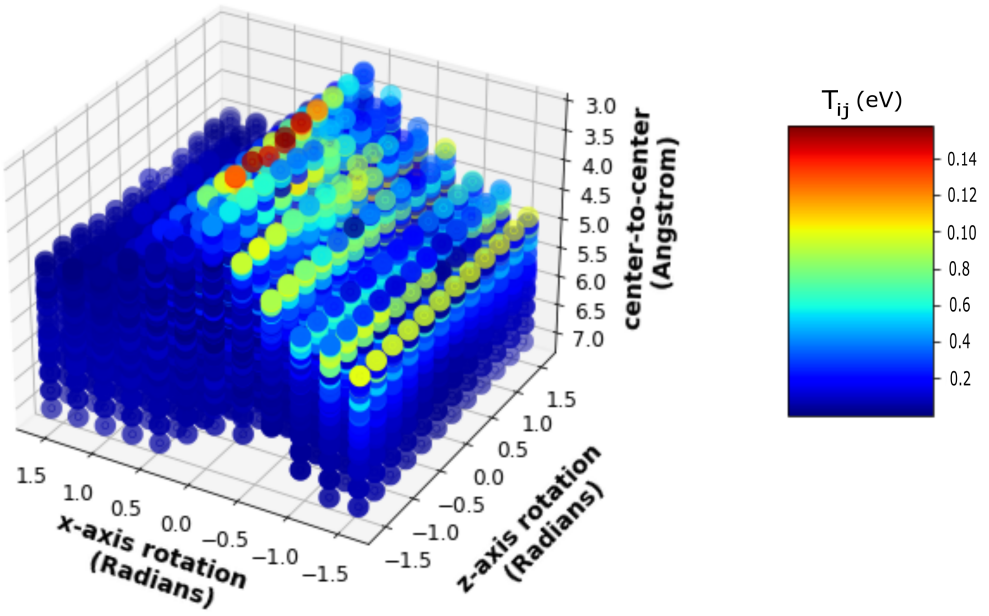
For the orientations found to be most and least optimal we calculate (shown in Figure 2.6), a T_{ij} of $.00005eV$ and $.15eV$ respectively. In the context of Equation 2.1 and Equation 2.5, the corresponding wait time for a charge to hop between to thiophenes in these orientations is $\tau = 2 \cdot 10^{-14}s$ and $\tau = 2 \cdot 10^{-7}s$. With latter case being two orders of magnitude slower than the lifetimes allowed in our kMC simulations, a hop would be highly unfavored for this orientation.

Creating the directory of `gsd`s and quantifying the T_{ij} between them took a wall time of $3.2h$ which averages to $\sim 2s$ per orientation.

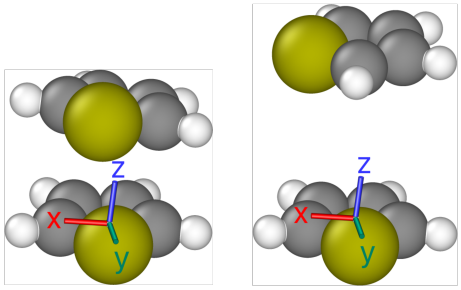
As evidenced by Figure 2.6((a)), for orientations resulting in a center-to-center distance of less than $0.5nm$ we calculate an average T_{ij} of $\sim 0.28eV$. These calculations match closely those calculated using more rigorous *ab initio* DFT methods[61], where realistic distances between thiophene rings in lamellar P3HT crystals between $0.38nm$ and $0.4nm$ gives T_{ij} values between $0.07eV$ and $0.1eV$.

In a similar work using a random forest machine learning to predict T_{ij} between thiophenes based on 9 input features, the authors found that the features of most importance for predicting T_{ij} was bonded vs non bonded, center-to-center distance and rotation about the y-axis [62].

Graduating these pairwise energetics to charge characteristics on the scale of MD simulations requires the use of an iterative algorithm. For this we employ kMC simulations using `MorphCT`.



(a) Individual dots correspond to 5680 distinct combinations of rotations and translations of the upper thiophene. Dots are colored based on the T_{ij} between thiophenes for the given x-axis rotation, z-axis rotation, and center to center distance from the reference thiophene.



(b) LEFT: The orientation (x-axis rotation $\sim$0.14, z-axis rotation ~ -0.14, center-to-center $\sim$3.2) with the highest T_{ij}. RIGHT: The orientation (x-axis rotation ~ -0.14, z-axis rotation ~ -1.5, center-to-center $\sim$5) with the lowest T_{ij}.

Figure 2.6

2.6.2 Charge Transport Calculation Verification

Having explored the performance of `pySCF` on the level of the molecule, we graduate these computations to the macroscopic scale with `MorphCT` to obtain charge mobility. In a previous work [27], researchers used `MorphCT` to predict charge mobility in P3HT. With `pySCF` newly integrated into `MorphCT` for reasons outlined in subsection 2.4.2, we verify our charge mobility against this work. The benchmark morphologies, which are the final frame of benchmark MD simulations, used to carry out this verification are public [63].

These benchmark simulations were carried out using coarse-graining techniques wherein molecular segments have been unified and treated as individual beads to lower computational cost and allow for larger and longer simulations. Using the Optimized Potentials for Liquid Simulations United Atom (OPLS-UA) force field, the researchers ran united-atom simulations, which do not explicitly keep track of the hydrogens in the simulation, but nevertheless accurately predict equilibrium geometries. This allowed the researchers to access length scales at which individual grains can form within the morphology. The orientation and boundaries of these grains effect charge transport and therefore provide a critical benchmark to compare against.

We first fined-grained MD data (hydrogens appended back into the morphology) and converted from `xml` to `gsd` format. At current, `MorphCT` is compatible with all-atom data and `gsd` format. As visualized in Figure 1.3, chromophores are taken to be individual monomers. Using `MorphCT`, for each chromophore, a unique object is instantiated and the energetics are obtained using `pySCF`. A kMC simulation is performed on the basis of these energetics with kMC temperature set to 300K. For each morphology, 10,000 individual charge carriers are injected (one at a time) into

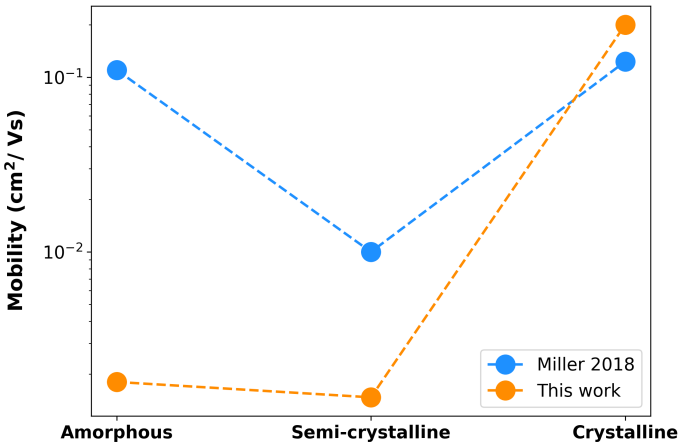


Figure 2.7: The results of mobility prediction on benchmark P3HT mobilities. We report the data in this way to show that the we found the same trend as previously reported using `ORCA` for QCCs.

the morphology, with 5,000 restricted to a lifetime of a tenth of a nanosecond and 5,000 restricted to a lifetime of a nanosecond. As outlined in subsection 2.3.1, from the difference in MSD between these sets of charge carriers, the zero-field charge mobility for the morphology is obtained.

