Multiscale Charge Transport Networks

Multiscale Charge Transport Networks: Quantitative Analysis and Structural Characterization

Zhongquan Chen

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Thesis title: Multiscale Charge Transport Networks: Quantitative Analysis and Structural Characterization / by Zhongquan Chen
Technische Universiteit Eindhoven, 2025. Proefschrift.

Cover design by Zhongquan Chen

A catalogue record is available from the Eindhoven University of Technology Library

ISBN 978-90-386-6409-5
Keywords: First-principle Multiscale Charge Transport; Markov Chain; Master Equation; Graph Random Walk; Trap; Spectral Clustering; Uncertainty Quantification

The work in this thesis has been sponsored by ICMS via project MPIPICMS2019001.

Multiscale Charge Transport Networks: Quantitative Analysis and Structural Characterization

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische Universiteit Eindhoven, op gezag van de rector magnificus prof.dr. S.K. Lenaerts, voor een commissie aangewezen door het College voor Promoties, in het openbaar te verdedigen op dinsdag 8 juli 2025 om 11:00 uur

door

Zhongquan Chen

geboren te Taishan, Guangdong Province, China

Dit proefschrift is goedgekeurd door de promotoren en de samenstelling van de promotiecommissie is als volgt:

voorzitter:	prof. dr. ir. S.C. Borst
promotor:	dr. B. Baumeier
co-promotor:	dr. W.L.F. van der Hoorn
Leden:	prof. dr. K.P. Veroy-Grepl
	prof. dr. L. Leppert (Universiteit Twente)
	prof. dr. F.C.R. Spieksma
	prof. dr. M.A. Peletier

Het onderzoek of ontwerp dat in dit proefschrift wordt beschreven is uitgevoerd in overeenstemming met de TU/e Gedragscode Weten- schapsbeoefening.

List of Publications

1. **Time-Dependent Density Functional Theory and Green's Functions Methods with the Bethe–Salpeter Equation**
   Wouter Scharpach, Zhongquan Chen, Vivek Sundaram, Björn Baumeier
   in Comprehensive Computational Chemistry (First Edition), edited by Manuel Yáñez, Russell J. Boyd, Elsevier, 84-115 (2024) (Chapter 2)

2. **A Graph Random Walk Method for Calculating Time-of-Flight Charge Mobility in Organic Semiconductors from Multiscale Simulations**
   Zhongquan Chen, Pim van der Hoorn, Björn Baumeier
   Accepted in Multiscale Modeling & Simulation, May, 2025 (Chapter 3 and 4)

3. **Uncertainty Quantification in Multiscale Models of Charge Transport in Organic Semiconductors: Influence of the Exhange-Correlation Functional**
   Zhongquan Chen, Pim van der Hoorn, Björn Baumeier
   arXiv preprint 2503.22356 (2025) (Chapter 5)

4. **Automatic Identification of Traps in Molecular Charge Transport Networks of Organic Semiconductors**
   Zhongquan Chen, Pim van der Hoorn, Björn Baumeier
   in Physical Review B, May, 2025 (Chapter 2 and Chapter 6)

Conference contribution

1. **Multi-Scale Modeling of Charge Dynamics in a Neuromorphic Device Based on PEDOT:PSS.**
   Zhongquan Chen, Björn Baumeier
   MRS Spring Meeting, Neuromorphic Computing and Biohybrid Systems -

Materials and Devices for Brain-Inspired Computing, Adaptive Biointerfacing and Smart Sensing, 2022

2. **Multiscale Modelling of Charge Dynamics in Neuromorphic Devices**
   Zhongquan Chen, Björn Baumeier
   5^{th} CCPBioSim/CCP5 Multiscale Modelling Conference, 2023

3. **Graph Random Walk for Time-of-flight Charge Mobility**
   Zhongquan Chen, Pim van der Hoorn, Björn Baumeier
   Dutch Computational Science (DUCOMS) Day, 2024

Contents

Chapter 1

Introduction

This chapter begins by introducing a material class, organic semiconductors, and the requirements for their application in electronic devices. We will briefly discuss the associated experimental pathways to obtain devices achieving optimal performance targets and highlight how computational modeling can guide in the rational design of organic semiconductors. Particular attention will be paid to multiscale modeling approaches and their respective methodological challenges related to accuracy and reliability, both from fundamental and numerical perspectives, leading to the three concrete problem statements to be tackled in this thesis.

1.1 Organic Semiconductors and Their Applications

Practically all modern electronics [1, 2, 3] are based on semiconducting materials [4]. Inorganic semiconductors, such as silicon, have traditionally dominated the field due to their well-understood properties and established manufacturing processes. Despite their successes, there is a growing interest in alternative semiconductors driven by the need for more versatile, cost-effective, and flexible materials.

Organic semiconductors are a class of materials whose building blocks are organic molecules, such as the ones shown in Fig. 1.1, instead of elementary atoms in case of inorganic semiconductors [5, 6]. As the interactions between the building blocks are much weaker than in regular atomic lattices, the result-

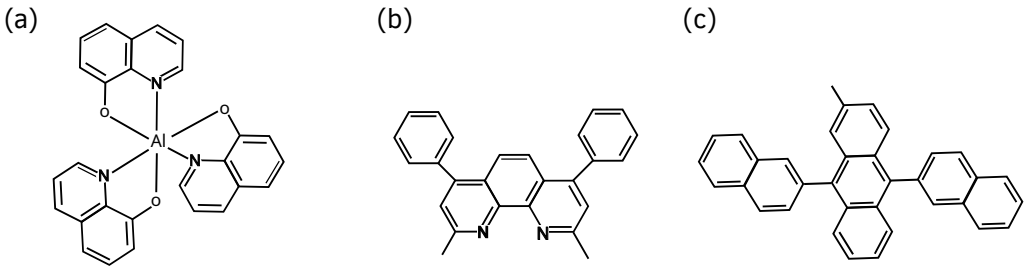

Figure 1.1: Chemical structures of three molecules commonly used in organic semiconductors: (a) Alq_3, (b) BCP, (c) MADN.

ing materials are often spatially disordered and even form amorphous structures at the mesoscopic scale [7]. They are also inherently lighter than their inorganic counterparts [8], which is beneficial for, e.g., portable and wearable devices. Their easy processing from solution using techniques like spin coating [9] or inkjet printing [10, 11] facilitates low-cost, large-area fabrication of electronic devices [12, 13] on flexible substrates [14, 15, 16]. However, such shape flexibility of organic semiconductors means that molecular structures are easily altered, significantly impacting their ability to conduct electrical charges and hence the performance of the electronic device they are used in.

Figure 1.2 shows a simplified schematic of a typical organic light-emitting diode (OLED) consisting of a layered structure, including a cathode, an anode, and intermediate layers comprising organic semiconductors that facilitate charge transport and light emission. When a voltage is applied, electrons are injected from the cathode into the electron transport layer, while holes are injected from the anode into the hole transport layer. These charge carriers migrate toward the central light-emitting layer, where they recombine to generate photons and emit light. The efficiency and stability of an OLED depends significantly (but not exclusively) on the charge transport properties in the electron and hole transport layers and is crucial for achieving high brightness, low power consumption, and extended device lifetime [17, 7, 18].

Charge transporting properties of the organic semiconductors also have similarly high importance for applications in other opto-electronic devices, such as

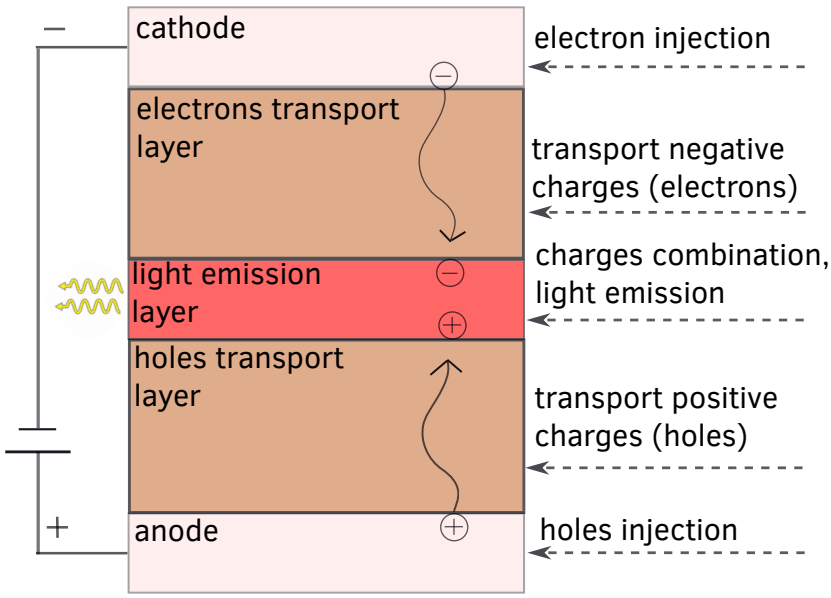

Figure 1.2: Schematic diagram of an OLED device structure. The device consists of multiple layers, including the anode, which extracts and transports positive charges (holes), and the cathode, which inserts electrons. The hole transport layer facilitates hole movement toward the emission layer, where charge recombination occurs, leading to light emission. The electron transport layer transports electrons toward the emission layer, ensuring efficient charge injection and recombination.

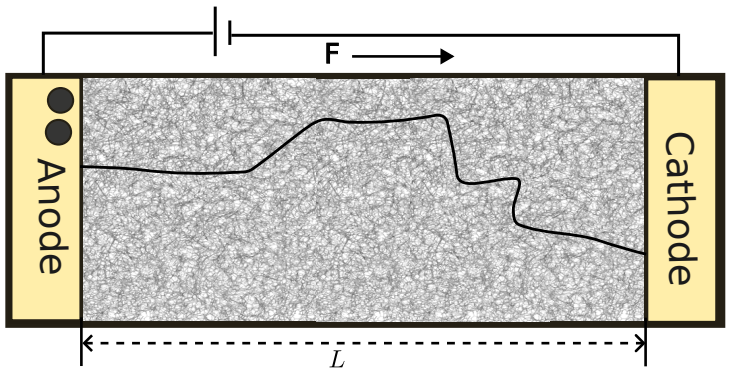

Figure 1.3: Schematic diagram of the Time-of-Flight experimental setup. The gray region between the anode and cathode electrodes represents the organic semiconductor, with solid circles on the anode indicating the injected charges. The electric field, denoted by **F**, is directed from the anode to the cathode. A solid curve within the organic semiconductor illustrates the potential charge transport pathway from the anode to the cathode.

organic photovoltaic cells [19, 20, 21], organic photodetectors [22, 23], sensors [24, 25, 26], organic field-effect transistors [27, 28].

1.2 Charge Mobility as Quantity of Interest

The primary property used to quantify charge transport in organic semiconductors is the charge mobility. While it is a material property, it is not a material constant, as in general the charge mobility is influenced by external factors such as temperature, charge carrier concentration, applied electric field (or voltage), or in some cases the size of the material. On the macroscale, charge transport can be seen as a drift-diffusion process and charge mobility additionally depends on whether the system is in a steady-state or not.

In the following we focus specifically on the non-steady-state situation, as it is more closely related to the operating conditions of organic semiconductors in devices. One of the ways in which charge mobility is measured experimentally is in a so-called time-of-flight setting [29, 30, 31], which is illustrated in Fig. 1.3. In this setup, two electrodes sandwich an organic semiconductor of length L. Charges

(holes or electrons) are injected at one side and migrate across the organic semiconductor towards the opposite electrode under the influence of an electric field of strength $|\mathbf{F}|$ which is oriented parallel to the normals of the electrodes. The time t taken for a charge to travel from one electrode to the other, once detected, is referred to as the time of flight (ToF). The charge mobility can be estimated as

$$\mu = \frac{\mathbf{v} \cdot \mathbf{F}}{|\mathbf{F}|^2} = \frac{L}{t|\mathbf{F}|}. \tag{1.1}$$

Using this time-of-flight mobility as the quantity-of-interest for the design of materials, experimental efforts focus either on improving the processing of the material to obtain, e.g., less disordered morphologies or systems with fewer defects and impurities, [32, 12, 33, 34] or on tuning the chemical details of the molecular building blocks themselves [35, 13, 36]. However, charge transport on the material scale is a result of an intricate interplay between structural features on micro- and mesoscale and the electronic properties of the molecular building blocks. While the tunability of organic semiconductors along these two dimensions (morphology and chemistry) holds in principle the potential to design of materials with tailored electronic properties to meet the requirements of various applications, disentangling the individual factors leading to the emerging observable material characteristic is practically impossible from an experimental perspective. Therefore, experimental approaches can be somewhat trial-and-error based, require substantial resources, and often lack a fundamental understanding of the underlying mechanisms. Consequently, even when a molecule with desirable conductive properties is identified, the principles governing its performance often remain obscure, hindering systematic design strategies for future materials.

1.3 The Role of Modeling

Computational approaches for modeling the charge transport processes in organic semiconductors can play in crucial role in guiding the experimental material design efforts by zooming into the explicit dynamics of the electronic and atomic degrees of freedom in the material, revealing fundamental processes and the in-

terplay between electronic structure of molecular building blocks and the local and global morphological details. Insight from such models should allow formulating design rules on the one hand, and *in silico* prescreening of materials on the other hand [37, 38, 39, 40].

Early efforts at modeling charge transport in organic semiconductors revolved around descriptive, qualitative models rather than predictive, quantitative ones. One of the seminal approaches is the Gaussian Disorder Model (GDM) [41]. It is built on the premise that, due to the amorphous and disordered molecular arrangement, charge carriers (electrons or holes) are localized [42], that is confined to specific regions within the material. Consequently, charge transport is considered a carrier hopping process among two localized states i and j (i.e., single molecules in organic semiconductors studied in this thesis) described by a transition rate ω_{ij}. Bässler further employs a cubic lattice to represent the localization points (often considered the centers-of-masses of molecules) for the electronic states and assigns them an energy E. The rate of transition between two states is then proportional to the exponential of the energy difference: $\omega_{ij} \sim \exp(E_i - E_j)$. Specifically, the energy E follows a Gaussian distribution, with the probability density function given by

$$p(E) = (2\pi\sigma)^{-1/2} \exp(-E^2/\sigma^2), \tag{1.2}$$

where σ is the energetic disorder parameter. The GDM allows simulation of charge transport as a sequence of individual transfer events and partially explains qualitatively the dependence of mobility on temperature, disorder, and charge concentration but shows discrepancies when compared to later experimental observations and lacks a fundamental interpretation of the underlying physics due to its empirical nature [43]. Later extensions considered an exponential density of states (DOS) where the energies follow an exponential distribution, or a correlated Gaussian disorder model when the state energies exhibit spatial correlations. While these extensions improve in some cases the description of experimental results, they do not consider microscopic factors such as molecular shape, relative orientation, and arrangement, which are crucial for optimizing organic semiconductors.

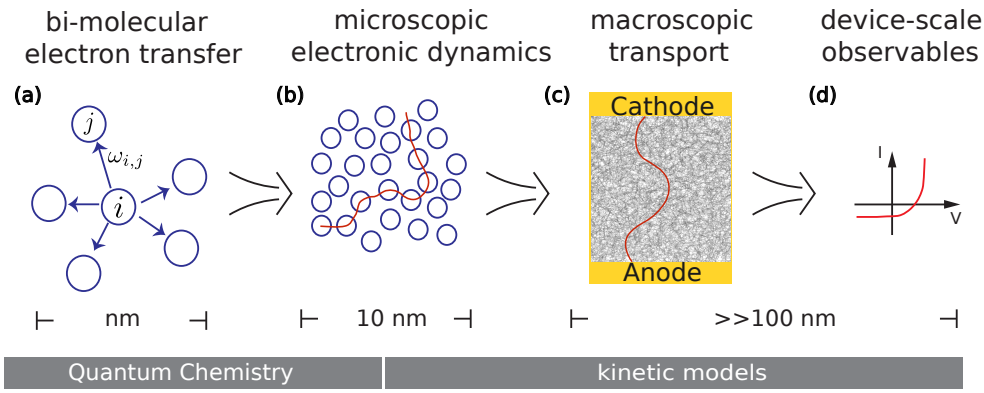

Figure 1.4: Schematic of the multiscale model of charge transport. (a) Circles represent localized states, and arrows indicate possible carrier transitions among molecules. (b) The red curve illustrates a sequence of carrier hops between molecules. (c) A simple organic semiconductor device with an anode and cathode, operating under the same conditions as in Fig. 1.3. The red curve represents a possible charge transfer trajectory from the anode to the cathode. (d) An example of a physical observable at the device scale.

A predictive model of charge transport in disordered organic semiconductors needs to start from the specific chemical and geometrical structure of the system and calculate the electronic dynamics and eventually the charge mobility from first principles. Due to the quantum nature of the electrons such an approach should be based on solving the time-dependent Schrödinger equations for the coupled electron-nuclear system. Obtaining a solution to this problem directly is practically impossible for material systems like organic semiconductors, even with commonly used approximate single-particle theories such as density-functional theory. Instead, it is possible to build a multiscale model based like the GDM and extensions on the idea of spatially localized electronic states and transitions between them. However, in contrast to the GDM, the approach, as sketched in Fig. 1.4 is based on explicit structural details of the material at atomistic level and explicit evaluation of the transition rates from first-principles theories [44]. The detailed theory and practice of this computational approach referred to as multiscale model in this thesis will be discussed in Chapter 2. Ultimately, the multiscale model, like the GDM, reduces the charge transport to the dynamics on a graph, whose vertices are the molecules' centers-of-masses and the transition

rates are used to define together with a distance criterion the weighted edges of the graph.

1.4 Problem Statements

Realizing that the graph model is the central element of the multiscale model, we refer to it as *molecular charge transport network*. One can distinguish three classes of challenges for obtaining accurate predictions of the charge mobility associated to different stages of the modeling process:

- the setup of the graph, which involves certain assumptions and approximations in solving the underlying classical and quantum models;
- the characterization and identification elements or regions in the network that influence the charge dynamics;
- the theoretical and numerical methods used to extract the charge mobility.

These overall challenges help to define three concrete problem statements that are at the core of the work presented in this thesis related to the following topics:

1.4.1 Accuracy of Charge Mobility Estimation

For a given molecular charge transport network, charge mobility is conventionally obtained from methods like Kinetic Monte Carlo (KMC) or the direct numerical solution of the master equation, a coupled system of ordinary differential equations for the time-dependent probablilities to find the graph in a specific state in its state space. Both methods will be fully described in Chapter 3. What both have in common is that they can be subject to convergence issues. For example, KMC simulations often face poor sampling convergence due to spending a long time in some low-energy states, leading to significant computational expenses and potentially large errors in mobility estimates. Numerically solving the master equation can have numerical stability problems. In both cases, these issues hamper the use of the mutiscale-modeled molecular charge transport networks in exploring

the fundamental aspects of charge transport and the accurate prediction of the quantity-of-interest. The first problem statement therefore is:

> **How to obtain accurate charge mobility predictions from molecular charge transport networks for a variety of different physical situation?**

1.4.2 Uncertainty of The Multiscale Model

As will be explained in detail in Chapter 2, the multiscale model framework used to set up the molecular charge transport network comprises a combination of different modeling techniques, ranging from classical molecular dynamics simulations to quantum and hybrid quantum-classical calculations. Each of these techniques requires or intrinsically contains certain approximations or parameter choices, which in turn lead to uncertainties in the details of the molecular charge transport network and, ultimately, in the predictions of the charge mobility. The uncertainty level might limit the usefulness of the computational model for materials design. Therefore, the work in this thesis focuses on the second problem statement:

> **Is it possible to quantify the uncertainty in the predictions of charge mobilities stemming from modeling choices in the construction of the molecular charge transport network?**

1.4.3 Identification of Traps in Charge Transport Networks

So-called *traps* are known to have a significant impact on the charge mobility. They are individual molecules or clusters of molecules from which it is difficult for the charge carriers to escape. The physico-chemical origin of such traps can be range from structural defects to impurities, but traps can also be the result of large energetic disorder with or without strong spatial correlations. It is typical that one observes a significant increase of the effective mobility depending on the number of charge carriers once all traps in the system are filled and the excess charge carriers are not subject to trapping effects [45, 46, 47, 48].

There are numerous experimental methods to determine the origin, density, and composition of charge carrier traps in organic semiconductors. For instance, space-charge limited current measurements [49] detect trap density based on device-level current measurements. The organic field-effect transistor method [50, 51] measures the overall density of trap states using gate voltage, which relates the trap to the molecular energy level of the organic semiconductors. For finer trap identification, electric force microscopy [52] and Kelvin probe force microscopy [53] techniques utilize electrons to make a picture of the topology. They can provide high spatial resolution images, correlating surface topography with traps.

However, these experimental techniques cannot link specific molecular structures to traps, and the equipment is often cost-prohibitive. Although multiscale modeling has been employed to simulate organic semiconductors, there is currently no computational method to detect traps in these models. Developing a computational method to identify trapping molecules in modeled organic semiconductors would be significantly insightful. The third problem statement is:

> **How can traps of different character uniquely be identified in a multiscale-modeled molecular charge transport network?**

1.5 Thesis Outline

With the above introduction, this thesis is organized as follows: Chapter 2 introduces the methodology of the multiscale model used to study charge transport in organic semiconductors. It covers the relevant theory, concepts, and models of multiscale modeling, followed by a description of the simulation procedures. Chapter 3 presents the theories for calculating mobility using the charge transition rates. Specifically, the master equation and kinetic Monte Carlo methods are introduced, along with a graph random walk (GRW) methodology, which calculates charge mobility from the expected hitting time of a continuous-time random walk, addressing the first problem statement. Chapter 4 verifies the GRW method by comparing it primarily with the KMC method. It also discusses the performance of the GRW method under various scenarios, followed by a comparative

analysis of the results and their implications for uncertainty quantification. Following this, Chapter 5 focuses on the uncertainty quantification of the multiscale-modeled charge mobility, based on quantities of interest obtained from the GRW method. The findings related to uncertainties in the model are presented, along with a discussion on their implications for the model's reliability and accuracy. Chapter 6 introduces a trap-identification method for multiscale-modeled organic semiconductors. The chapter begins by defining traps in organic semiconductors, discussing their significance, and reviewing existing trap-detection methods. The development of the method based on spectral clustering is then presented, followed by the results and robustness of the trap-identification approach. Finally, Chapter 7 concludes the thesis by summarizing the main findings and contributions to the field. It also suggests potential directions for future research based on the results.

Chapter 2

Multiscale Model of Charge Transport

Adapted from the book chapter "*Time-Dependent Density Functional Theory and Green's Functions Methods with the Bethe–Salpeter Equation* by W. Scharpach, Z. Chen, V. Sundaram, B. Baumeier, in Comprehensive Computational Chemistry (First Edition), edited by Manuel Yáñez, Russell J. Boyd, Elsevier, 84-115 (2024)" and the paper "*Automatic Identification of Traps in Molecular Charge Transport Networks of Organic Semiconductors*, Z. Chen, P. van der Hoorn, B. Baumeier, arXiv preprint 2411.07136 (2025)".

This chapter provides a concise summary of the methodological background of the elements of the multiscale model of charge transport that is at the center of this thesis. As charge transport involves electrons, we start from the time-dependent many-electron Schrödinger equation as the most general quantum theory for the dynamics of the coupled electron-nuclei system. We introduce typical approximations, such as the separation of nuclear and electronic degrees of freedom, density-functional theory as an effective single-particle theory, and classical molecular dynamics (MD), and discuss the differences between adiabatic and non-adiabatic dynamics. For the latter, we revisit Fermi's Golden Rule for time-dependent perturbations in application to electron transfer processes to derive the Marcus rate expression. In the second part of the chapter, we will focus on practical aspects of the application of various numerical methods in the overall multiscale modeling framework used in this thesis.

2.1 Multiscale Model: Theory

In the following, Hartree atomic units are used, i.e., with a reduced Planck constant $\hbar = 1$, all charges in units of the elementary charge, positions in units of the Bohr radius, and masses in units of the electron mass. We consider a molecular system comprising M nuclei and N electrons. The coordinates $\mathbf{R}_\alpha$ of the individual nuclei with charges Z_α and masses M_α and $\mathbf{r}_i$ of the individual electrons are combined into the variables $\overline{\mathbf{R}} = (\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M)$ and $\overline{\mathbf{r}} = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$, respectively.

2.1.1 Many-electron Schrödinger Equation

Quantum mechanics describes the state of a system by the wavefunction $\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t)$. With the wavefunction, the probability of finding the system in a state is

$$|\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t)|^2 d^{3N}\overline{\mathbf{r}} d^{3M}\overline{\mathbf{R}} \tag{2.1}$$

given that

$$\int_{\mathbb{R}^{3(M+N)}} |\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t)|^2 d^{3N}\overline{\mathbf{r}} d^{3M}\overline{\mathbf{R}} = 1. \tag{2.2}$$

Denote the non-relativistic many-body Hamiltonian as

$$\begin{aligned}\hat{H} = &\underbrace{-\frac{1}{2}\sum_{\alpha=1}^{M} M_\alpha \Delta_{\mathbf{R}_\alpha}}_{\hat{T}_{\text{nuc}}} + \underbrace{\frac{1}{2}\sum_{\substack{\alpha,\beta=1,\\ \alpha\neq\beta}}^{M} \frac{Z_\alpha Z_\beta}{|\mathbf{R}_\alpha - \mathbf{R}_\beta|}}_{\hat{V}_{\text{nuc-nuc}}} \underbrace{-\frac{1}{2}\sum_{i=1}^{N} \Delta_{\mathbf{r}_i}}_{\hat{T}_{\text{el}}} + \underbrace{\frac{1}{2}\sum_{\substack{i,j=1,\\ i\neq j}}^{N} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}_{\text{el-el}}} \\ &\underbrace{-\sum_{\alpha=1}^{M}\sum_{i=1}^{N} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|}}_{\hat{V}_{\text{nuc-el}}},\end{aligned} \tag{2.3}$$

where $\hat{T}$ and $\hat{V}$ are the respective operators for the kinetic and potential energies involving the nuclear (nuc) and electronic (el) subsystems. The time evolution of

the many-body wave function $\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t)$ is obtained by solving the time-dependent Schrödinger equation [54]

$$\hat{H}\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t) = \mathrm{i}\frac{\partial}{\partial t}\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t) \tag{2.4}$$

and can be used to calculate the expectation value of an observable described by an operator $\hat{A}$ as

$$A(t) = \langle \hat{A} \rangle_{\Psi} = \int \Psi^*(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t)\hat{A}\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t)\mathrm{d}^{3N}\overline{\mathbf{r}}\mathrm{d}^{3M}\overline{\mathbf{R}}. \tag{2.5}$$

This allows – formally – to extract information about structural, electronic and optical properties of the system, as well as the dynamic response to an external perturbation (described by an additional potential term in Eq. (2.3)) as measured, e.g., by injecting external charges.

In practice however, Eq. (2.4) is exactly solvable only for $M = N = 1$, which is a hydrogen atom, so we will need to explore several approximations to make the problem tractable. The standard method of solving a partial differential equation such as Eq. (2.4) is the method of separation of variables in which one makes a product function *Ansatz*, i.e., $\Psi(\overline{\mathbf{r}}, \overline{\mathbf{R}}, t) = \Phi(\overline{\mathbf{r}}, \overline{\mathbf{R}})\Upsilon(t)$. If the Hamiltonian in Eq. (2.3) is not explicitly time-dependent, its expectation value, the total molecular energy E, is constant according to Eq. (2.5), and the time evolution of the wave function is given by $\Upsilon(t) = C\exp(-\mathrm{i}Et)$. The spatial component $\Phi(\overline{\mathbf{r}}, \overline{\mathbf{R}})$ of the wave function and the total energy are obtained as solutions of the stationary Schrödinger equation

$$\hat{H}\Phi(\overline{\mathbf{r}}, \overline{\mathbf{R}}) = E\Phi(\overline{\mathbf{r}}, \overline{\mathbf{R}}). \tag{2.6}$$

Note that in Eq. (2.6) both $\overline{\mathbf{r}}$ and $\overline{\mathbf{R}}$ are explicit dynamic variables. Using the Born-Oppenheimer separation and the adiabatic approximation [55], the nuclei are considered as fixed and generate a static external potential in which the electrons are moving. To express this situation in formal terms, we consider a *fixed arrangement* of nuclei $\overline{\mathbf{R}}$. The Hamiltonian $\hat{H}_{\mathrm{el}}$ representing the electronic system

that interacts with the fixed nuclear configuration is

$$\hat{H}_{\text{el}} = \hat{H}_{\text{el}}(\overline{\mathbf{R}}) = \underbrace{\hat{T}_{\text{el}} + \hat{V}_{\text{nuc-el}}(\overline{\mathbf{R}})}_{\text{1-electron operator}} + \underbrace{\hat{V}_{\text{el-el}}}_{\text{2-electron operator}} . \tag{2.7}$$

In this situation, $\overline{\mathbf{R}}$ is no longer a variable of the electronic system, but a fixed parameter for the electronic degrees of freedom. The corresponding stationary electronic Schrödinger equation is given by

$$\hat{H}_{\text{el}}(\overline{\mathbf{R}})\Phi_\nu(\overline{\mathbf{r}};\overline{\mathbf{R}}) = E_\nu(\overline{\mathbf{R}})\Phi_\nu(\overline{\mathbf{r}};\overline{\mathbf{R}}), \tag{2.8}$$

where $\left\{\Phi_\nu(\overline{\mathbf{r}};\overline{\mathbf{R}})\right\}$ is a set of adiabatic electronic wave functions with state ν. Those can be used as a basis to expand the molecular wave function $\Phi(\overline{\mathbf{r}},\overline{\mathbf{R}})$ according to

$$\Phi(\overline{\mathbf{r}},\overline{\mathbf{R}}) = \sum_\nu \Xi_\nu(\overline{\mathbf{R}})\Phi_\nu(\overline{\mathbf{r}};\overline{\mathbf{R}}). \tag{2.9}$$

Substituting these Born-Oppenheimer expanded wave function into Eq. (2.6) yields [56] a coupled set of equations for the coefficients $\left\{\Xi_\nu(\overline{\mathbf{R}})\right\}$:

$$E\Xi_\nu(\overline{\mathbf{R}}) = \left(E_\nu(\overline{\mathbf{R}}) + \hat{T}_{\text{nuc}} + \hat{V}_{\text{nuc-nuc}}\right)\Xi_\nu(\overline{\mathbf{R}}) + \sum_\mu A_{\nu\mu}\Xi_\mu(\overline{\mathbf{R}}). \tag{2.10}$$

The integrals $A_{\nu\mu}$ are matrix elements of the transition between electronic states ν and μ induced by the dynamics of the nuclei. In the adiabatic approximation, it is assumed that $A_{\mu\nu} = 0$, i.e., there are no transitions between different electronic states, and the nuclear motion for each electronic state ν is determined by

$$E\Xi_\nu(\overline{\mathbf{R}}) = \left(E_\nu(\overline{\mathbf{R}}) + \hat{T}_{\text{nuc}} + \hat{V}_{\text{nuc-nuc}}\right)\Xi_\nu(\overline{\mathbf{R}}). \tag{2.11}$$

This describes the motion of the nuclei in a potential energy surface (PES):

$$U_\nu(\overline{\mathbf{R}}) = E_\nu(\overline{\mathbf{R}}) + \hat{V}_{\text{nuc-nuc}}(\overline{\mathbf{R}}). \tag{2.12}$$

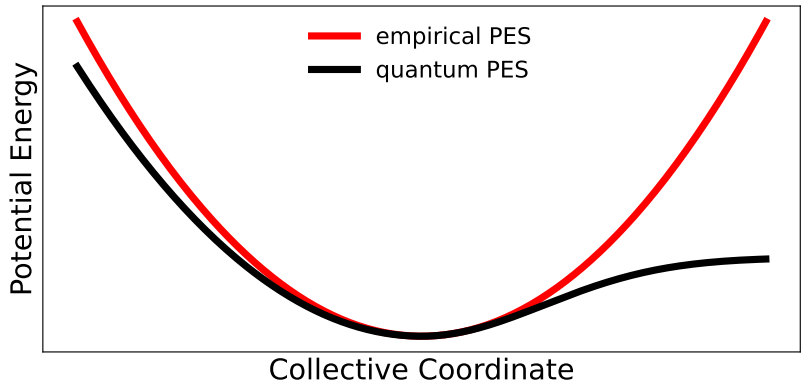

Figure 2.1: Harmonic empirical approximation (red) of the quantum potential energy surface (black).

2.1.2 Molecular Dynamics

Instead of solving the nuclear Schrödinger equation Eq. (2.11), it is possible to approximate the dynamics of the nuclei within the framework of classical mechanics. For a given quantum PES U_ν as the one sketched in Fig. 2.1, one can construct a classical Hamiltonian

$$\mathcal{H}(\overline{\mathbf{R}}(t), \overline{\mathbf{P}}(t)) = \sum_{\alpha=1}^{M} \frac{\mathbf{P}_\alpha(t)\mathbf{P}_\alpha(t)}{2M_\alpha} + U_\nu(\overline{\mathbf{R}}(t)), \tag{2.13}$$

where the $\overline{\mathbf{P}}(t)$ are the momenta of the nuclei. Specifying the initial conditions, i.e., the initial positions and momenta of a collection of particles, determines the system's evolution over time. For each particle α, the time-dependent trajectories $\mathbf{R}_\alpha(t)$ and $\mathbf{P}_\alpha(t)$ are governed by Hamilton's equations of motion (dropping the explicit time-dependence for clarity):

$$\begin{aligned} \frac{d\mathbf{R}_\alpha}{dt} &= \frac{\partial \mathcal{H}(\overline{\mathbf{R}}, \overline{\mathbf{P}})}{\partial \mathbf{P}_\alpha} \\ \frac{d\mathbf{P}_\alpha}{dt} &= -\frac{\partial \mathcal{H}(\overline{\mathbf{R}}, \overline{\mathbf{P}})}{\partial \mathbf{R}_\alpha}. \end{aligned} \tag{2.14}$$

For details on how to numerically solve these equations, including discussions about the different statistical ensembles, we refer the reader to Ref. [57]. Using the quantum PES $U_\nu(\overline{\mathbf{R}})$ in $\mathcal{H}(\overline{\mathbf{R}}, \overline{\mathbf{P}})$ is referred to as *ab-initio molecular dynamics*. However, even with the approximations to the full Schrödinger equation that will be introduced in the following sections, the computational cost for such simulations is enormous, as the quantum PES needs to be recalculated at each time step. To reduce the cost, empirical approximations can be made to the quantum PES as sketched also in Fig. 2.1. This step involves to expand the full PES of the complete set of dynamical variables into two- (bonds), three- (angles), four-body (dihedrals), and higher-order terms, and approximating each of them with a model function, such as harmonic potentials for bonds and angles. The collection of all such model functions is referred to as a *force field*, and the evolution of the Hamiltonian system as *classical molecular dynamics*.

2.1.3 Density-Functional Theory

Density-functional theory (DFT) offers an alternative to the electronic Schrödinger equation by reformulating the problem in terms of the electron density $n(\mathbf{r})$ rather than the many-body wave function. The electron density (suppressing the parametric dependence on $\overline{\mathbf{R}}$) is

$$n(\mathbf{r}) = N \int_{\mathbb{R}^{3N-3}} |\Phi(\bar{\mathbf{r}})|^2 d\mathbf{r}_2 \cdots d\mathbf{r}_N. \tag{2.15}$$

The foundation of DFT lies in the two Hohenberg–Kohn theorems [58], which relate the ground state to the electron density:

1. The density n_0, which minimizes the ground-state energy, uniquely determines the external potential $\hat{V}_{\text{ext}}(\mathbf{r})$ acting on the electronic system. The ground state Φ_0 is a one-to-one functional of the particle density $n(\mathbf{r})$.
   (Note that in this theorem, external potentials are considered equivalent if they only differ by an additive constant.)