It is known about P3HT that it can have vastly different crystallinities based on how it has been processed. We predict charge mobility in three morphologies with crystallinities that vary from amorphous to highly crystalline. These morphologies have been coerced into various levels of crystallinity through a simulated annealing process. We refer to these morphologies as amorphous, semi-crystalline, and crystalline.

Results from our work compare well to a predecessor of our work, which implemented ZINDO/s, another semi-empirical QCC method [64][16]. Their work utilized an earlier version of `MorphCT` which used the QCC software `ORCA` [65] to provide the quantum chemical approximations. `ORCA`'s proprietary licensing was prohibitive in

the efforts to containerize `MorphCT` for use on cluster and for creating reproducible results. This motivated the restructuring of `MorphCT` for use with `pySCF`.

The results in 2.7 show the charge mobility reported in the prior work, which used `ORCA` for the QCC calculation, and the charge mobility obtained using using the current workflow, which uses `pySCF`. The previous work found that charge mobility is highest for the crystalline morphology, followed by the amorphous, and finally the semicrystalline. This seemingly anomalous behavior can be explained. While the semicrystalline morphology has more ordered high speed highways of transport, the anisotropic movement inhibits average displacement. Our work replicates this trend.

In the previous work on these P3HT morphologies, 7 lifetimes were chosen and a linear regression was performed to estimate the slope of the MSD. The current work found that the mobility can be appropriately reproduced with an appropriate choice of 2 lifetimes beyond the ballistic transport timescale. It was found that with the squared displacement averaged over 5000 holes at $10^{-9}s$ and again $10^{-10}s$ the slope of the MSD can be reproduced with 1000's less individual holes having to be simulated.

2.7 Charge Transport Sensitivity Analysis

The sensitivity of our kMC simulation analysis to various parameters was performed on the benchmark crystalline P3HT morphology. As the overarching goal is to connect the morphological features to charge mobility, it is critical to be explicit about how each parameter can affect the resulting value of charge mobility.

By pressure testing the algorithm with a range of values for our input parameter, we explore the robustness of the algorithm to individual inputs. Sensitivity analysis can also help us understand where to invest resources into dialing the accuracy of any given parameter. It also gives motivation to be meticulous about keeping these

Table 2.2: d_{cut} Sensitivity

d_{cut} Å	4	6	8	10	12	14	16
pairs	318	22000	49000	96000	113026	113315	113315
μ_0 (cm^2/Vs)	$-2.17 \cdot 10^{-6}$	$6.13 \cdot 10^{-4}$	.01	.17	.17	.22	.22

parameters constant across analysis of the same material under different processing conditions. For example, if a higher charge mobility is discovered for a given morphology after some simulated annealing, it should be confirmed that it is not because the researcher used a lower reorganization energy for example.

We perform sensitivity analysis for 4 parameters: (1) neighbor cutoff distance (d_{cut}), (2) chromophore reorganization energy, (3) kMC temperature, and (4) choice of carrier lifetimes.

2.7.1 Neighbor Cutoff (d_{cut})

Voronoi analysis allows for the computationally efficient partitioning of space into polyhedra chromophore cells. As shown in 2.4, chromophore i is the neighbor of chromophore j if the voronoi cells constructed around their geometric center share a boundary with one another.

With each chromophore pair requiring a relatively costly QCC, after narrowing down the chromophore pairs with voronoi, it could be computationally preferable to calculate the distances between all pairs and remove neighbors more than d_{cut} apart.

Table 2.2 shows the effect of cutoff distance on value of calculated mobility. We can see in table 2.2, that with $d_{cut} = 10$ we get comparable mobilities to the $d_{cut} = 12$ simulation with 10^5 less pairs, which could suggest that beyond $d_{cut} = 10$ there is a diminishing returns on mobility prediction with the additional chromophores.

However, we found for the materials currently under investigation, `pySCF` is speedy

enough such that introducing d_{cut} adds more ambiguity into the workflow than is necessary given that the average time per QCC dimer calculation is .036s for P3HT and 1.2s for ITIC. Furthermore, with `MorphCT` acting on static atomistic orientations, these calculations are only necessarily performed once per morphology. With that, our workflow defaults this cutoff distance to half the length of the simulation box, rendering it effectively moot.

If, in the future, more computationally expensive methods are incorporated into the QCC step, or chromophores in other organic materials are a heavier lift `pySCF`, it could be beneficial to reintroduce this cutoff distance. The optimal d_{cut} will vary depending on the material and before doing large sweeping analysis on a new material it is at a discount to do some preliminary analysis to determine an appropriate value.

2.7.2 Reorganization energy

In the context of section 2.2, reorganization energy, λ_{ij}, constitutes the energy required to distort the dimer's equilibrium geometry with a charge on chromophore i into the dimers equilibrium geometry with charge on chromophore j. Reorganization energy consist of the energy change associated with the distortion of the dimers geometry, and the distortion of the surrounding medium in response the movement of the charge. It can be written as follows:

$$\lambda_{total} = \lambda_{internal} + \lambda_{external}. \tag{2.9}$$

$\lambda = 0.3eV$ is chosen to be the default reorganization energy ($\lambda_{internal} = 0.1eV$ and $\lambda_{external} = .02eV$) as others have done with P3HT [16] and a flourene-triphenylamine copolymer, TFB [40].