2. The energy functional

$$E[n(\mathbf{r})] = \int v_{\text{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r} + \langle \Phi | \hat{T} + \hat{V}_{\text{el-el}} | \Phi \rangle \tag{2.16}$$

obeys a variational principle with respect to the particle density $n(\mathbf{r})$ and is minimal for the ground-state density $n_0(\mathbf{r})$:

$$E[n_0(\mathbf{r})] \leq E[n(\mathbf{r})]. \tag{2.17}$$

These two theorems restrict DFT to studies of the ground state. From the Hohenberg–Kohn theorems, it follows that the exact ground-state energy and density can be found by minimizing a universal energy functional $E[n(\mathbf{r})]$ under the constraint that $\int n(\mathbf{r}) \mathrm{d}^3\mathbf{r} = N$, so

$$\frac{\delta}{\delta n(\mathbf{r})} \left[E[n(\mathbf{r})] - \mu \left(\int n(\mathbf{r}) \mathrm{d}^3\mathbf{r} - N \right) \right] = 0 \tag{2.18}$$

for $n(\mathbf{r}) = n_0(\mathbf{r})$. The Lagrange multiplier μ can physically be interpreted as a chemical potential. However, the expression of the functional in Eq. (2.16) is unsuitable for this purpose, as terms for the kinetic energy and the electron-electron interaction are not expressed as functionals of the density. To work around this problem, Kohn and Sham proposed two approximations. First, they introduced a fictitious system of N non-interacting electrons reproducing the exact ground-state density, described by single-electron wave functions $\phi_i(\mathbf{r})$. The kinetic energy $T_{\text{s}}[n(\mathbf{r})]$ for such a non-interacting electron system is

$$T_{\text{s}}[n(\mathbf{r})] = -\frac{1}{2} \sum_{i=1}^{N} \int \phi_i^*(\mathbf{r}) \Delta \phi_i(\mathbf{r}) \mathrm{d}^3\mathbf{r}, \tag{2.19}$$

For the electron-electron interaction, one splits off the classical Coulomb interaction, called the Hartree energy

$$E_{\text{H}}[n(\mathbf{r})] = \int \frac{n(\mathbf{r}) n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \mathrm{d}^3\mathbf{r} \mathrm{d}^3\mathbf{r}', \tag{2.20}$$

so that the Kohn–Sham (KS) energy functional reads

$$E^{\text{KS}}[n(\mathbf{r})] = T_{\text{s}}[n(\mathbf{r})] + \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r})\mathrm{d}\mathbf{r} + E_{\text{H}}[n(\mathbf{r})] + E_{\text{xc}}[n(\mathbf{r})]. \tag{2.21}$$

Here, the *exchange-correlation functional* $E_{\text{xc}}[n(\mathbf{r})]$ combines the differences of the true electron-electron interaction energy with all quantum effects and the classical Hartree energy as well as errors made in the kinetic energy expression:

$$E_{\text{xc}}[n(\mathbf{r})] = (T[n(\mathbf{r})] - T_{\text{s}}[n(\mathbf{r})]) + (E_{\text{el-el}}[n(\mathbf{r})] - E_{\text{H}}[n(\mathbf{r})]), \tag{2.22}$$

where $T[n(\mathbf{r})]$ is the true kinetic energy functional, and $E_{\text{el-el}}[n(\mathbf{r})]$ is the true electron-electron interaction energy. Instead of finding the ground-state energy via variation with respect to the density Eq. (2.18), one can now perform a variation of Eq. (2.21) to find a set of equations to determine the orbitals ϕ_i such that the density $n(\mathbf{r})$ minimizes $E^{\text{KS}}[n]$. This yields a set of effective single-particle equations known as the Kohn–Sham equations [59]

$$\left\{-\frac{1}{2}\Delta + v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}[n(\mathbf{r})] + \frac{\delta E_{\text{xc}}[n(\mathbf{r})]}{\delta n(\mathbf{r})}\right\}\phi_i^{\text{KS}}(\mathbf{r}) = \varepsilon_i^{\text{KS}}\phi_i^{\text{KS}}(\mathbf{r}), \tag{2.23}$$

where ϕ_i^{KS} is the Kohn–Sham orbital with the effective Kohn–Sham Hamiltonian:

$$\hat{H}_{\text{KS}} = -\frac{1}{2}\Delta + v_{\text{KS}}[n(\mathbf{r})] = -\frac{1}{2}\Delta + v_{\text{ext}}(\mathbf{r}) + v_{\text{H}}[n(\mathbf{r})] + \frac{\delta E_{\text{xc}}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \tag{2.24}$$

Note that the $\varepsilon_i^{\text{KS}}$ are formally only Lagrangian multipliers used to introduce the constraints that $\int \phi_i^*(\mathbf{r})\phi_i(\mathbf{r})\mathrm{d}^3\mathbf{r} = 1$ in the minimization, but are often interpreted as effective single-particle energies.

Two points are noteworthy about Eq. (2.23). First, the variation of the exchange-correlation energy functional defines a multiplicative exchange-correlation potential

$$v_{\text{xc}}[n(\mathbf{r})] = \frac{\delta E_{\text{xc}}[n(\mathbf{r})]}{\delta n(\mathbf{r})}. \tag{2.25}$$

As the exact form of $E_{\text{xc}}[n(\mathbf{r})]$ is unknown, one has to resort to physically motivated approximations with varying accuracy [60, 61, 62]. Second, both $v_{\text{H}}[n(\mathbf{r})]$

and $v_{\text{xc}}[n(\mathbf{r})]$, needed to determine the ϕ_i, depend on the density, which is in turn determined from the ϕ_i, requiring a self-consistent procedure to find the solutions.

2.1.4 Marcus Rate Theory

Despite the use of DFT for the calculation of molecular energies and wavefunctions, the coupled dynamics of electrons and nuclei in charge transfer present significant computational challenges due to the disparity in their respective time scales. A potential solution to this issue involves treating the charge transfer as transition events between the localized states. This section will introduce the Marcus rate theory [63, 64] starting from two-level quantum systems and time-dependent perturbation theory.

We start out from a model of a coupled donor-acceptor complex and its adiabatic PES. Using bra-ket notation, let $|+\rangle$ and $|-\rangle$ be two eigenstates of the coupled system with energies ε_+ and ε_-, respectively. The donor-acceptor Hamiltonian $\hat{H}^{\text{DA}}$ can then be written in its adiabatic representation as

$$\hat{H}^{\text{ad}} = \varepsilon_+ \, |+\rangle \langle +| \, + \, \varepsilon_- \, |-\rangle \langle -| \, . \tag{2.26}$$

However, as these adiabatic states of the coupled system are often not known, one instead considers two states of the uncoupled system, $|\text{D}\rangle$ and $|\text{A}\rangle$, which are solutions to $\hat{H}_{\text{D}} |\text{D}\rangle = \varepsilon_{\text{D}} |\text{D}\rangle$ and $\hat{H}_{\text{A}} |\text{A}\rangle = \varepsilon_{\text{A}} |\text{A}\rangle$, respectively. With these the Hamiltonian of the two level system can be written in its diabatic form as

$$\hat{H}^{\text{dia}} = \varepsilon_{\text{D}} |\text{D}\rangle \langle \text{D}| + \varepsilon_{\text{A}} |\text{A}\rangle \langle \text{A}| + J_{\text{DA}} |\text{D}\rangle \langle \text{A}| + J_{\text{AD}} |\text{A}\rangle \langle \text{D}| \, , \tag{2.27}$$

where $J_{\text{DA}} = J_{\text{AD}}^{\star}$ is the coupling between state $|1\rangle$ and state $|2\rangle$, which also equals $J_{\text{DA}} = \langle \text{D}|\hat{H}^{\text{DA}}|\text{A}\rangle$. The energies of the coupled system can be obtained from Eq. (2.27) with the Ansatz $|\pm\rangle = c_{\text{D}}^{\pm} |\text{D}\rangle + c_{\text{A}}^{\pm} |\text{A}\rangle$, yielding

$$\varepsilon_{\pm} = \frac{1}{2} \left(\varepsilon_{\text{D}} + \varepsilon_{\text{A}} \pm \sqrt{(\varepsilon_{\text{D}} - \varepsilon_{\text{A}})^2 + 4|J_{\text{DA}}|^2} \right) . \tag{2.28}$$

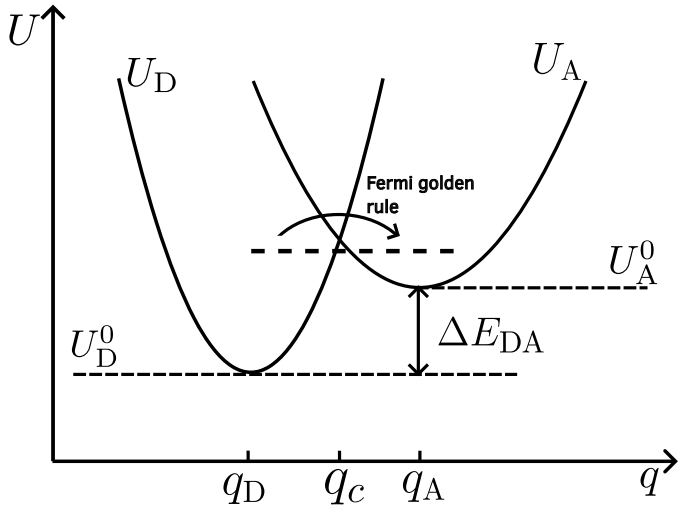

Figure 2.2: Potential energy surfaces of the donor and acceptor as a function of collective variables. The transfer rate due to tunneling near the crossing point is calculated using the Fermi golden rule.

The diabatic representation also allows invoking Fermi's Golden Rule to estimate the rate for a transition between the two diabatic states, according to:

$$\omega_{\mathrm{DA}} = \frac{2\pi}{\hbar}|J_{\mathrm{DA}}|^2\delta(\varepsilon_{\mathrm{D}} - \varepsilon_{\mathrm{A}}). \tag{2.29}$$

Figure 2.2 shows the extension of the idea of two-level system to include some (vibrational) degree of freedom q, generalizing the energies in Eq. (2.28) to potential energy surfaces, here specifically showing the two PES $U_{\mathrm{D}}(q)$ and $U_{\mathrm{A}}(q)$ of the diabatic states, respectively. The adiabatic PES are then, cf. Eq. (2.27)

$$U_{\pm}(q) = \frac{1}{2}\left(U_{\mathrm{D}}(q) + U_{\mathrm{A}}(q) \pm \sqrt{(U_{\mathrm{D}}(q) - U_{\mathrm{A}}(q))^2 + 4|J_{\mathrm{DA}}(q)|^2}\right). \tag{2.30}$$

The value q_{c} is the crossing point at which the two PES intersect. In this picture, for electron transfer to occur, the system must approach from the donor state the crossing point. The inclusion of a vibrational degree of freedom also needs to be reflected in Fermi's Golden Rule.

Marcus theory [63, 65] proceeds based on the following assumptions:

1. The charge transfer process is non-adiabatic, meaning the molecular vibrational motion is much faster than the electron's motion.

2. The temperature T is high: $k_{\mathrm{B}}T \gg \hbar\nu_{\mathrm{vib}}$, where ν_{vib} is the vibrational frequency, and k_{B} is the Boltzmann constant. This condition allows a classical description of vibrational modes and a continuous density of states.

3. Averaging is performed over a large ensemble of donor-acceptor complexes (states), corresponding to the thermal equilibrium distribution function.

The potential energy surfaces of the donor, $U_{\mathrm{D}}(q)$, and acceptor, $U_{\mathrm{A}}(q)$, can be approximated as parabolic harmonic functions of the collective variable q (in vibrational coordinates), along which charge transfer occurs:

$$U_m(q) = U_m^0 + \frac{1}{2}\nu_{\mathrm{vib}}^2(q - q_m)^2, \tag{2.31}$$

where ν_{vib} is the vibrational frequency, and $m = \mathrm{D}, \mathrm{A}$. Averaging over the Boltzmann distribution, $f(q) = Z^{-1}e^{-U_{\mathrm{D}}(q)/k_{\mathrm{B}}T}$, where the partition function $Z = \int_q e^{-U_{\mathrm{D}}(q)/k_{\mathrm{B}}T}$, yields:

$$\omega_{\mathrm{DA}} = \frac{2\pi}{\hbar}\int_q f(q)|J_{\mathrm{DA}}|^2\delta(U_{\mathrm{D}}(q) - U_{\mathrm{A}}(q))\,dq. \tag{2.32}$$

Due to the δ-function, the crossing point q_{c} must be determined to evaluate the integral. Expanding

$$\begin{aligned} U_{\mathrm{D}}(q) - U_{\mathrm{A}}(q) &= U_{\mathrm{D}}^0 - U_{\mathrm{A}}^0 + \frac{\nu_{\mathrm{vib}}}{2}\left((q - q_{\mathrm{D}})^2 - (q - q_{\mathrm{A}})^2\right) \\ &= \Delta E_{\mathrm{DA}} - \nu_{\mathrm{vib}}(q_{\mathrm{D}} - q_{\mathrm{A}})q + \frac{\nu_{\mathrm{vib}}}{2}(q_{\mathrm{D}}^2 - q_{\mathrm{A}}^2) \end{aligned} \tag{2.33}$$

where $\Delta E_{\mathrm{DA}} = U_{\mathrm{D}}^0 - U_{\mathrm{A}}^0$. The crossing point is then:

$$q_{\mathrm{c}} = \frac{\Delta E_{\mathrm{DA}} + \frac{\nu_{\mathrm{vib}}}{2}(q_{\mathrm{D}}^2 - q_{\mathrm{A}}^2)}{\nu_{\mathrm{vib}}(q_{\mathrm{D}} - q_{\mathrm{A}})}. \tag{2.34}$$

Assuming further that J_{DA} does not depend on q, the electron transfer rate becomes:

$$\omega_{\text{DA}} = \frac{2\pi}{\hbar} \frac{|J_{\text{DA}}|^2}{\sqrt{2\pi k_{\text{B}} T \nu_{\text{vib}} (q_{\text{D}} - q_{\text{A}})^2}} \exp\left(-\frac{\nu_{\text{vib}} (q_c - q_{\text{D}})^2}{2 k_{\text{B}} T} \right). \tag{2.35}$$

Defining the reorganization energy as:

$$\Lambda_{\text{DA}} = \frac{\nu_{\text{vib}}}{2} (q_{\text{A}} - q_{\text{D}})^2, \tag{2.36}$$

and substituting Λ_{DA} into equation Eq. (2.35), the Marcus rate for charge transfer from donor to acceptor is:

$$\omega_{\text{DA}} = \frac{|J_{\text{DA}}|^2}{\hbar} \sqrt{\frac{\pi}{k_{\text{B}} T \Lambda_{\text{DA}}}} \exp\left(-\frac{(\Delta E_{\text{DA}} - \Lambda_{\text{DA}})^2}{4 \Lambda_{\text{DA}} k_{\text{B}} T} \right). \tag{2.37}$$

With the transfer rate theory described in this section is the central aspect of the multiscale model of charge transport. The energies U_{D}^0 and U_{A}^0, the reorganization energy Λ_{DA}, and the electronic coupling element J_{DA} can be calculated for given chemical and structural arrangements in a material system from the quantum and classical methods introduced in this section. Having introduced the theory of the multiscale model, the next section will introduce the specific implementation of the multiscale model.

2.2 Multiscale Model: Practice

This section briefly introduces the practical aspects of the multiscale model for charge transport in organic semiconductors. As indicated in the Introduction (Chapter 1), the multiscale model maps the molecular system of an organic semiconductor into a weighted, directed graph, facilitating the computational model of charge transport. Denoting the graph as $\mathbf{G}(\mathbf{V}, \mathbf{W})$, then the set of nodes $\mathbf{V}$ represents the molecules, and the adjacency matrix $\mathbf{W}$ has its element ω_{ij} being the Marcus rate from molecule i to molecule j (note that these indices replace the D and A in the previous section). As shown in Fig. 2.3, the multiscale model consists of the following computational procedures:

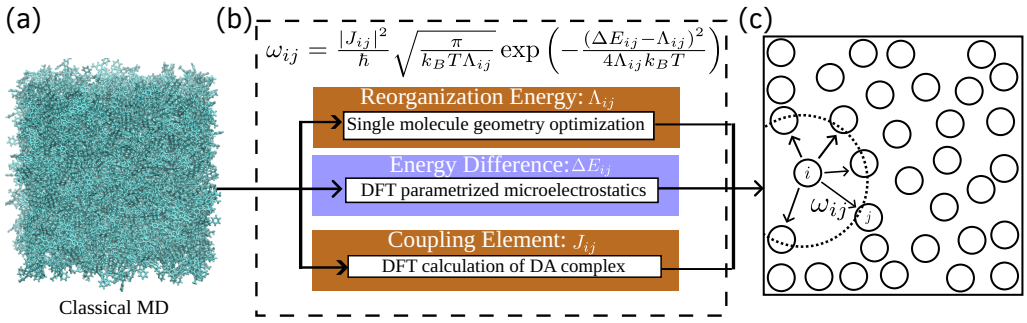

Figure 2.3: Summary of the multiscale model that maps a molecular structure into a directed graph. (a) Molecular structure generated from classical MD simulation. (b) Components and corresponding calculation scheme of the Marcus rate. (c) Output of the multiscale model, represented as a graph with nodes depicted by circles. Each node is connected to neighboring nodes whose closest-contact-distance is within a cutoff value, as indicated by the dashed circle. The charge transfer rate from molecule i to molecule j, ω_{ij}, is calculated using the methodologies of the multiscale model. For clarity, the rest of the directed edges and their corresponding weights are omitted. The arrow indicates that the multiscale model maps a molecular structure uniquely into a directed graph.

1. Classical MD simulations with empirical force fields are employed to generate the morphology of the organic semiconductor.

2. DFT calculation is performed on a single molecule to obtain the ground state electronic properties including the reorganization energy. This procedure is called the geometry optimization of a single molecule.

3. Electrostatic potential calculations based on the DFT parametrized electronic properties are performed to obtain the energy difference.

4. DFT calculation on the DA complexes to obtain the coupling elements.

5. Calculation of Marcus rates and construct the graph, where each node represents a molecule (localized state) and the edges represent the Marcus rate between the molecules. Since charge transfer only happens between the pair of molecules whose closest contact distance is close, each node is connected to other nodes within a cutoff distance.

The integration of the above methodologies enables the multiscale model to capture the complex interplay between molecular electronic structures, spatial arrangements, and morphological inhomogeneities. By mapping these factors into a graph, the multiscale model facilitates the detailed study and modeling of the charge carriers transport using the kinetic models [66, 67, 68]. Below we will focus on the computational details in determining the molecular morphology and the physical quantities $\Lambda_{ij}, \Delta E_{ij}$, and J_{ij} as they enter the Marcus rate.

2.2.1 Computational Details of Molecular Dynamics

While having different molecular building blocks, the explicit molecular materials we will use as study systems in Chapters 5 and 6 both are amorphous bulk materials. To obtain structural information with classical molecular dynamics, the identical computational procedure is used in both cases. Atomistic molecular dynamics simulations are conducted using the GROMACS software package [69], employing a gromos54a7 type force field obtained via the tool Automated Topology Builder [70]. Initially, 1000 molecules are randomly placed in a cubic cell with a side length of 10 nm. Periodic boundary conditions are applied throughout in all three spatial directions. After energy minimization, the system is simulated for 1 ns at a constant temperature of 300 K and a constant pressure of 1 bar in the NpT ensemble using the V-rescale thermostat [71] with the coupling time constant 0.1 ps and the Parrinello-Rahman barostat [72] with a time constant for pressure coupling 2 ps. The leap-frog algorithm [73] is used to integrate the motion equation Eq. (2.14) with a time step of 1 fs.

In the next phase, we employ simulated annealing to first increase the temperature to 800 K during a period of 0.5 ns, i.e., well above the glass transition temperature of the material. The system is maintained at the temperature for 1 ns before cooling back down to 300 K during a period of 0.5 ns. Such heating-cooling cycle is repeated three times. After this simulated annealing, a production run is conducted for 2 ns using the NpT ensemble. The final configuration is chosen for the rate calculation using multiscale model methodologies.

2.2.2 Computational Details of Marcus Rate

The Marcus rate depends on the physical quantities $\Lambda_{ij}, \Delta E_{ij}$, and J_{ij}, which are all related to quantum-level calculations in the framework of DFT in this thesis.

Since the molecules are in an environment consisting of many other molecules, classical mechanics methods are required to evaluate electrostatic-related forces using all the atomic charges obtained from the charge density due to DFT calculations.

Solving the Kohn–Sham equations Eq. (2.23) provides information about the effective single-electronic wave functions $\phi_l(\mathbf{r})$ and associated energies ε_l, and with that the charge density is determined as $n(\mathbf{r}) = \sum_{l=1}^{N} |\phi_l(\mathbf{r})|^2$. A self-consistent solution to the Kohn–Sham equations corresponds to finding the ground-state density ρ_0 that minimized the total energy of the system $U[n(\mathbf{r})] = E^{\text{KS}}[n(\mathbf{r})]$ as given by Eq. (2.21).

To solve Eq. (2.23) self-consistently, the Kohn-Sham orbitals $\phi_l(\mathbf{r})$ are expanded in terms of a chosen basis set $\{\varphi(\mathbf{r})\}$:

$$\phi_l(\mathbf{r}) = \sum_{\alpha} c_{l\alpha} \varphi_\alpha(\mathbf{r}) \tag{2.38}$$

where $c_{l\alpha}$ are the expansion coefficients, and $\varphi_\alpha(\mathbf{r})$ are the basis functions. Substituting the basis set expansion into the Kohn-Sham equation and projecting onto a basis function φ_β, we obtain:

$$\sum_{\alpha} \langle \varphi_\beta | \hat{H}_{\text{KS}} | \varphi_\alpha \rangle = \varepsilon_l \sum_{\alpha} c_{l\alpha} \langle \varphi_\beta | \varphi_\alpha \rangle. \tag{2.39}$$

This can be written in matrix form as: $\sum_\alpha H_{\beta\alpha} c_{l\alpha} = \varepsilon_l \sum_\alpha S_{\beta\alpha} c_{l\alpha}$, or

$$\underline{\mathbf{H}} \mathbf{c}_l = \varepsilon_l \underline{\mathbf{S}} \mathbf{c}_l, \tag{2.40}$$

where $\underline{\mathbf{H}} : H_{\beta\alpha} = \langle \varphi_\beta | \hat{H}_{\text{KS}} | \varphi_\alpha \rangle$ is the Hamiltonian matrix, and $\underline{\mathbf{S}} : S_{\beta\alpha} = \langle \varphi_\beta | \varphi_\alpha \rangle$ is the overlap matrix element.

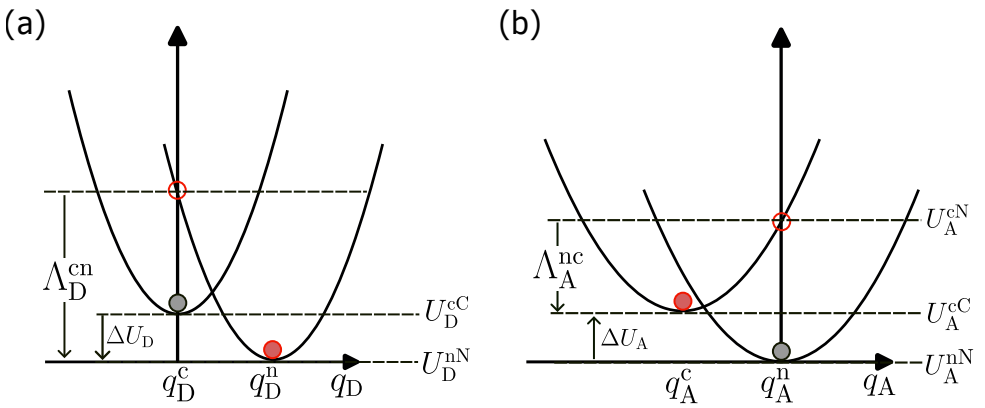

Figure 2.4: Potential energy surfaces of (a) the donor and (b) the acceptor in both charged and neutral states, denoted by the superscripts c and n. The superscripts N and C refer to the neutral and charged structures of the donor/acceptor, respectively. After the charge state is altered, both molecules relax their nuclear coordinates. If all vibrational modes are treated classically, the reorganization energy is the sum of the donor's discharge energy, $\Lambda_{\mathrm{D}}^{\mathrm{cn}}$, and the acceptor's charge energy, $\Lambda_{\mathrm{A}}^{\mathrm{nc}}$: $\Lambda_{\mathrm{DA}} = \Lambda_{\mathrm{D}}^{\mathrm{cn}} + \Lambda_{\mathrm{A}}^{\mathrm{nc}}$.

Reorganization Energy To calculate the reorganization energy Λ_{ij}, Equations (2.21) and (2.23) can be solved for different total charge states x = n,c and corresponding equilibrium geometries X = N,C, where n and N stand for "neutral" and c and C for "charged". The respective total energies will be denoted in the following as U^{xX}, dropping the explicit mention of the functional dependency on $n(\mathbf{r})$ for compactness. According to Fig. 2.4, the reorganization energy Λ_{ij} for charge transfer from molecule i to molecule j as it enters the Marcus rate equation Eq. (2.37) can be calculated as:

$$\Lambda_{ij} = U_i^{\mathrm{nC}} - U_i^{\mathrm{nN}} + U_j^{\mathrm{cN}} - U_j^{\mathrm{cC}}. \tag{2.41}$$

Site energy difference The term $\Delta E_{ij} = E_i - E_j$ entering the Marcus rate is the energy difference between two sites i and j (which refer to localized states in the material). This energy difference is crucial as it directly influences the rate of charge transfer between these sites. Each site energy E_i is defined as the difference between the total energies of the system in which molecule i is

carrying a charge and the system in which molecule i is not carrying a charge. Mathematically, this can be expressed as:

$$E_i = E_i^{\mathrm{c}} - E_i^{\mathrm{n}}, \tag{2.42}$$

where E_i^{c} represents the total energy of the system when molecule i is charged, and E_i^{n} represents the total energy when molecule i is neutral. This difference corresponds to the ionization potential in the case of hole transport and the negative of the electron affinity in the case of electron transport.

The total energy of a charged or neutral molecule, E_i^{x} (where x denotes either the charged state c or the neutral state n), is composed of several contributions from different physical mechanisms. These contributions can be written as:

$$E_i^{\mathrm{x}} = U_i^{\mathrm{xX}} + E_i^{\mathrm{x,el}} + E_i^{\mathrm{x,polar}}, \tag{2.43}$$

where each term represents a specific energy component:

- U_i^{xX} is the intramolecular contribution from single molecules. This term includes the energy associated with the electronic structure of the molecule and changes due to the addition or removal of a charge.
- $E_i^{\mathrm{x,static}}$ is the contribution from static intermolecular interactions. These interactions arise due to the fixed positions and orientations of neighboring molecules.
- $E_i^{\mathrm{x,polar}}$ is the contribution from polarization effects. When a molecule is charged, it induces a polarization in the surrounding medium, which in turn affects the energy of the charged molecule. This term accounts for the energy change due to this polarization.

These detailed components of E_i^{x} is essential for accurately modeling charge transport in organic semiconductors. The intramolecular contribution U_i^{xX} can be calculated using Eq. (2.21), while the static and polarization contributions often require classical mechanics methods which consider the contributions due to inhomogeneous molecular environments such as packing, orientation, and correlation.

In the multiscale model here, we employ a point charge representation [74] for the electrostatic potential of charged and neutral molecules. Specifically, the atomic charge is derived from the DFT potential energy surfaces according to the CHELPG (Charges from electrostatic potentials using a grid-based method) program. The main goal of the CHELPG method is to determine atomic charges, q_i, that best reproduce the electrostatic potential (ESP) calculated from the molecular wavefunction. This is done by minimizing the difference between the calculated ESP and the ESP generated by the fitted charges at a set of grid points around the molecule. Using the determined atomic charges, the electrostatic energy contribution is

$$E_i^{\mathrm{x,el}} = \frac{1}{4\pi\epsilon_0} \sum_{a_i} \sum_{b_k, k\neq i} \frac{q_{a_i}^{\mathrm{x}} q_{b_k}^{\mathrm{n}}}{|\mathbf{R}_{a_i} - \mathbf{R}_{b_k}|} \tag{2.44}$$

where ϵ_0 is the vacuum permittivity, a_i, b_k denotes the atoms in molecule i, k, $q_{a_i}^{\mathrm{x}}$ are the partial charge of atom a when molecule i is in state x.

To account for the effects of polarization, and to evaluate $E_i^{\mathrm{x,polar}}$, we use the model of distributed atomic dipole polarizabilities (Thole model) [75]. The details of this method are as follows. Firstly, evaluate the electric field at atom a of molecule i, $\vec{F}_{a_i}^{(0)}$ created by all atomic partial charges, then the induced dipole moment $\vec{\mu}_{a_i}^{(0)}$ are computed. In this first step, intramolecular interactions are excluded. Secondly, induced dipole moments are iteratively refined as $\vec{\mu}_{a_i}^{(k+1)} = w\vec{F}_{a_i}^{(k)}\alpha_{a_i} + (1-w)\vec{\mu}_{a_i}^{(k)}$. Here α_{a_i} is the isotropic atomic polarizability and w is a damping constant ($w = 0.5$ is used in this work) for successive relaxation. The following electric fields are then computed with the induced dipole moment which now includes the intra-molecular interactions. The process is repeated until $\sum_{a_i} |\vec{\mu}_{a_i}^{(k+1)} - \vec{\mu}_{a_i}^{(k)}| < 10^{-6}\mathrm{Debye}$. In conclusion, $E_i^{\mathrm{x,el}} + E_i^{\mathrm{x,polar}}$ is a function of all atomic partial charges $q_{a_i}^x$ and isotropic atomic polarizability α_{a_i}. For a given molecular structure, the sum of the electrostatic energy and the polarization energy, $E_i^{\mathrm{x,el}} + E_i^{\mathrm{x,polar}}$, is uniquely determined, so as E_i. Practical calculations of the site energies are performed using the VOTCA software [39, 76, 77, 78].

Electronic Coupling Element The coupling element between two molecules i and j describes the coupling between two localized states, approximated by

monomer single-particle wavefunctions $|\phi_i\rangle$ and $|\phi_j\rangle$. Throughout the thesis, the frontier orbital of interest is the HOMO because the charge carriers are holes. The frontier orbitals of a dimer, that is the adiabatic energy surfaces, result from the interaction of the frontier orbitals of monomers and can be expanded in terms of $|\phi_i\rangle$ and $|\phi_j\rangle$. The expansion coefficients $\mathbf{C}$ can be determined by solving the generalized secular equation

$$(\underline{\mathbf{H}} - E\underline{\mathbf{S}})\mathbf{C} = 0 \tag{2.45}$$

where the Hamiltonian matrices $\underline{\mathbf{H}}$ and the overlap matrices $\underline{\mathbf{S}}$ are written as:

$$\begin{aligned} \underline{\mathbf{H}} &= \begin{pmatrix} e_i & J_{ij}^0 \\ J_{ij}^0 & e_j \end{pmatrix} \\ \underline{\mathbf{S}} &= \begin{pmatrix} 1 & S_{ij} \\ S_{ij} & 1 \end{pmatrix} \end{aligned} \tag{2.46}$$

with $e_i = \langle\phi_i|\hat{H}_{\text{KS}}^{\text{D}}|\phi_i\rangle$, $e_j = \langle\phi_j|\hat{H}_{\text{KS}}^{\text{D}}|\phi_j\rangle$, $J_{ij}^0 = \langle\phi_i|\hat{H}_{\text{KS}}^{\text{D}}|\phi_j\rangle$ and $S_{ij} = \langle\phi_i|\phi_j\rangle$ with bra-ket notation. The Hamiltonian of the dimer $\hat{H}_{\text{KS}}^{\text{D}}$ (see Eq. (2.24)) is diagonal in its eigenbasis $\{|\phi_k^D\rangle\}$ with eigenvalues $\{\varepsilon_k^D\}$, so $\hat{H}_{\text{KS}}^{\text{D}} = \text{diag}(\varepsilon^D)$. It is noteworthy that the monomer orbitals ϕ_i and ϕ_j are not orthonormal in general, so that $\underline{\mathbf{S}} \neq \underline{\mathbf{1}}$.

To facilitate the computations, it is necessary to transform Eq. (2.45) into a standard eigenvalue problem of the form $\underline{\mathbf{H}}^{\text{eff}}\mathbf{C} = E\mathbf{C}$. Such a transformation [79, 80, 81] is achieved by:

$$\underline{\mathbf{H}}^{\text{eff}} = \underline{\mathbf{S}}^{-1/2}\underline{\mathbf{H}}\underline{\mathbf{S}}^{-1/2} \tag{2.47}$$

where $\underline{\mathbf{S}}^{-1/2}$ can be evaluated [81] as:

$$\underline{\mathbf{S}}^{-1/2} = \underline{\mathbf{U}} \begin{pmatrix} s_1^{-\frac{1}{2}} & & 0 \\ & \ddots & \\ 0 & & s_N^{-\frac{1}{2}} \end{pmatrix} \underline{\mathbf{U}}^{\text{T}} \tag{2.48}$$

where the columns of the matrix $\underline{\mathbf{U}}$ contain the N eigenvectors of the matrix $\underline{\mathbf{S}}$, and $s_1, \ldots, s_N$ are the associated eigenvalues. Applying this decomposition to Eq. (2.45) yields an effective Hamiltonian matrix $\underline{\mathbf{H}}^{\text{eff}}$ in an orthonormalized basis, and its entries can be directly identified with the coupling elements:

$$\underline{\mathbf{H}}^{\text{eff}} = \begin{pmatrix} \varepsilon_i & J_{ij} \\ J_{ij} & \varepsilon_j \end{pmatrix}. \tag{2.49}$$

Applying Eq. (2.47) to the elements of $\underline{\mathbf{H}}^{\text{eff}}$ yields the formula for the coupling elements:

$$J_{ij} = \frac{J_{ij}^0 - \frac{1}{2}(e_i + e_j)S_{ij}}{1 - S_{ij}^2}. \tag{2.50}$$

Until this point, the terms in Eq. (2.50) can be calculated by projecting the monomer functions onto the dimer eigenbasis (This method is called Dimer-Projection Method [82]). Using this method, the vector $\mathbf{p}_i : p_{ik}$ has entries $p_{ik} = \langle \phi_i | \phi_k^{\text{D}} \rangle$ and $\mathbf{p}_j : p_{jk}$ has entries $p_{jk} = \langle \phi_j | \phi_k^{\text{D}} \rangle$, J_{ij}^0 can be calculated as $J_{ij}^0 = \mathbf{p}_i^{\text{T}} \text{diag}(\varepsilon^{\text{D}}) \mathbf{p}_j$. Similarly, $e_{i(j)} = \mathbf{p}_{i(j)}^{\text{T}} \mathbf{p}_{i(j)}$ and $S_{ij} = \mathbf{p}_i^{\text{T}} \mathbf{p}_j$. All of these operations are performed in the basis set representation of the Kohn–Sham wave functions as implemented in VOTCA [76, 78].

Chapter 3

Methodology: Random Walk on Graph

Adapted from the paper "*A Graph Random Walk Method for Calculating Time-of-Flight Charge Mobility in Organic Semiconductors from Multiscale Simulations*, Z. Chen, P. van der Hoorn, B. Baumeier, arXiv preprint 2405.15836 (2025)".

Random walks in random environments are often used to model physical processes [83], where the random walker represents the motion of a particle and the random environment is a model for a system that is characterized by some type of disorder. Charge transport processes in organic semiconductors are examples of random walks in an environment that is random in the sense that the electronic structures/Marcus rates have various degrees of disorder. The dynamics of charge carriers in organic semiconductors are influenced by myriad factors, including the disordered nature of the molecular structure and the dynamic environment in which they operate. A comprehensive understanding of the dynamics necessitates a robust theoretical framework, one that can capture the stochastic nature of charge movement. The multiscale model maps the organic semiconductor to a graph, where each node corresponds to a molecular site, and the edges denote possible transitions between the sites. This representation naturally lends itself to a Markov chain framework, where the charge carriers perform exclusive random walks on the graph due to the Pauli exclusion. The transition probability from one node to another depends essentially on the energy landscape and spatial configuration of all the molecules. By analyzing these random walks, one can gain

insights into the charge transport properties and predict the dynamics of charge carriers over time.

The computation of charge mobility necessitates the solution of the master equation (MEq), a mathematical framework that describes the temporal evolution of the probability distribution of the charge carrier's position. Solving the MEq provides the probability distribution of locating the charge carrier at a specific node at any given time. However, the multiscale characteristics of charge dynamics introduce significant stability challenges in the solution of the MEq.

As an alternative, the KMC method approximates the solution of MEq and generates trajectories of charge carriers' random walks on a graph by probabilistically selecting transitions between nodes. This method captures the stochastic nature of the process and provides detailed insights into the system's temporal evolution. However, KMC simulations can be computationally expensive, particularly for large systems with low charge carrier concentrations. Moreover, a large number of simulations are often required to obtain accurate estimates of charge mobility.

Starting from this point, we propose the ToF (defined in the Chapter 1) evaluation based on the expected first hitting time of continuous-time random walks, and we call this method **GRW**. With this background, the theory and practice of random walks on graph will be introduced in this chapter. It begins with introducing Markov chain theory for the description of random walks on graph. Subsequently, the MEq method for solving the Markov chain is introduced followed by the KMC method for simulating the random walk on graph and the GRW method, which is based on the matrix formula for ToF calculation. Finally, the practice of random walk on a graph will be introduced, including the specification of the Markov chain, methods of solving the MEq, KMC algorithm, and the implementation of GRW method for evaluation of the ToF charge mobility.

3.1 Random Walk on Graph: Theory

The theory used for describing the random walks on graph is the Markov chain theory on a discrete, finite state space. Markov chains serve as a fundamental model for a wide range of phenomena featuring in various scientific fields. A Markov chain is a stochastic process $\{\xi_t\}_{t\geq 0}$ that satisfies the Markov property, meaning that the future state of the process depends only on the present state and not on the past states. Depending on the time t, the Markov chain is discrete-time Markov chain (DTMC) when $t \in \mathbb{N}$, and continuous-time Markov chain (CTMC) when $t \in \mathbb{R}$. This work will focus on the CTMC.

3.1.1 Continuous Time Markov Chain

The definition of a CTMC is related to a generator matrix, which can be defined as:

Definition 3.1.1. *Let S be a finite and countable set. A generator matrix $\underline{\mathbf{\Omega}} = (\Omega_{ij} : i, j \in S)$ is a matrix satisfying the following condition:*

- $0 \geq \Omega_{ii} > -\infty$ *for all i;*
- $\Omega_{ij} \geq 0$ *for all $i \neq j$;*
- $\sum_{j\in S} \Omega_{ij} = 0$ *for all i.*

For any generator matrix $\underline{\mathbf{\Omega}}$ the series $\sum_{k=0}^{\infty} \underline{\mathbf{\Omega}}^k/k!$ converges componentwise and this limit is denoted by $e^{\underline{\mathbf{\Omega}}}$. Related to the generator matrix, a CTMC is defined as follows (as a reference, see [84] Theorem 2.8.2):

Definition 3.1.2. *Let $\{\xi_t\}_{t\geq 0}$ be a continuous time-dependent stochastic process taking the values in a finite state space S, and denote the transition probabilities, the probability from state i to state j in a time period t as $p_{ij}(t) := p(\xi_t = j|\xi_0 = i)$. Then ξ_t is a* ***continuous-time Markov chain*** *if there exists a generator matrix $\underline{\mathbf{\Omega}}$ such that: The transition probabilities $p_{ij}(t)$ can be calculated from the generator*

matrix $\underline{\mathbf{\Omega}}$ *using the matrix exponential formula:*

$$p_{ij}(t) = [e^{\underline{\mathbf{\Omega}}t}]_{ij} \tag{3.1}$$

where $e^{\underline{\mathbf{\Omega}}t}$ *is the matrix after the exponential of matrix* $\underline{\mathbf{\Omega}}t$*, and the subscript* i, j *means the* i*th row and* j*th column of the matrix* $e^{\underline{\mathbf{\Omega}}t}$*.*

A CTMC has the **memoryless** property, that is, the probability of finding the future state j, $p(\xi_{t_{n+1}} = j)$ depends only upon the present state at time t_n, and does not depend on the past. That is,

$$p(\xi_{t_{n+1}} = j \,|\, \xi_{t_1} = i_1, \xi_{t_2} = i_2, \ldots, \xi_{t_n} = i_n) = p(\xi_{t_{n+1}} = j \,|\, \xi_{t_n} = i_n) \tag{3.2}$$

for all $j, i_1, i_2 \ldots i_n \in S$ and any sequence of time $t_1 < t_2 < \cdots < t_n < t_{n+1}$.