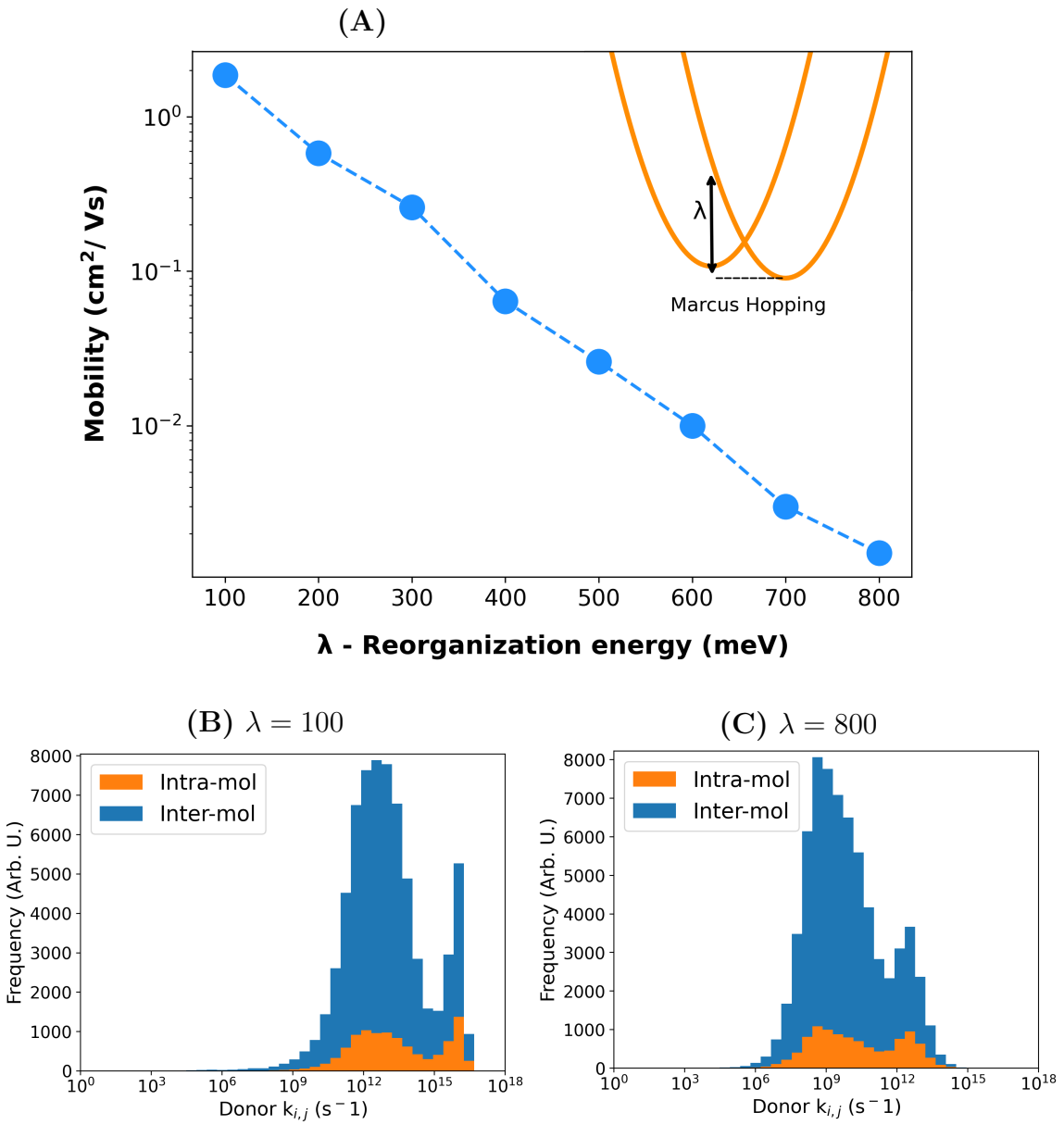


Figure 2.8: Figure (A) shows how kMC simulated mobility values scale with the prescribed reorganization energy values, λ. Figures (B) and (C) show the distribution of hopping rates with $\lambda = 100eV$ and $\lambda = 800eV$ respectively. We see that the exponential decay of charge mobility as reorganization energy increases is a result of a shift in the distribution of hop rates throughout the simulation.

In our workflow, reorganization energy is set as an attribute to the chromophore object. It is defaulted to $\sim 300meV$ as for all chromophore objects. To test the sensitivity of the algorithm to this value we ran 8 simulations with chromophores assigned reorganization energies of $100 - 800meV$. In `MorphCT`, this is as easy as retrieving the pickled crystalline system, looping through the chromophores, setting the reorganization energy, and rerunning the kMC simulation.

The results, shown in Figure 2.8, are expected from the inspection of Equation 2.1. Because these simulation were run on the same morphology, the variation in distributions of k_{ij} values, shown in Figure 2.8(B)(C), is solely due to the choice of λ. The cumulative outcome of this is the exponential decay in mobility as λ is increased.

2.7.3 Temperature

To test the sensitivity of our implementation to temperature, 15 kMC simulations from $100K$ to $800K$ were run on the benchmark P3HT crystalline morphology. It is clear from the results in figure 2.9(A) that the mobilities trend upward with temperature. This should be expected from the assumptions of the model outlined in section 2.2. With relatively weak electronic coupling (T_{ij}) between chromophores, electron transfer proceeds nonadiabatically [8]. With this weak coupling, the temperature in the Gibbs free energy of activation term dominates the effect that temperature has on the hop rate value calculated with Equation 2.1.

An interesting result is that increasing the temperature of the kMC simulation also increases the wall time of the kMC simulations. As an illustration of why this is the case, the distribution of hop rates is plotted for $100K$ and $800K$ in figure 2.9(C)(D). With the distribution of hop rates skewed drastically higher at $800K$, each charge carrier will experience orders of magnitude more hops during its specified lifetime.