A Markov chain models the time evolution of the system states according to the transition probability among the states. Time characteristics are central in understanding Markov chains, including hitting time [85] which is the time a specified state is reached starting from a given initial state, cover time [86] which is the expected time it takes for the process to visit every state at least once, and mixing time [87, 88] which is the time it takes for the chain to become close to its stationary distribution, etc. In the Source-Sink setting, some states are *Source* states where the charge dynamics begins and some states are *Sink* states where the charge dynamics stops. The ToF from the Source state to the Sink state is related to the hitting time of a CTMC. The expected first hitting times are essential for characterizing the dynamic behavior of the system, as they provide insights into how quickly or slowly certain states are attained [89, 90, 85]. The first hitting time to the Sink states of a CTMC is a random variable written as:

$$\inf\{t \geq 0 : \xi_t \in \text{Sink}\},$$

Starting from state i, that is, $\xi_{t_1=0} = i$, then the expected first hitting time to Sink state is

$$\tau_i = \mathbb{E}[\inf\{t \geq 0 : \xi_t \in \text{Sink}\} | \xi_0 = i\,], \tag{3.3}$$

where $i \in S$ is the initial state. The quantity τ_i is related to the time evolution of the state's occupation probability and is typically governed by the MEq, whose theory will be introduced in the next section. The process of hitting the Sink states from individual Source states is reminiscent of a parallel electric network of capacitors [91], and accordingly, we evaluate the ToF as:

$$\tau = N_{\text{Source}} \left[\sum_{i \in \text{Source}} (\tau_i)^{-1} \right]^{-1}, \tag{3.4}$$

with N_{Source} the number of source states, i.e., as the harmonic mean of the respective τ_i.

3.1.2 Master Equation

To model the charge transport dynamics, the position of the charge carriers as a function of time must be obtained. That is, one needs to obtain the probability of finding the state i at time t: $p_i(t), i \in S$. This can be achieved by leveraging the definition of the CTMC, wherein the transition probability $p_{ij}(t)$ from state i to state j in time t is derived from the generator matrix $\underline{\mathbf{\Omega}}$. The entries of $\underline{\mathbf{\Omega}}$ represent the charge transfer rates, which govern the dynamics of the system.

Denote the transition probability matrix by $\underline{\mathbf{P}}(t) = (p_{ij}(t) : i, j \in S)$. According to Definition 3.1.2, the transition probability $p_{ij}(t)$ can be expressed as $[e^{\underline{\mathbf{\Omega}} t}]_{ij}$, where $\underline{\mathbf{P}}(t) = e^{\underline{\mathbf{\Omega}} t}$. Taking the component-wise derivative with respect to t yields the forward equation:

$$\frac{d\underline{\mathbf{P}}(t)}{dt} = \underline{\mathbf{\Omega}}\, \underline{\mathbf{P}}(t) \tag{3.5}$$

Denote the vector $\underline{\mathbf{p}}(t) : p_i(t)$ whose ith entry is the probability of finding the state i at time t, this probability vector evolves according to: $\underline{\mathbf{p}}(t) = \underline{\mathbf{P}}(t)\mathbf{p}(0)$, where $\mathbf{p}(0)$ is determined by the Markov chain initial state $\xi_0 = i$, that is, $p_i(0) = 1$ and $p_j(0) = 0$ for all $j \neq i$. To derive the master equation, differentiate $\mathbf{p}(t)$ with

respect to t yields the master equation:

$$\begin{aligned}\frac{d\mathbf{p}(t)}{dt} &= \frac{d}{dt}(\underline{\mathbf{P}}(t)\,\mathbf{p}(0)) \\ &= \underline{\mathbf{\Omega}}\,\underline{\mathbf{P}}(t)\mathbf{p}(0) \\ &= \underline{\mathbf{\Omega}}\,\mathbf{p}(t)\end{aligned} \tag{3.6}$$

The solution of the MEq Eq. (3.6) depends on both the boundary conditions and initial value $\mathbf{p}(0)$. In the ToF setting for calculating charge mobility, we consider absorbing boundary conditions: certain states are designated as *Source*, representing the initial states at time $t = 0$, and others as *Sink*, where the random walk terminates upon reaching these states. The CTMC always starts from a specific state $i \notin$ Sink, so the initial value is: $p_i(0) = 1$ and $p_{j\neq i}(0) = 0$.

Under this Source-Sink setting, one is interested in the expected first hitting time. The expected first hitting time in Eq. (3.3) can be seen as the average time that the walker survives before entering the Sink state, since the dynamics stops once the Markov chain reaches one of the absorbing states and stays there forever. In such case, $p_{ki}(t) = 0$ for all $t \geq 0$ and $k \in$ Sink. Solving Eq. (3.6) gives the time evolution of $\mathbf{p}(t)$, and the expected first hitting time with the initial value $p_i(0) = 1$ can by calculated as (as a reference, see [92] Chapter VI Eq. (7.5)):

$$\tau_i = \sum_{j\in\text{Sink}} \int_0^\infty t\,\frac{dp_j(t)}{dt}dt, \tag{3.7}$$

where the subscript i denotes the initial value $p_i(0) = 1$, and the right hand side of Eq. (3.7) depends on $p_i(0) = 1$ due to the solution of $p_j(t)$ depending on $\mathbf{p}(0)$. Analytical methods can solve Eq. (3.6) and further calculate τ_i via Eq. (3.7). To do so, one can use the eigenvalue and eigenvector of $\underline{\mathbf{\Omega}}$.

Denote the eigenvalues of $\underline{\mathbf{\Omega}}$ as $\lambda_1, \lambda_2, \ldots, \lambda_n \leq 0$, and the corresponding eigenvectors as $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n \in \mathbb{R}^n$, where n is the number of states. When the eigenvalues are real, the matrix $\underline{\mathbf{\Omega}}$ is diagonalizable [93], $\underline{\mathbf{\Omega}} = \underline{\mathbf{U}}\,\underline{\mathbf{\Lambda}}\,\underline{\mathbf{U}}^{-1}$ where $\underline{\mathbf{\Lambda}} \in \mathbb{R}^{n\times n}$ is the diagonal matrix consisting of the eigenvalues λ_i, and $\underline{\mathbf{U}} = [\mathbf{u}_1\ \mathbf{u}_2\ \cdots\ \mathbf{u}_n]$. With this, Eq. (3.6) can be written as a system of decoupled

linear ODE:

$$\frac{d(\underline{\mathbf{U}}^{-1}\,\mathbf{p}(t))}{dt} = \underline{\mathbf{\Lambda}}(\underline{\mathbf{U}}^{-1}\mathbf{p}(t)), \tag{3.8}$$

which has a unique solution for the vector $\mathbf{p}(t)$:

$$\underline{\mathbf{U}}^{-1}\mathbf{p}(t) = [c_1 e^{\lambda_1 t} \ \ldots \ c_n e^{\lambda_n t}]^{\mathrm{T}},$$

where $c_k, k = 1, \ldots, n$ are coefficients determined by the initial value: $c_k = \sum_{j=1}^{n} [\,\underline{\mathbf{U}}^{-1}\,]_{kj}\, p_j(0)$. So

$$\mathbf{p}(t) = \sum_{k=1}^{n} c_k e^{\lambda_k t}\mathbf{u}_k, \tag{3.9}$$

where k loops over all states. Using this eigenvector representation of $\mathbf{p}(t)$ the expected first hitting time with initial value $p_i(0) = 1$ becomes:

$$\begin{aligned} \tau_i &= \int_0^{\infty} \frac{d}{dt}\Big(\sum_{j\in\text{Sink}} \Big[\sum_{k=1}^{n} c_k\, e^{\lambda_k t}\mathbf{u}_k\Big]_j\Big)tdt \\ &= \int_0^{\infty} \sum_{j\in\text{Sink}} \Big[\sum_{k=1}^{n} c_k\lambda_k\, e^{\lambda_k t}\mathbf{u}_k\Big]_j tdt. \end{aligned} \tag{3.10}$$

Here $\sum_{k=1}^{n} c_k\, e^{\lambda_k t}\mathbf{u}_k$ is a vector, and the subscript notation $\Big[\sum_{k=1}^{n} c_k\, e^{\lambda_k t}\mathbf{u}_k\Big]_j$ denotes the jth component of the vector. Using $\sum_{k=1}^{n} c_k\lambda_k\, e^{\lambda_k t} = \sum_{k=1,\lambda_k\neq 0}^{n} c_k\lambda_k\, e^{\lambda_k t}$, and integration by part, Eq. (3.10) can be formulated as:

$$\begin{aligned} \tau_i &= \int_0^{\infty} \sum_{j\in\text{Sink}} \Big[\sum_{k=1,\lambda_k\neq 0}^{n} c_k\lambda_k\, e^{\lambda_k t}\mathbf{u}_k\Big]_j tdt \\ &= \sum_{j\in\text{Sink}} \Big[\sum_{k=1,\lambda_k\neq 0}^{n} c_k\mathbf{u}_k\, e^{\lambda_k t}\Big]_j t\Big|_0^{\infty} - \sum_{j\in\text{Sink}} \Big[\int_0^{\infty} \sum_{k=1,\lambda_k\neq 0}^{n} c_k\mathbf{u}_k e^{\lambda_k t}dt\Big]_j \\ &= \sum_{j\in\text{Sink}} \Big[\sum_{k=1,\lambda_k\neq 0}^{n} \frac{c_k}{\lambda_k}\mathbf{u}_k\Big]_j \end{aligned} \tag{3.11}$$

where the second line use the fact that $\lambda_k\, e^{\lambda_k t}$ has a primitive function $e^{\lambda_k t}$. The third line is the integration by part using the fact that $\lambda_k \leq 0$. While the MEq theory provides a deterministic framework for modeling random walks on graphs by describing the time evolution of occupation probabilities, the above analytical method and direct numerical solution introduced in the later section can become computationally prohibitive for large systems due to its complexity. To overcome this challenge, the KMC theory introduced in the next section offers a stochastic alternative by explicitly simulating individual charge carrier trajectories.

3.1.3 Kinetic Monte Carlo

Instead of solving the full set of coupled differential equations in MEq theory, KMC samples transition events according to the underlying rate matrix, dynamically propagating the system in discrete steps. This event-driven approach enables efficient modeling of charge transport while preserving the statistical properties described by the MEq.

The KMC method approximates the solution of the MEq by simulating the jumping process of the charge carriers using Monte Carlo algorithm [94, 95]. The elementary step of KMC is that at time t_n, node i becomes occupied by a charge carrier, at a later time t_{n+1}, the node i is unoccupied and the carrier jumps to a different node $j \neq i$. The period $T_i = t_{n+1} - t_n$ during which the node i changes from being occupied by the charge carrier to being unoccupied is called the **sojourn time** of node i. The time T_i is exponentially distributed [96], and the expected sojourn time is $\mathbb{E}[T_i] = \frac{1}{\sum_j \Omega_{ij}}$ where $\sum_j \Omega_{ij}$ is the sum of all the possible transition rates out of node i. So to simulate the jumping processes, whenever a hopping event is realized, the time should be updated with an increment Δt selected from an exponential distribution. Suppose the at time $t = 0$, the dynamics starts at state i. After many jumping steps, the first hitting time t_i^{f} is recorded if one of the Sink node is occupied by charge carrier. One needs to perform many KMC simulations to obtain multiple t_i^{f} as random variables, and the expected first hitting time from state i is calculated as the average of those iid random variables.

The KMC method operates on a simple principle, but its accuracy depends on fulfilling some key criteria [96]:

1. The KMC should run long enough such that when the system is in a steady-state setting, the detailed balance is satisfied. Specifically, the relationship $\tilde{\pi}_i \Omega_{ji} = \tilde{\pi}_j \Omega_{ij}$ must hold for all states i and j, where $\tilde{\pi}_i$ is the occupation probability of state i obtained from the Monte Carlo samples.

2. The independence of various hopping events must be effectively ensured.

Those criteria require that a sufficient number of KMC trajectories is needed to achieve convergence in the ToF calculations. Although KMC solutions are inherently approximate and lack the exactness and elegance of analytical approaches like the MEq, they correctly implement the underlying statistical principles that define the Markov chain. In KMC theory, the number of Monte Carlo samples to achieve convergence is usually unknown, and inefficient samples usually lead to deviated results. Instead of simulating the charge transport trajectories to obtain the charge mobility, the ToF can be calculated from the hitting time theory of CTMC with matrix equations.

3.1.4 ToF Calculation with Matrix Formula

The GRW method evaluates the expected first hitting time τ_i via matrix calculation and further calculates ToF τ using Eq. (3.4) Let $\boldsymbol{\tau} : \tau_i \in \mathbb{R}_+^n$ be the vector of expected first hitting times of a CTMC: $\{\xi_t\}$ with generator matrix $\underline{\mathbf{\Omega}}$. The vector element τ_i denotes the expected first hitting time to sink states starting with initial value $p_i(0) = 1$ for $i \in S$, as in Eq. (3.3). Denote $\mathcal{P}_{ij} = \Omega_{ij}/D_i$ where $D_i := \sum_{j \neq i} \Omega_{ij}$. It is noteworthy that $\mathcal{P}_{ij}$ is time-independent, and is different from the $p_{ij}(t)$ used in the Markov chain Definition 3.1.2. The time-dependent probability $p_{ij}(t)$ represents the probability of the transitioning from state i to state j in a time interval t while $\mathcal{P}_{ij}$ is the ratio between Ω_{ij} and the total rate out of state i.

With those settings, the vector τ_i can be calculated as the minimal non-negative solution to the following recursive equation (see, for example, Theorem 3.3.3 in [84]):

$$\tau_i = \begin{cases} \frac{1}{D_i} + \sum_{j \neq i} \mathcal{P}_{ij}\, \tau_j & \text{if } i \text{ is not a sink state,} \\ 0 & \text{else.} \end{cases} \tag{3.12}$$

Equation (3.12) can be written in matrix form by considering the transition matrix $\underline{\boldsymbol{\mathcal{P}}} \in [0,1]^{n\times n}$, whose entry in ith row and jth column is:

$$\begin{cases} 0 & \text{if } j \neq j \text{ and } i \text{ is a sink state,} \\ 1 & \text{if } i = j \text{ is a sink state,} \\ \mathcal{P}_{ij} & \text{else.} \end{cases} \tag{3.13}$$

By separating the sink states and the non-absorbing states, the matrix $\underline{\boldsymbol{\mathcal{P}}}$ can be rearranged into a block matrix:

$$\underline{\boldsymbol{\mathcal{P}}} = \begin{pmatrix} \underline{\boldsymbol{\mathcal{P}}}_1 & \underline{\boldsymbol{\mathcal{P}}}_2 \\ \mathbf{0} & \underline{\mathbf{I}} \end{pmatrix} \tag{3.14}$$

Here, $\underline{\mathbf{I}}$ is the identity matrix, $\underline{\boldsymbol{\mathcal{P}}}_1$ is the submatrix restricted to the non-sink states, and $\underline{\boldsymbol{\mathcal{P}}}_2$ is the submatrix of the probability of transition from non-sink vertices to sink vertices. The Markov chain under study is irreducible since there is a path, a sequence of transitions, from any state to any other state with positive probability. This ensures that the matrix $\underline{\boldsymbol{\mathcal{P}}}$ is irreducible, In this case, the Perron-Frobenius theorem guarantees a unique largest eigenvalue with a strictly positive eigenvector [97]. Since the row sum of matrix $\underline{\boldsymbol{\mathcal{P}}}$ is 1, its largest eigenvalue is 1. The matrix $(\underline{\mathbf{I}} - \underline{\boldsymbol{\mathcal{P}}})$ is not invertible, but $(\underline{\mathbf{I}} - \underline{\boldsymbol{\mathcal{P}}}_1)$ is invertible.

Now, let $\boldsymbol{\tau}^*$ be the part of the expected first hitting time vector for non-sink states. Equation 3.12 can be written as $\tau_i - \sum_{j\neq i} \mathcal{P}_{ij}\tau_j = \frac{1}{D_i}$ for non-sink state i, its a matrix form yields:

$$(\underline{\mathbf{I}} - \underline{\boldsymbol{\mathcal{P}}}_1)\boldsymbol{\tau}^* = \boldsymbol{\tau}^0, \tag{3.15}$$

where the ith element of $\boldsymbol{\tau}^0$ is $\frac{1}{D_i}$. The ToF can be obtained by substituting the vector $\boldsymbol{\tau}^*$ into Eq. (3.4). Based on the above theoretical background of CTMC, MEq and KMC, the next section will introduce the specification of the system setting: state space and the generator matrix.

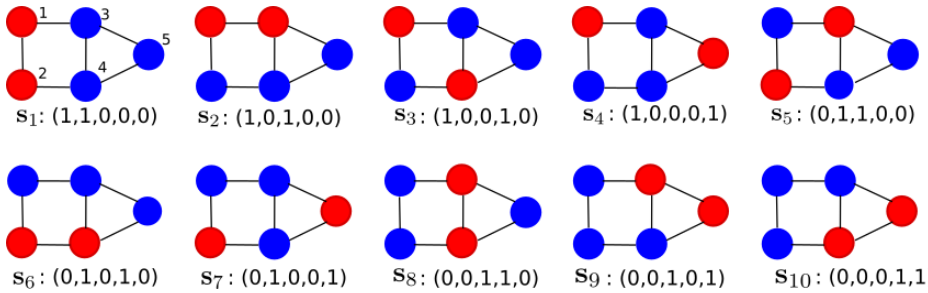

Figure 3.1: The ten states of a connected 5-site system with blue circles representing unoccupied sites and red circles for occupied sites. The numbers on the top left figure give the site indices. The state notations are the brackets containing 0 and 1 below each state plot. If site 5 is the Sink, then the Sink states are $\mathbf{s}_4, \mathbf{s}_7, \mathbf{s}_9, \mathbf{s}_{10}$.

3.2 State Space and Generator Matrix

Numerically solving the MEq with Source-Sink setting, implementation of KMC method and ToF calculation with matrix formula require the definition of the state space, transition rates, and the generator matrix.

As mentioned in Chapter 2, the molecular system is mapped into a graph denoted as $\mathbf{G}(\mathbf{V}, \mathbf{W})$, where the vertices are the sites $\mathbf{V} = \{1, 2, \ldots, n\}$ and the rate matrix $\mathbf{W}$ reads $\mathbf{W} = (\omega_{ij} : i, j \in \mathbf{V})$. When there is one charge carrier, the state space consists of the carrier position at one of the n nodes. The generator matrix $\underline{\mathbf{\Omega}} : \Omega_{ij}$ is related to the transition rate via:

$$\Omega_{ij} = \begin{cases} \omega_{ij} & \text{for } i \neq j \\ -\sum_{k \neq i}^{n} \omega_{ik} & \text{for } i = j \end{cases} \tag{3.16}$$

That is, the off-diagonal elements are the charge transfer rates, and the diagonal consists of the negative sum of the transfer rates.

When there are multiple charge carriers, the charge transport dynamics refers to a number N_c of charge carriers of equal sign moving through the molecular system. The Pauli exclusion implies that each site can only be occupied by no more than one charge carrier. This leads to dependencies between the dynamics

of every individual charge carrier. In this situation, the full dynamics are modeled as a continuous-time random walk on a larger graph $\mathbf{G}_S(\mathbf{V}_S, \mathbf{W}_S)$, where each node represents a state describing the position of all carriers in the material. The state space of the CTMC, that is, the vertex set $\mathbf{V}_S$ is the countable ordered set $\mathbf{V}_S = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{\binom{n}{N_c}}\}$. Each vertex is a configuration of charge carriers on the sites of the system. That is, a vertex $\mathbf{s}$ is a set $\mathbf{s} = (s_1, s_2, \ldots, s_n)$, whose element s_i is the occupancy of site i:

$$s_i = \begin{cases} 1 & \text{if } i \text{ is Occupied} \\ 0 & \text{if Unoccupied} \end{cases} \tag{3.17}$$

Figure 3.1 visualizes the 10 possible states for a 5-site system with two charge carriers. In the ToF setting for charge mobility calculation, some sites are designated as *Sources* corresponding to the molecules where charges are injected. Some sites are designated as *Sinks* where the carriers are detected and the charge dynamics will terminate if one of the Sinks is occupied. A state $\mathbf{s}$ is called a *source state* if all the carrier-occupied sites are source sites. That is, if $s_i = 1$, i has to be the source site.

Now we need the connections between the states to obtain the generator matrix. In our model, there is a connection from state $\mathbf{s}$ to $\mathbf{s}'$ when starting from the former configuration, the latter is obtained by one charge carrier moving over an edge in the graph $\mathbf{G}$ representing the material. Based on the above reasoning, we define the arcs between the states. There is an arc $(\mathbf{s}, \mathbf{s}')$ if and only if:

1. $s_k = s'_k = 1$ holds for exactly $N_c - 1$ indices k, and
2. there exists two indices i, j such that $\omega_{ij} > 0$, $s_i = 1$ and $s'_i = 0$, while $s_j = 0$ and $s'_j = 1$

In this case, we say that $\mathbf{s}$ is *connected* to $\mathbf{s}'$. While if any of these conditions fail, we say that $\mathbf{s}$ is *not connected* to $\mathbf{s}'$. For example, in the system displayed in Figure 3.1, state $\mathbf{s}_1$ is connected to $\mathbf{s}_3$ and $\mathbf{s}_5$, but is not connected to $\mathbf{s}_2$ and $\mathbf{s}_4$. For every pair of connected states $(\mathbf{s}, \mathbf{s}')$, there exist unique indices i and j such that $s_i = 1$ and $s'_i = 1 - s_i$, as well as $s'_j = 0$ and $s_j = 1 - s'_j$. In this case, we say that $\mathbf{s}$ is connected to $\mathbf{s}'$ via the pair (i, j).

So the generator matrix for the continuous-time random walk on this larger graph $\mathbf{G}_S(\mathbf{V}_S, \mathbf{W}_S)$ is:

$$\Omega_{\mathbf{ss}'} = \begin{cases} 0 & \text{if } \mathbf{s} \neq \mathbf{s}' \text{ and } \mathbf{s} \text{ is not connected to } \mathbf{s}' \\ \omega_{ij} & \text{if } \mathbf{s} \neq \mathbf{s}' \text{ and } \mathbf{s} \text{ is connected to } \mathbf{s}' \text{ due to } (i,j) \\ -\sum_{\mathbf{t} \neq \mathbf{t}'} \omega_{\mathbf{tt}'} & \text{if } \mathbf{s} = \mathbf{s}' \end{cases} \tag{3.18}$$

A typical multiscale modeled system has several hundreds of nodes, while each node is connected to less than 20 nodes, which has a contact distance within a cutoff value. So the matrix $\underline{\mathbf{\Omega}}$ is usually sparse. When $N_c = 1$, the matrix Eq. (3.18) is reduced to the matrix in the form Eq. (3.16). With this setup of the state space and the generator matrix, the ToF and charge mobility can be evaluated from the theory discussed in the above sections. The next session will introduce the implementation of those methods in practice.

3.3 Random Walk on Graph: Practice

Based on the theory of random walk on graph, and the state space and generator matrix of CTMC, the practical aspects of random walk on a graph will be introduced in this section.

3.3.1 Solving the Master Equation

In practice, numerical schemes are needed to obtain the ToF τ from the solutions $\mathbf{p}(t)$ of the Master Equation Eq. (3.6). Within the analytical expression for the solution based on Eq. (3.9), the eigenvalues and eigenvectors of $\underline{\mathbf{\Omega}}$ need to be calculated, and numerical schemes for dense matrices typically have a complexity of $\mathcal{O}(n^3)$, where n is the dimension of the matrix. For large systems, the computational costs for this step becomes significant.

Alternative direct numerical discretization methods of solving Eq. (3.6) are often prone to stability issues. To illustrate this, consider the application of the

Runge-Kutta 4 method, such as shown in [98] Chapter 12.5, in the form:

$$\mathbf{p}(t+h) = \mathbf{p}(t) + \frac{1}{6}h(\mathbf{y}_1 + \mathbf{y}_2 + \mathbf{y}_3 + \mathbf{y}_4), \tag{3.19}$$

where $\mathbf{y}_1 = \underline{\mathbf{\Omega}}\mathbf{p}(t)$, $\mathbf{y}_2 = \underline{\mathbf{\Omega}}(2\mathbf{p}(t) + \mathbf{y}_1)$, $\mathbf{y}_3 = \underline{\mathbf{\Omega}}(2\mathbf{p}(t) + \mathbf{y}_2)$, $\mathbf{y}_4 = \underline{\mathbf{\Omega}}(\mathbf{p}(t) + \mathbf{y}_3)$. To analyze the numerical stability, it is more convenient to write Eq. (3.19) with the stability function:

$$\mathbf{p}(t+h) = \left(\underline{\mathbf{I}} + \underline{\mathbf{\Omega}}h + \underline{\mathbf{\Omega}}^2\frac{h^2}{2} + \underline{\mathbf{\Omega}}^3\frac{h^3}{6} + \underline{\mathbf{\Omega}}^4\frac{h^4}{24}\right)\mathbf{p}(t). \tag{3.20}$$

For numerical stability it is required that:

$$\left|1 + h\lambda_i + \frac{(h\lambda_i)^2}{2} + \frac{(h\lambda_i)^3}{6} + \frac{(h\lambda_i)^4}{24}\right| \leq 1 \tag{3.21}$$

for all eigenvalues λ_i of $\underline{\mathbf{\Omega}}$. To get an idea of the practical implications, consider further that in the $\underline{\mathbf{\Omega}}$ obtained from the multiscale model described in Chapter 2, the non-zero negative eigenvalues λ_i typically range from -10 to -10^{14}. From this, it is necessary that $h \approx 10^{-14}\,\mathrm{s}$ to ensure stability of the numerical solution. The time-of-flight for materials with very efficient charge transport is around $10^{-8}\,\mathrm{s}$, necessitating 10^6 steps in the RK4 method to resolve. For other materials, this number can increase by orders of magnitudes more, leading ultimately to prohibitively long computation times.

Due to the above-mentioned challenges, the ToF is usually evaluated using the KMC method. The continuous-time random walk can be simulated according to the transition probability in the definition of the Markov chain, and the time from the Source state to the Sink state can be evaluated as random variables. An advantage of KMC method is that the sparsity is naturally utilized. The implementation of KMC will be introduced in the next section.

3.3.2 Implementation of Kinetic Monte Carlo

Here, we briefly summarize the KMC algorithm used in this work [44], which focuses on the jumping events between sites and accounts for the Pauli exclusion

principle at the same time. This setting means that KMC does not have to deal with the large number of states due to multiple carriers, and it is an advantage for implementation. Denote the clock time as t when the KMC starts. This clock time will be updated whenever a carrier takes a transition to a different site. First, only processes involving an occupied site i are considered, and the total escape rate from site i is calculated as:

$$\omega_i = \sum_j \omega_{ji} \tag{3.22}$$

where the sum is over all sites j connected to site i. A particular charge carrier is then selected using the *First Reaction Method* [99], which involves choosing the charge carrier k with the smallest waiting time $t_k^{\mathrm{w}} = \min(t_i^{\mathrm{w}})$ where the minimum is taken over the indexes of nodes occupied by a carrier. This waiting time t_i^{w} is exponentially distributed with parameter ω_i^{-1}, and in KMC, is sampled using a uniformly distributed random number $U_1 \in (0, 1]$:

$$t_i^{\mathrm{w}} = \omega_i^{-1} \ln(U_1^{-1}), \tag{3.23}$$

and the system time is advanced to the new time $t := t + t_k^{\mathrm{w}}$. Next, a destination site for the transition is selected according to the variable step size method (VSSM). Specifically, the largest j is chosen such that:

$$\omega_k^{-1} \sum_{i=1}^{j} \omega_{ki} \leq U_2, \tag{3.24}$$

where $U_2 \in (0, 1]$ is the second uniformly distributed random number. If the transition process is prohibited due to Pauli repulsion, the charge carrier remains at site k and we return to the step of sampling $U_2 \in (0, 1]$. If the transition is allowed, the charge carrier moves to site j. A new waiting time for the charge carrier at site j is then calculated as: $\omega_j^{-1} \ln(U_3^{-1})$ where $U_3 \in (0, 1]$ is another uniformly distributed random number. The charge carrier k with the shortest waiting time t_k^{w} is selected again, and the KMC cycle repeats. When a carrier reaches the Sink, the final clock time t is recorded as the first hitting time t^{f}. The expected first hitting time is the averaged t^{f} over multiple KMC runs. The KMC algorithm is summarized with pseudo-code in Algorithm 1.

The KMC simulation procedure demonstrates that the number of states, arising from the quantity of carriers and sites, does not pose a computational challenge. A significant advantage of the KMC method lies in its capacity to model the behavior of systems with increased complexity caused by the number of charge carriers, as long as the random walk rules of the processes are well-defined. However, the primary challenge in utilizing the KMC method to evaluate mobility is related to the convergence issue, partially due to the presence of trapping states or regions. Specifically, the system includes states that are infrequently sampled by the Monte Carlo algorithm. When these states are eventually sampled, the first hitting time t^{f} is substantially larger compared to other t^{f} values, thus significantly impacting the averaged t^{f}. There are also trap regions where the random walk processes will stay for a long time before visiting other states. This behavior will increase the computation burden for each KMC simulation. Besides those computational challenges, one can not define convergence criteria before performing the KMC simulations.

To overcome those challenges, we propose the evaluation of ToF and charge mobility by solving the expected hitting time in Eq. (3.12) via matrix methods. In the Source-Sink setting with absorbing states, only the expected first hitting time τ_i is related to the charge mobility. This quantity can be in fact calculated as the expected first hitting time of CTRW without solving the MEq or simulating the KMC trajectories, circumventing the stability and the convergence issues involved in those methods. The theory of this approach is not new, although it is not commonly seen in literature for studying charge dynamics in organic semiconductors. The implementation of this method will be introduced in the next section.

3.3.3 Calculation of Expected First Hitting Time

In practice, the sparsity of $\underline{\boldsymbol{\mathcal{P}}}_1$ allows for efficiently solving Eq. (3.15) using iterative techniques. In particular, we first apply the quasi-minimal residual method (QMR) [100] to solve Eq. (3.15). In a general context, the QMR method is an iterative algorithm used to solve nonsymmetric linear systems utilizing the Krylov subspace (the linear space span by $\{\boldsymbol{\tau}^0, (\underline{\mathbf{I}} - \underline{\boldsymbol{\mathcal{P}}}_1)\boldsymbol{\tau}^0, (\underline{\mathbf{I}} - \underline{\boldsymbol{\mathcal{P}}}_1)^2\boldsymbol{\tau}^0, \dots\}$). QMR method is designed to provide a smoother and more reliable convergence than

Algorithm 1 Kinetic Monte Carlo Algorithm

Initialize system time $t \leftarrow 0$
Initialize carrier positions and calculate initial escape rates ω_i
for each occupied site i **do**
    Calculate total escape rate $\omega_i \leftarrow \sum_j \omega_{ji}$
Generate random number $U_1 \in (0, 1]$
for each carrier k **do**
    Calculate waiting time $t_k^{\mathtt{w}} \leftarrow \omega_k^{-1} \ln(U_1^{-1})$
Select carrier k with smallest waiting time $t_k^{\mathtt{w}}$ (first reaction)
Advance system time $t \leftarrow t + t_k^{\mathtt{w}}$
Generate random number $U_2 \in (0, 1]$
Select destination site l using Variable Step Size Method (VSSM):
  Find largest j such that $\omega_k^{-1} \sum_{i=1}^{j} \omega_{ki} \leq U_2$
if transition to site j is allowed **then**
    Move carrier to site j
else
    Carrier remains at site k and repeat from step 11
if carrier k reaches the Sink **then**
    Record first hitting time $t^{\mathtt{f}} \leftarrow t$
else
    Repeat the algorithm from Step 3
Compute average first hitting time $\frac{1}{N_{\mathtt{KMC}}} \sum_{i=1}^{N_{\mathtt{KMC}}} t^{\mathtt{f}}$ over a total of $N_{\mathtt{KMC}}$ runs

other Krylov subspace methods such as the biconjugate gradient method [100]. Due to the sparse matrix involved, the computational complexity of each iteration of the QMR method is proportional to the number of non-zeros in the matrix. The iterative process is considered to be converged when $\|\boldsymbol{\tau}^0 - (\underline{\mathbf{I}} - \underline{\boldsymbol{\mathcal{P}}}_1)\boldsymbol{\tau}^*\|_2 \leq \epsilon^\star \cdot \|\boldsymbol{\tau}^0\|_2$, where the convergence threshold selected and used throughout the thesis is $\epsilon^\star = 10^{-5}$. When the iterative steps exceed 10^4 and this criterion is not yet achieved, the bi-conjugate gradient stabilized method [101] is used to solve Eq. (3.15) using the outcome $\boldsymbol{\tau}^*$ of the previous method as the initial guess. The convergence criterion remains the same. Compared to QMR, the bi-conjugate gradient stabilized method can be more efficient and converge faster in practice, especially for large-scale problems.

To summarize, we have presented the solution of MEq, implementation of KMC and calculating expected first hitting time with matrix formula. The KMC simulation only has events where a single walker jumps and the clock time advances by an exponentially distributed time. In contrast, solving the MEq with the boundary condition of Source-Sink setting and the calculation of expected first hitting time requires the definition of the state space and the generator matrix. The next chapter will present the results for verifying the GRW method of evaluating the ToF and charge mobility.

Chapter 4

Random Walk and KMC Comparison

Adapted from the paper "*A Graph Random Walk Method for Calculating Time-of-Flight Charge Mobility in Organic Semiconductors from Multiscale Simulations*, Z. Chen, P. van der Hoorn, B. Baumeier, arXiv preprint 2405.15836 (2025)".

This chapter presents the results of verifying the GRW method of evaluating ToF charge mobility as mentioned in Chapter 3. The GRW approach is validated by conducting numerical studies of charge dynamics of single and multiple carriers in diffusive and drift-diffusive (due to an external electric field) regimes using a surrogate lattice model of a realistic material, i.e., an amorphous phase of tris(8-hydroxyquinoline)aluminum (Alq_3), which is an organic semiconductor with high energetic disorder and commonly used in OLED as an electron-transport material and emitting layer material. The properties of this material have been simulated within a multiscale model framework combining quantum-mechanical and molecular-mechanics methods. The surrogate model allows varying types and strengths of energetic disorder from the reference baseline and verifies the GRW approach for various material classes. A comparison with results from the MEq confirms the theoretical equivalence of both approaches also in numerical implementations. We further present that KMC results show substantial deviations due to inadequate sampling. All in all, the result suggests that the GRW method provides a powerful alternative to the more commonly used methods without sampling issues and with the benefit of making use of sparse matrix methods.

In what follows, the surrogate model and the system setting are introduced, followed by the results of verifying the GRW method for the ToF calculation. The Poole-Frankel behavior of the organic semiconductor is studied based on the proposed method. Finally, the scalability and the performance of the GRW method will be presented.

4.1 Surrogate Model

The material is modeled as a directed graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$. As detailed in Section 3.1.1, when there is one charge carrier, the carrier dynamics can be modeled as a continuous time random walk on this graph with transition rates between the connected site i and j given by the transfer rate $\omega_{ij} > 0$. In the case of multiple charge carriers, the full charge transport dynamics are due to a large number N_c of charge carriers of equal sign, and each site can only be occupied by no more than one charge carrier. Due to these dependencies between the continuous time random walks of every individual charge carrier, the full dynamics can be modeled as a continuous time random walk on a large graph $\mathbf{G}_S = (\mathbf{V}_S, \mathbf{E}_S)$ with the generator matrix given by Eq. (3.18).

The parameters required to describe the charge dynamics via the continuous time random walks are the positions $\mathbf{r}_i$ of all the site i and the transfer rates ω_{ij} (Eq. (2.37)) obtained from the multiscale model. The external electric field contributes to the Marcus rates by acting as drift force. Considering the effect of the drifted electric field and that each pair of site i and site j constitute a DA complex, Eq. (2.37) becomes:

$$\omega_{ij} = \frac{|J_{ij}|^2}{\hbar}\sqrt{\frac{\pi}{k_\mathrm{B}T\Lambda_{ij}}}\exp\left(-\frac{(\Delta E_{ij} + q\mathbf{F}\cdot\mathbf{r}_{ij} - \Lambda_{ij})^2}{4\Lambda_{ij}k_\mathrm{B}T}\right), \tag{4.1}$$

where the external electric field $\mathbf{F}$ (in $\mathrm{V/m}$) and the charge of the carrier q (in e) can be considered as parameters of the simulation.

The surrogate model allows easy definition of different scenarios, specifically for the important energy difference ΔE_{ij} in the exponential in Eq. (4.1). The surrogate model is based on a regular lattice model in three dimensions, filling

the cubic simulation box $[0, L] \times [0, L] \times [0, L] \subset \mathbb{R}^3, L > 0$. With lattice constant a, and N lattice points per dimension, $L = (N-1)a$, and the position of vertex i is $\mathbf{r}_i^T = (x_i, y_i, z_i) = (i_x, i_y, i_z)a$ with $i_x, i_y, i_z = 0, \ldots, N-1$. For the ToF setup, the source and sink regions are located at $x = 0$ and $x = L$, respectively, i.e., we study the ToF along the x-direction. While we consider all states involving a site with $x = L$ (i.e. $i_x = N-1$) as the set of sink states, we will specify different choices for the source states from the sites at $x = 0$ in Section 4.2. As soon as one of the charge carriers reaches the sink regions at $x = L$, the charge dynamics terminates and the ToF is recorded. The sink state can be interpreted as those states containing an occupied site at the sink region. With this setup, the connection vector is denoted as $\mathbf{r}_{ij}^T = \mathbf{r}_i^T - \mathbf{r}_j^T = (x_{ij}, y_{ij}, z_{ij})$, where $x_{ij} = x_i - x_j$. For the perpendicular directions y and z, we assume cyclic boundary conditions, i.e., for the y-component of the vector connecting i and j

$$y_{ij} = \begin{cases} y_j - y_i & \text{if } |y_j - y_i| < L/2, \\ L + (y_j - y_i) & \text{if } |y_j - y_i| \geq L/2 \text{ and } y_i > y_j \\ -L + (y_j - y_i) & \text{if } |y_j - y_i| \geq L/2 \text{ and } y_i < y_j \end{cases} \tag{4.2}$$

and z_{ij} are defined analogously. We use the so defined entries of the connection vector $\mathbf{r}_{ij}$ to determine the distance metric $|\mathbf{r}_{ij}| = \sqrt{(x_{ij})^2 + (y_{ij})^2 + (z_{ij})^2}$ and an edge is assigned to all vertex pairs (i, j) for which $|\mathbf{r}_{ij}| < 2a$. For simplicity, we set $a = 1\,\mathrm{nm}$. In the current setup of the lattice model, the total number of sites n is related to number of lattice points N by $n = N^3$. To ensure computational feasibility for both the MEq, KMC, and GRW methods, we select $N = 8$. This choice balances the requirements of numerical implementation across all methods for the purpose of comparison and aligns with typical molecular system sizes used in first-principles multiscale modeling.

As in the explicit multiscale model, we use a single reorganization energy for all pairs, i.e., $\Lambda_{ij} = \Lambda = 0.23\,\mathrm{eV}$. We model the coupling elements mimicking its known exponential distance dependence by $|J_{ij}(|\mathbf{r}_{ij}|)|^2 = J_0 \exp(-(|\mathbf{r}_{ij}| - a))$ with $J_0 = 1.79 \cdot 10^{-4}\,(\mathrm{eV})^2$. In the multiscale model, the site energies ϵ_i follow a Gaussian distribution with mean $\bar{\epsilon}$ and variance σ^2. In the first class of site-energy models, the ϵ_i are taken as independent and identically distributed (iid) Gaussian samples. We refer to this model as *uncorrelated*. However, in Alq_3 (and several

other materials) the site energies show spatial correlation. In a *correlated* site-energy model, we adopt a moving average procedure as in [66], making use of the invariance properties of the normal distribution with respect to convolution. First, with three sequences of iid random variables $M_i^{(a)}, M_i^{(b)}, M_i^{(c)} \sim \mathcal{N}(0, \sigma^2)$, every site is assigned the 4-tuple $(V_i, M_i^{(a)}, M_i^{(b)}, M_i^{(c)})$ to which we want to allocate a random site energy ϵ_i. If $V_i^{(1)}$,$V_i^{(2)}$,$\cdots$,$V_i^{(l)}$ are the l nearest neighbors of V_i with corresponding random variables $M_i^{(b),(j)}, M_i^{(c),(j)}$, $j = 1, 2, \cdots, l$ the spatially correlated energies are evaluated as:

$$\epsilon_i = \sqrt{\eta_a} M_i^{(a)} + \sqrt{\frac{\eta_b}{l_b}} \sum_{j=1}^{l_b} M_i^{(b),(j)} + \sqrt{\frac{1-\eta_a-\eta_b}{l_c}} \sum_{j=1}^{l_c} M_i^{(c),(j)} + \bar{\epsilon}, \tag{4.3}$$

where $\eta_a, \eta_b \geq 0$ ($\eta_a + \eta_b \leq 1$) are the weights for the individual components and $l_b, l_c > 0$ for some integers. Equation (4.3) develops the spatially correlated ϵ_i as a superposition of three independent energy landscapes. From the explicit Alq_3 data we choose $\bar{\epsilon} = -0.76\,\mathrm{eV}$, $\sigma = 0.19\,\mathrm{eV}$, $\eta_a = 0.2$, $\eta_b = 0.4$, as well $l_b = 9$ and $l_c = 280$.

With this surrogate model, the main physical quantity of interest studied is the mobility (or equivalently ToF) in the system representing the molecular material. Both of these come down to the first time a charge carrier hits one of the Sink states. In terms of random walk description, this is called the *hitting time* of the sink states. Our proposed method as detailed in Chapter 3.1.4 will be used for computing the hitting time. After this, we compare it to two other known methods: master equation (see Chapter 3.1.2) and KMC (see Chapter 3.1.3).

4.2 Results

we show the results for the ToF τ and the dependence of the mobility μ on the strength of the externally applied electric field $\mathbf{F}$ from the GRW method for different settings and comparisons to MEq and KMC methods. The base model we use is a lattice model as introduced in Section 4.1 with $N = 8$. This system size is typical for medium-scale molecular systems in multiscale modeling studies (examples can be found in [44, 102, 103, 104]). It also ensures that the numerical

solution of the master equation Eq. (4.4), remains computationally feasible, enabling a direct comparison between our proposed method and the true dynamics of Eq. (3.6). While based on properties of the explicit multiscale model for Alq_3 from [44, 66], the choice of a lattice model allows studying different settings for the site energies on an equal footing: the no-disorder case ($\epsilon_i = 0$); uncorrelated, Gaussian distributed disorder; and spatially correlated Gaussian disorder as modeled by Eq. (4.3). To study different *amounts* of disorder, we introduce a scaling parameter $k \in [0, 1]$ into the site energy term in the exponential of the Marcus rate, i.e., we let $\Delta E_{ij} \to k\Delta E_{ij}$ in Eq. (4.1), effectively smoothly interpolating from $\sigma = 0.0\,\mathrm{eV}$ to $\sigma = 0.19\,\mathrm{eV}$ using the same reference random samples of the site energies. Further, we consider the case of $N_\mathrm{c} = 1$ and $N_\mathrm{c} = 2$ charge carriers.

4.2.1 Comparison of MEq and GRW

In Chapter 3.1.2 and Chapter 3.1.4 we have seen that the MEq and GRW approaches provide formally equivalent predictions of expected hitting times for the defined sink regions and with that of the charge carriers' ToF. In practice, the MEq method serves as a valuable tool for investigating charge transport phenomena within disordered organic materials [105, 106, 107, 108]. Here, we aim to validate our numerical implementation of the GRW by comparing its predictions of τ to those obtained from the MEq. We employ the analytic form of the solution of the initial value problem in Eq. (3.9) with the eigenvalues $\{\lambda_\alpha\}$ and corresponding eigenvectors $\{\mathbf{u}_\alpha\}$ of the matrix $\underline{\mathbf{\Omega}}$ as

$$\mathbf{p}(t) = \sum_{\alpha=1}^{\binom{n}{N_\mathrm{c}}} c_\alpha \mathbf{u}_\alpha e^{\lambda_\alpha t}, \tag{4.4}$$

where $\mathbf{c} = \mathbf{U}^{-1}\mathbf{p}(0)$ and $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_{\binom{n}{N_\mathrm{c}}})$. As initial value for Eq. (3.6), we set

$$\mathbf{p}(0) = \begin{cases} N_\mathrm{c}/N_\mathrm{Source} & \text{for } \mathbf{s} \in \text{Source} \\ 0 & \text{else} \end{cases},$$

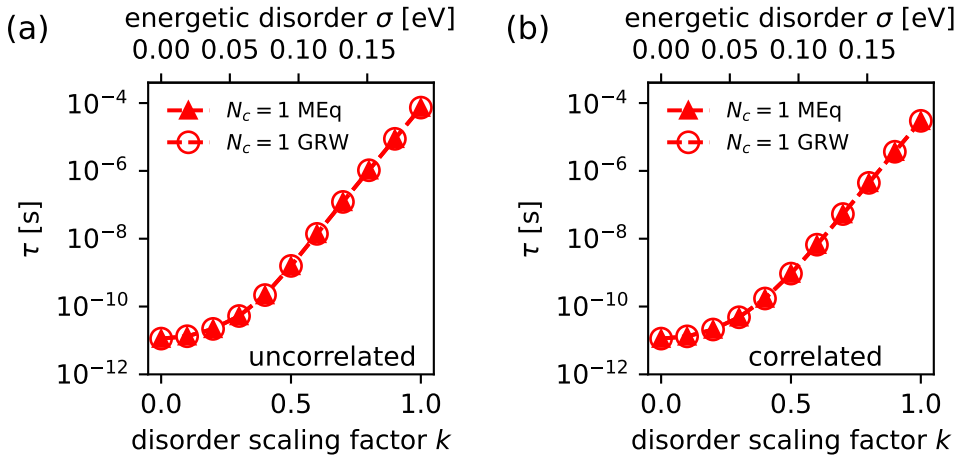

Figure 4.1: Calculated ToF τ (in s) for $N_c = 1$ depending on disorder strength in systems with uncorrelated (a) and spatially correlated (b) site energies, obtained from MEq (filled symbols) and GRW (open symbols), respectively. Each data point τ represents the sample average of the ten realizations of the Gaussian distributed energy landscapes.

where N_{source} is the number of source states defined. The form of Eq. (4.4) allows analytic evaluation of the integral in Eq. (3.7) to obtain τ. For the purpose of comparing the two methods, we consider a single source state for both $N_c = 1$ ($\mathbf{s} = (1, 0, \ldots, 0)$) and $N_c = 2$ ($\mathbf{s} = (1, 1, 0, \ldots, 0)$), respectively. Specifically, the site s_1 is located at $\mathbf{r}_1^T = (0, 0, 0)$ and s_2 at $\mathbf{r}_2^T = (0, 0, a)$. In Fig. 4.1 we show the results of the obtained dependence of τ on k for the systems with uncorrelated and correlated disorder and $N_c = 1$, respectively. No external electric field is applied, i.e., $\mathbf{F} = \mathbf{0}$, so the dynamics of the charge carriers are purely diffusive. For both MEq and GRW methods, the hitting time τ is calculated by averaging ten systems each with the same $\bar{\epsilon}, \sigma$. It is visually clear that there is numerically excellent agreement between both approaches, as was expected from the theoretical remarks in Chapter 3.1.2 and 3.1.4. The relative errors are consistently smaller than $0.1\,\%$, as shown in Fig. 4.2.

This comparison confirms the validity of the GRW method we propose, at least for $N_c = 1$. As convenient as the form in Eq. (4.4) is at first glance, its application

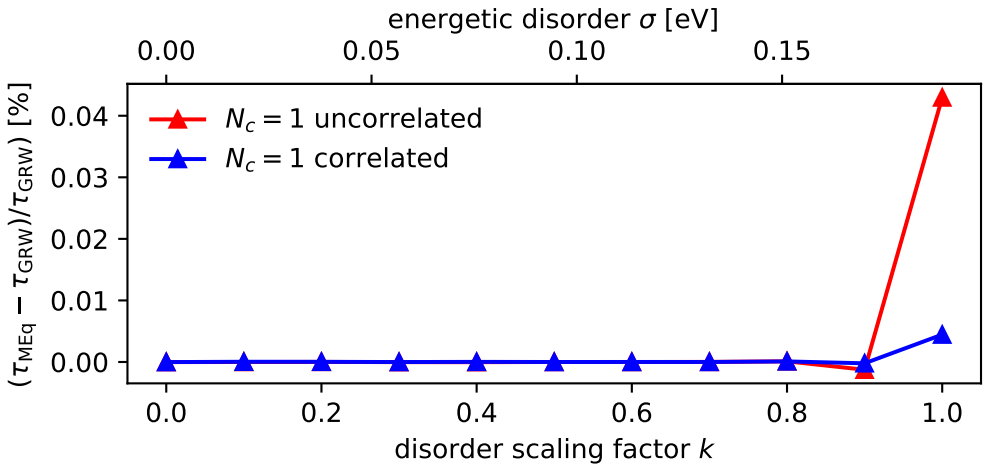

Figure 4.2: Relative difference $(\tau_{\mathrm{MEq}} - \tau_{\mathrm{GRW}})/\tau_{\mathrm{GRW}}$ (in %) for different disorder scaling factors k in systems with uncorrelated and spatially correlated disorder and different numbers of charge carriers, respectively.

to larger-scale problems is problematic numerically: it requires the full eigendecomposition of $\underline{\mathbf{\Omega}}$, which cannot be obtained with sparse matrix methods. In the present study, the $N_{\mathrm{c}} = 2$ case with $n = 512$ requires numerical diagonalization of a full matrix with dimension 130816, even though $\underline{\mathbf{\Omega}}$ is extremely sparse. This highlights one of the advantages of the GRW method as it can make use of efficient sparse matrix implementations throughout. Alternative approaches to solve Eq. (3.6) that make use of numerical discretization schemes like forward and backward Euler methods or higher-order Runge–Kutta methods, could be formulated in sparse forms. However, the details of the matrix $\underline{\mathbf{\Omega}}$ make time-stepping methods inefficient. For instance, in one realization of the $k = 1$ uncorrelated disorder case, the non-zero eigenvalues of $\underline{\mathbf{\Omega}}$ range from -10^{13} to -10^{2}. With the associated time-scales of the dynamic modes ranging from $10^{-13}\,\mathrm{s}$ to $10^{-2}\,\mathrm{s}$, and the average hitting time of $\sim 10^{-3}\,\mathrm{s}$, a very small time step ($10^{-14}\,\mathrm{s}$) and many such steps (10^{11}) might be required in an extreme case to resolve τ reliably. In addition, the discretization scheme must also allow for integration of Eq. (3.7). All in all, using time-stepping methods that resolve the actual dynamics of the probabilities to obtain τ seems cumbersome.

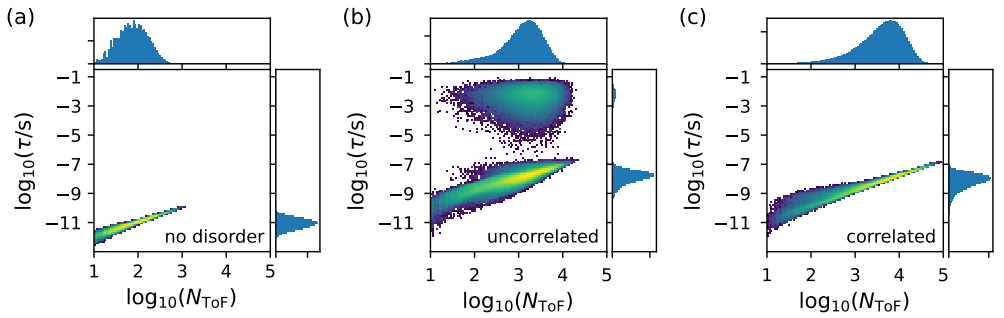

Figure 4.3: Distributions of KMC steps until absorption to sink (N_{ToF}) and the corresponding first hitting time τ obtained via KMC for a single realization of the lattice model with no site-energy disorder (a), Gaussian-distributed site-energies without (b) and with spatial correlations (c). The density histogram consists of 10^7 KMC sample points grouped into 100 bins.

4.2.2 Comparison of GRW and KMC

Besides the MEq method, many literature report the KMC method from studying transport processes in organic semiconductors [109, 107, 110, 111, 112]. However, the KMC method often requires long computations for each simulation, with a single simulation sometimes requiring hours or even days to calculate the quantities of interest. Moreover, many simulations need to be conducted to ensure that the results indeed correspond to the true average values of the stochastic model. We begin the discussion of the results by highlighting the potentially problematic convergence behavior of KMC simulations as an alternative to solving the explicit dynamics of Eq. (3.6). For a single realization of the site energies with no, uncorrelated, and spatially correlated disorder with $\sigma = 0.19\,\mathrm{eV}$, respectively, we performed 10^7 KMC simulations. We recorded for each of these simulations the individual first hitting time of the sink region (τ) and the associated number of KMC steps (N_{ToF}). The resulting distributions are depicted in Fig. 4.3. One can see that the no disorder case in panel (a) is largely unremarkable. There is very little noticeable spread in the distributions of the ToF and number of steps recorded, with the maximum number of steps needed being around 1000. In contrast, the data for the uncorrelated disorder case shown in Fig. 4.3(b), illustrate one of the key challenges for KMC: the majority of τ values cluster around $10^{-8}\,\mathrm{s}$, while only

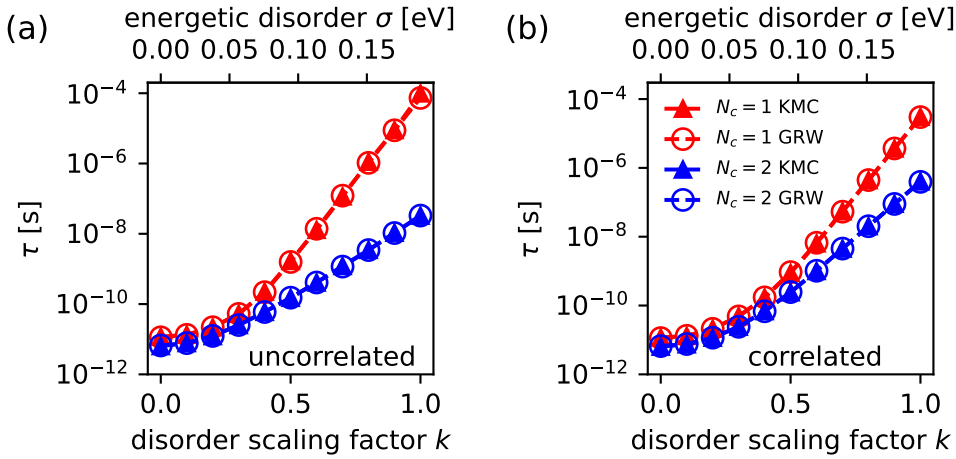

Figure 4.4: Calculated ToF τ (in s) depending on disorder strength on systems with uncorrelated (a) and spatially correlated (b) site energies, obtained from KMC (filled symbols) and GRW (open symbols), respectively. Data points in red (blue) show the results for $N_c = 1$ ($N_c = 2$). Each data point τ represents the sample average of the ten realizations of the Gaussian distributed energy landscapes. Each KMC simulation contains 1000 runs.

a small fraction falls near 10^{-2} s, which is the main contribution to the average hitting time ($4.3 \cdot 10^{-4}$ s). The comparatively rare occurrences of long ToFs can be attributed to the presence of isolated sites with low energies that are not always visited. Most of the KMC simulations seem to require 10000 steps or fewer. In spatially correlated disorder case (Fig. 4.3(c)), we do not observe the influence of apparently rarely sampled sites. Instead, due to the spatial correlation, low-energy sites are not isolated, and several sites are likely to form *regions* of relatively low energy. As the relative energy difference among the sites inside these regions is small, the random walker in the KMC simulations spends a significant number of steps in these regions before it progresses to another of such regions or the sink. Consequently, some KMC simulations require up to 10^5 steps to finish. Overall, it is clear that depending on the characteristics of the energetic disorder in the material, the results of ToF simulations from KMC simulation might be subject to significant fluctuations, with convergence requiring numerous samples, and in general potentially long simulation times per sample.

Table 4.1: Calculated mean ToF τ and standard error of the mean (in s) depending on disorder strength on systems with uncorrelated and spatially correlated site energies, obtained from GRW and KMC, respectively. Each data point τ represents the sample average of the ten realizations of the Gaussian distributed energy landscapes. Each KMC simulation contains 1000 runs.

	uncorrelated		**correlated**	
k	**GRW**	**KMC**	**GRW**	**KMC**
		$\mathbf{N_c = 1}$		
0.0	$(1.13 \pm 0.00) \cdot 10^{-11}$	$(1.18 \pm 0.00) \cdot 10^{-11}$	$(1.13 \pm 0.00) \cdot 10^{-11}$	$(1.18 \pm 0.00) \cdot 10^{-11}$
0.1	$(1.33 \pm 0.02) \cdot 10^{-11}$	$(1.35 \pm 0.02) \cdot 10^{-11}$	$(1.32 \pm 0.05) \cdot 10^{-11}$	$(1.31 \pm 0.05) \cdot 10^{-11}$
0.2	$(2.19 \pm 0.08) \cdot 10^{-11}$	$(2.18 \pm 0.08) \cdot 10^{-11}$	$(2.10 \pm 0.23) \cdot 10^{-11}$	$(2.12 \pm 0.26) \cdot 10^{-11}$
0.3	$(5.25 \pm 0.75) \cdot 10^{-11}$	$(5.32 \pm 0.77) \cdot 10^{-11}$	$(4.82 \pm 1.29) \cdot 10^{-11}$	$(4.92 \pm 1.39) \cdot 10^{-11}$
0.4	$(2.18 \pm 0.95) \cdot 10^{-10}$	$(2.25 \pm 1.02) \cdot 10^{-10}$	$(1.72 \pm 0.86) \cdot 10^{-10}$	$(1.71 \pm 0.86) \cdot 10^{-10}$
0.5	$(1.58 \pm 1.16) \cdot 10^{-9}$	$(1.68 \pm 1.26) \cdot 10^{-9}$	$(9.37 \pm 6.65) \cdot 10^{-10}$	$(8.83 \pm 6.14) \cdot 10^{-10}$
0.6	$(1.39 \pm 1.21) \cdot 10^{-8}$	$(1.29 \pm 1.17) \cdot 10^{-8}$	$(6.66 \pm 5.57) \cdot 10^{-9}$	$(6.41 \pm 5.33) \cdot 10^{-9}$
0.7	$(1.22 \pm 1.12) \cdot 10^{-7}$	$(1.16 \pm 1.07) \cdot 10^{-7}$	$(5.34 \pm 4.79) \cdot 10^{-8}$	$(5.49 \pm 4.93) \cdot 10^{-8}$
0.8	$(1.04 \pm 0.97) \cdot 10^{-6}$	$(1.12 \pm 1.05) \cdot 10^{-6}$	$(4.43 \pm 4.10) \cdot 10^{-7}$	$(4.40 \pm 4.06) \cdot 10^{-7}$
0.9	$(8.82 \pm 8.32) \cdot 10^{-6}$	$(8.97 \pm 8.46) \cdot 10^{-6}$	$(3.67 \pm 3.44) \cdot 10^{-6}$	$(3.65 \pm 3.43) \cdot 10^{-6}$
1.0	$(7.34 \pm 6.95) \cdot 10^{-5}$	$(9.87 \pm 9.35) \cdot 10^{-5}$	$(2.99 \pm 2.82) \cdot 10^{-5}$	$(3.05 \pm 2.88) \cdot 10^{-5}$
		$\mathbf{N_c = 2}$		
0.0	$(6.63 \pm 0.00) \cdot 10^{-12}$	$(6.37 \pm 0.00) \cdot 10^{-12}$	$(6.63 \pm 0.00) \cdot 10^{-12}$	$(6.37 \pm 0.00) \cdot 10^{-12}$
0.1	$(7.74 \pm 0.08) \cdot 10^{-12}$	$(7.76 \pm 0.08) \cdot 10^{-12}$	$(7.62 \pm 0.25) \cdot 10^{-12}$	$(7.47 \pm 0.25) \cdot 10^{-12}$
0.2	$(1.24 \pm 0.04) \cdot 10^{-11}$	$(1.23 \pm 0.04) \cdot 10^{-11}$	$(1.16 \pm 0.11) \cdot 10^{-11}$	$(1.16 \pm 0.11) \cdot 10^{-11}$
0.3	$(2.53 \pm 0.15) \cdot 10^{-11}$	$(2.53 \pm 0.15) \cdot 10^{-11}$	$(2.41 \pm 0.54) \cdot 10^{-11}$	$(2.43 \pm 0.55) \cdot 10^{-11}$
0.4	$(5.95 \pm 0.54) \cdot 10^{-11}$	$(6.00 \pm 0.56) \cdot 10^{-11}$	$(6.81 \pm 2.84) \cdot 10^{-11}$	$(6.89 \pm 2.88) \cdot 10^{-11}$
0.5	$(1.52 \pm 0.20) \cdot 10^{-10}$	$(1.51 \pm 0.20) \cdot 10^{-10}$	$(2.46 \pm 1.48) \cdot 10^{-10}$	$(2.43 \pm 1.43) \cdot 10^{-10}$
0.6	$(4.11 \pm 0.78) \cdot 10^{-10}$	$(4.11 \pm 0.81) \cdot 10^{-10}$	$(1.02 \pm 0.75) \cdot 10^{-9}$	$(1.01 \pm 0.73) \cdot 10^{-9}$
0.7	$(1.17 \pm 0.31) \cdot 10^{-9}$	$(1.21 \pm 0.34) \cdot 10^{-9}$	$(4.51 \pm 3.67) \cdot 10^{-9}$	$(4.51 \pm 3.67) \cdot 10^{-9}$
0.8	$(3.44 \pm 1.21) \cdot 10^{-9}$	$(3.48 \pm 1.23) \cdot 10^{-9}$	$(2.02 \pm 1.73) \cdot 10^{-8}$	$(2.02 \pm 1.74) \cdot 10^{-8}$
0.9	$(1.05 \pm 0.46) \cdot 10^{-8}$	$(1.08 \pm 0.47) \cdot 10^{-8}$	$(8.95 \pm 7.94) \cdot 10^{-8}$	$(9.04 \pm 8.04) \cdot 10^{-8}$
1.0	$(3.30 \pm 1.71) \cdot 10^{-8}$	$(3.28 \pm 1.69) \cdot 10^{-8}$	$(3.92 \pm 3.55) \cdot 10^{-7}$	$(4.04 \pm 3.66) \cdot 10^{-7}$

Against this background, combined with the potential numerical challenges of MEq mentioned in the previous section, the GRW method is expected to provide a significant advantage in not being affected by these and related convergence problems. We demonstrate the quality of the predictions of τ obtained by solving Eq. (3.15) compared to a reference from extensive KMC calculations. As in the

previous section, the hitting time τ is calculated by averaging ten systems each with the same $\bar{\epsilon}, \sigma$, with the tuned amount of disorder. The full set of results are listed in Table 4.1. Figure 4.4 shows the obtained dependence of τ on k from the two methods for uncorrelated and correlated disorder and for one and two charge carriers, respectively. Again, no external electric field is applied, i.e., $\mathbf{F} = \mathbf{0}$.

In the case of KMC, 1000 runs are performed for each realization of the site energies, with carriers populating the source state. Overall there is a very good agreement between the results obtained with the GRW method as compared to KMC. Qualitatively, the data indicate that the more disorder is in the system, the larger the ToF becomes. We also observe in all settings a reduction of τ as the number of carriers increases, and that τ decreases more in systems with uncorrelated site energies as compared to those with correlated site energies. This observation aligns with the findings reported in [113]. According to [43], one contributing factor to this phenomenon is the nature of the isolated low-energy sites in the uncorrelated case vs the possible existence of regions (or small clusters) of low-energy sites in the spatially correlated case. In the former, a single carrier can be trapped in such an isolated low-energy site, and for $N_\mathrm{c} = 2$, due to the exclusion, the second carrier can "freely" diffuse within the rest of the system. This would correspond in Fig. 4.3(b) to the elimination of the KMC trajectories with the large τ as discussed above. From the logarithmic scale on the y-axes in Fig. 4.4, it is difficult to assess the differences in the results from GRW and KMC in detail. We therefore show in Fig. 4.5 the relative difference $\Delta_\mathrm{rel}\tau = (\tau_\mathrm{KMC} - \tau_\mathrm{GRW})/\tau_\mathrm{GRW}$ depending on the disorder scaling factor k, for all four cases studied. Except for the case $N_\mathrm{c} = 1$ in uncorrelated disorder with $k = 1$, the results differ at most by $7\,\%$ ($N_\mathrm{c} = 1$ uncorrelated), $4\,\%$ ($N_\mathrm{c} = 2$ uncorrelated), $6\,\%$ ($N_\mathrm{c} = 1$ correlated), and $4\,\%$ ($N_\mathrm{c} = 2$ correlated), respectively. For the case $N_\mathrm{c} = 1$ with uncorrelated disorder and $k = 1$, a $30\,\%$ relative error arise due to the inadequate KMC samplings. This error highlights the convergence limitations of KMC, attributed to insufficient sampling due to the low-energy regions of the molecular system.

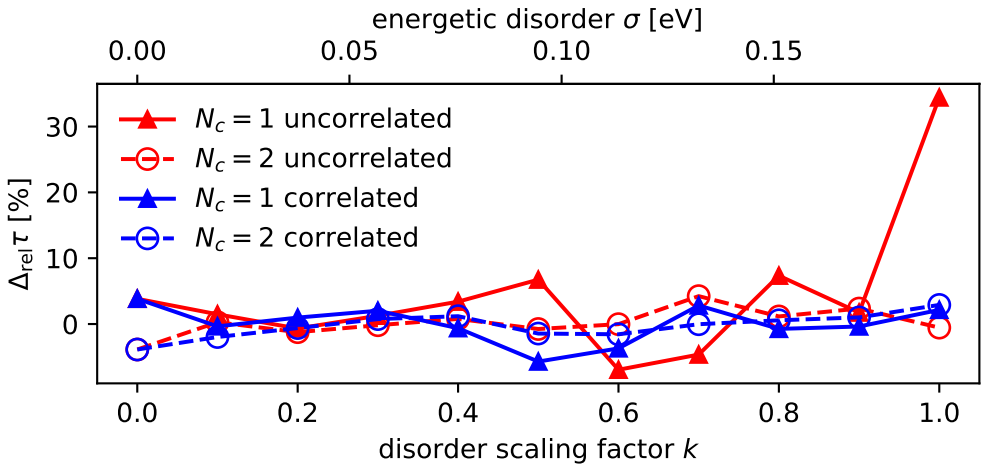

Figure 4.5: Relative difference $\Delta_{\text{rel}}\tau = (\tau_{\text{KMC}} - \tau_{\text{GRW}})/\tau_{\text{GRW}}$ (in %) for different disorder scaling factors k in systems with uncorrelated and spatially correlated disorder and different numbers of charge carriers, respectively.

4.2.3 Electric Field Dependent of ToF Mobility from GRW

With the advantages of the GRW method over MEq and KMC approaches as discussed in the previous two sections for the case of purely diffusive charge carrier dynamics, we now turn to the application of the GRW framework to study the electric-field dependent mobility of one and two charge carriers, whose dynamics now correspond to a drift-diffusion process. Recall that the mobility is defined as $\mu = \frac{\mathbf{v}\cdot\mathbf{F}}{|\mathbf{F}|^2}$. In the following, we set $\mathbf{F}^T = (F, 0, 0)$, such that we can evaluate the mobility as a function of F as $\mu(F) = \frac{L}{\tau(F)F}$. We use $\tau(F)$ here to emphasize that the ToF itself also depends on the value of the electric field, through the Marcus rates Eq. (4.1). Of course, as seen before for the diffusive carrier dynamics, the ToF also depends on the type and amount of disorder in the system.

From experiments, it is empirically known that the charge carrier mobility of many disordered organic semiconductors is approximately $\mu(F) = \mu_0 \exp(\beta\sqrt{F})$, for $F \in \left[10^7\,\text{V/m}, 10^8\,\text{V/m}\right]$. Poole and Frenkel also predicted this electric-field dependence in a model describing the mechanism of trap-assisted electron trans-

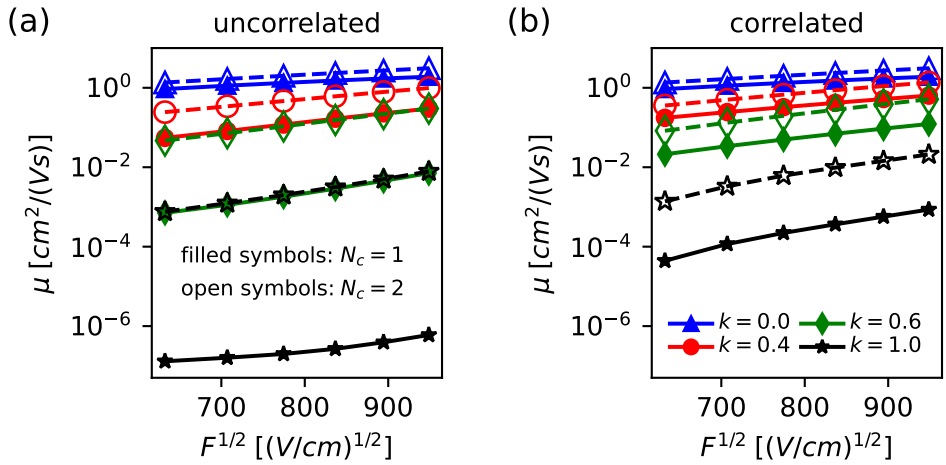

Figure 4.6: Electric-field dependence of the mobility μ in the $N = 8$ multiscale model system for (a) spatially uncorrelated and (b) spatially correlated disorder for $N_c = 1$ (filled symbols) and $N_c = 2$ (open symbols) and varying strength of energetic disorder via scaling factor k, as obtained from the GRW method.

port for insulators and semiconductors [114]. It is therefore common to plot the mobility μ against $\sqrt{F}$ in a so-called *Poole–Frenkel plot*. We show in Fig. 4.6 such

a plot as resulting from GRW calculations for our system with (a) spatially uncorrelated and (b) correlated disorder of different strengths indicated by the values of k. With the logarithmic y-axis one can observe indeed an ideal linear dependence of the mobility on $\sqrt{F}$ for most of the cases. Exceptions can be noted for $k = 1$ with $N_c = 1$ in (a) and $k = 1$ for both one and two charge carriers in (b), where the shown field-dependence deviates from the Poole–Frenkel model at low F. Nevertheless, for all scenarios studied in this section, we have extracted the Poole–Frenkel parameters μ_0 and β from the results in Fig. 4.6 and summarize them in Table 4.2. For $k = 0$, there is by construction no difference between the uncorrelated and correlated cases. With an increasing amount of disorder in the system, the mobility increases because the respective ToF increases as discussed in the field-free cases before, which is reflected in the values for μ_0.

Table 4.2: Poole–Frenkel parameters μ_0 (in $\mathrm{cm}^2/(\mathrm{Vs})$) and β (in $\sqrt{\mathrm{cm/V}}$) extracted from GRW simulations for $N_\mathrm{c} = 1$ and $N_\mathrm{c} = 2$.

	uncorrelated		**correlated**	
k	μ_0	β	μ_0	β
		GRW for $\mathbf{N_c = 1}$		
0.0	$2.4{\cdot}10^{-1}$	$2.2{\cdot}10^{-4}$	$2.4{\cdot}10^{-1}$	$2.2{\cdot}10^{-4}$
0.4	$1.9{\cdot}10^{-3}$	$5.3{\cdot}10^{-4}$	$1.5{\cdot}10^{-2}$	$3.9{\cdot}10^{-4}$
0.6	$2.5{\cdot}10^{-4}$	$7.8{\cdot}10^{-4}$	$8.1{\cdot}10^{-4}$	$5.3{\cdot}10^{-4}$
1.0	$2.2{\cdot}10^{-9}$	$5.8{\cdot}10^{-4}$	$4.5{\cdot}10^{-7}$	$8.0{\cdot}10^{-4}$
		GRW for $\mathbf{N_c = 2}$		
0.0	$2.7{\cdot}10^{-1}$	$2.6{\cdot}10^{-4}$	$2.7{\cdot}10^{-1}$	$2.6{\cdot}10^{-4}$
0.4	$1.6{\cdot}10^{-2}$	$4.4{\cdot}10^{-4}$	$2.7{\cdot}10^{-2}$	$4.1{\cdot}10^{-4}$
0.6	$1.1{\cdot}10^{-3}$	$5.9{\cdot}10^{-4}$	$2.5{\cdot}10^{-3}$	$5.6{\cdot}10^{-4}$
1.0	$4.6{\cdot}10^{-6}$	$7.8{\cdot}10^{-4}$	$2.1{\cdot}10^{-5}$	$7.3{\cdot}10^{-4}$

For each value of $k > 0$, mobilities for the spatially uncorrelated site energies are lower than those in the correlated case. From the inverse relation of μ and τ, this seems at first glance at odds with the ToF for $F = 0$ in Fig. 4.4, where for $N_\mathrm{c} = 2$ shorter τ are recorded for systems with uncorrelated site-energies than for correlated ones, which would indicate a higher mobility for the former. However, the additional drift component of the charge carriers' dynamics in the direction of an applied electric field has strong effects on the ToF that are different for each of the scenarios studied in the work. One can see an indication of this in Fig. 4.6 comparing the uncorrelated and correlated cases at low fields. In the correlated case, there is the already mentioned deviation from the ideal linear relation between μ and $\sqrt{F}$ in the Poole–Frenkel plot. This is not visible in the uncorrelated case.

We therefore show in Fig. 4.7 the dependence of the mobility on the electric field for much smaller field strengths in the interval $[4 \cdot 10^4\ \mathrm{V/m}, 4 \cdot 10^7\ \mathrm{V/m}]$ for the case of $N_\mathrm{c} = 2$ and $k = 1$. One can clearly see a crossover between the values of the mobility in the uncorrelated and correlated disorder cases for low fields,

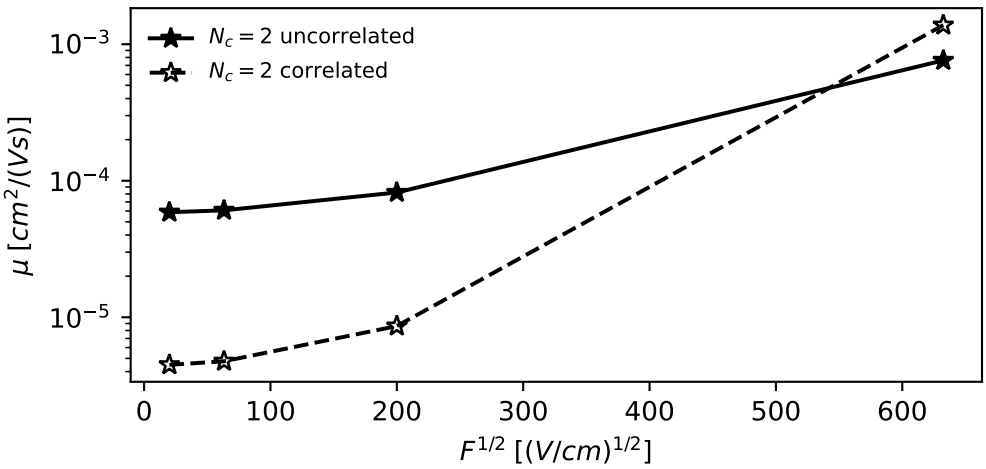

Figure 4.7: Electric-field dependence of the mobility μ for small field strengths in the $N = 8$ multiscale model system for spatially uncorrelated and spatially correlated disorder for $N_c = 2$ and $k = 1$, as obtained from the GRW method.

agreeing qualitatively with what is expected from the ToF data in the diffusive regime discussed in Section 4.2. This observation of the in general not ideal Poole–Frenkel behavior for small fields emphasizes that the extracted values of μ_0 as listed in Table 4.2 should not be interpreted as the true mobilities at $F = 0$.