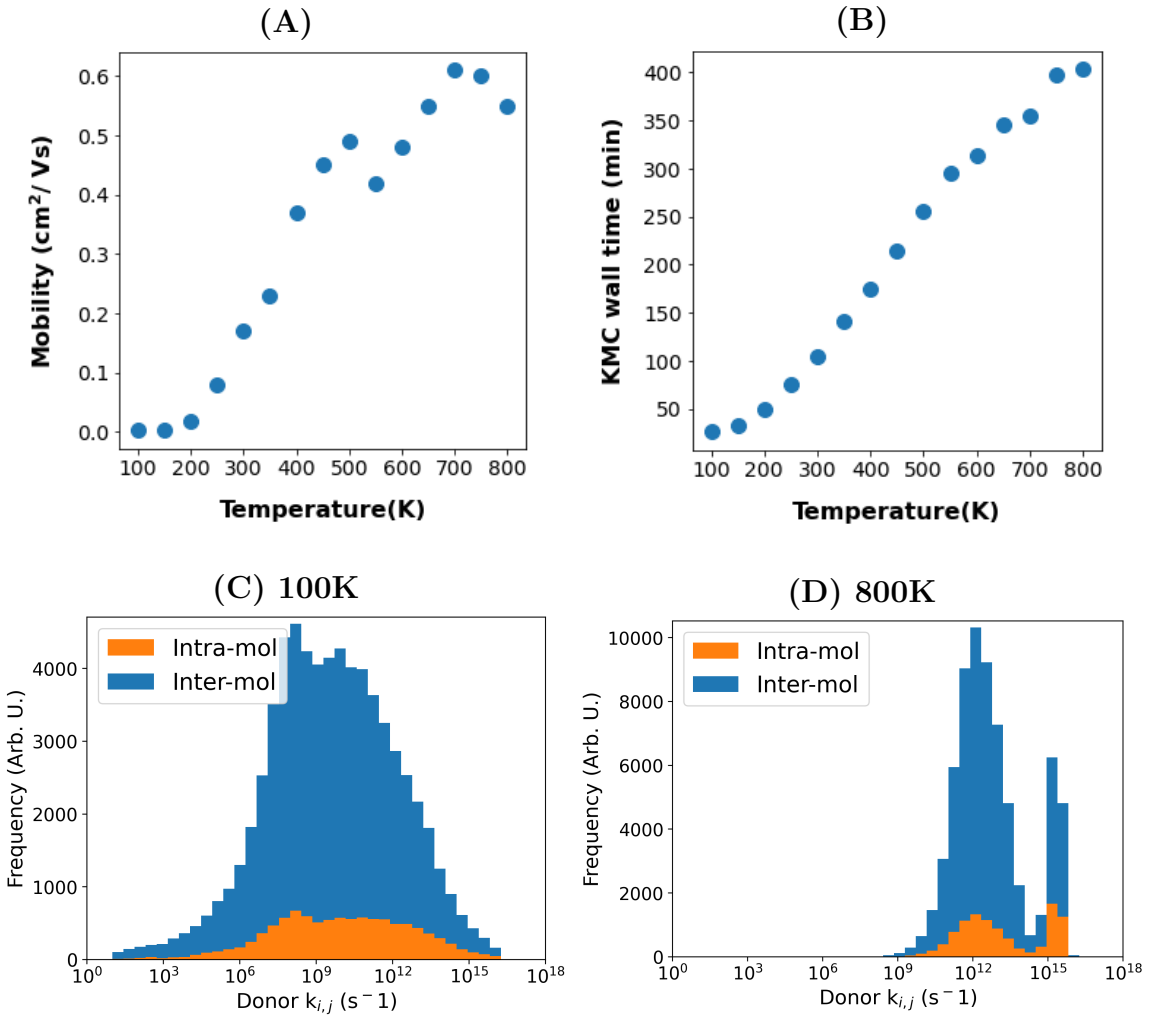


Figure 2.9: The resulting mobility (A) and kMC wall time (B) of 15 kMC simulations from $100K$ to $800K$. The hop rate distribution for the lowest (C) and highest (D) temperature kMC simulations are provided as context for the relationships observed in (A) and (B). With each hop modeled as a thermally activated process, an increase in temperature increases average hop rate and mobility. Orders of magnitude faster hops means orders of magnitude more hops to track in computer memory over a charge carrier's lifetime, which we see results in longer kMC simulation wall times.

2.7.4 MSD (lifetimes)

Introduced in subsection 2.3.1, the 'carrier lifetimes' chosen for a given simulation can effect the analysis of the slope of the MSD as time goes to infinity. Including MSD data for an extremely short lifetime can inflate the estimation. Including MSD data for extremely long lifetimes wastes computation and could introduce unnecessary noise into the data [42]. For example, in an attempt to simulate out to the physical limit, a simulation with a microsecond($10^{-6}s$) lifetime resulted in a single hole hopping for 9 wall time hours.

In real systems, free charge carrier lifetime is subject to a complex interplay between geminate recombination, non-geminate recombination, charge trapping, temperature, and charge density. These dynamics play out across a picosecond to microsecond timescales and vary wildly from material to material and from microstructure to microstructure for a given material [66].

Testing the sensitivity of this choice was not as one-to-one as it was for the parameters tested above. We could choose as many lifetimes across whatever length scales we please. We saw that the choice of two lifetimes is sufficient for estimating slope in section 2.6. So we test the sensitivity to setting the first lifetime progressively shorter.

To do so, we set the second lifetime comfortably in the linear region of the MSD. In a previous work on P3HT, the slope becomes linear around a tenth of a nanosecond [16]. With that, we take the second life time to be two orders of magnitude beyond that at ten nanoseconds. 6 simulations were run with progressively shorter first lifetimes. The first lifetime was set to $10^{-9}s$, $10^{-10}s$, $10^{-11}s$, $10^{-12}s$, and $10^{-13}s$ respectively. The results are plotted in figure 2.10.

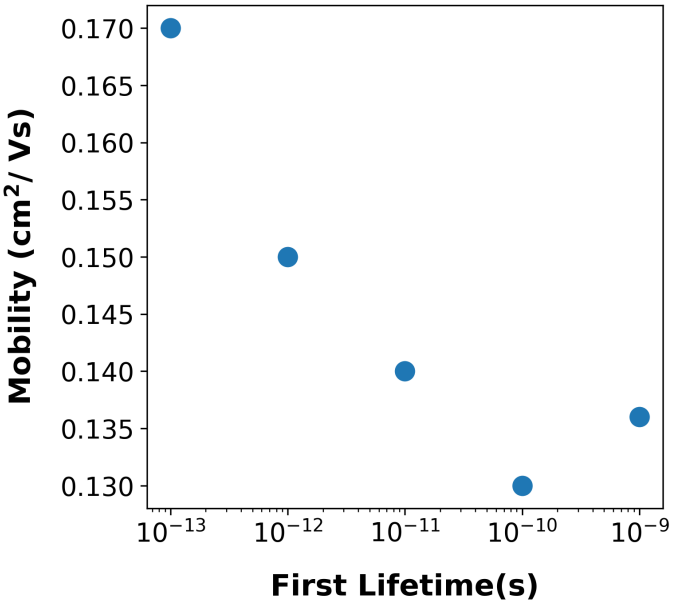


Figure 2.10: The results of running 5 kMC simulations with the first lifetimes as described in the text. It can be seen that below the ballistic timescale ($\sim 10^{-10} s$), the resulting mobility increases.

As expected, as the first lifetime progresses in the ballistic transport timescale, the resulting mobility increases. If the starting lifetime is even shorter the workflow breaks down because as can be seen from the hop rate distribution in figure 2.9(D), even at extreme temperatures, holes need more time that that to hop even once.

Interestingly, the algorithm seems to be quite robust against choice of lifetimes. As can be seen in the figure, order of magnitude differences in lifetime choices results in less than 2X difference in the resulting mobility. Furthermore, as we saw on section 2.6, fitting the slope of the MSD from only two lifetimes results in a satisfactory charge prediction. This suggests that going lifetime crazy is a waste of computation. What is more important then is reporting the lifetimes used in the study for comparison across multiple studies.

2.8 ITIC

Our pipeline is meant to facilitate the computational screening of OPVs. Here we use the pipeline to predict charge mobility in ITIC. This is a first foray into the extensibility of the complete pipeline to a material not named P3HT.

The ITIC morphology was simulated using Plankton-Flow [45] on Fry, a high performance computing cluster at Boise State University. Using `planckton-flow`, a 1000 molecule morphology of ITIC was equilibrated over a 10^7 step MD simulation at room temperature. From the results of the $\sim 200nm^3$ MD simulation, the last frame of the atomic trajectories is taken to represent an accurate equilibrium geometry of ITIC.

To apply the hopping model to this atomistic morphology requires the delineation of segments within the morphology upon which charges can delocalize along LUMO (or HOMO for donors). The LUMO of ITIC delocalizes along the backbone of the molecule, with negligible electron density in the side chains. This makes the backbone, composed of the fused-ring core and end groups, the obvious choice of chromophore. This has been quantified and well visualized using *ab initio* DFT at the level of the molecule by Han et al. [67]. We have visualized this at the nanometer scale in figure 2.11 using the openly available visualization tool OVITO [49].

A single molecule of ITIC has 186 atoms, with the backbone consisting of 70 atoms. We deployed two different approaches to the delineation of chromophores within the ITIC morphology. We first take the backbone to be the chromophores. In another simulation, we take the whole molecule to be a chromophore.

Including the whole molecule necessarily requires more heavy lifting from `pySCF` but is trivially easy from an indexing perspective. Similar to the results of the d_{cut}

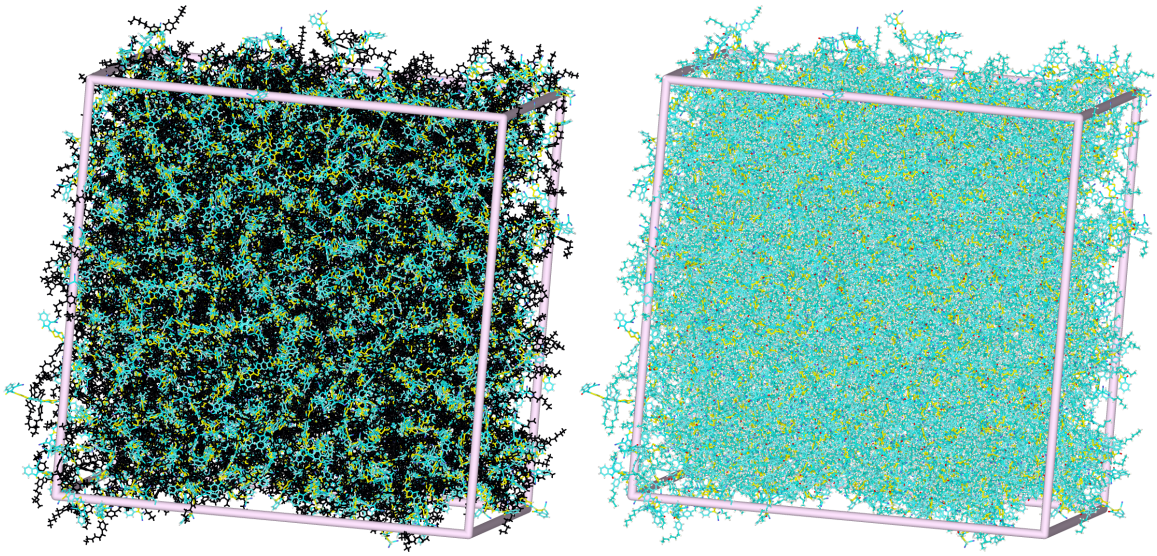


Figure 2.11: 1000 **molecule ITIC morphology. Colored atoms were included in the QCCs for the backbone chromophore simulation (LEFT) and whole molecule chromophore simulations (RIGHT).**

investigation, this implementation of `pySCF`, combined with clever pickling of system objects at various stages of the workflow, suggest that the laborious delineation of chromophores can be largely circumvented in macromolecular systems like ITIC. We found that, while the 70 atom chromophores took $1.2s$ per dimer while, the whole molecule dimer calculations with 186 atoms per chromophore took on average $3.3s$. This is a substantial increase across a hundred thousand pair calculation. However, as we have seen in subsection 2.4.2, organizing a workflow that ensures that this step only be performed once minimizes the computational blow.

We found comparable mobilities of $(1.019 \pm 0.001) \cdot 10^{-3}$ for the backbone chromophores and $(1.275 \pm 0.001) \cdot 10^{-3}$ for the whole molecule. This slight increase in mobility comports with the reality, as including the side chains adds some electron density off the axis of the backbone which could facilitated hopping pathways unexplored by the backbone only simulations. While the contribution from the electron

density along the side chains is minimal when discussing mobility values that vary orders of magnitude, the contribution is not zero.

As discussed in the sensitivity analysis, our mobility calculations are relatively sensitive to the choice of reorganization energy. For ITIC, $\lambda_{internal}$ has been well investigated and is widely reported as $0.15eV$ [67]. The external reorganization is harder to estimate. We take $\lambda_{total} = 0.3eV$ as we did for P3HT. With the fused backbone resulting in a higher internal contribution and the lack of long range order resulting in a lower external contribution.

The reported experimental electron mobility of ITIC varies depending on how it was processed and how it was measured. Time-of-flight electron mobilities on the order of 10^{-4} [68] and field effect mobilities on the order of 10^{-2} [69] have been reported. Another computational study, that also used Marcus hopping and kMC, found an electron mobility of $7 \cdot 10^{-4}$ [70] in a pure ITIC crystal.

CHAPTER 3:

CONCLUSION

As others have argued [71][40][16], the kind of workflow outlined in this thesis can allow for cheaper and more expansive screening of OPV materials across varied chemistries and processing conditions.

We have shown that it is theoretically possible and practically achievable to simulate OPVs. Through the publication of workflows, code development, and data we have outlined a Transparent, Reproducable, Usable, Extensible pipeline through which any researcher can quantify material properties in organic semiconductors. We have verified that our implementation is satisfactorily prediction charge mobility. We have performed sensitivity analysis on our pipeline using benchmark P3HT morphologies. Through our study on ITIC, have also seen that the pipeline is readily extensible to other OPV's.