4.3 Scalability and Performance

In the previous section, we demonstrated that the GRW method proposed in this work yields quantitatively accurate estimates of the ToF, or the charge mobility in ToF settings, without the need to explicitly calculate the dynamics of charge carriers in multiscale systems and without common sampling and convergence problems of state-of-the art methods, such as KMC. In particular, low charge concentration typically poses the most significant computational challenges for the KMC method, as discussed in Section 4.2.2 (cf. Fig. 4.3(b)) and in previous studies [115, 104]. Long computational times and inadequate sampling are preva-

lent issues in most first-principles multiscale systems. As shown in Section 4.2.2 Fig. 4.5, a 30% error was observed in our test case with uncorrelated energies and $k = 1$. That is, the KMC method reported ToF values that are 30% larger than the true ToF value. These challenges become more pronounced as the system size and disorder increases. Here, we now turn to a discussion of the scaling and convergence behavior of the GRW method, by considering model systems with different sizes $\binom{n}{N_c}$ of the state space for different amounts of spatially uncorrelated energetic disorder. Specifically, we vary the extension of the lattice in the x-direction, such that $[0, L_x] \times [0, L] \times [0, L] \subset \mathbb{R}^3$ where $L = (N-1)a$ with $N = 8$ as used in the previous section, and $L_x = (N_x - 1)a$ for $N_x = 10, 20, 30, 40$. This corresponds to systems with $n = 640, 1280, 1920, 2560$, all typical system sizes accessible to first-principle multiscale models. Analogously to the previous section, sites with $x = 0$ are source sites, and sites with $x = L_x$ are sink sites, with source and sinks states determined accordingly.

Using the quasi-minimal residual method and the biconjugate gradients stabilized method for solving the matrix equation Eq. (3.15), the convergence of the solution is quantified using the metric $\epsilon^\star := \|\boldsymbol{\tau}^0\|_2 / \|\boldsymbol{\tau}^0 - (\underline{\mathbf{I}} - \underline{\mathbf{P}}_1)\boldsymbol{\tau}^*\|_2$. The convergence of the GRW method is studied by monitoring $\epsilon^\star$ as a function of the number of iterative steps N_{iter}. As described in Chapter 3.3.3, convergence is achieved when $\epsilon^\star \leq 10^{-5}$. Figure 4.8 illustrates that for $N_c = 1$, the number of iterative steps required for convergence increases with both the disorder parameter k and the system size n. Specifically, as the system size increases from 640 to 2560, the number of required iterations grows approximately linearly, from 100 to 430.

Figure 4.9 illustrates a similar trend of increasing iteration steps required for convergence. As the system size increases from 640 to 2560, the iteration steps rise substantially from 600 to 3280. The convergence becomes increasingly challenging with higher disorder and larger n, as evidenced by the pronounced oscillations. This nonlinear growth arises from the combinatorial increase in states, $\binom{n}{N_c}$, for $N_c = 2$ carriers. In larger systems with $N_c > 2$, the proposed method faces computational challenges due to the exponential growth in state space. At the same time, the KMC method can also have convergence difficulties in large systems even if there are multiple carriers.

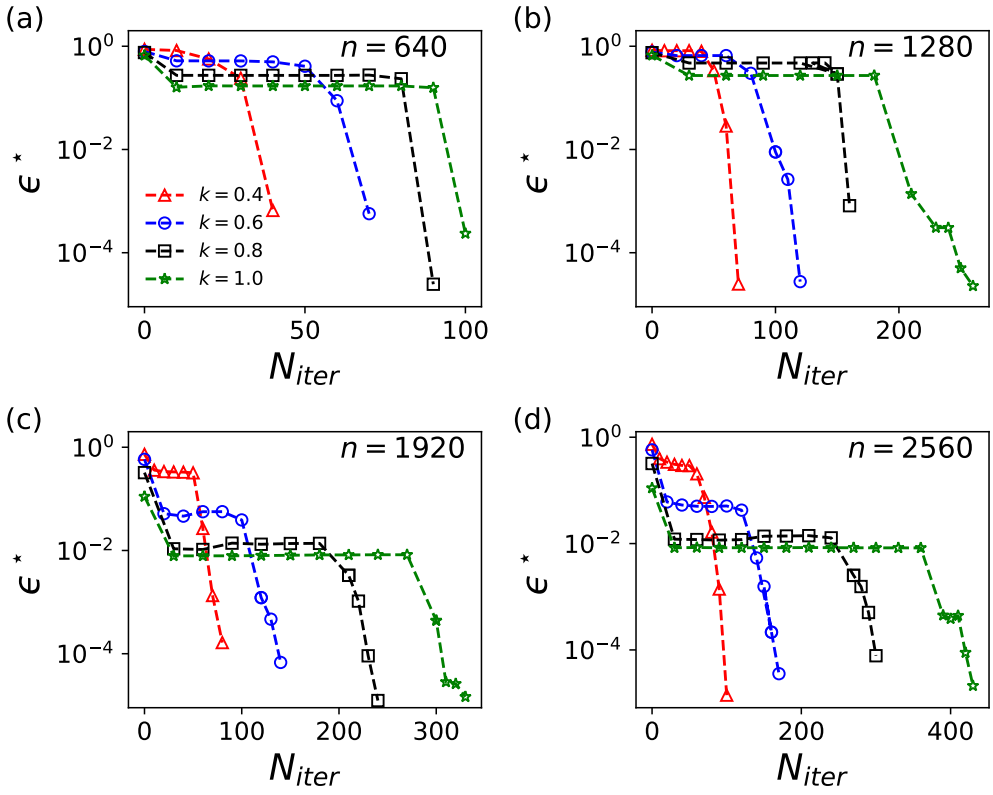

Figure 4.8: Convergence behavior of the ToF calculation using the GRW method for $N_c = 1$.

From a practical perspective, the time required to obtain the convergence in GRW method, the time-to-solution t_{sol}, is more relevant than the number of iterations needed. In Fig. 4.10 we show this time for $N_c = 1$ and $N_c = 2$ and different numbers of sites with varying energetic disorder, as recorded with our proof-of-concept implementation in Python on an Intel(R) Xeon(R) Gold 5120 CPU @ 2.20GHz with 28 threads. When there is one carrier, the GRW method takes less than 1 s to achieve converged ToF, even for the largest system with $n = 2560$. In contrast, a single KMC simulation, i.e., for one random seed from one sink state, using production code for the uncorrelated system with size $n = 512$ can take more than 1 minute to complete on the same system, and the ToF obtained from

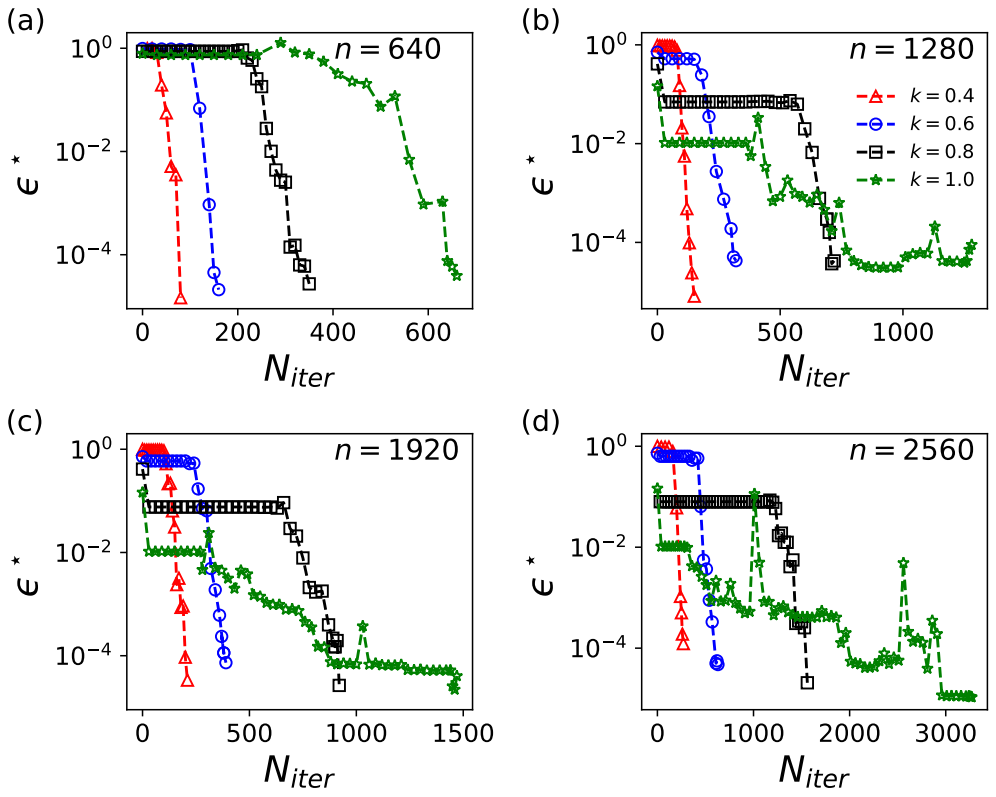

Figure 4.9: Convergence behavior of the ToF calculation using the GRW method for $N_c = 2$.

1000 KMC simulations stills has 30% relative error. Even though the KMC simulations can be trivially parallelized, these aspects highlight the advantages of the GRW method in computational time. When $N_c = 2$, one can see that the required computational time increases exponentially. This exponential time increment for multiple carrier systems is a challenge for the GRW method and requires further exploration.

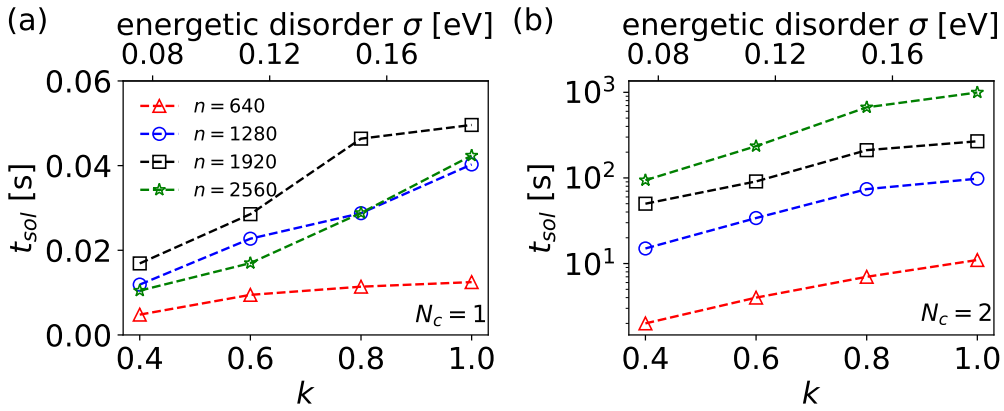

Figure 4.10: The time-to-solution (in s) using the GRW method for systems with (a) $N_c = 1$ and (b) $N_c = 2$ charge carriers and different number of sites depending on the energetic disorder, as recorded with our Python implementation on an Intel(R) Xeon(R) Gold 5120 CPU @ 2.20GHz with 28 threads.

4.4 Conclusion

Using the lattice surrogate models, this chapter validates the accuracy of ToF evaluation with the GRW method. Based on graph theory, the GRW method avoids numerically cumbersome calculations of the explicit dynamics of the carriers, and is therefore not prone to discretization problems or sampling issues as commonly used master equation or KMC approaches. It also shows that the ToF obtained from the GRW method has better accuracy compared to that obtained via KMC. We have shown that it allows for accurate and reliable predictions of the effective ToF (in the diffusive regime) and the field-dependent charge mobility (in the drift-diffusive regime) for a wide range of scenarios covering vastly different possible material properties. The GRW method will be used in the next chapter to calculate the quantity of interest (i.e., ToF) for uncertainty quantification in a multiscale-modeled organic semiconductor, MADN.

Chapter 5

Uncertainty Quantification in Multiscale Charge Transport Models

Adapted from the paper "*Uncertainty Quantification in Multiscale Models of Charge Transport in Organic Semiconductors: Influence of the Exhange-Correlation Functional*, Z. Chen, P. van der Hoorn, B. Baumeier, arXiv preprint 2503.22356 (2025)".

This chapter presents the impact of exchange-correlation functional choices on the predictive accuracy of multiscale models for charge transport in organic semiconductors. A hybrid functional approach is applied to analyze uncertainties in key parameters influencing charge mobility, focusing on the Hartree–Fock exchange fraction. Using 2-methyl-9,10-bis(naphthalen-2-yl)anthracene (MADN) as a test system, first-principle based multiscale modeling methods as introduced in Chapter 2 are used to compute reorganization energies, site energies, and electronic coupling elements. Monte Carlo sampling quantifies the uncertainty propagation, revealing that site energy variations dominate transport property uncertainties, while coupling elements contribute minimally. The findings underscore the need for accurate parameter determination and functional selection, with implications for enhancing the reliability of first-principles-based multiscale modeling frameworks in organic semiconductor design. This chapter is organized as follows: Section 5.1 introduce the source of uncertainty of our investigation, Section 5.2 presents the explicit results from the model using different values of α_{HFX}, before the results of uncertainty quantification and sensitivity analysis via Monte

Carlo sampling are shown in Section 5.3. A discussion and summary conclude this chapter.

5.1 Source of Uncertainty

In computational science, quantifying the uncertainties, or UQ, of a model sheds light on the robustness and the sensitivity of the model, and generates new insights into the results [116, 117, 118, 119, 120, 121]. As detailed in Chapter 2, one important source of uncertainties in the multiscale models is the approximation of the DFT exchange-correlation functionals, which plays an fundamental role in multiscale model methodologies.

The single molecular geometry optimization for obtaining the reorganization energy, the mixed quantum-classical methods for obtaining the site energies with the help of microelectrostatic methods [122, 123], and the dimer projection method for determining coupling elements [124] fundamentally depend on the DFT calculations. Therefore, the dependence of DFT calculations on the choice of an exchange-correlation functional raises the question of how sensitive the simulated charge transport is to this choice and how certain predictions of material properties are.

Many common uncertainty quantification studies focus on models with (partial) differential equations, e.g., drift-diffusion equations in which the diffusivity is a parameter, and assume a certain distribution for the values of the parameter(s). For the multiscale model of charge transport, it is not straightforward to cast the large variety of available exchange-correlation functionals into the role of a model parameter with some distribution. To keep the problem tractable, we focus here instead specifically on the exchange part in hybrid functionals [125, 126], in which a DFT model for the exchange is mixed using a weighting factor α_{HFX} with a Hartree–Fock type exchange, i.e.,

$$E_{\mathrm{x}} = \alpha_{\mathrm{HFX}} E_{\mathrm{x}}^{\mathrm{HF}} + (1 - \alpha_{\mathrm{HFX}}) E_{\mathrm{x}}^{\mathrm{DFT}}. \tag{5.1}$$

More specifically, we take as the basis the PBE0 functional [127] and scrutinize (i) how the predictions of the multiscale model of charge transport are affected by variation of α_{HFX} as a proxy for uncertainties in the choice of DFT functionals, (ii) what the level of confidence is in quantitative predictions, and (iii) what are the most sensitive quantities in the model. In this sense, Eq. (5.1) is deceptively simple. For each value of α_{HFX}, the graph $\mathbf{G}$ generated by the multiscale model is constructed using the respective value of the reorganization energy, the N_{mol} site energies, and the N_{pair} coupling elements, and the dimensionality of the problem from the perspective of UQ is $N_{\text{UQ}} = 1+N_{\text{mol}}+N_{\text{pair}}$, which can easily be on the order of $10^4 - 10^5$. We consider the simulated ToF and the associated mobility as the quantity of interest, which then are subject to uncertainties in these N_{UQ} parameters *stemming* from the variation in α_{HFX}. As a prototypical system, we will study hole transport in an amorphous morphology of 2-methyl-9, 10-bis(naphthalen-2-yl)anthracene (MADN), a wide-gap semiconductor that is used extensively as an ambipolar host material in organic light-emitting diodes [128, 129]. The next section will present the modeling results of this material.

5.2 Explicit Results from the Multiscale Model

In this section, we present and analyze the explicit results of the multiscale model of charge transport in amorphous MADN as obtained for different values of α_{HFX}.

5.2.1 Molecular Parameters

The molecular parameters are presented first, as they are used in different ways in the multiscale model. Figure 5.1 shows the adiabatic ionization potential, reorganization energy, neutral state dipole moment, and isotropic polarizability in neutral and cationic (hole) states, respectively. The adiabatic ionization potential in Fig. 5.1(a) contributes to the site energy, but as it is determined per molecule-type in the system, it does not affect the site-energy difference ΔE_{ij} in Eq. (4.1). It is nevertheless interesting to see that it increases almost linearly over the shown range of α_{HFX}. In contrast, the reorganization energy as shown in Fig. 5.1(b) ap-

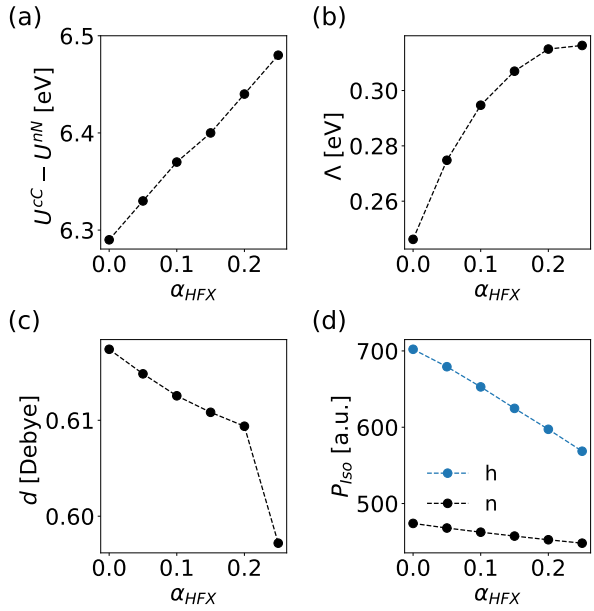

Figure 5.1: Dependence of molecular parameters as used directly in the multiscale model or in its paramaterization phase on the amount of Hartree–Fock type exchange in the PBE0-based hybrid functional α_{HFX}. (a) The adiabatic ionization potential, (b) reorganization energy, (c) dipole moment of the neutral molecule, and (d) isotropic molecular polarizability in neutral and charged states, respectively.

pears to saturate for α_{HFX} after an initially close to linear increase. In total, Λ is found to be in an interval between 0.25 eV and 0.33 eV. Panels (c) and (d) of Fig. 5.1 show the dipole moment of the neutral MADN molecule and the isotropic polarizability of the neutral and cationic (hole) states, respectively, both as electrostatic properties that enter indirectly the parameterization of the microelectrostatic model. As is visible, the dipole moment is rather independent on α_{HFX} (note that the jump of the last shown data point appears more pronounced because of the very small scale on the y-axis). The isotropic polarizabilities in panel (d) exhibit a linear decrease with increasing value of α_{HFX} which can be attributed to an increasingly attractive effective potential from stronger the Hartree–Fock-like exchange term and consequently more strongly bound electrons as one can also see from the increasing ionization potential in panel (a).

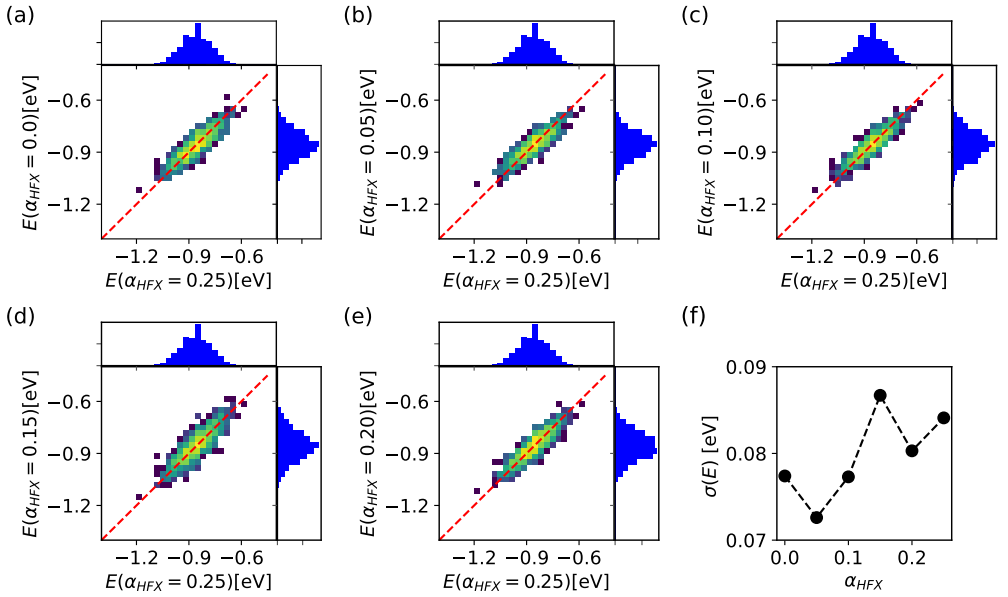

Figure 5.2: Panels (a)-(e): Scatter plot of site energies of the MADN system for different values of α_{HFX} vs the PBE0 ($\alpha_{\mathrm{HFX}} = 0.25$) reference. (f) Site energy disorder σ in eV for all values of α_{HFX}.

5.2.2 Site Energies

Figure 5.2(a)-(e) presents scatter plots comparing the distributions of the calculated site energies of MADN molecules for different α_{HFX} values, using $\alpha_{\mathrm{HFX}} = 0.25$ (PBE0 functional) as a reference in all cases. Visually, the distributions look very similar as most data points cluster near the diagonal line. In Table 5.1, we report the minimum (E_{min}) and maximum (E_{max}) values of the energies and the standard deviation σ of the individual distributions, which are also shown in Fig. 5.2(f). All noted variations in these quantities appear small. The data in Fig. 5.3 and 5.4 show that the observations on the total site energies also hold for the individual contributions of electrostatic and polarization energies. This is also reflected by the small variations in the respective standard deviations σ_{el} and σ_{polar} listed in Table 5.1.

Table 5.1: Characteristics of site energies E (all in eV) obtained for the different values of α_{HFX} : minimum (E_{min}) and maximum (E_{max}), the overall standard deviation and the contributions for static and polar calculations, as well as the largest absolute variation $D_{\alpha_{\mathrm{HFX}}}$ with respect to the PBE0 data ($\alpha_{\mathrm{HFX}} = 0.25$).

α_{HFX}	E_{min}	E_{max}	σ	σ_{el}	σ_{polar}	$D_{\alpha_{\mathrm{HFX}}}$
0	-1.11	-0.59	0.08	0.10	0.04	0.13
0.05	-1.07	-0.63	0.07	0.09	0.04	0.12
0.10	-1.13	-0.61	0.08	0.10	0.04	0.13
0.15	-1.16	-0.59	0.09	0.11	0.05	0.15
0.20	-1.12	-0.63	0.08	0.10	0.05	0.12
0.25	-1.17	-0.60	0.08	0.11	0.04	—

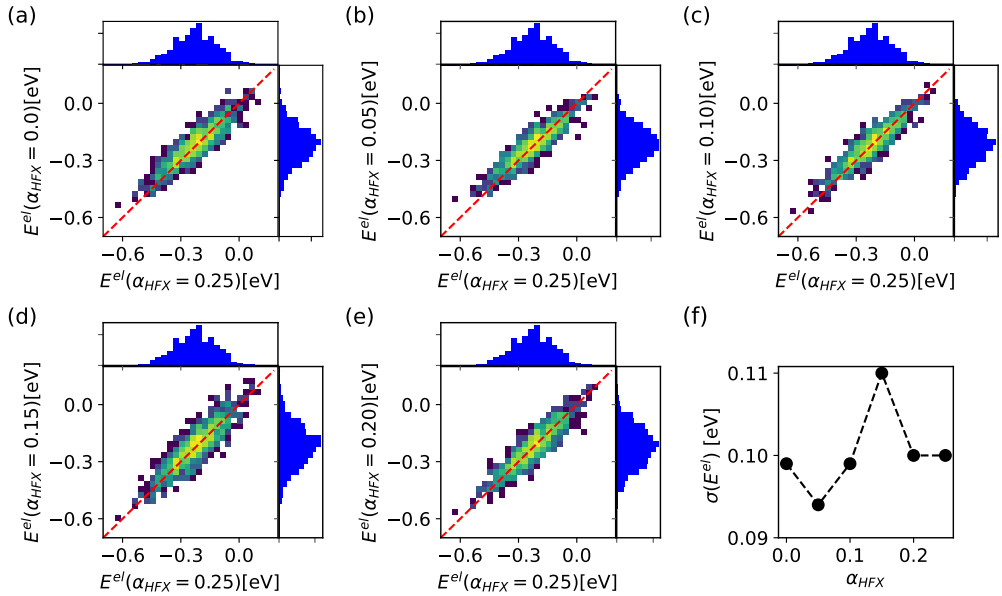

Figure 5.3: Panels (a)-(e): Scatter plot of the electrostatic contribution to site energies of the MADN system for different values of α_{HFX} vs the PBE0 ($\alpha_{\mathrm{HFX}} = 0.25$) reference. (f) Site energy disorder σ in eV for all values of α_{HFX}.

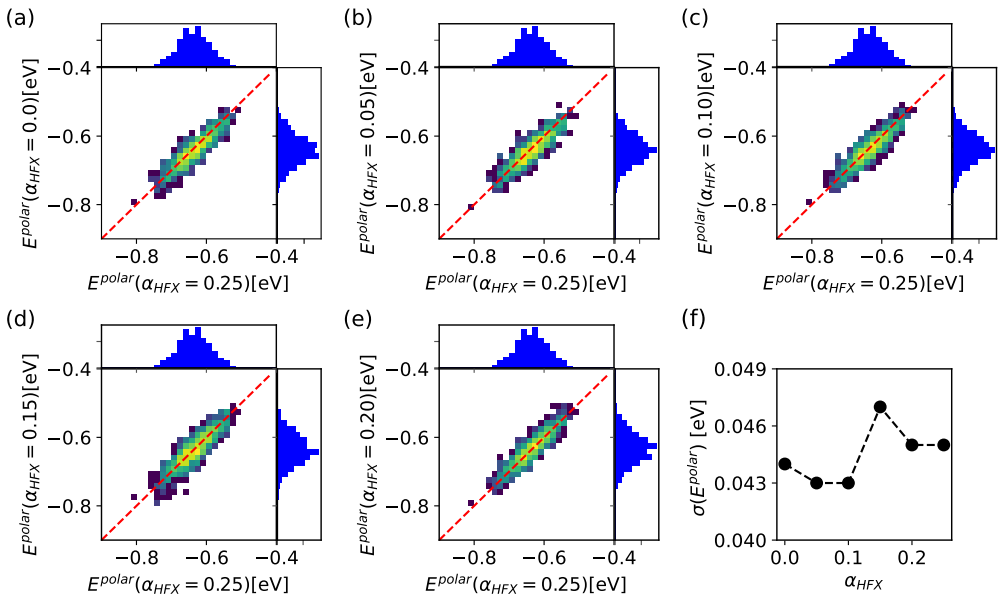

Figure 5.4: Panels (a)-(e): Scatter plot of polarization contribution to site energies of the MADN system for different values of α_{HFX} vs the PBE0 ($\alpha_{\mathrm{HFX}} = 0.25$) reference. (f) Site energy disorder σ in eV for all values of α_{HFX}.

Interestingly, we find the biggest variation of 0.1 eV in the values for E_{min}. This minimal energy (or rather all the energies in the low-energy region of the distributions) can have a very significant influence on the charge transport properties, depending on how much the site energy differences are influenced in the charge transport network. To further characterize the effect of different α_{HFX} values, we consider the largest absolute variation in the site energies depending on α_{HFX} by calculating the Chebyshev distance $D_{\alpha_{\text{HFX}}} = \max_i(|E_i(\alpha_{\text{HFX}} = 0.25) - E_i(\alpha_{\text{HFX}})|)$. From the results listed in Table 5.1, one can see that on individual level the site energies can vary substantially, with more than 0.1 eV. This variation is larger than the recorded values of σ and with all α_{HFX} yielding the same mean site energy of -0.86 eV this could indicate non-neglibible effects on charge transport.

5.2.3 Distributions of Electronic Couplings

Distributions and correlations compared with the PBE0 reference for the electronic coupling elements calculated with different values for α_{HFX} are shown in Fig. 5.5. The individual distributions appear very similar, with a peak of $\log_{10}[(J_{ij}/\text{eV})^2]$ between -5 to -6 , and a long tail of the distribution towards more negative values. The comparison with the PBE0 reference shows that while there is a clear correlation between the results for different values of α_{HFX}, the spread in the order of magntitudes of J_{ij}^2 can be very large especially for the lower coupling regions. Overall, the squared electronic coupling elements are found in a very wide range from 10^{-2} to $10^{-15}\,(\text{eV})^2$ due to its exponential distance dependence and sensitivity to mutual orientation of the two involved molecules. Whereas the site energy distributions discussed in the preceding section are well-defined in the sense that each energy is unambiguously associated to a physical entity – a molecule in the morphology, the coupling elements are evaluated for a neighborlist with the closest contact distance less than $0.5\,\text{nm}$. Clearly, if this cutoff is chosen to be large, a lot of hopping pairs with very small coupling elements will be considered that may not be relevant at all (or even unphysical) for charge transport. Therefore, the increasing deviations for the most negative values in Fig. 5.5 may not be relevant either. To determine the range of $\log_{10}[(J_{ij}/\text{eV})^2]$ that are significant for the charge dynamics, a percolation analysis is performed to find a critical threshold value for the squared electronic coupling below which

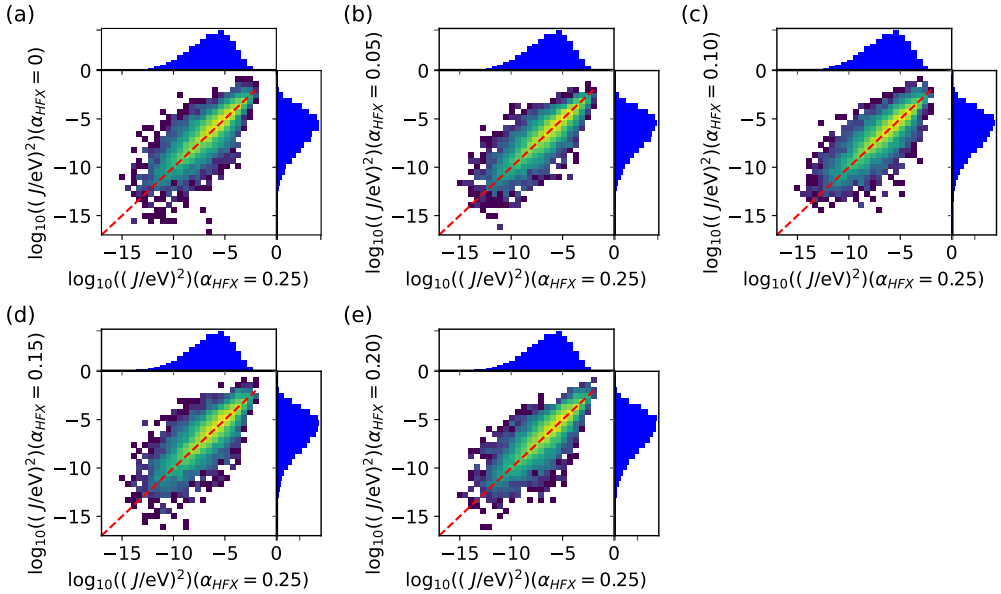

Figure 5.5: Scatter plot of MADN coupling element $\log_{10} J^2$ calculated from different HFX, compared to the polarization energy calculated from HFX=0.25 (The PBE0 functional). The brighter color near the diagonal lines indicates a denser population of the molecules. The top and right histograms show the energy distributions.

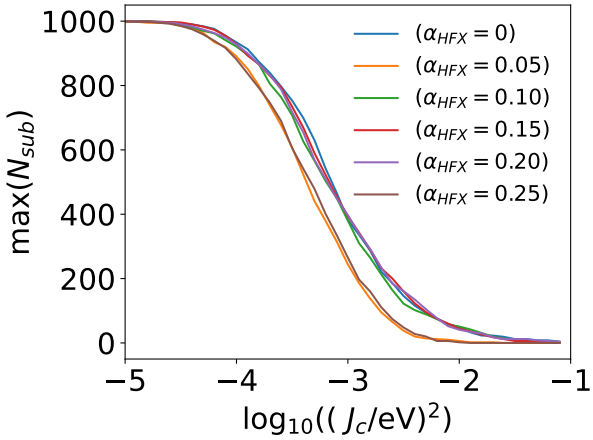

Figure 5.6: The maximum size of the subgraphs obtained in the percolation algorithm as a function of critical value J_c.

the largest connected subgraph is identical to the full graph. This is achieved by removing for a given value for J_c from the full graph the edges with $J_{ij}^2 < J_\mathrm{c}^2$ followed by the determination of the number of vertices in the larges connected subgraph $\max(N_\mathrm{sub})$. Figure 5.6 shows the resulting dependence of $\max(N_\mathrm{sub})$ on the critical value J_c. It is apparent that for $\log_{10} J_\mathrm{c}^2 = -5$, all vertices are in the largest connected subgraph. The charge transport network's overall connectivity is mostly affected by coupling elements larger than this threshold, and the charge transport properties are expected to be more sensitive to deviations for the associated edges. We also note that overall, the results in Fig. 5.6 again seem to be very similar for all α_HFX studied in the work.

5.2.4 ToF Calculations

From the analysis of reorganization energies, site energies, and electronic coupling elements for different values of α_HFX, it is not clear how the mostly on distribution-level observed variations impact the overall charge transport properties. To scrutinize the dependence of such a material property, we now calculate the ToF τ and report the respective values in Table 5.2. Here, we refer

Table 5.2: Simulated ToF (in s) for different values of α_{HFX} in the full graph, with an additional cutoff on J^2, and for the case with no energetic disorder, respectively.

α_{HFX}	full	$J_{ij}\mathbb{1}(\log_{10} J_{ij}^2 > -5)$	$\Delta E_{ij} = 0$
0.00	$6.4 \cdot 10^{-9}$	$7.5 \cdot 10^{-9}$	$1.9 \cdot 10^{-10}$
0.05	$7.9 \cdot 10^{-9}$	$8.1 \cdot 10^{-9}$	$4.1 \cdot 10^{-10}$
0.10	$1.6 \cdot 10^{-8}$	$1.6 \cdot 10^{-8}$	$4.0 \cdot 10^{-10}$
0.15	$3.0 \cdot 10^{-8}$	$3.1 \cdot 10^{-8}$	$4.0 \cdot 10^{-10}$
0.20	$2.1 \cdot 10^{-8}$	$2.5 \cdot 10^{-8}$	$4.5 \cdot 10^{-10}$
0.25	$9.5 \cdot 10^{-8}$	$9.7 \cdot 10^{-8}$	$7.2 \cdot 10^{-10}$

with "full" to the ToF obtained for the as-calculated charge transport network. One can see that τ varies by roughly one order of magnitude between $\alpha_{\mathrm{HFX}} = 0$ ($6.3 \cdot 10^{-9}$ s) and $\alpha_{\mathrm{HFX}} = 0.25$ ($9.5 \cdot 10^{-8}$ s), with an almost monotonous increase. When the squared coupling elements with values below $10^{-5}\,(\mathrm{eV})^2$ are set to zero ($J_{ij}\mathbb{1}(\log_{10} J_{ij}^2 > -5)$), one obtains only minimally, but consistently, larger τ, corroborating the notion that the very small coupling elements are of little relevance for charge transport. Finally, we also consider the case in which the energetic disorder is ignored ($\Delta E_{ij} = 0$). Here, one also can see (next to the generally shorter ToF) a consistent increase in τ, however only by a factor of about 3.8 , a combined effect of the increased reorganization energy and variations in coupling elements.

5.3 Uncertainty Quantification and Sensitivity Analysis

The previous section shows that different HFX affect the calculated ToF. In this section, we use the Monte Carlo method to estimate the range of the ToF given a confidence level, followed by a sensitivity analysis to determine which parameter contributes more to the variance of ToF.

5.3.1 ToF of Diffusion

As discussed the ToF is calculated from the graph with weighted edges defined by the electronic structure data $(\Lambda, E_1, \ldots, E_N, J_1^2, \ldots, J_{N_\mathrm{p}}^2)$, where N is the number of molecules (vertices) and N_p the number of molecule pairs (edges) in the neighbor list. We can therefore consider that $\tau = \tau(\mathbf{x})$ with

$$\mathbf{x} = (\Lambda, E_1, \ldots, E_N, \log_{10}(J_1^2), \ldots, \log_{10}(J_{N_\mathrm{p}}^2))$$

as dependent on $M = 1 + N + N_\mathrm{p}$ parameters with uncertainty, in this case stemming from the possible choices of α_HFX in the exchange-correlation functional of the multiscale model. Note that due to the data range for the electronic coupling element as discussed in the previous section, we consider $\log_{10}(J^2)$ instead of J^2. For MADN, we have $N = 1000$ and $N_\mathrm{p} = 8764$ so that $M = 9765$.

Because the precise distribution of the uncertainties of the M parameters is not known, we consider the maximum amount of uncertainties from a maximum likelihood estimation based on the data obtained from the explicit results for the different values of α_HFX and the assumption of normal distributions. In other words, each of the parameters x_i for $i = 1, \ldots, M$ is assigned a normal distribution with $\mathcal{N}(\mathbb{E}(x_i), \mathbb{V}(x_i))$, with $\mathbb{E}(x_i)$ ($\mathbb{V}(x_i)$) the mean (variance) of the respective data. Then, Monte-Carlo sampling is used to obtain $N_\mathrm{MC} = 50000$ different realizations of $\mathbf{x}$, the respective ToF is calculated from these samples, and the resulting distribution $P(\log_{10}(\tau))$ is statistically analyzed. Specifically, we consider four different settings: in three settings, only one of the parameters blocks $(\Lambda, \{E\}, \{\log_{10}(J^2)\})$ is sampled while the values of the other blocks are set to their respective mean values, and we denote the resulting distributions as $P_\Lambda(\log_{10}(\tau))$, $P_E(\log_{10}(\tau))$, and $P_J(\log_{10}(\tau))$, respectively. In the fourth setting, all parameters are sampled at the same time, yielding $P_\mathbf{x}(\log_{10}(\tau))$. As these distributions do not necessarily follow any specific analytic form, we estimate a 99% confidence interval around the median using the equal-tailed percentile method.

The four respective distributions of τ in the MADN system are shown in Fig. 5.7 together with the indicated confidence intervals obtained from Monte-Carlo sampling. Black dots at the x-axes are the data from the explicit calculations for the six α_HFX values discussed before. Considering first the three distributions with single

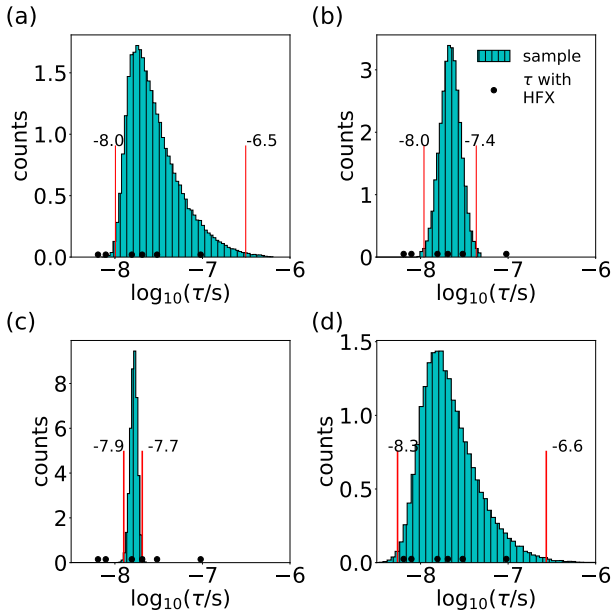

Figure 5.7: Distribution of ToFs in the MADN system from Monte Carlo sampling with a sample size $N_{\mathrm{MC}} = 50000$. The red vertical lines indicate the lower bound and upper bound of the 99% confidence interval around the median. The black circles indicate the ToF obtained using explicit α_{HFX} values. Uncertainties considered are (a) in the reorganziation energy, (b) in the site energy, (c) in the coupling elements, and (d) in all simultaneously.

property sampling in panels (a)-(c), we find the confidence intervals for $\log_{10}(\tau)$ of $[-8.0, -7.4]$ for uncertain Λ, $[-8.0, -6.5]$ for uncertain E, and $[-7.9, -7.7]$ for uncertain J, respectively. Interestingly, all three share very similar lower limits of the confidence interval, while the upper limits and the shape of the distributions differ significantly. While $P_\Lambda(\log_{10}(\tau))$ and $P_J(\log_{10}(\tau))$ appear relatively symmetric, indictaing possible a log-normal distribution of τ, $P_E(\log_{10}(\tau))$ is distinctly asymmetric which might point to a normal distribution of τ for uncertain energies. From the three distributions and the confidence intervals one can see that the effect of the uncertainty in the coupling elements on the ToF is small (factor of 2 on τ). The effect of uncertainty in the reorganization energy is only slightly higher (factor of four). Uncertainty in the site energies, in contrast, has a much bigger impact on the estimates of τ. From the distribution of $P_{\mathbf{x}}(\log_{10}(\tau))$

in Fig. 5.7(d), one can see than when all parameters are sampled, the result is dominated by the effects from the uncertainty in the site energies, with an overall similar shape and a confidence interval of $[-8.3, -6.6]$, corresponding to a range from about 5 ns to about 250 ns in τ.

We note that the above analysis from sampling via independent distributions of the uncertain parameters assumes that the parameters themselves are independent or uncorrelated. Correlations in organic semiconductors are typically found among the site energies when the molecules are polar [39, 130]. MADN has a very low dipole moment, and we find practically no spatial correlations between the site energies, making the sampling approach suitable for this material system.

Besides the inspection on the distributions for various uncertain parameter (sets) as above, one can also quantify the sensitivity of the time-of-fight to the respective parameters by evaluating their contributions to the overall observed variance in τ. One way of decomposing the variance of the model output into fractions attributed to input parameters is the variance-based sensitivity analysis using Sobol indices [131].

To measure the parameter x_i's contribution to $\tau = \tau(\mathbf{x})$ including all variance caused by its interaction with other parameters $\{x_k, k \neq i\}$, the total Sobol index $S_{T,i}$ is calculated as

$$S_{T,i} = \frac{\mathbb{E}_{\mathbf{x}_{\sim i}}[\mathbb{V}_{x_i}(\tau|\mathbf{x}_{\sim i})]}{\mathbb{V}(\tau)}. \tag{5.2}$$

Here, $\mathbf{x}_{\sim i}$ denotes the vector of all entries of $\mathbf{x}$ but x_i. The variance of τ given a set of $\mathbf{x}_{\sim i}$ taken over x_i is $\mathbb{V}_{x_i}(\tau|\mathbf{x}_{\sim i})$ and $\mathbb{E}_{\mathbf{x}_{\sim i}}[\cdot]$ denotes the mean of argument $(\cdot)$ taken over all factors but x_i. We then consider based on the structure of the vector $\mathbf{x}$ the Sobol index for the reorganization energy as $S_{T,\Lambda} = S_{T,1}$, for the site energies as $S_{T,E} = \sum_{i=2}^{N+1} S_{T,i}$, and for the coupling elements as $S_{T,E} = \sum_{i=N+2}^{N+1+N_\mathrm{p}} S_{T,i}$. Using the quasi Monte Carlo method[132] with a sample size $N_\mathrm{QMC} = 1000$ to calculate $S_{T,i}$, we find that $S_{T,\Lambda} = 0.097$, $S_{T,E} = 0.950$ and $S_{T,J} = 0.028$. The Sobol indices corroborate the observation from the confidence intervals that the uncertainty in the site energies gives the dominant contribution to the variance of the simulated ToF. Contributions from the reorganization energy and the coupling

elements are both small in relation, with the one of Λ slightly larger than the one of J. This is also in line with the general observations on the distributions in Fig. 5.7.

5.3.2 Drift-diffusion Charge Mobility

When an external electric field, $\mathbf{F}$, is applied to the system, the transport is not just diffusive but a drift-diffusion process. The material property of interest is then the drift mobility evaluated as $\mu = \frac{\mathbf{v}\cdot\mathbf{F}}{|\mathbf{F}|^2}$ where $\mathbf{v} = \mathbf{L}/\tau$ is the effective velocity of the charge carrier and $\mathbf{L}$ is the vector connecting the initial and final positions of the charge carrier. Here, we perform calculations for six different field directions (positive and negative Cartesian directions) with the source and sink conditions set accordingly. The electric field is taken into account in the Marcus rate via an additional term in ΔE_{ij} given by $e\mathbf{F}\cdot\mathbf{r}_{ij}$, where $\mathbf{r}_{ij}$ connects the center-of-masses (COMs) of molecules i and j (see Eq. (4.1)). Finally, we define for a certain field strength $F = |\mathbf{F}|$ the ToF drift mobility $\mu(F)$ is as the average over the six directions.

We perform such mobility calculations for the six explicit α_{HFX} models as well as the Monte-Carlo sampled uncertainty for six different values of F. The results are shown in Fig. 5.8. While the distributions shift depending on the strength of the applied field, the overall shape and the width of the confidence intervals is very similar. The width of the confidence intervals in $\log_{10}(\mu)$ decreases from 0.9 at low fields ($4\cdot10^7\,\mathrm{V/m}$) to 0.7 at a field of $9\cdot10^7\,\mathrm{V/m}$, corresponding broadly to a width of close to one order of magnitude in μ. A stronger drift component due to a larger external electric field seems to make the mobility slightly less sensitive to uncertainties.

Figure 5.9 shows the plot of the charge mobility as a function of $\sqrt{F}$ to reveal the electric field dependence of the type $\mu(F) = \mu_0 \exp(\beta\sqrt{F})$ as predicted by Poole and Frenkel [133]. We show the explicit results for the six chosen α_{HFX} values together with the confidence interval estimated from the uncertainty quantification via Monte-Carlo sampling. In Table 5.3 we also summarize the parameters μ_0 and β extracted from fits of the explicit data to the Poole–Frenkel expression.

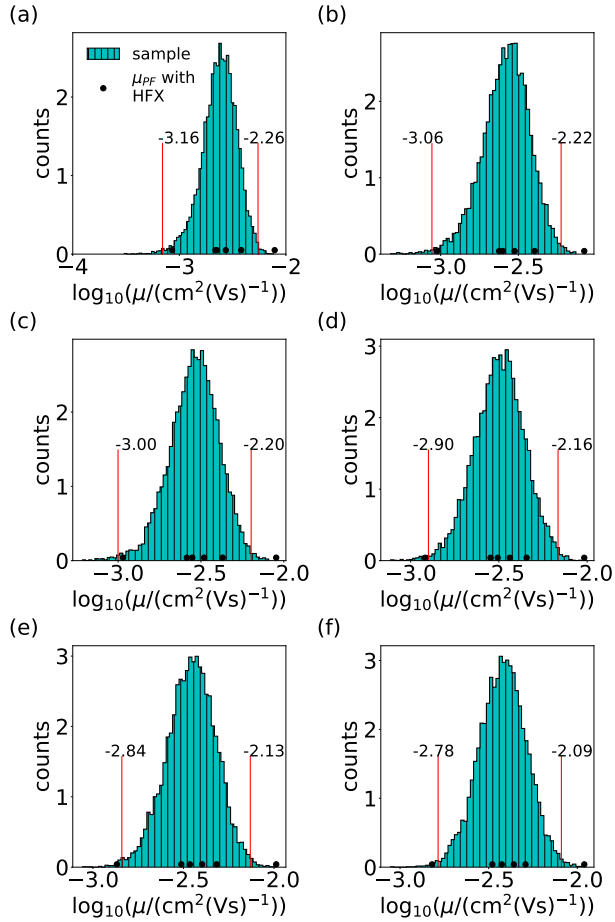

Figure 5.8: Distribution of the mobility μ for different values of the externally applied electric field F as obtained from uncertainty quantification using Monte Carlo sampling with a sample size $N_{\mathrm{MC}} = 10000$. The red vertical lines indicate the lower bound and upper bound of the 99% confidence interval around the median. (a) $|\mathbf{F}| = 4 \cdot 10^7 \mathrm{V/m}$, (b) $|\mathbf{F}| = 5 \cdot 10^7 \mathrm{V/m}$, (c) $|\mathbf{F}| = 6 \cdot 10^7 \mathrm{V/m}$, (d) $|\mathbf{F}| = 7 \cdot 10^7 \mathrm{V/m}$, (e) $|\mathbf{F}| = 8 \cdot 10^7 \mathrm{V/m}$, (f) $|\mathbf{F}| = 9 \cdot 10^7 \mathrm{V/m}$.

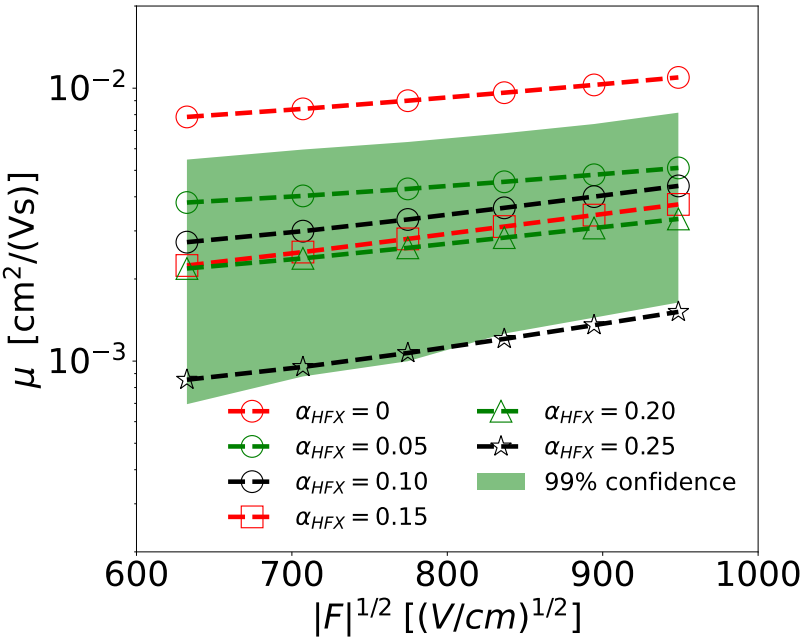

Figure 5.9: Electric-field dependence of the mobility μ in the multiscale modeled MADN system. The dash lines are Poole–Frenkel plots obtained with specific α_{HFX} values, and the green shaded area indicates the 99% confidence interval estimated from the MC sampling with a sample size of $N_{\mathrm{MC}} = 10000$.

Table 5.3: Poole-Frenkel parameters μ_0 (in cm^2/(Vs)) and β (in $\sqrt{\mathrm{cm/V}}$) of the multiscale modeled MADN, calculated by the six different α_{HFX} values.

α_{HFX}	μ_0	β
0	$4.0 \cdot 10^{-3}$	$1.0 \cdot 10^{-3}$
0.05	$2.1 \cdot 10^{-3}$	$9.3 \cdot 10^{-4}$
0.10	$1.0 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$
0.15	$8.0 \cdot 10^{-4}$	$1.6 \cdot 10^{-3}$
0.20	$9.3 \cdot 10^{-4}$	$1.3 \cdot 10^{-3}$
0.25	$2.6 \cdot 10^{-4}$	$1.8 \cdot 10^{-3}$

The values for μ_0 vary from $2.6 \cdot 10^{-4}\,\mathrm{cm}^2/(\mathrm{Vs})$ to $4.0 \cdot 10^{-3}\,\mathrm{cm}^2/(\mathrm{Vs})$, roughly one order or magnitude depending on the chosen α_{HFX}, in line with the uncertainty from the Monte-Carlo sampling. The Poole–Frenkel slope β in contrast is found to be much less sensitive to uncertainties, with recorded values between $9.3 \cdot 10^{-4}\,\sqrt{\mathrm{cm/V}}$ and $1.8 \cdot 10^{-3}\,\sqrt{\mathrm{cm/V}}$.

5.4 Discussion and Conclusion

Results in this chapter show that, by systematically varying the Hartree–Fock exchange fraction in hybrid functionals, the site energy uncertainties have a dominant influence on the variability of the ToF and mobility predictions. In contrast, the uncertainties in reorganization energies and electronic coupling elements are shown to have a relatively minor effect. This suggests that further efforts to enhance the accuracy of site energy calculations – potentially through improved parameterization or advanced quantum mechanical methods – could significantly improve the reliability of multiscale modeling predictions. Interestingly, the results also demonstrate that the uncertainties in electronic coupling elements, particularly those with very small magnitudes, are less relevant for overall charge transport. This aligns with the percolation analysis, which indicates that coupling elements below a critical threshold contribute minimally to the connectivity of the charge transport network. Thus, simplifying approximations in handling these low-value couplings could be justified in specific scenarios to reduce computational complexity without compromising accuracy.

Another notable finding is the robustness of field-dependent mobility predictions. Despite uncertainties in the underlying molecular parameters, the electric field dependence of mobility exhibits relatively consistent Poole–Frenkel behavior across the tested range of α_{HFX} values. This robustness suggests that multiscale models retain predictive value for field-dependent trends, even when parameter uncertainties are present. However, the results also highlight the challenge of achieving quantitative accuracy in charge mobility predictions, given that the uncertainty range spans an order of magnitude. This raises important considerations for interpreting simulation results in the context of experimental data. For

instance, small differences in predicted mobility between two materials or designs may not be statistically significant given the intrinsic uncertainties.

Chapter 6

Trap Identification in Charge Transport Networks

Adapted from the paper "*Automatic Identification of Traps in Molecular Charge Transport Networks of Organic Semiconductors*, Z. Chen, P. van der Hoorn, B. Baumeier, arXiv preprint 2411.07136 (2025)".

In this chapter, we will present the method that automatically identifies traps or regions of traps in a multiscale-model charge transport network. Our proposed method builds on the random walk model of charge dynamics on a directed, weighted graph, the molecular transport network. It comprises an effective heuristic to determine the number of traps or trap clusters based on the eigenvalues and eigenvectors of the random walk Laplacian matrix and uses subsequent spectral clustering techniques to identify these traps. In contrast to currently available methods, ours enables the identification of trap molecules in organic semiconductors without having to explicitly simulate the charge dynamics.

Using the charge transport network resulting from a multiscale model of an amorphous morphology of bathocuproine (BCP) [134], a molecular material with known high energetic disorder [135] and complex charge trapping behavior, we demonstrate that our approach successfully identifies both a single trap, multiple distributed traps, and a combination of a single-molecule trap and trap regions on an equal footing. We also find a strong relation between the cost function

associated with the normalized cut and the charge-carrier dynamics simulated in a time-of-flight setup, as well as the physical characteristics of the trap (regions).

This chapter is organized as follows: the traps characteristics in organic semiconductors and potential identification methods in the literature are briefly reviewed in Section 6.1. Section 6.2 introduced the graph-theoretic decomposition (GD) method used in this work, and Section 6.3 gives the details of the spectral clustering-based trap identification method we propose in this work, including the determination of the cluster number and K-means clustering. The results of the application of this method to a molecular network (consisting of BCP molecules) and its modifications to cover different trap types are presented and discussed in Section 6.5 to 6.8. Section 6.9 presents the application of the proposed method on the systems where traps can not be found purely based on the lowest energy.

6.1 Traps in Organic Semiconductors

Depending on the materials, traps can be defect-like single molecules or clusters of several neighboring ones, and can have a significant impact on the dynamics of charge carriers. On a macroscopic level, traps are often considered in the literature in terms of the energy density of state (DOS) $p(E)$, typically assumed to be Gaussian curves or exponential. In equilibrium, the mean energy of a charge carrier in the DOS is

$$E_\infty = \frac{\int_{-\infty}^{\infty} E g(E) p(E) dE}{\int_{-\infty}^{\infty} g(E) p(E) dE}. \tag{6.1}$$

Here $g(E) = [\exp(\frac{E-E_\mathrm{F}}{k_\mathrm{B}T}) + 1]^{-1}$ is the Fermi-Dirac distribution with the Fermi energy E_F determined by $\int_{-\infty}^{\infty} g(E)p(E)dE = N_\mathrm{c}$, with N_c being the number of charge carriers. A visualization that qualitatively shows the trap energies' location is shown in Fig. 6.1. Molecules with energies much lower than E_∞ are then considered as (deep) traps. However, such a qualitative criterion is insufficient to identify traps in a molecular charge transport network for several reasons: First, the estimate of E_∞ is based on a chosen model DOS, which has some assumed continuous distribution. A realistic material, even on the scale 100 nm, will, how-

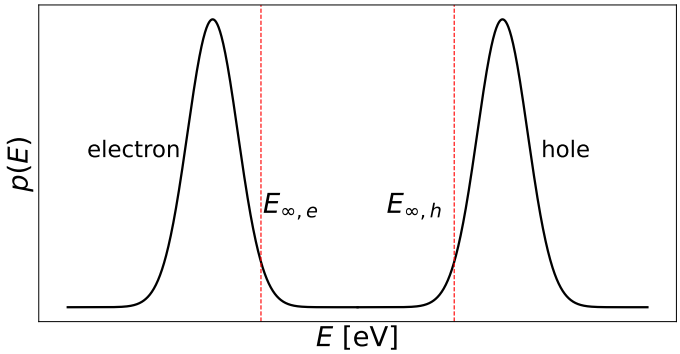

Figure 6.1: Density of state function of OSC assuming a Gaussian model. The bell curve annotated as "hole/electron" indicates the range of transport energies for hole/electron. The red dashed lines indicate the location of equilibration energy $E_{\infty,h}$ $E_{\infty,e}$ for hole and electron.

ever, not exhibit such a continuous DOS. Second, a discrete version of Eq. (6.1) depends on the number of molecules in the system, and the equilibrium energy in such a discrete DOS is dependent on system size [136]. Third, focusing on the DOS alone ignores other contributing factors to the charge dynamics, or the features of the transport network, such as electronic coupling elements between pairs of neighboring molecules, structural details of the material and or spatial correlations. These details are connected to the variety of physical sources for traps, e.g., interfacial effects, defects in molecular packing, or chemical impurities. This makes it difficult to provide a quantitative definition of traps that can be used for identification. The bi-molecular transfer rate (such as the Marcus rate which is introduced in Chapter 2) cannot be used to identify traps. If multiple connected molecules form a trap region, the rates within the region can be very large although the rates jumping out of the region are small, resulting in the charge carrier spending a long time within the region [137].

At present, there are no methods for the identification of traps in molecular charge transport networks that perform reliably for all different trap types. Few attempts have been reported in identifying trap regions, or clusters, based on analyzing the actual simulated dynamics, e.g., via kinetic Monte Carlo (KMC) [104, 77].

Qualitatively, once entered into such a trap region, the random walk (representing the charge dynamic of a single carrier) transitions mostly within it and escaping it is a rare event, making such KMC simulations very time-consuming. Two methods to accelerate KMC simulations which indirectly involve trap identification have previously been discussed. One is based on the (stochastic) watershed algorithm filling regions ("basins") in the spatially resolved energy distribution [104]. This purely energy-based criterion does, however, not consider additional details of the factors influencing the molecular charge transport network. The second method [138] is based on a graph-theoretic decomposition and makes use of the fact that in the presence of trapping regions the Markov chain on the molecular charge transport network is nearly completely decomposable [139], allowing the associated graph to be partitioned into subgraphs. While this method takes the full information of the hopping-type dynamics into account, it is sensitive to the choice of parameters (related to, e.g., graph connectivity properties or transition rate ratios) and is not successful in identifying single trap nodes in the graph (as will be also discussed in later sections). The next section will introduce the details of the GD method as implemented in this work.

6.2 Graph-theoretic Decomposition Method

The GD method proposed in [138] is used to speed up the KMC simulation by identifying and grouping regions of traps (problem regions) where the walkers become stuck. This accelerated KMC is named aggregated Monte Carlo simulation [115]. The GD method used in identifying the problem regions is a potential method to identify traps and calculate mobility with less computational effort. The GD algorithm begins by taking a vertex of minimal vertex degree in $\mathbf{G}$ and uses this vertex as the basis of a cluster. Then each node v' adjacent to v is checked, and added to $\bar{\mathbf{G}}$ if certain criteria are satisfied. The criteria for including a node into a cluster are:

1. Either a completeness criterion or a fullness criterion. 1) The completeness criterion requires that $\frac{R(\bar{\mathbf{G}}\cup\{v'\})}{R(\bar{\mathbf{G}})} > \alpha$ for some $\alpha > 0$, where $R(\bar{\mathbf{G}})$ is the ratio of the number of edges in the graph $\bar{\mathbf{G}}$ to the number of edges that $\bar{\mathbf{G}}$ would have if it were complete. In the case of $\bar{\mathbf{G}}$ consist of only one node,

Algorithm 2 Graph-Theoretic Decomposition

1: put $\mathbf{C} = \mathbf{V}$, set $j = 1$.
2: Put $\mathbf{S} = \mathbf{S}' = \emptyset$
3: Choose from $\mathbf{C}$ a vertex v of minimal degree, mark it and add it to $\mathbf{S}$.
4: Move to $\mathbf{S}'$ all vertices adjacent to v.
5: Choose a vertex s' in $\mathbf{S}'$.
6: If the fullness or connectivity criterion is satisfied and the threshold criterion is satisfied, then move s' to $\mathbf{S}$ and add to $\mathbf{S}'$ all vertices in $\mathbf{C}$ adjacent to s'. Otherwise, move s' to $\mathbf{C}$.
7: If $\mathbf{S}' \neq \emptyset$, repeat from step 5.
8: Put aside the vertices in $\mathbf{S}$ as superstate V_j.
9: If $\mathbf{C} \neq \emptyset$, set $j = j + 1$ and repeat from step 2.

$R(\bar{\mathbf{G}}) = 1$ since $\bar{\mathbf{G}}$ is complete. 2) The fullness criterion requires that the external node v' be adjacent to at least a proportion β of nodes in the cluster $\bar{\mathbf{G}}$ for some $\beta > 0$.

2. A threshold criterion. This requires that at least one transition probability from v' into a node in $\bar{\mathbf{G}}$ be bigger than γ and that at least one transition probability from $\bar{\mathbf{G}}$ to v' be bigger than γ for some $\gamma > 0$.

The process is repeated until no more nodes can be added to $\bar{\mathbf{G}}$. At that stage, the nodes of subgraph $\bar{\mathbf{G}}$ are removed from $\mathbf{G}$ and classified as a cluster. The algorithm begins again until all nodes are classified.

Starting from a graph $\mathbf{G} = (\mathbf{V}, \mathbf{E})$ The GD algorithm is summarized in Algorithm 2

6.3 Spectral Clustering Method

In this section, we present the theoretical background of spectral clustering for trap identification, including a recapitulation of details about the random walk Laplacian matrix, the graph partitioning based on its eigenvalues and eigenvec-

tors, and the K-means clustering algorithm. We will also propose our heuristics for determining the number of traps, within this framework.

6.3.1 Random Walk Laplacian Matrix and Graph Partitioning

Identifying traps within a multiscale modeled molecular charge transport network corresponds to finding the regions of nodes in the corresponding graph in which the random walk process takes a long time, while the transition between the regions is rare.

In spectral clustering theory, see for example [140] Proposition 5, finding such regions of nodes corresponds to cutting through the graph such that the resulting partitions $\mathbf{G}_1, \cdots, \mathbf{G}_k$ minimize the objective function

$$\text{NCut}(\mathbf{G}_1, \cdots, \mathbf{G}_k) := \sum_{i=1}^{k} \frac{\text{cut}(\mathbf{G}_i, \bar{\mathbf{G}}_i)}{\text{vol}(\mathbf{G}_i)}, \tag{6.2}$$

where $\text{cut}(\mathbf{G}_i, \bar{\mathbf{G}}_i) = \frac{1}{2}(W_{\mathbf{G}_i, \bar{\mathbf{G}}_i} + W_{\bar{\mathbf{G}}_i, \mathbf{G}_i})$, with $W_{\mathbf{A},\mathbf{B}} := \sum_{i \in \mathbf{A}, j \in \mathbf{B}} \omega_{ij}$, and $\text{vol}(\mathbf{G}_i)$ the volume $\mathbf{G}_i$, calculated by summing up the weights of the edges within $\mathbf{G}_i$.

In the context of trap identification, the partition that minimizes Eq. (6.2) will correspond to the traps in the system and the remaining molecules. That is, we have $K = k - 1$ traps and one element $\mathbf{G}_i$ will represent all non-trap nodes.

The problem is that solving Eq. (6.2) is known to be an NP-hard problem. However, a relaxation of it can be solved using the so-called *random walk Laplacian* matrix. For a weighted graph $\mathbf{G} = (\mathbf{V}, \mathbf{W})$ define the out-degree matrix by $\underline{\mathbf{D}} := D_{i,i} = \sum_j \omega_{ij}$, and the Laplacian matrix by $\underline{\mathbf{L}} = \underline{\mathbf{D}} - \mathbf{W}$. The random walk Laplacian matrix of a graph $\mathbf{G}$ is given by

$$\underline{\mathbf{L}}_{\text{rw}} = \underline{\mathbf{I}} - \underline{\mathbf{D}}^{-1}\mathbf{W}, \tag{6.3}$$

So $\underline{\mathbf{L}}_{\text{rw}}$ is the Laplacian matrix normalized by $\underline{\mathbf{D}}$.

Note that the charge transport graph is directed and hence the Laplacian matrix is not symmetric in general. Nevertheless, the random walk Laplacian $\underline{\mathbf{L}}_{\text{rw}}$

has a real-valued spectrum. This follows from the fact that

$$\underline{\mathbf{L}}_{\text{rw}} = \underline{\mathbf{I}} - \underline{\mathbf{D}}^{-\frac{1}{2}}(\underline{\mathbf{I}} - \underline{\mathbf{L}}_{\text{sym}})\underline{\mathbf{D}}^{-\frac{1}{2}}, \tag{6.4}$$

where $\underline{\mathbf{L}}_{\text{sym}} = \underline{\mathbf{I}} - \underline{\mathbf{D}}^{-\frac{1}{2}}\mathbf{W}\underline{\mathbf{D}}^{-\frac{1}{2}}$ is the normalized symmetric Laplacian. So $\underline{\mathbf{L}}_{\text{rw}}$ is similar to $\underline{\mathbf{L}}_{\text{sym}}$, and since the latter has a real-valued spectrum, $\underline{\mathbf{L}}_{\text{rw}}$ has N real eigenvalues: $\lambda_1 = 0 \leq \lambda_2 \leq \lambda_3 \leq \cdots \leq \lambda_N$. The fact that $\underline{\mathbf{L}}_{\text{rw}}$ has real eigenvalues enables the use of spectral clustering methods for our directed charge transport network.

The idea behind spectral clustering is to consider the relaxation of the NP-hard discrete minimization problem for NCut Eq. (6.2), which tries to find $\underline{\mathbf{T}} \in \mathbb{R}^{N\times k}$ that minimizes

$$\min_{\underline{\mathbf{T}}\in\mathbb{R}^{N\times k}} \text{Tr}(\underline{\mathbf{T}}'\underline{\mathbf{D}}^{-\frac{1}{2}}\underline{\mathbf{L}}_{\text{rw}}\underline{\mathbf{D}}^{-\frac{1}{2}}\underline{\mathbf{T}}) \tag{6.5}$$

subject to the constraint that $\underline{\mathbf{T}}'\underline{\mathbf{T}} = \underline{\mathbf{I}}$. The solution $\underline{\mathbf{T}}^*$ to Eq. (6.5) is formed by the first k eigenvectors of random walk Laplacian matrix $\underline{\mathbf{L}}_{\text{rw}}$.

It should be noted that solving Eq. (6.5) does not yield a partition of the graph, as this is a relaxation of the true discrete optimization problem we wish to solve. To construct a partition in practice a K-means algorithm, introduced in the next subsection, is performed on the rows of solution $\underline{\mathbf{T}}^*$ to Eq. (6.5).

6.3.2 K-means Clustering Algorithm

In general, the K-means clustering method partitions a dataset consisting of N data points into distinct clusters, minimizing the distance between points in each cluster. Let $k \geq 2$ and consider N data points: $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$ where $\mathbf{x}_i \in \mathbb{R}^d$. For any partition $\mathcal{C}_1, \cdots, \mathcal{C}_k$ of $\{1, 2, \ldots, N\}$ we define the cost function

$$Z(\mathcal{C}_1, \cdots, \mathcal{C}_k) = \sum_{l=1}^{k} \frac{1}{2|\mathcal{C}_l|} \sum_{i,j\in\mathcal{C}_l} ||\mathbf{x}_i - \mathbf{x}_j||_2^2. \tag{6.6}$$

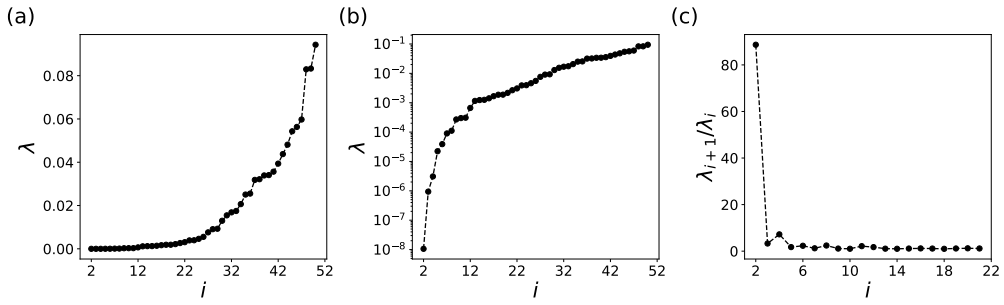

Figure 6.2: The first fifty eigenvalues of the random walk Laplacian matrix $\underline{\mathbf{L}}_{\mathrm{rw}}$ (a) without log scale and (b) with log scale. The first eigenvalue is zero and is not shown in the plot. (c) Eigenvalue ratio λ_{i+1}/λ_i as a function of the index.

The objective of K-means clustering is to find the partition $\mathcal{C}_1^*, \cdots, \mathcal{C}_k^*$ that minimizes Eq. (6.6). That is, the partition minimizes the pairwise squared distance within each cluster normalized by the cluster size. A practical approach to this is via a local search algorithm such as Lloyd's algorithm [141]: In this work, the input N data points are the N rows of the matrix $\underline{\mathbf{T}}^*$ that is the solution to Eq. (6.5). The resulting k clusters are then understood to correspond to $K = k - 1$ traps in the system and the remaining group of nodes representing the rest of the system.

Algorithm 3 Lloyd's algorithm for K-means clustering

1: Input k and $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$, $\mathcal{C}_1, \cdots, \mathcal{C}_k$ are randomly initialized
2: **while** not converged **do**
3: Compute $c_l = \frac{1}{|\mathcal{C}_l|} \sum_{i \in \mathcal{C}_l} \mathbf{x}_i$ for $l = 1, 2, \cdots, k$
4: Update $\mathcal{C}_1, \cdots, \mathcal{C}_k$ by assigning each point $\mathbf{x}_i$ to the cluster whose centroid c_l is closest to.
5: If the cluster assignment did not change, convergence is achieved.
6: Return cluster assignment $\mathcal{C}_1, \cdots, \mathcal{C}_k$.

6.3.3 Determination of the Cluster Number

For the description of the K-means clustering algorithm, it follows that one needs to provide an input k for the number of clusters. The main problem, as is also the case for our intended application, is that it is often unclear what k should be. Therefore, this subsection will introduce an algorithm to determine the number of clusters, which is based on the more commonly applied eigengap heuristic.

Starting from spectral clustering theory, we know that the multiplicity of the smallest eigenvalue 0 of a graph Laplacian equals the number of connected components in the graph $\mathbf{G}$. If $\mathbf{G}_1, \cdots, \mathbf{G}_k$ are the connected components, then the eigenspace corresponding to eigenvalue 0 is spanned by the vectors $\{\mathbf{1}_{\mathbf{G}_i}\}_{i=1}^{k}$. Here $\mathbf{1}_{\mathbf{G}_i}$ is the vector where elements corresponding to the nodes in cluster $\mathbf{G}_i$ are one and the rest are 0. This shows that both the eigenvalues as well as the corresponding eigenvectors contain relevant information about the possible number of clusters.

Of course, in many applications, the graph is a single connected graph and the interesting clusters do not reside in disconnected components. Hence, we cannot simply use the multiplicity of the first eigenvector. Instead, a variety of ways to determine the number of clusters have been discussed previously. For instance, Fraley and Raftery proposed finding the number of clusters based on the log-likelihood of the data [142], while Still and Bialek suggested information-theoretic criteria based on the ratio of within-cluster and between cluster similarity [143]. Tibshirani *el at.* used gap statistics on general data points to find the cluster number [144]. Specifically, in the heuristic eigengap method based on perturbation theory, k is chosen such that $\lambda_1, \cdots, \lambda_k$ are very small and λ_{k+1} is relatively large [140]. In practice, it can however be difficult to implement this heuristic as it is not well-defined what "relatively large" is. This problem can clearly be seen in the eigenvalues of $\underline{\mathbf{L}}_{\text{rw}}$ for the BCP model system, as shown on linear and log scale in Fig. 6.2(a,b), respectively.

To overcome the non-obvious challenge of finding a good k, we notice that the ratio of the eigenvalue λ_{i+1}/λ_i is large for the small indices i, as shown in Fig. 6.2(c). As the index i becomes large, λ_{i+1}/λ_i becomes small. This prompts

us to utilize the index ℓ yielding the maximum $\lambda_{\ell+1}/\lambda_\ell$ as a good first guess for the number of clusters.

Additionally, the eigenvectors also contain important information about possible clusters. For example, when considering 2 clusters, the entries below 0 in the second eigenvectors will correspond to nodes in one cluster while those above 0 correspond to nodes in the other cluster. From this, the idea is that we should only consider those nodes i such that the i-th entry in the ℓ-th eigenvector is large in absolute value.

All together, our method first chooses the index ℓ such that $\lambda_{\ell+1}/\lambda_\ell$ is the largest. Then using the eigenvectors corresponding to the first ℓ eigenvalues, we consider the elements sufficiently distinct from zero. This is controlled via a parameter a. For each of the first ℓ eigenvectors, we will look at the induced subgraph on those nodes whose entries are larger than a in absolute value. We then collect the disconnected components of this induced subgraph as separate graphs. After this, we have a collection of small subgraphs from which we will remove all subgraphs that are themselves a subgraph of another graph in this collection. This procedure leaves us with a list of disjoint subgraphs given by the first ℓ eigenvectors and we take k to be the number of these graphs plus 1. The full procedure to determine the cluster number k is summarized in Algorithm 4.

Let us make a few remarks about this algorithm. Firstly, to select ℓ such that $\lambda_{\ell+1}/\lambda_\ell$ is maximum, the algorithm starts from $i = 2$ since the first eigenvalue is always 0. This is also the only eigenvalue equal to 0 as we only work with connected graphs. We then normalize the eigenvectors corresponding to the selected eigenvalues as $|\mathbf{u}_i|/\max(|\mathbf{u}_i|)$. The purpose of this normalization is to make the algorithm more robust and less dependent on the parameter a, which is used to indicate the entries of eigenvector elements distinct from zero. In particular, we can take $0 < a \leq 1$. In general, the value of a should not be too close to 1 to avoid missing relevant subgraphs. It should also not be too close to zero because then it will not be selective enough. In this work we pick different increasing values of a starting at 0.9, to check how robust the analysis is for a closer to 1. Finally, note that when counting the total number of induced subgraphs, if the subgraphs are contained in other subgraphs, those do not contribute to the value of k.