Throughout this work, the most critical resource has been collaboration with other developers. This collaboration took place in real time with active developers and asynchronously through the trails of bread crumbs left behind by predecessors in the form of documentation, searchable public communication, and tutorials. This is unsurprising, because learners that are new to an area of research, or to an application programming interface, will often experience similar pitfalls. Open source software development provides a scaffolding around which we can take note of these pitfalls and

actively work to smooth the path for the next researcher to expand the boundaries of the research further.

Through this lens, we have seen that the frayed edges, the pure scientific theory and the pure data/computer science, of scientific software development can entangle an aspiring researcher. For this particular project, that meant learning quantum theory and data science simultaneously. We have attempted to tie these ends together by creating tutorials that expose a new user to the `MorphCT` code base and to the scientific theory underlying it. Maintaining these tutorials as part of the code base allows future researchers to modify and expand on them. Tutorials for using `MorphCT` to explore the energetics of an organic compound (commit 8695a81), to delineate chromophores within an atomistic morphology (commit 8695a81), and to perform a mobility prediction for a morphology (commit 29d6b33) are maintained on the github repository for `MorphCT` [72]. Jupyter notebooks containing the code used to generate the figures for this work can be found on the github repository for this thesis [73].

As for the algorithm employed with `MorphCT`, we have learned that `PySCF` is quantitatively and computationally sufficient for providing QCCs across large atomistic morphologies. We found that the efficacy of PySCF allows for the simplification of our algorithms for neighbor listing and chromophore delineation.

We have provided a proof of concept of the extensibility of our pipeline in our investigation of ITIC. However, the next step in this research will be to improve the extensibility of `MorphCT` through the development of `MorphCT-Flow` for the containerization and dataspace management for large scale screening across molecules and state spaces.

REFERENCES

[1] Hongliang Chen, Weining Zhang, Mingliang Li, Gen He, and Xuefeng Guo. Interface Engineering in Organic Field-Effect Transistors: Principles, Applications, and Perspectives. *Chemical Reviews*, 120(5):2879–2949, 2020.

[2] Christopher J. Bettinger and Zhenan Bao. Biomaterials-based organic electronic devices. *Polymer International*, 59(5):563–567, 2010.

[3] Clara Santato (Editor) Fabio Cicoira (Editor). *Organic Electronics: Emerging Concepts and Technologies*. Wiley, 2013.

[4] Christoph J. Brabec, Andreas Distler, Xiaoyan Du, Hans Joachim Egelhaaf, Jens Hauch, Thomas Heumueller, and Ning Li. Material Strategies to Accelerate OPV Technology Toward a GW Technology. *Advanced Energy Materials*, 10(43):1–10, 2020.

[5] Qishi Liu, Yufan Jiang, Ke Jin, Jianqiang Qin, Jingui Xu, Wenting Li, Ji Xiong, Jinfeng Liu, Zuo Xiao, Kuan Sun, Shangfeng Yang, Xiaotao Zhang, and Liming Ding. 18% Efficiency organic solar cells. *Science Bulletin*, 65(4):272–275, 2020.

[6] Stefano Baroni. first dielectric for Si. 33(10), 1986.

[7] John George and Nandhibatla V. Sastry. Densities, viscosities, speeds of sound, and relative permittivities for water + cyclic amides (2-pyrrolidinone, 1-methyl-

2-pyrrolidinone, and 1-vinyl-2-pyrrolidinone) at different temperatures. *Journal of Chemical and Engineering Data*, 49(2):235–242, 2004.

[8] Tracey M. Clarke and James R. Durrant. Charge Photogeneration in Organic Solar Cells. *Chemical Reviews*, 110(11):6736–6767, nov 2010.

[9] C. W. Tang. Two-layer organic photovoltaic cell. *Applied Physics Letters*, 48(2):183–185, 1986.

[10] Michael A. Fusella, YunHui L. Lin, and Barry P. Rand. *Organic photovoltaics (OPVs): Device physics*. Elsevier Ltd., 2 edition, 2019.

[11] Markus C. Scharber. On the Efficiency Limit of Conjugated Polymer:Fullerene-Based Bulk Heterojunction Solar Cells. *Advanced Materials*, 28(10):1994–2001, 2016.

[12] Jiayu Wang and Xiaowei Zhan. Fused-Ring Electron Acceptors for Photovoltaics and beyond. *Accounts of Chemical Research*, 54(1):132–143, 2021.

[13] Leandro Benatto and Marlus Koehler. Effects of Fluorination on Exciton Binding Energy and Charge Transport of -Conjugated Donor Polymers and the ITIC Molecular Acceptor: A Theoretical Study. *Journal of Physical Chemistry C*, 123(11):6395–6406, 2019.

[14] P. Vanlaeke, A. Swinnen, I. Haeldermans, G. Vanhoyland, T. Aernouts, D. Cheyns, C. Deibel, J. D'Haen, P. Heremans, J. Poortmans, and J. V. Manca. P3HT/PCBM bulk heterojunction solar cells: Relation between morphology and electro-optical characteristics. *Solar Energy Materials and Solar Cells*, 90(14):2150–2158, 2006.

[15] David P McMahon, David L Cheung, Ludwig Goris, Javier Dacuña, Alberto Salleo, and Alessandro Troisi. Relation between Microstructure and Charge Transport in Polymers of Different Regioregularity. *The Journal of Physical Chemistry C*, 115(39):19386–19393, oct 2011.

[16] Matthew Lewis Jones and Eric Jankowski. Computationally connecting organic photovoltaic performance to atomistic arrangements and bulk morphology. *Molecular Simulation*, 43(10-11):1–18, mar 2017.

[17] C.W. Voet, D. and Voet, J.G. and Pratt. *Fundamentals of Biochemistry: Life at the Molecular Level, 4th Edition: Life at the Molecular Level*. Wiley, 2011.

[18] Zhen Wang, Xuncheng Liu, Haiying Jiang, Xiaobo Zhou, Lianjie Zhang, Feilong Pan, Xianfeng Qiao, Dongge Ma, Wei Ma, Liming Ding, Yong Cao, and Junwu Chen. Organic Solar Cells Based on High Hole Mobility Conjugated Polymer and Nonfullerene Acceptor with Comparable Bandgaps and Suitable Energy Level Offsets Showing Significant Suppression of Jsc–Voc Trade-Off. *Solar RRL*, 3(7):1–9, 2019.

[19] Jonathan A. Bartelt, David Lam, Timothy M. Burke, Sean M. Sweetnam, and Michael D. McGehee. Charge-Carrier Mobility Requirements for Bulk Heterojunction Solar Cells with High Fill Factor and External Quantum Efficiency >90%. *Advanced Energy Materials*, 5(15):1–10, 2015.