Algorithm 4 Determination of cluster number k

Input: $\mathbf{G}$, $\underline{\mathbf{L}}_{\mathrm{rw}}$ a.
`Calculate` $(\lambda_i, \mathbf{u}_i)$ `for` $i = 1, \cdots, N$, `such that` $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$.
for $i = 2, 3, \cdots, N-1$ **do**
    $\mathbf{u}_i \leftarrow |\mathbf{u}_i| / \max(|\mathbf{u}_i|)$
`Calculate` $\ell = \mathtt{argmax}(\frac{\lambda_{i+1}}{\lambda_i})$
Denote empty set $B = \emptyset$
for $i = 1, 2, \cdots, \ell$ **do**
    `Set node list` $Q = \emptyset$
    for $j = 1, 2, \cdots, N$ **do**
        if $\mathbf{u}_{ij} > a$ **then**
            `add` j `to` Q
    `Let` $H_1, \ldots, H_M$ `denote the disconnected components of the induced subgraph in` G `on the nodes in` Q.
    `Update` $B = B \cup \{H_1, \ldots, H_M\}$
for $H, H' \in B$ **do**
    if $H \subseteq H'$ **then**
        `remove` H `from` B
$k = |B| + 1$

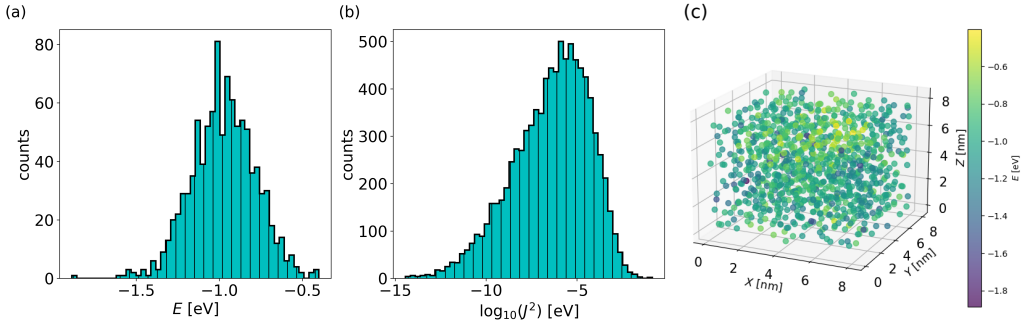

Figure 6.3: (a) The distribution of site energies. (b) The distribution of coupling elements. (c) The spatial resolution of molecule energies in 3-dimension, where each solid circle represents a molecule.

6.4 Trap identification Based on ToF

The BCP morphology is obtained by the classical molecular dynamic simulation with computational details in Chapter 2.2.1. Performing the multiscale model in Chapter 2, the Marcus rates and the material graph are evaluated with computational details in Chapter 2.2.2 using the functional BHANDHLYP for the exchange-correlation functional [145]. The electronic structure parameters of the BCP molecule calculated from the multiscale model are shown in Fig.6.3.

We begin with a short semi-quantitative analysis of the effect of traps in a molecular charge transport network. As the name suggests, once a charge carrier encounters a trap, it will spend a significant amount of time in it and the observed charge dynamics will be slow (large ToF). However, if there are more carriers in the system than traps, one can expect that once all traps are filled, the remaining charge carriers are very mobile. Denote as N_t the number of traps, with N_c the number of charge carriers as in Chapter 3.1.1 and $\tau(N_\mathrm{c})$ the ToF depending on the number of charge carriers, evaluated, e.g., by Eq. (3.15) and Eq. (3.4). Then one expects the ratio $\tau(N_\mathrm{c})/\tau(N_\mathrm{c}+1)$ to be large for $N_\mathrm{c} = N_\mathrm{t}$. Analysis of this ratio then provides a qualitative indication of the *effective number of traps* (depending on a definition of "large") in the transport network but not their location. In

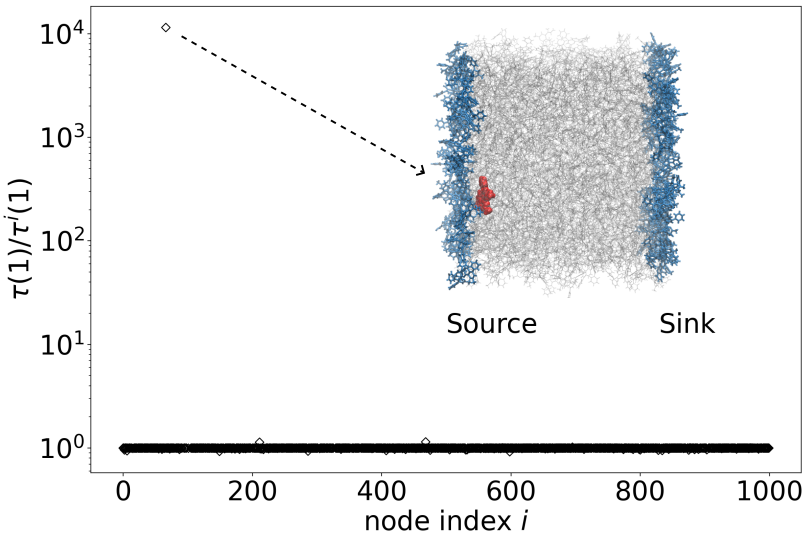

Figure 6.4: The scatter plot of $\tau(1)/\tau^i(1)$ for each node with index i. The only point with a value greater than 100 is the trap node whose index is 66. Inset: A box of 1000 BCP molecules simulated from MD. The Source and Sink molecules are highlighted by the blue color, and the molecule corresponding to trap node having large $\tau(1)/\tau^i(1)$ is highlighted by the red color. The other BCP molecules are in gray color.

addition, the evaluation of $\tau(N_\mathrm{c})$ for a large number of carriers is computationally cumbersome [146].

To avoid such expensive calculations and still gain insight into which molecules correspond to the N_t traps, one can consider all different subsets $Q(N_\mathrm{t})$ of size N_t of molecules in the network. We then consider the ToF of a single charge carrier $\tau(1)$ in the full system and in a system in which the subset $Q(N_\mathrm{t})$ has been removed from the network, $\tau^Q(1)$. The removal of the nodes is motivated by the fact that in the case of very strong traps, carriers will not easily escape them and thus they will for $N_\mathrm{c} > N_\mathrm{t}$ be largely inaccessible to the mobile carriers. In this scenario, one can inspect the ratio $\tau(1)/\tau^Q(1)$ and identify traps from its "large" values. In the special case with only a single trap $N_\mathrm{t} = 1$ the set $Q(1)$ is given by a single node index i, and we can inspect $\tau(1)/\tau^i(1)$. This is shown for all i for the simulated BCP system in Fig. 6.4. It is clear that there is a single molecule node

whose presence in the molecular charge transport network affects the ToF by four orders of magnitude. While it is intuitive to call this particular node a trap in this specific case, it is less obvious in general and the precise definition of a threshold value here is hardly possible. Another drawback of using the inspection of the ratio $\tau(1)/\tau^i(1)$, or its more general form, is the need for many cumbersome ToF calculations of the $\binom{N}{N_\mathrm{t}}$ different scenarios. This shows that a general method that identifies traps in molecular charge transport networks based on the graph structure alone without the need to actually calculate the dynamical properties is of great usefulness.

6.5 Identification of Single-molecule Trap

From the above analysis of the ToF sensitivity to the various nodes, we have identified node 66 is being a trap in the simulated BCP system, with site energy -1.89 eV. We will refer to this node as *the trap node* and will denote it as v_trap. In the following this trap node is used as the reference to scrutinize approaches for trap identification in molecular charge transport networks that solely rely on network properties.

6.5.1 Results from GD Method

Before turning to the spectral clustering method as proposed in Section 6.3, we first show the results of the GD method [138], whose technical details are summarized in Section 6.2. The method contains three adjustable parameters, α, β, γ. The first two are related to details of the connectivity of the graph, whereas γ is used to define a threshold involving the ratios of transition rates. In application to multi-node traps or trap regions, a choice of $\alpha, \beta = 0.02$ and $\gamma = 0.2$ has been reported before [104]. We apply the GD method to BCP and calculate the number of clusters n_cl and the total number of molecules in the cluster containing the trap node, denoted as n_m for different choices of the parameters. The parameter α characterizes the change in the ratio of the number of edges in the cluster to the number of edges that G would have if it were complete. In our BCP graph,

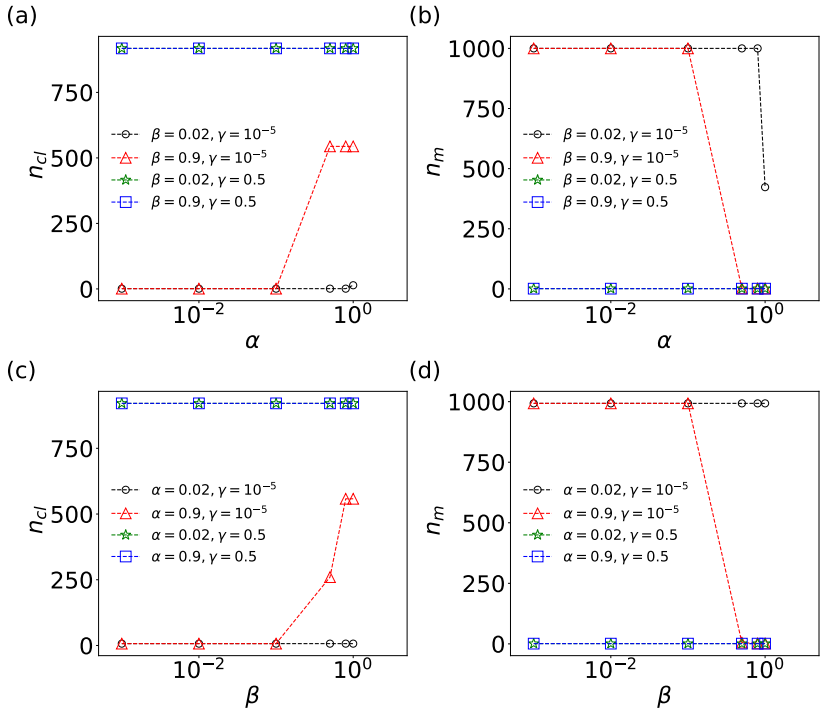

Figure 6.5: (a) The number of clusters after performing the GD method and (b) the size of the cluster containing the trap node with $\beta = 0.02, 0.9$ and $\gamma = 0.5, 10^{-5}$, as a function of α. (c) The number of clusters after performing the GD method and (d) the size of the cluster containing the trap node with $\alpha = 0.02, 0.9$ and $\gamma = 0.5, 10^{-5}$, as a function of β.

this change is negligible. And β characterizes the number of edges between a subgraph and an adjacent node. We find the results of the GD method to be mostly insensitive to the choice of α and β as shown in Fig. 6.5. Figure 6.5 shows the results of using different α, β combinations to identify traps in the BCP structure. Small and large γ are used with varying α, β. This figure shows that at both small $\gamma = 10^{-5}$ and large $\gamma = 0.5$, various values of α, β combination can not identify node v_1 as a trap.

The parameter γ related to the rate ratios has, however, a significant impact on the obtained clustering, as is shown in Fig. 6.6(a) for the range of $10^{-5} \leq \gamma \leq 2 \cdot 10^{-1}$. For very small γ, we obtain a single large cluster equal to the

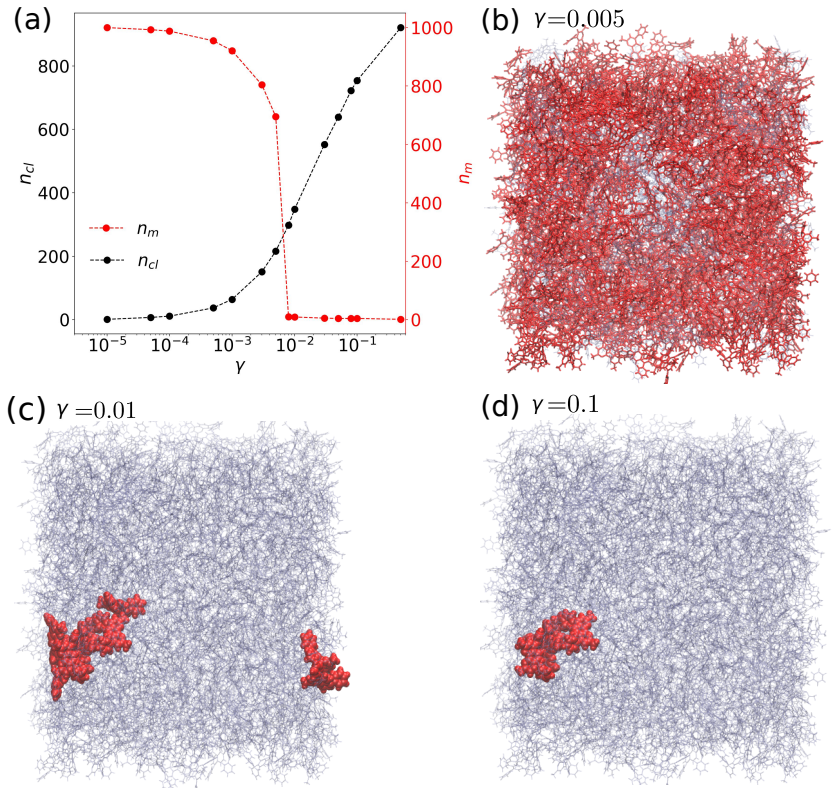

Figure 6.6: Results of GD method on the BCP charge transport network. (a) The total number of clusters (n_{cl}) after performing the GD method, and the size of the cluster containing the trap node (n_{m}), as a function of GD parameter γ. For three values of γ, the BCP structures are visualized: (b) $\gamma = 0.005$, the red molecules indicate the cluster containing the trap node with more than 600 molecules. (c) $\gamma = 0.01$, the 9 molecules in red indicate the cluster containing the trap node, and the molecules in blue color contain 348 clusters (d) $\gamma = 0.1$ with 4 molecules in red forming the cluster containing the trap node.

whole system. With increasing γ the GD method yields more clusters. Between $\gamma = 10^{-3}$ and 10^{-2} one can see a rather sharp transition in the size of the cluster that contains the previously identified trap node. At the onset of the transition ($\gamma = 5 \cdot 10^{-3}$), see Fig. 6.6(b), the cluster containing this node contains more than 600 molecules. After the transition ($\gamma = 10^{-2}$, Fig. 6.6(c)), this size is massively reduced. For the previously recommended value of $\gamma = 0.2$, each molecule is a

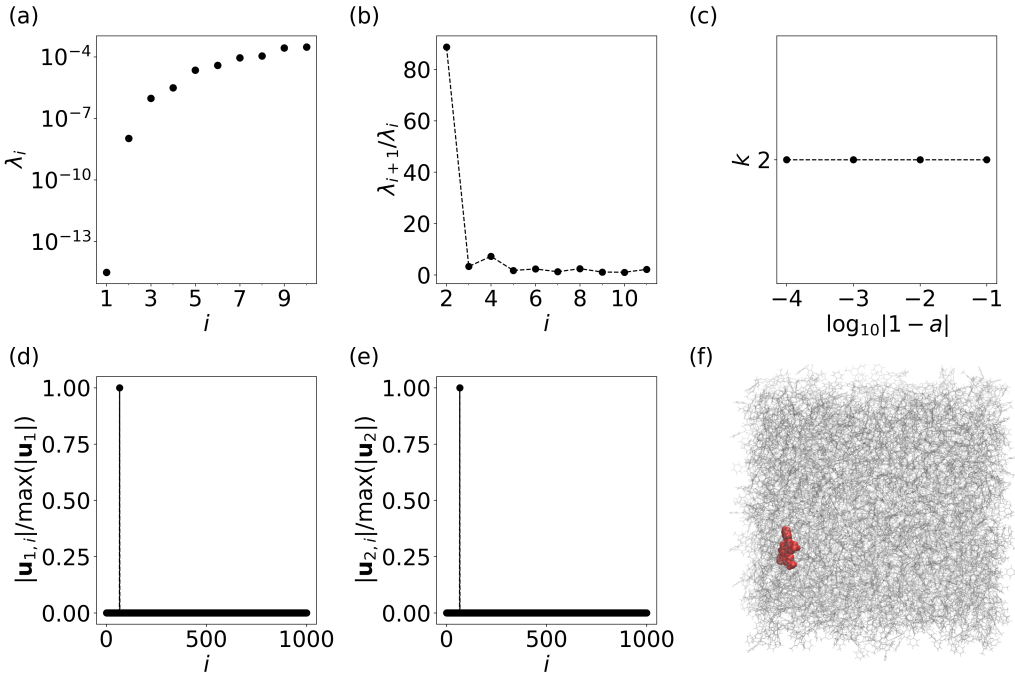

Figure 6.7: Results of the spectral clustering method for multiscale modeled BCP system: (a) The first ten eigenvalues of $\underline{\mathbf{L}}_{\mathrm{rw}}$. (b) λ_{i+1}/λ_i as a function of i. (c) The number of clusters determined from Algorithm 4 as a function of a. (d)-(e) The first and second normalized eigenvector elements of $\underline{\mathbf{L}}_{\mathrm{rw}}$ as a function of node indexes. (f) The BCP system where molecules in red and grey color are first and second clusters, respectively, as identified using the K-means clustering with $k = 2$.

cluster by itself. Tuning the parameter values of α, β does not help in identifying the trap node and GD cannot directly detect the single trap node that leads to large ToF, although the GD method should group molecules into clusters where random walk jumps are more frequent compared to jumps outside the clusters.

6.5.2 Results from Spectral Clustering Method

Turning to the performance of the spectral clustering method in application to BCP, we first show in Fig. 6.7(a) the ten smallest eigenvalues of the random-walk

Laplacian. We note that the first eigenvalue is numerically very close to zero, as expected by the theory. Focusing on the next two eigenvalues, one can see a strong increase from λ_2 to λ_3, while the preceding eigenvalues increase less quickly. These observations lead to a behavior of the ratio λ_{i+1}/λ_i as shown in Fig. 6.7(b) for $2 \leq i \leq 10$. The maximum of λ_{i+1}/λ_i is larger than 80 when $i = 2$, while all other values are below 5. Our procedure in Algorithm 4 will thus take $\ell = 2$ in line 5. Figure 6.7(c) shows the outcome of the algorithm for different values of the threshold parameter a ranging from 0.9 to 0.9999. We note it robustly yields the value $k = 2$ for the K-means clustering step, meaning that there should be one trap in the system. The reason for the observed robustness is clear from the inspection of the normalized elements of the first and second eigenvectors in Fig. 6.7(d) and (e), respectively, which show a single entry (corresponding to the trap node) being 1. Accordingly, the K-means clustering step is performed with $k = 2$ yielding one cluster that only contains the previously identified trap node, while the other cluster is the rest of the system, visually indicated in Fig. 6.7(f).

6.5.3 Relation between Cost function and ToF

The spectral clustering method very clearly identified the trap node in the BCP system. This is also reflected by the fact that the value of the cost function Eq. (6.6) for the proposed cut is $8.8 \cdot 10^{-12}$, while the cost of cutting any other single node is 0.99. We thus see a clear indication that a low cost is a qualitative signal of a possible trap. On the other hand, it is known that the energy of a trap node qualitatively influences the recorded ToF. This raises whether there is also a relation between the site energy of single-molecule traps and the cost of the corresponding cut, and hence between the cost function and the ToF. To investigate this, we take the simulated BCP system and modify the site energy E_{trap} of the trap node, ranging from a small energy value (-2.4 eV) to a relatively large value (-1.6 eV) and determine the cost of cutting according to our method this node and the ToF in the system. We note that when the original trap node has a site energy above (-1.6 eV) it is no longer defined as a trap by our algorithm, which explains why we use this as the upper bound on the energy for this experiment. The results are shown in Fig. 6.8(a). For low values of the site energy, one simultaneously observes a large ToF and small cost, indicating both a clear trap characteristic and

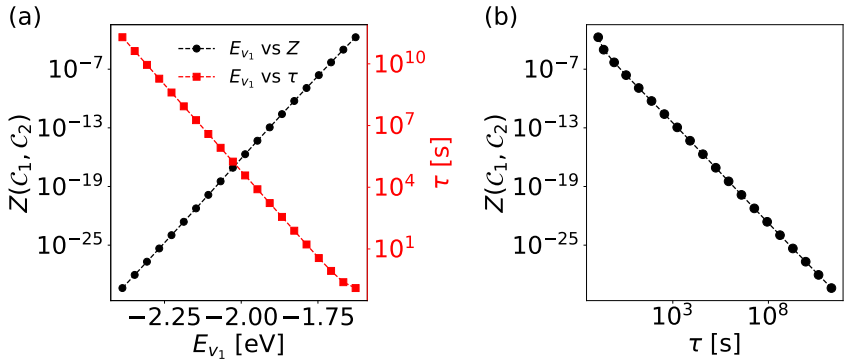

Figure 6.8: In a BCP system, the trap node's energy E_{trap} is varied from -2.4 eV to -1.6 eV. (a) Dependence of the cutting cost $Z(\mathcal{C}_1, \mathcal{C}_2)$ and the ToF τ on E_{trap}. (b) Dependence of $Z(\mathcal{C}_1, \mathcal{C}_2)$ on τ.

straightforward identification via our spectral clustering method. Increasing the energy leads to a decrease in ToF and an increase in the cost of cutting the node from the graph. Moreover, when plotting the pairs of cost function as a function of the ToF, as in Fig. 6.8(b), we observe a power-law dependence between the two quantities. Together these results show that indeed the cost function associated with cutting out a single-molecule trap and the ToF of the system are intricately related.

6.5.4 Results With an Applied Electric Field

To further validate the robustness of our spectral clustering method, we investigate its performance under non-equilibrium conditions by introducing an external electric field. This scenario is particularly relevant for practical applications, as organic semiconductor devices often operate under bias, leading to non-equilibrium charge transport dynamics. To illustrate this, we apply an electric field to the BCP system and analyze the resulting charge transport dynamics quantified by the ToF and clusters identified by the spectral clustering method. The electric field we applied is $\mathbf{F} = [F_x\ 0\ 0]$ where F_x has values in the range 0 to $1 \times 10^8 \mathrm{V/m}$.

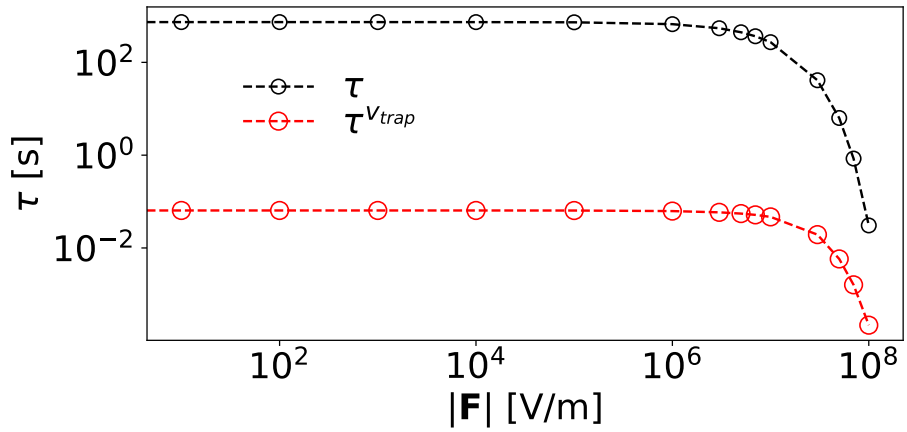

Figure 6.9: (a) The 1-carrier ToF τ, and the ToF $\tau^{v_{\text{trap}}}$ when removing the node v_{trap} as a function of the electric field. (b) The number of clusters and the partition cost function Z as a function of the electric field.

Figure 6.9 shows that as the electric field strength increases, the ToF decreases significantly, indicating a faster charge transport along the direction of $\mathbf{F}$. Using the proposed method, the node v_{trap} is found as a single cluster across the range of the studied electric field. As shown in Fig. 6.9, removing the v_{trap} results in a significant decrease in ToF as indicated by the gap between τ and $\tau^{v_{\text{trap}}}$. Furthermore, as the electric field $|\mathbf{F}| > 10^6\,\mathrm{V/m}$, the gap between τ and $\tau^{v_{\text{trap}}}$ shows signs of reducing. This phenomena suggests that the electric field, having the direction drift force for the charge dyanmics, can potentially reduce the trap effects. When the electric field $|\mathbf{F}| > 10^8\,\mathrm{V/m}$, the trapping effect of v_{trap} is totally suppressed.

These results demonstrate that, even under non-equilibrium conditions induced by the electric field, the trap node v_{trap} identified at equilibrium continues to exhibit a strong trapping effect, as evidenced by the cost function and the change in ToF upon its removal. Our proposed method effectively identifies such traps in non-equilibrium conditions, where the occupancy of these nodes is influenced by external forces.

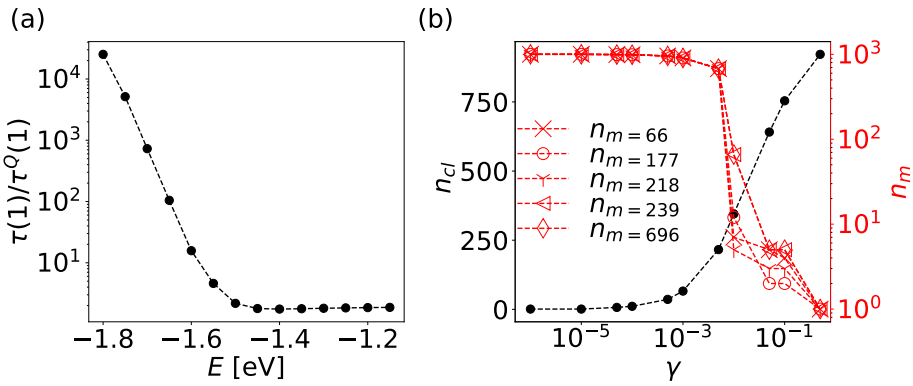

Figure 6.10: Sensitivity of ToF and performance of the GD method for multiple distributed traps. (a) The ratio $\tau(1)/\tau^Q(1)$ as a function of the energies of the nodes in $Q = \{v_1, v_2, \cdots, v_5\}$. (b) The number of clusters n_{cl} after performing the GD method when the nodes in Q have energies $E = -1.8\,\text{eV}$, and the size of the cluster containing a specific node n_{m_i}, as a function of GD parameter γ.

6.6 Identification of Multiple Distributed Traps

The previous section shows that the spectral clustering method can reliably identify a single trap node as it occurs in the modeled BCP system. In the following, we will scrutinize if the same holds in a system with multiple trap molecules that are not connected with each other – a scenario we refer to as *multiple distributed traps*. Within the spectral clustering method, we expect that having multiple distributed traps will increase the cluster number k for the K-means clustering steps as given by Algorithm 3. As the simulated BCP system does not have multiple distributed traps, we modify it by taking the trap node we found and 4 other nodes ($Q = \{v_1 = v_{\text{trap}}, v_2 \cdots, v_5\}$), which are not connected by an edge in the charge transport network and set their site energies all to $E = -1.8\,\text{eV}$, which is close to the site energy of the original trap node. It can be seen from the dependence of the ratio $\tau(1)/\tau^Q(1)$ on E in Fig. 6.10(a) that such a value points indeed to a very pronounced trapping effect of the charge carrier.

First, we consider the predictions from the GD method in Fig. 6.10(b). We again show the resulting number of clusters n_{cl} and the number of molecules n_{m_i}

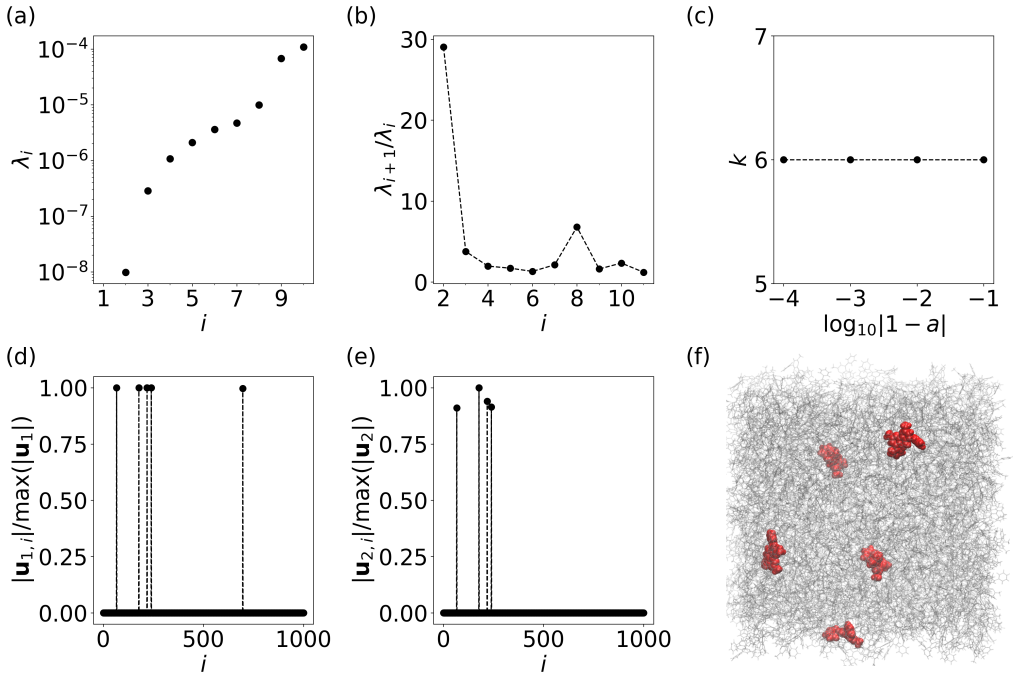

Figure 6.11: Results of the trap identification by spectral clustering methods for the BCP systems with multiple distributed traps. (a) The first ten eigenvalues of $\underline{\mathbf{L}}_{\mathrm{rw}}$. (b) λ_{i+1}/λ_i as a function of i. (c) The number of clusters determined from Algorithm 4 as a function of a. (d)-(e) The first and second eigenvector elements of $\underline{\mathbf{L}}_{\mathrm{rw}}$ as a function of node indices. (f) The BCP system where molecules in red color are molecules consisting of nodes $v_1, v_2, \cdots, v_5$. Each red molecule is partitioned as one cluster. The grey molecules are one cluster, as identified using the K-means clustering with $k = 6$.

in each of the clusters containing one of the 5 prepared trap nodes v_i, depending on the parameter γ in the GD method. Qualitatively, one can observe the same behavior as for the single trap case: the method only yields isolated traps if nearly each molecule is its own cluster, or the whole system is a single cluster. Evidently, the method fails to correctly characterize the multiple distributed traps situation in the charge transport network.

Turning to our spectral clustering method, we see in Fig. 6.11(a) and (b) that the largest value for the ratio is still at $i = 2$, as was the case for the single trap sit-

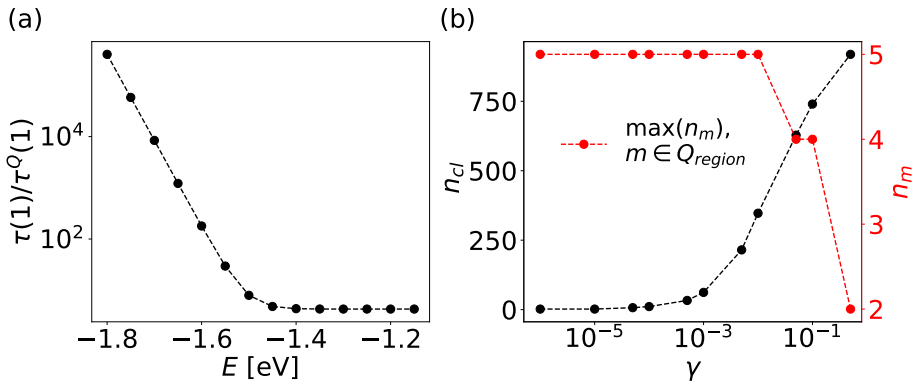

Figure 6.12: Sensitivity of ToF and performance of the GD method for trap regions. (a) The ratio $\tau(1)/\tau^Q(1)$ as a function of the energies of the nodes in $Q = \{v_1, v_2, \cdots, v_6\}$. (b): The number of clusters after performing the GD method when the nodes in Q have energies $E = -1.8\,\text{eV}$, and the maximum size of the cluster containing at least one specific node in Q, as a function of GD parameter γ.

uation. However, unlike that situation, when inspecting the first two normalized eigenvectors in Fig. 6.11(d,e), one can clearly see five and four relevant non-zero elements, respectively. Overall, the determined number of unique induced subgraphs is $k = 6$, which indeed implies that there are 5 traps in the system shown in Fig. 6.11(f): the five individual distributed trap nodes and one cluster that is the rest of the system. This shows the necessity of the addition step in Algorithm 3. The result $k = 6$ is also relatively robustly with respect to the choice of the method's parameter a as shown in Fig. 6.11(c). Only when $a > 0.9$ is used, the spectral clustering method yields $k = 2$. So again, using Algorithm 4 with $a = 0.9$ finds the correct number of clusters $k = 6$, and also the right trap nodes.

6.7 Identification of Trap Region

It is known that for some materials the site energies are correlated in space, e.g., due to strong permanent dipole interactions as in amorphous Alq_3 [39, 66, 104]. This leads to a situation in which several sites with low energy are connected,

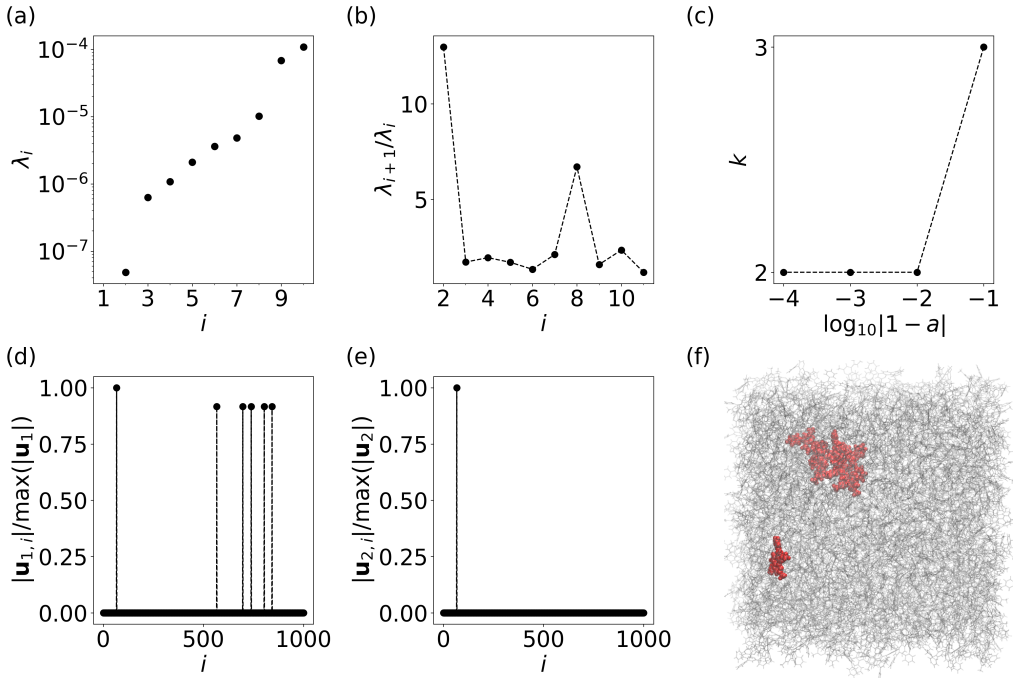

Figure 6.13: Trap identification results using spectral clustering methods on the BCP system with trap regions. (a) The first ten eigenvalues of $\underline{\mathbf{L}}_{\text{rw}}$. (b) λ_{i+1}/λ_i as a function of i. (c) The number of clusters determined from Algorithm 4 as a function of a. (d)-(e) The first and second eigenvector elements of $\underline{\mathbf{L}}_{\text{rw}}$ as a function of node indices. (f) The BCP system where molecules in red color show v_1 as a single molecule cluster and the connected region $Q = \{v_2, \cdots, v_6\}$ is a second cluster. The gray molecules form the third cluster, as identified using K-means clustering with $k = 3$.

forming a trap region instead of single-molecule traps as studied in the previous two sections. A random walk process that enters such a region tends to jump among those nodes with only a small probability of escaping the region. This scenario was the original motivation for the GD method as it aims at partitioning the graph into clusters that the random walk spends a significant amount of time in.

We investigate whether our proposed spectral clustering methods can identify such trap regions in the BCP system. Because the BCP system does not

have strong spatial correlations, there are no such trap regions in the as-modeled molecular charge transport network. We therefore select six nodes $Q = \{v_1 = v_{\mathrm{trap}}, v_2, \ldots, v_6\}$ consisting of the original trap node and a distant node v_2 disconnected from $v_1 = v_{\mathrm{trap}}$ plus its four closest neighbors v_3, v_4, v_5, v_6. We then set the site energies of the nodes in Q to a value E so the v_{trap} is a single-molecule trap while $Q_{\mathrm{region}} := \{v_2, \ldots, v_6\}$ is trap region. In Fig. 6.12(a), we show $\tau(1)/\tau^Q(1)$ for different values of E and observe a clear trapping effect on the ToF for $E < -1.6\,\mathrm{eV}$. In the following we set $E = -1.8\,\mathrm{eV}$.

Figure 6.12(b) shows the result of the GD method to identify the traps: the total number of clusters n_{cl} and the maximum cluster size n_{m} when m is one of the nodes in Q_{region} for different values of γ. When $\gamma = 10^{-6}$, $n_{\mathrm{m}} = 5$, and Q_{region} is successfully identified as a single cluster and the rest of the 995 nodes as another cluster. As in Section 6.5, the isolated trap node is not identified as a separate cluster. As γ increases, more clusters appear, and when $\gamma > 10^{-2}$, the five low-energy nodes in Q_{region} are no longer partitioned into one cluster.