[20] Cephas E. Small, Sai Wing Tsang, Song Chen, Sujin Baek, Chad M. Amb, Jegadesan Subbiah, John R. Reynolds, and Franky So. Loss mechanisms in thick-film low-bandgap polymer solar cells. *Advanced Energy Materials*, 3(7):909–916, 2013.

[21] Peter T Cummings. Computational Screening of Soft Materials Systems with Application to Nano-Lubrication Systems. In *Telluride Science Research Center Workshop: Molecular engineering of soft matter: Spanning small molecules to macromolecules*, Telluride, CO, 2017.

[22] Eric Jankowski, Neale Ellyson, Jenny W. Fothergill, Michael M. Henry, Mitchell H. Leibowitz, Evan D. Miller, Mone't Alberts, Samantha Chesser, Jaime D. Guevara, Chris D. Jones, Mia Klopfenstein, Kendra K. Noneman, Rachel Singleton, Ramon A. Uriarte-Mendoza, Stephen Thomas, Carla E. Estridge, and Matthew L. Jones. Perspective on coarse-graining, cognitive load, and materials simulation. *Computational Materials Science*, 171:109129, jan 2020.

[23] Eric Jankowski, Neale Ellyson, Jenny W Fothergill, Michael M Henry, Mitchell H Leibowitz, Evan D Miller, Mone't Alberts, Samantha Chesser, Jaime D Guevara, Chris D. Jones, Mia Klopfenstein, Kendra K Noneman, Rachel Singleton, Ramon A Uriarte-Mendoza, Stephen Thomas, Carla E Estridge, and Matthew L Jones. Perspective on Coarse-Graining , Cognitive Load , and Materials Simulation. *Computational Materials Science*, 169(109129):109129, jan 2020.

[24] Carl Poelking, Kostas Daoulas, Alessandro Troisi, and Denis Andrienko. Morphology and Charge Transport in P3HT: A Theorist's Perspective. In *Advances in Polymer Science*, pages 139–180. Springer-Verlag Berlin Heidelberg, 2014.

[25] Jana Zaumseil. P3HT and other Polythiophene field-effect transistors. *Advances in Polymer Science*, 265(June):107–138, 2014.

[26] Huitao Bai, Yifan Wang, Pei Cheng, Jiayu Wang, Yao Wu, Jianhui Hou, and Xiaowei Zhan. An electron acceptor based on indacenodithiophene and 1,1-dicyanomethylene-3-indanone for fullerene-free organic solar cells. *Journal of Materials Chemistry A*, 3(5):1910–1914, 2015.

[27] Evan Miller, Matthew Jones, Michael Henry, Paul Chery, Kyle Miller, and Eric Jankowski. Optimization and Validation of Efficient Models for Predicting Polythiophene Self-Assembly. *Polymers*, 10(12):1305, nov 2018.

[28] J E Jones and Proc R Soc Lond A. On the determination of molecular fields. —II. From the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(738):463–477, 1924.

[29] William G. Hoover. Canonical Dynamics: Equilibrium Phase-space Distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[30] Glenn J. Martyna, Douglas J. Tobias, and Michael L. Klein. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*, 101(5):4177–4189, 1994.

[31] J. Cao and G. J. Martyna. Adiabatic path integral molecular dynamics methods. II. Algorithms. *The Journal of Chemical Physics*, 104(5):2028–2035, feb 1996.

[32] Michael M. Henry, Matthew Lewis Jones, Stefan D. Oosterhout, Wade A. Braunecker, Travis W. Kemper, Ross E. Larsen, Nikos Kopidakis, Michael F. Toney, Dana C. Olson, and Eric Jankowski. Simplified Models for Accelerated Structural

Prediction of Conjugated Semiconducting Polymers. *The Journal of Physical Chemistry C*, 121(47):26528–26538, nov 2017.

[33] Qiang Cui and Marcus Elstner. Density functional tight binding: Values of semi-empirical methods in an ab initio era. *Physical Chemistry Chemical Physics*, 16(28):14368–14377, 2014.

[34] Walter Thiel. Semiempirical quantum-chemical methods. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(2):145–157, 2014.

[35] Jingsong Huang and Miklos Kertesz. Validation of intermolecular transfer integral and bandwidth calculations for organic molecular materials. *Journal of Chemical Physics*, 122(23), 2005.

[36] Wei-Qiao Deng and William A. Goddard. Predictions of Hole Mobilities in Oligoacene Organic Semiconductors from Quantum Mechanical Calculations. *The Journal of Physical Chemistry B*, 108(25):8614–8621, jun 2004.

[37] J. L. Bredas, J P Calbert, D A da Silva Filho, and J Cornil. Organic Semiconductors: A Theoretical Characterization of the Basic Parameters Governing Charge Transport. *Proceedings of the National Academy of Sciences*, 99(9):5804–5809, apr 2002.

[38] Vyas Ramasubramani, Bradley D. Dice, Eric S. Harper, Matthew P. Spellings, Joshua A. Anderson, and Sharon C. Glotzer. freud: A software suite for high throughput analysis of particle simulation data. *Computer Physics Communications*, 254:107275, 2020.

[39] Jon Louis Bentley, Bruce W. Weide, and Andrew C. Yao. Optimal Expected-Time Algorithms for Closest Point Problems. *ACM Transactions on Mathematical Software (TOMS)*, 6(4):563–580, 1980.

[40] Sai Manoj Gali, Gabriele D'Avino, Philippe Aurel, Guangchao Han, Yuanping Yi, Theodoros A. Papadopoulos, Veaceslav Coropceanu, Jean-Luc Brédas, Georges Hadziioannou, Claudio Zannoni, and Luca Muccioli. Energetic fluctuations in amorphous semiconducting polymers: Impact on charge-carrier mobility. *The Journal of Chemical Physics*, 147(13):134904, oct 2017.

[41] Waldemar Kaiser, Johannes Popp, Michael Rinderle, Tim Albes, and Alessio Gagliardi. Generalized kinetic Monte Carlo framework for organic electronics. *Algorithms*, 11(4), 2018.

[42] Edward J. Maginn, Richard A. Messerly, Daniel J. Carlson, Daniel R. Roe, and J. Richard Elliott. Best Practices for Computing Transport Properties 1. Self-Diffusivity and Viscosity from Equilibrium Molecular Dynamics [Article v1.0]. *Living Journal of Computational Molecular Science*, 1(1):6324, 2018.

[43] Akshay Kokil, Ke Yang, and Jayant Kumar. Techniques for characterization of charge carrier mobility in organic semiconductors. *Journal of Polymer Science, Part B: Polymer Physics*, 50(15):1130–1144, 2012.

[44] Baijun Chen, Chun Sing Lee, Shuit Tong Lee, Patrick Webb, Yan Cheong Chan, William Gambling, He Tian, and Weihong Zhu. Improved time-of-fiight technique for measuring carrier mobility in thin films of organic electroluminescent materials. *Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers*, 39(3 A):1190–1192, 2000.

[45] https://github.com/cmelab.

[46] Thomas Kluyver, Benjamin Ragan-kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter Notebooks - A Publishing Format for Reproducible Computational Workflows. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90, 2016.

[47] Christoph Klein, Andrew Z. Summers, Matthew W. Thompson, Justin Gilmer, Clare McCabe, Peter T. Cummings, Janos Sallai, and Christopher R. Iacovella. Formalizing Atom-typing and the Dissemination of Force Fields with Foyer. *arXiv*, 167(May):1–38, 2018.

[48] Joshua A. Anderson, Jens Glaser, and Sharon C. Glotzer. HOOMD-blue: A Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Computational Materials Science*, 173(April 2019):109363, 2020.

[49] Alexander Stukowski. Visualization and analysis of atomistic simulation data with OVITO-the Open Visualization Tool. *Modelling and Simulation in Materials Science and Engineering*, 18(1), 2010.

[50] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry*, 30(13):2157–2164, oct 2009.

[51] Qiming Sun, Timothy C. Berkelbach, Nick S. Blunt, George H. Booth, Sheng

Guo, Zhendong Li, Junzi Liu, James D. McClain, Elvira R. Sayfutyarova, Sandeep Sharma, Sebastian Wouters, and Garnet Kin Lic Chan. PySCF: the Python-based simulations of chemistry framework. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(1), 2018.

[52] Carl S. Adorf, Paul M. Dodd, and Sharon C. Glotzer. signac - A Simple Data Management Framework. pages 1–14, nov 2016.

[53] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, feb 1996.

[54] Junmei Wang, Romain M. Wolf, James W. Caldwell, Peter A. Kollman, and David A. Case. Development and testing of a general Amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[55] Gregory M Kurtzer, Vanessa Sochat, and Michael W Bauer. Singularity: Scientific containers for mobility of compute. *PLOS ONE*, 12(5):1–20, 2017.

[56] Dirk Merkel. Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux J.*, 2014(239), mar 2014.

[57] Jürgen Cito and Harald C. Gall. Using docker containers to improve reproducibility in software engineering research. *Proceedings - International Conference on Software Engineering*, pages 906–907, 2016.

[58] Christoph Klein, János Sallai, Trevor J. Jones, Christopher R. Iacovella, Clare McCabe, and Peter T. Cummings. A Hierarchical, Component Based Approach to Screening Properties of Soft Matter. In *Foundations of Molecular Modeling and*

Simulation, Molecular Modeling and Simulation: Applications and Perspectives, pages 79–92. 2016.

[59] S. I. Gorelsky and A. B.P. Lever. Electronic structure and spectra of ruthenium diimine complexes by density functional theory and INDO/S. Comparison of the two methods. *Journal of Organometallic Chemistry*, 635(1-2):187–196, 2001.

[60] Juan Aragó, Pedro M. Viruela, Enrique Ortí, Reyes Malavé Osuna, Barbara Vercelli, Gianni Zotti, Víctor Hernández, Juan T. López Navarrete, John T. Henssler, Adam J. Matzger, Yoshitake Suzuki, and Shigehiro Yamaguchi. Neutral and oxidized triisopropylsilyl end-capped oligothienoacenes: A combined electrochemical, spectroscopic, and theoretical study. *Chemistry - A European Journal*, 16(18):5481–5491, 2010.

[61] Yi-kang Lan and Ching-i Huang. A Theoretical Study of the Charge Transfer Behavior of the Highly Regioregular Poly-3-hexylthiophene in the Ordered State. *The Journal of Physical Chemistry B*, 112(47):14857–14862, nov 2008.

[62] E. Jankowski. Machine Learning Predictions of Electronic Couplings for Charge Transport Calculations of P3HT. In *AIChE Annual Meeting*, 2019.

[63] Evan D Miller, Matthew Lewis Jones, Michael M Henry, and Eric Jankowski. Molecular Dynamics Data for Optimization and Validation of Modeling Techniques for Predicting Structures and Charge Mobilities of P3HT, 2018.

[64] Evan Miller, Matthew Jones, and Eric Jankowski. Tying Together Multiscale Calculations for Charge Transport in P3HT: Structural Descriptors, Morphology, and Tie-Chains. *Polymers*, 10(12):1358, dec 2018.

[65] F. Neese. The ORCA Program System. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(1):73–78, 2012.

[66] Frédéric Laquai, Denis Andrienko, Ralf Mauer, and Paul W.M. Blom. Charge carrier transport and photogeneration in P3HT:PCBM photovoltaic blends. *Macromolecular Rapid Communications*, 36(11):1001–1025, 2015.

[67] Guangchao Han, Yuan Guo, Lu Ning, and Yuanping Yi. Improving the Electron Mobility of ITIC by End-Group Modulation: The Role of Fluorination and π-Extension. *Solar RRL*, 3(1):1–5, 2019.

[68] Natalie A. Mica, Stuart A.J. Thomson, and Ifor D.W. Samuel. Electron mobility of non-fullerene acceptors using a time of flight method. *Organic Electronics*, 63(August):415–420, 2018.

[69] Youngrak Park, Canek Fuentes-Hernandez, Xiaojia Jia, Felipe A. Larrain, Junxiang Zhang, Seth R. Marder, and Bernard Kippelen. Measurements of the field-effect electron mobility of the acceptor ITIC. *Organic Electronics*, 58(February):290–293, 2018.

[70] Guangchao Han, Yuan Guo, Xiaoxian Song, Yue Wang, and Yuanping Yi. Terminal π-π Stacking determines three-dimensional molecular packing and isotropic charge transport in an A-π-A electron acceptor for non-fullerene organic solar cells. *Journal of Materials Chemistry C*, 5(20):4852–4857, 2017.

[71] David R. Evans, Hyunwook S. Kwak, David J. Giesen, Alexander Goldberg, Mathew D. Halls, and Masahito Oh-E. Estimation of charge carrier mobility

in amorphous organic materials using percolation corrected random-walk model. *Organic Electronics*, 29:50–56, 2016.

[72] https://github.com/cmelab/morphct/tree/master/examples.

[73] https://github.com/JimmyRushing/thesis.