Turning now to the trap identification by spectral clustering, we show the eigenvalues of $\underline{\mathbf{L}}_{\mathrm{rw}}$ and the ratio λ_{i+1}/λ_i in Fig. 6.13(a,b), respectively. We again observe that the maximum of the ratio is found at $i = 2$. The full algorithm yields a value $k = 3$, indicating the presence of two traps. As can be seen from Fig. 6.13(c), the determination of k is very robust with respect to the choice of the parameter a. Looking at the elements of the first two normalized eigenvectors (see Fig. 6.13(d,e)) we clearly see a single large entry and a group of five other large entries, representing Q_{region}. Finally, K-means clustering with $k = 3$ yields the three correct clusters as shown in Fig. 6.13(f): in red the isolated trap node v_{trap} and the trap region formed by Q_{region} and in gray the third cluster containing all other molecules.

It should be noted that there is a second large jump for the eigenvalues at $i = 4$. To further check the performance of our method we also perform the spectral clustering method on the first four eigenvectors, instead of the first two. This gives $k = 4$ with the only difference being that there is now a fourth cluster that comprises a single node v_7. However, when checking the effect of this node on the ToF via the ratio $\tau(1)/\tau^Q(1)$ with node v_7 added to Q, we find that the

ratio is ≈ 1, i.e., node v_7 does not have the system characteristics we associated with a trap. This supports the choice of $\ell = 2$ yielded by our Algorithm 4.

6.8 Spectral Clustering for Networks Without Traps

To further demonstrate the applicability of the proposed method, we present the results of performing spectral clustering on two systems devoid of traps. The first system is the commonly used BCP device, with all site energies set to be equal, representing an amorphous system without energy disorder but with coupling element disorder. The second system is a $10 \times 10 \times 10$ cubic lattice, where all site energies are zero and each node is connected to its nearest neighbors with equal coupling elements, resulting in uniform Marcus rates.

In the zero-energy BCP system, the eigenvalues of the corresponding $\underline{\mathbf{L}}_{\text{rw}}$ and their ratios are shown in Fig. 6.14(a)(b). It is observed that $i = 3$ has the maximum ratio $\frac{\lambda_{i+1}}{\lambda_i}$. The first eigenvector is a constant vector $\mathbf{1}$ because the graph is symmetric and the first eigenvalue is zero. Consequently, the first eigenvector does not provide significant information about the clustering structure. Applying Algorithm 4 to the second and third eigenvectors identifies the number of clusters as $k = 3$.

Performing a K-means clustering using $k = 3$ finds the following results: one cluster contains 15 nodes $Q_1 = \{v_1, v_2, \cdots, v_{15}\}$ and the second cluster contains 20 nodes $\{v_{16}, v_2, \cdots, v_{35}\}$. These two clusters are indicated by the red regions in Fig. 6.14. Those clusters are not traps because changing the energy of the nodes in those clusters will not significantly change the ToF. The third cluster consists of the rest of the 965 nodes. Here the first two clusters found through the proposed method contain much more nodes than the eigenvector elements above $a = 0.9$. Since the system has no trap, the method tries to cut the graph to solve the minimization problem Eq. (6.2), so the clusters are relatively large ($Z_{C_1,C_2,C_3} = 7.6 \times 10^{-3}$).

For the $10 \times 10 \times 10$ lattice system, the eigenvalues of its $\underline{\mathbf{L}}_{\text{rw}}$ and their ratios are shown in Fig. 6.15(a)(b). The first eigenvector is a constant vector and the

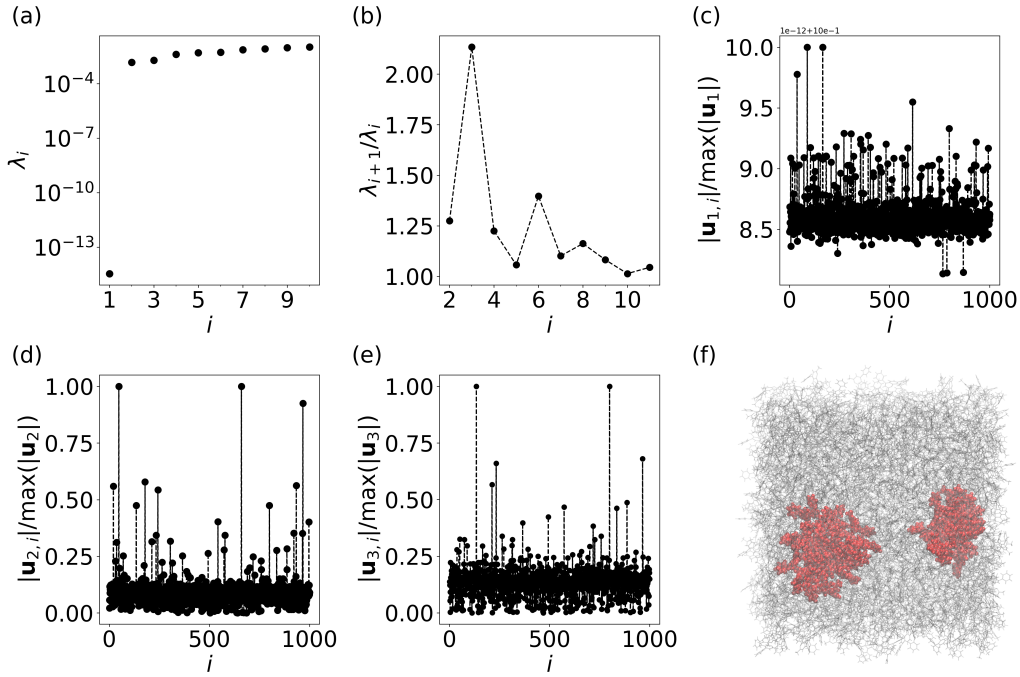

Figure 6.14: Spectral clustering results on the BCP device graph where all energies are set to be 0. (a): The first ten eigenvalues of $\underline{\mathbf{L}}_{\text{rw}}$. (b): $\frac{\lambda_{i+1}}{\lambda_i}$ as a function of i. (c)-(e): The first, second and third eigenvector elements of $\underline{\mathbf{L}}_{\text{rw}}$ as a function of node indexes. (f): The BCP device where molecules in red color form two large clusters. And the gray molecules are one cluster, as identified using the K-means clustering with $k = 3$.

entries of the second eigenvector exhibit a stepping feature, while the third eigenvector shows no clear boundary between its elements. Applying Algorithm 4 to the second eigenvector reveals that $k = 2$. Performing a K-means clustering using $k = 3$, the K-means clustering algorithm partitions the first 500 nodes (nodes 1-500) into one cluster, with the remaining 500 nodes forming the second cluster as shown in Fig. 6.15. This bisected partitioning minimizes the objective function described in Eq. (6.5).

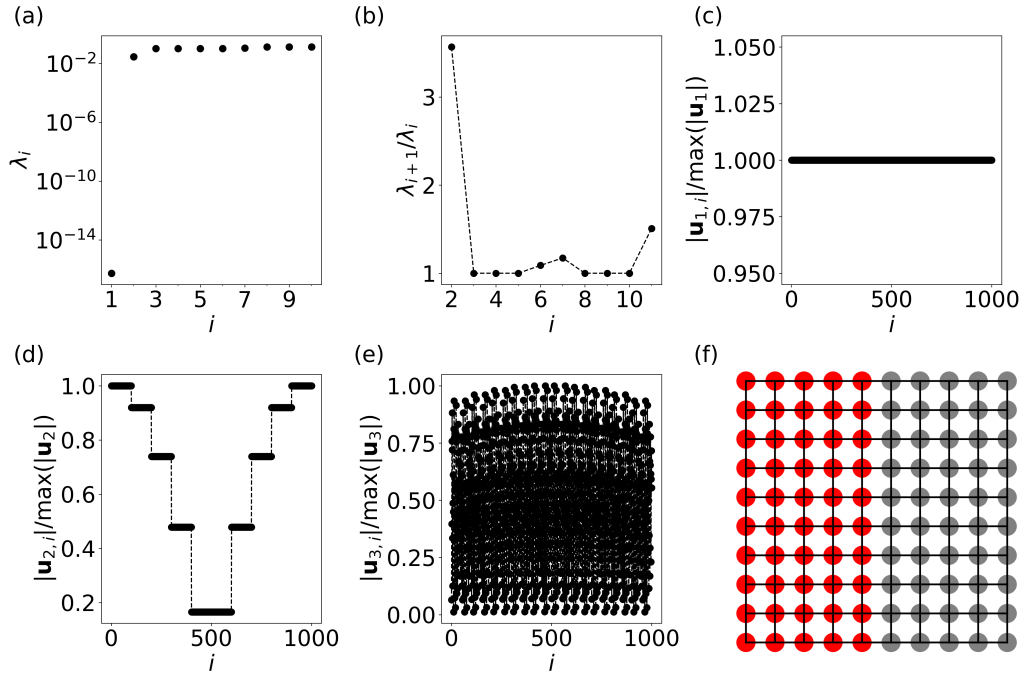

Figure 6.15: Spectral clustering results on the graph of a cubic lattice with 1000 sites, where all energies are set to be 0 and the connection to nearest neighbors are equal. (a): The first ten eigenvalues of $\underline{\mathbf{L}}_{\mathrm{rw}}$. (b): $\frac{\lambda_{i+1}}{\lambda_i}$ as a function of i. (c-e): The first, second and third eigenvector elements of $\underline{\mathbf{L}}_{\mathrm{rw}}$ as a function of node indexes. (f): A top view of a cubic lattice with 1000 nodes, where the nodes in red are partitioned as one cluster and the nodes in gray color are in a second cluster.

6.9 Comparison with Energy-Based Trap Identification

As mentioned in Section 6.1, the energy-based identification of traps is straightforward in many scenarios, but it relies heavily on the assumption that traps are simply the sites with the lowest energy in the system. This assumption holds under equilibrium conditions where the occupancy of traps follows Fermi-Dirac statistics, and the transition rates between sites obey detailed balance. However, there are several situations where the energy-based method fails to correctly identify traps, particularly when the dynamics of charge transport are influenced by factors beyond just the site energies. The bi-molecular charge transition rate (such

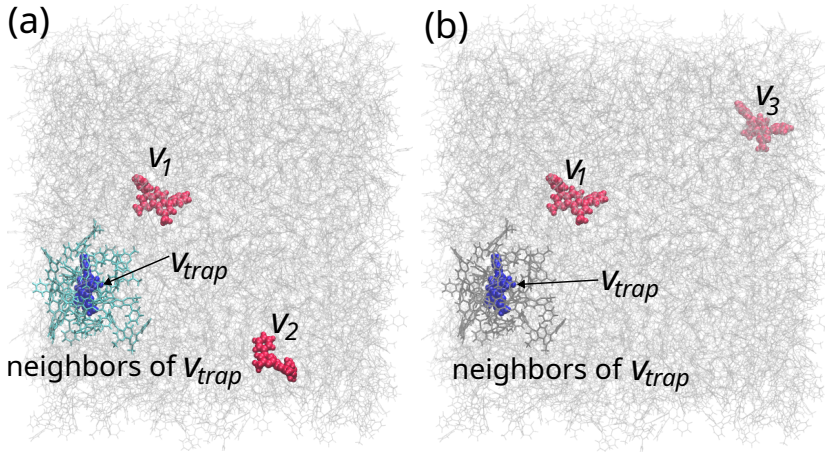

Figure 6.16: (a) Visualization of the node v_{trap} which is colored in blue and has the lowest energy $E_{\text{trap}} = -1.80$. The neighbors of v_{trap} are colored in green, representing different molecular types. The reorganization energies between v_{trap} and its neighbors are non-symmetric: $\Lambda_{v_{\text{trap}}\, j} = 0.1\,\text{eV}$ and $\Lambda_{j\, v_{\text{trap}}} = 0.2\,\text{eV}$. Two BCP molecules in red color, v_1 and v_2, are two single clusters identified using the K-means clustering with $k = 3$. (b) Visualization of the node v_{trap} which is colored in blue and has the lowest energy $E_{\text{trap}} = -1.80$. The neighbors of v_{trap} are colored in light-dark, representing different molecular types. The coupling element between v_{trap} and its neighbors are not large, having values in range $10^{-9}\text{eV}^2 > J_{ij}^2 > 10^{-14}\text{eV}^2$. Two BCP molecules in red color, v_1 and v_2, are two single clusters identified using the K-means clustering with $k = 3$.

as the Marcus rate used in the draft) is also related to coupling elements J_{ij} and reorganization energy λ_{ij}. So focusing on the energy alone ignores other factors. If the charge transfer rate into the lowest energy site is small, then the occupation probability of this site will be small and has no trap effect. In this section, we demonstrate the limitations of the energy-based method and highlight the advantages of our spectral clustering approach through two illustrative models.

6.9.1 Non-Symmetric Reorganization Energy

In the first example, we consider a system where the reorganization energy Λ_{ij} for charge transfer between molecules is non-symmetric. This can occur in materials

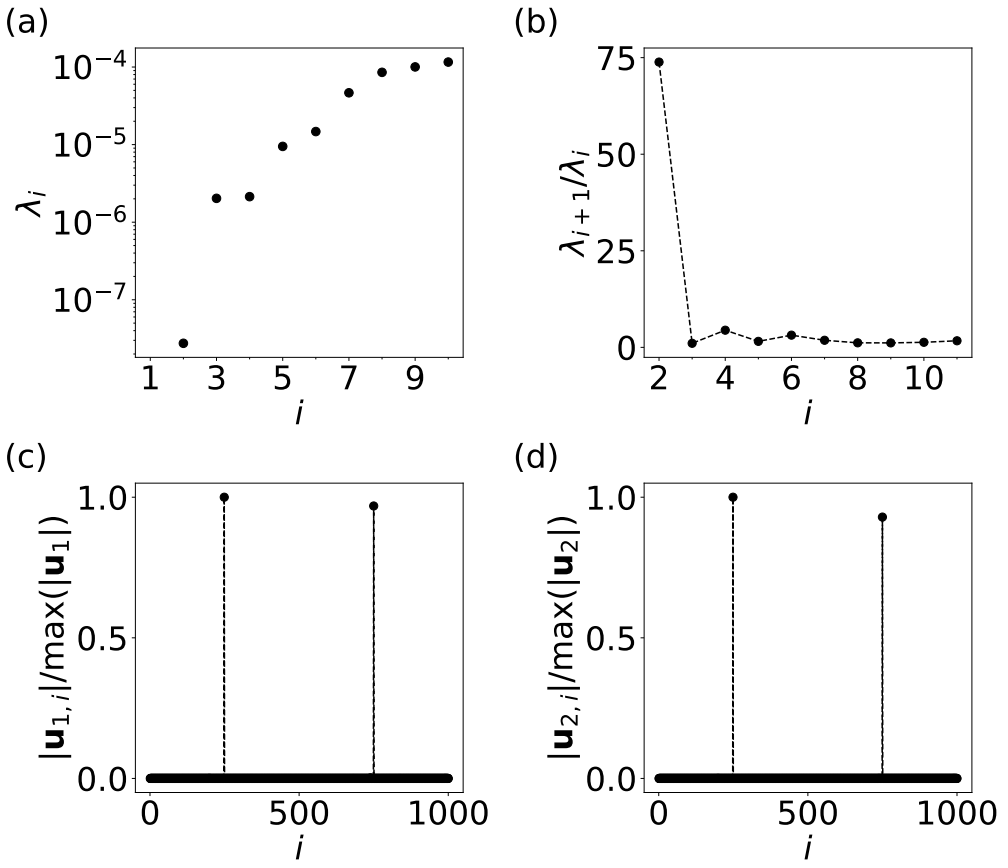

Figure 6.17: Trap identification results using spectral clustering methods on the BCP system where the site v_{trap} and its neighbors have nonsymmetric reorganization energy. (a) The first ten eigenvalues of L_{rw}. (b) λ_{i+1}/λ_i as a function of i. (c)-(d) The first and second normalized eigenvector elements of L_{rw} as a function of node indexes.

with different types of molecules and complex molecular geometries, where the energy required to reorganize the environment during charge transfer depends on the direction of the transfer. In such cases, the lowest energy site may not necessarily act as a trap, even if it has a significantly lower site energy compared to the rest of the network.

To illustrate this, we modify the BCP system by introducing a non-symmetric reorganization energy for specific pairs of molecules. We select the node v_{trap} with a low site energy $E_{v_{\text{trap}}} = -1.80\,\text{eV}$ and adjust the reorganization energy Λ such that $\Lambda_{v_{\text{trap}}\,j} = 0.1\,\text{eV}$ and $\Lambda_{j\,v_{\text{trap}}} = 0.2\,\text{eV}$. We show this setup in Fig. 6.16(a), where the node v_{trap} is colored in blue and its neighbors are in light-black color. The node v_{trap} has the lowest energy.

Due to the non-symmetric reorganization energy, $\Lambda_{v_{\text{trap}}\,j} \neq \Lambda_{j\,v_{\text{trap}}}$, the transition rate from this molecule to its neighbors is not small. This results in a situation where the low-energy site is not a trap because charge carriers can easily escape from it due to the high forward transition rate. In contrast, the spectral clustering method, which considers the full graph structure and transition rates, correctly identifies that this site is not a trap. The energy-based method, however, would incorrectly classify this site as a trap solely based on its low energy.

The result of the proposed spectral clustering methods is shown in Fig. 6.17. The method identifies nodes v_1 and v_2 with site energy $E_{v_1} = E_{v_2} = -1.70\text{eV}$ (in red color shown in Fig. 6.16(a)) as the single traps in this system, putting those two nodes in their own cluster and the other 998 nodes in another. The ToF of the system without removing any node is $\tau = 0.5684\,\text{s}$, after removing v_1, v_2, we obtain $\tau^{(v_1,v_2)} = 1.420 \times 10^{-3}\,\text{s}$. But after removing v_{trap} we obtain $\tau^{v_{\text{trap}}} = 0.5541\,\text{s}$, which is very close to $\tau = 0.5684\,\text{s}$. We clearly see that removing the lowest energy site v_{trap} has no effect of the ToF, and the nodes identified from our proposed method influence the ToF significantly.

6.9.2 Small Coupling Element with Neighboring Sites

In the second example, we consider a scenario where the lowest energy site v_{trap} has a very small electronic coupling element $J_{v_{\text{trap}}\,j}$ with its neighboring sites j. This can occur in materials with poor orbital overlap between molecules or in systems with significant spatial disorder. In such cases, even if a site has a very low energy, the probability of charge carriers transitioning to or from this site is extremely low, resulting in low occupancy and hence no trapping effect.

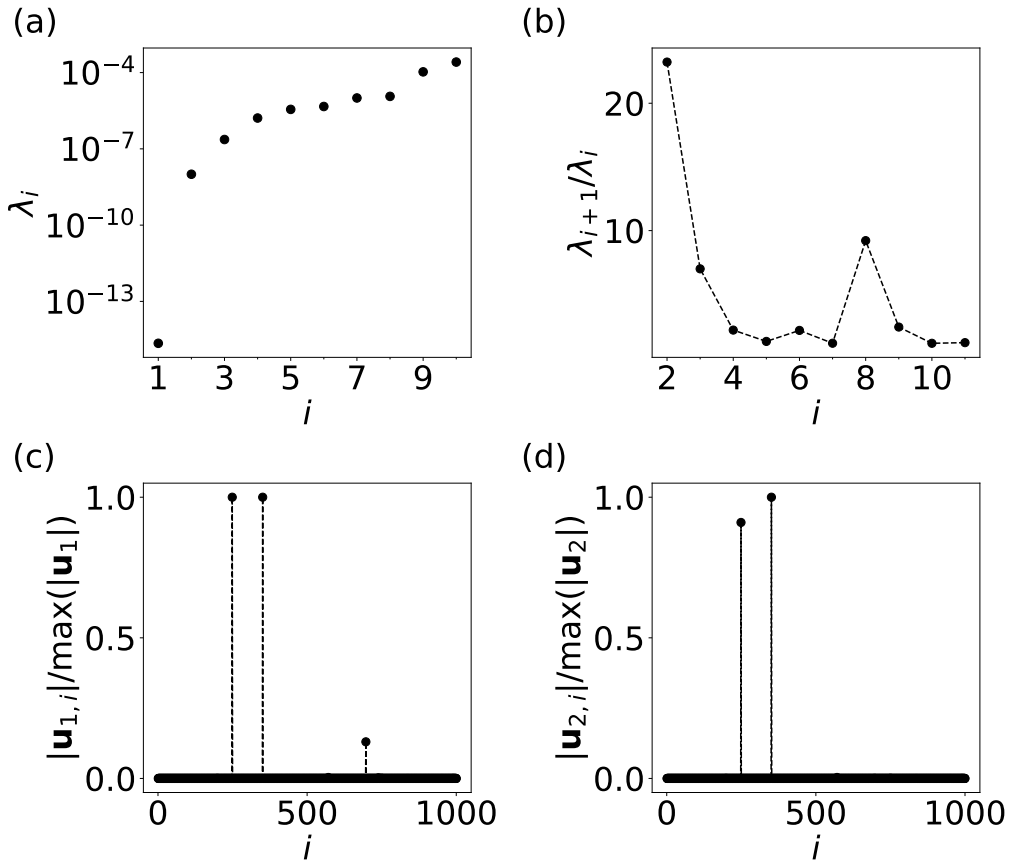

Figure 6.18: Trap identification results using spectral clustering methods on the BCP system where the node v_{trap} and its neighbors have weak coupling elements. (a) The first ten eigenvalues of L_{rw}. (b) λ_{i+1}/λ_i as a function of i. (c)-(d) The first and second normalized eigenvector elements of L_{rw} as a function of node indexes.

To demonstrate this, we modify the BCP system by selecting a molecule with the lowest site energy and reducing its electronic coupling elements with all neighboring molecules to a relatively small value. Specifically, the electronic coupling elements $J_{v_{\text{trap}}\,j}$ between the lowest energy site v_{trap} (highlighted in blue in Fig. 6.18) and its neighbors j has range: $10^{-9}\text{eV}^2 > J^2_{v_{\text{trap}}\,j} > 10^{-14}\text{eV}^2$. Such a system is visualized in Fig. 6.16(b), where the lowest energy site v_{trap} is in color blue and its neighbors are in light-dark color. As a result, the transition rates to

and from this site become negligible, and charge carriers rarely occupy this site. While the energy-based method would still classify this site as a trap due to its low energy, it does not exhibit any trapping effect in practice, as carriers are unlikely to be captured or released from this site. The ToF of this system is $\tau = 0.5968\,\mathrm{s}$. After removing v_{trap}, the ToF is $\tau^{v_{\text{trap}}} = 0.5968\,\mathrm{s}$, which is the same as the ToF without removing any nodes.

Applying our proposed spectral clustering method, the ratio of the eigenvalues λ_{i+1}/λ_i and the analysis of the first two eigenvectors suggest that the K-means clustering should be performed with $k = 3$. The clustering results identify v_1 and v_3 as two single clusters while the remaining 998 nodes form a single large cluster. Removing v_1, v_3, the ToF is $\tau^{(v_1,v_3)} = 6.342 \times 10^{-2}\,\mathrm{s}$, indicating the real trapping effect of v_1, v_3.

These two models demonstrate the limitations of the energy-based trap identification method and highlight the advantages of the spectral clustering approach. While the energy-based method is effective in simple scenarios where transition rates are solely dependent on their energy, it fails in more complex situations where the dynamics of charge transport are influenced by factors such as non-symmetric reorganization energies or low electronic coupling elements. In contrast, the spectral clustering method, which considers the full graph structure and transition rates, can correctly identify traps in these complex scenarios.

The added value of the spectral clustering method lies in its ability to capture the interplay between site energies, transition rates, and the overall network structure. This makes it a more robust and general approach for identifying traps in organic semiconductors, particularly in systems with complex charge transport dynamics. By leveraging the spectral properties of the graph Laplacian, our method provides a more comprehensive understanding of trap behavior, which is crucial for optimizing the design and performance of organic electronic devices.

6.10 Conclusion

Traps play an important role in defining the electrical properties of organic semiconductors and their functionality in electronic devices. They significantly influence the charge transport mechanism in organic semiconductors and allow for a tuning of the charge mobility as indicated by variations in the ToF measurements. Understanding and eventual control of the charge mobility is often sought via multiscale modeling of the materials which leads to its representation as a molecular charge transport network. Here we have developed and implemented a novel method based on spectral clustering to identify traps within such molecular charge transport network of multiscale-modeled organic semiconductors. Our approach avoids the complexities associated with calculating multiple-carrier ToF and overcomes the need for system-dependent parameters, making it a robust method for general multiscale model system. Leveraging the eigenvalue ratio and eigenvector entries allows the application of K-means clustering on the eigenvector elements from the Laplacian matrix, reinforcing the robustness of our spectral clustering approach.

The effectiveness of our method is demonstrated through the analysis of the multiscale-model BCP device, where we successfully identified a single-molecule trap, distributed traps, and region of traps coexisting with single-molecule traps. This capability to identify and quantify traps without extensive parameter tuning marks a significant advancement over other candidate methods, such as the watershed algorithm and GD. Our proposed method and the findings from this study can potentially enhance our understanding of trap dynamics in organic semiconductors and also offer a powerful tool for optimizing the design and performance of these materials in practical applications.

Chapter 7

Conclusion

The multiscale model of charge dynamics in organic semiconductors results in molecular charge transport networks. As mentioned in the Introduction Chapter 1, there are three problem statements in this theme. The first problem is related to the accurate extraction of charge mobility from the charge transport networks. This thesis proposes the GRW method to obtain accurate charge mobility predictions from the molecular charge transport network. Relating the ToF to the expected first hitting time of a CTMC, the key aspect of this method is the expected first hitting time calculation using a matrix formula. By numerical studies on a surrogate model of an Alq_3 device, this method is validated by comparison to the MEq and KMC method. The charge mobility calculation with GRW method can avoid the numerical stability issue in MEq and the KMC convergence issue in certain network scenarios.

Then, the next problem statement in the molecular charge transport networks is due to the setup of the graph, whose construction involves assumptions and approximations associated with the quantum and classical models. The modeling choice to deal with those assumptions and approximations raises the problem of the possibility to quantify the uncertainty in the multiscale model. To investigate this problem, numerical studies are performed on an MADN charge transport network, where the ToFs are calculated using the GRW method. The ToFs are calculated with five different values of the α_{HFX} in the DFT functional PBE0. Using Monte Carlo sampling of the electronic parameters, the ToF uncertainties associated with various Hartree-Fock fractions in the DFT exchange-correlation functional are quantified within a specific confidence level. The sensitivity of the

reorganization energy Λ, site energy E, and the coupling element J to the ToF is analyzed through Sobol indexes calculation, which shows that ToFs are most sensitive to Es and least sensitive to Js. Our UQ investigation sheds light on interpreting the computational results of organic semiconductors.

The third problem regards the trap identification in the molecular charge transport network. This thesis proposes a method based on spectral clustering to successfully identify traps in networks containing a single trap, distributed traps, and a combination of a single trap and a trap region. From an input of the molecular network, this method identifies traps without relying on system-specific parameters or requiring a large number of ToF calculations. This method improves the computational robustness and efficiency for trap identification in molecular charge transport networks.

Although the problems are addressed in this thesis, there are limitations in the proposed methods and numerical studies, and future research can be performed to potentially address those limitations.

Firstly, when evaluating the charge mobility using the GRW method, the computational challenges of the MEq and KMC are circumvented partially because all the states are considered simultaneously when calculating the ToF. For systems with a large number of charge carriers, the state space is so large that the matrix equations are difficult to solve. So, the proposed GRW method is only suitable for situations with low carrier density. (Those situations are usually the ones for which KMC has sampling issues, so our proposed method is a complementary method for KMC.) Developing computational models/methods for charge mobility evaluation without suffering from the MEq's discretization problems, KMC's convergence issue, and the GRW method's dimension problem will make the computational models applicable to a wider range of physical situations.

The second limitation is that the uncertainty quantification is only performed for the MADN molecular network, whose site energy is practically uncorrelated. Different material models have other parameters that this thesis has not considered yet, such as the energy correlation and coupling element correlation, since the Monte Carlo sampling scheme ignores those model parameters. Performing uncertainty quantification that takes those parameters into account will make

quantitative analysis more complete and rigorous, and facilitate the investigation of a broader range of material models.

Another limitation regarding the uncertainty quantification is the representation of the uncertainties. This thesis considers the uncertainty generated by different α_{HFX}. Other model settings, such as the empirical force field for approximating the PES, and the types of basis sets used for solving the Kohn-Sham equation, also represent important sources of uncertainties. If those uncertainties can be quantified, the predictive ability of the multiscale model will be improved.

In addition to addressing these limitations, there are additional recommendations for future work building on the foundations laid in this thesis.

First of all, machine learning related methods have been used for setting up the molecular networks, such as the calculation of the terms entering the Marcus rate expression [147, 148, 149, 150]. Integration with machine learning can potentially reduce the need for computational resources. Further, utilizing machine learning methods for Marcus rate calculation in the multiscale model can generate more accurate results. For example, deep neural networks have been used to achieve nearly exact solutions of the electronic Schrödinger equation for molecules with up to 30 electrons [151]. When the amount of electrons in a localized state is less than this number, and the electronic structure parameters (Λ, E, J) can be evaluated with such deep neural network method, then the molecular charge transport network can be set up more accurately, yielding a more accurate charge mobility evaluation.

A second future direction is to understand the physical mechanism of the exponential dependence of the partition cost function on the energy of traps, as shown in Chapter 6. While the spectral clustering based method works well in trap identification, such exponential dependence is not yet explored. The partition cost function depends on all the nodes/edges in the network, while the change of a single molecule's energy surprisingly leads to an exponential change in the partition cost function. So, understanding its exponential dependence on the energy of the traps sheds light on the interaction between the molecules, and the way by which such interaction contributes to the charge transport in the molecular networks.

Furthermore, the proposed spectral clustering method works very well for characterizing traps, but this method completely ignores the trajectory of charge transport. This is largely because we want to develop an effective method for finding traps, thus ignoring the transport paths that are usually determined by high-dimensional parameters. Undoubtedly, if the charge transport path can be further reflected in the proposed method, it will be more helpful in understanding charge dynamics and aiding the optimization of organic semiconductors.

A last recommendation is to study the electronic structure parameters, charge mobility, and trap characterization for multiple molecular structures sampled from the MD simulation. Those investigations will show the charge mobility's time series behavior, such as the correlation between the electronic parameters, charge mobility as a function of time and the traps' evolution. The time correlation between the electronic structures, such as Es or Js, reveals the variation of electronic structures as the atom coordinates change along the PES. Molecular structures are not constant at all the times. Either with an NpT condition or NVT condition described in Chapter 2, the coordinates of all molecules in a molecular network vary at different times. Thus, the temporal evolution of networks' properties provides a more comprehensive understanding of the charge transport with various molecular topologies.

Summary

This thesis investigates the charge transport dynamics in amorphous organic semiconductors based on a first principle multiscale model, focusing on three primary objectives. First, it aims to develop a method for accurately evaluating the ToF in organic semiconductors, overcoming computational challenges associated with the MEq and KMC methods. Second, it seeks to quantify uncertainties in the multiscale model, particularly those related to exchange-correlation functionals. Third, it tries to create an efficient method for identifying traps in the multiscale modeled molecular network of organic semiconductors.

The thesis presents the theoretical foundation and implementation of the multiscale model, which links charge dynamics to microscopic factors such as molecular electronic structures, spatial arrangements, and morphological inhomogeneities. Subsequently, related theories and methodologies within the graph random walk framework are introduced, including the MEq, KMC, and the ToF calculation based on the matrix formula in the GRW method.

By leveraging these theories, the thesis verifies the accuracy of the proposed ToF evaluation using the GRW method. This method facilitates the quantification of uncertainties in the multiscale model due to exchange-correlation functionals and the development of a trap-identification method employing spectral clustering techniques. The uncertainty quantification provides a confidence range for charge mobility prediction in the multiscale model, fostering a closer connection between charge transport and the microstructure of organic semiconductors. The trap-identification method contributes to detecting key molecules in controlling the switching behavior of organic semiconductors, showing potential applications in the optimization of innovative organic devices. In summary, this research and the methodologies developed herein advance the understanding of charge dynamics in organic semiconductors through computational studies.

Acknowledgement

Looking back on my long doctoral studies journey, I realize there are many people and things for which I want to express my gratitude.

First and foremost, I wish to extend my heartfelt gratitude to my family members — my parents and sister — for their support and understanding of my decision to pursue a doctorate. Leaving my hometown in a small city in the south of China and settling in the Netherlands has not only been a significant step in my academic journey but also a challenging transition for the entire family. Despite the distance, their constant encouragement has always been a source of strength and motivation for me.

My parents have played an indispensable role in shaping the person I am today. From an early age, they nurtured my curiosity and instilled in me a love for learning. I still vividly remember the moments when they patiently taught me how to write Chinese characters and grasp basic mathematics. As I grew older, they expanded my understanding by introducing me to practical engineering skills through daily tasks and simple scientific concepts. These early lessons laid the foundation for my academic interests and cultivated a mindset of inquiry and perseverance.

Beyond academics, my parents have also guided me in developing strong moral principles. They taught me the importance of respecting the elderly, caring for the young, being helpful, and showing kindness to others. More importantly, they emphasized the need to strive for comprehensive personal growth — to not only excel in my studies but also to become a well-rounded individual with empathy, integrity, and resilience.

Throughout my journey, my parents have consistently expressed their positive support. No matter the challenges I faced, they were always there to listen, offer comfort, and help me find solutions in every possible way. Their boundless care and dedication have been a constant source of inspiration, reminding me that I am never alone in this journey. For all these invaluable lessons and endless support, I am profoundly grateful. It is their love and guidance that have shaped me into the person I am today and have empowered me to pursue my dreams with confidence and determination.

I also wish to express my sincere gratitude to my supervisor, Dr. Bjorn Baumeier, for giving me this wonderful opportunity to pursue a PhD at CASA, in the Department of Mathematics and Computer Science. I still vividly remember the first time I came across the project's advertisement — a moment filled with both excitement and a touch of nervousness as I reached out to Dr. Baumeier, hopeful yet uncertain. When I was offered the PhD position, I was deeply grateful, eager to dive into the project and make meaningful progress.

Dr. Baumeier has been a continuous source of inspiration, not only in academic pursuits but also in problem-solving and personal growth. His philosophy of "learning by doing" has profoundly influenced me, demonstrating that true understanding often comes from direct engagement and hands-on experience. Working with him has left a lasting impression on me, particularly his ability to transform abstract concepts into concrete examples, making complex ideas more approachable and insightful.

My PhD project involved extensive simulations, with each step building upon models derived from the previous one. Whenever I encountered obstacles along this challenging journey, Dr. Baumeier was always there with thoughtful advice and innovative solutions. His patient guidance kept me moving forward, and without his support, I would not have been able to overcome these hurdles and achieve the results I did.

Beyond his academic guidance, Dr. Baumeier has also been a role model in how to communicate with empathy and respect. He has shown me the importance of considering others' perspectives and responding with patience and understanding — a quality I strive to embody in both professional and personal interactions.

A particularly memorable moment that I will always cherish happened in February 2020, when I returned to the Netherlands from my hometown at the onset of the Covid-19 pandemic. At that time, it was recommended that anyone traveling from China undergo a week of home quarantine. However, Dr. Baumeier trusted me completely, believing in my honesty about my health and allowing me to return to the office without hesitation. His trust and respect during such a tense period truly touched me and reinforced the strong sense of mutual respect. For all his invaluable mentorship, constant support, and the profound impact he has had on both my academic and personal development, I am sincerely grateful. His guidance has not only shaped my research but has also inspired me to grow as a thoughtful and compassionate individual.

I am also deeply grateful to my co-supervisor, Dr. Pim van der Hoorn, for his insightful guidance and expertise in probability theory and stochastic processes. His clear and precise explanations have greatly enriched my understanding of these complex subjects, providing me with essential tools to tackle the mathematical challenges of my research. I am especially thankful for his thoughtful feedback on my papers, which has significantly improved the quality of my academic writing. Furthermore, his valuable suggestions on structuring and refining my thesis have been instrumental in shaping it into a coherent and well-argued piece of work. Dr. Pim van der Hoorn support and dedication have played a crucial role in my academic growth, and I truly appreciate his patience and encouragement throughout this journey.

I would like to express my sincere gratitude to Pan Gao for her warm companionship, thoughtful invitations to explore beautiful parks, local superstores, and various charming spots in the community. I deeply value our meaningful conversations about life, mental health, career choices, well-being, academia, and personal relationships. Her support and friendship have been a source of comfort and inspiration.

I am also thankful to Yankun Hong for the memorable gatherings and engaging discussions about life and career paths. These moments of connection and shared experiences have enriched my journey, and I truly appreciate their kindness and encouragement. We are both from Guangdong and studied in Sun Yat-sen University during our Bachelor's. It is not easy to meet someone in the

Netherlands who comes from the same hometown with the same schooling. With such a coincidence, isn't it something to be grateful for the continuation of the good things in the past?

I am grateful to my colleagues in CASA. When I began my PhD, I am thankful for Vivek Sundaram, Onur Caylak, Gianluca Tirimbò, and Wouter Scharpach for the inspiring academic discussions and memorable hangouts during the challenging times of the Corona lockdown. Our conversations, not only about research but also about life and career choices, have been truly meaningful. I would also like to say my thanks to Ruben for his generous support in programming, especially with GitHub and C++. Additionally, I am thankful to Haoxiang Zhao and Qin Zhou for the relaxing and engaging gatherings, as well as the eye-opening conversations that introduced me to many fascinating aspects of life and culture in other provinces of China beyond my hometown. Thank you for the wonderful and comfortable time that we spent together in the office and outside the campus! I wish all my colleagues could have their dream career and future.

I would like to express my heartfelt gratitude to the people I met during my time on the ICMS student outreach team. This experience taught me invaluable skills in task prioritization and, more importantly, how to support friends with a devoted and sincere heart. With a strong spirit of service, our team organized a variety of activities aimed at helping everyone build meaningful connections within the academic community and find joy beyond their research work.

Organizing academic events and networking activities across multiple departments was undoubtedly challenging. However, it allowed me to grow in many ways, particularly in communicating effectively with individuals from diverse academic backgrounds, embracing different perspectives — often broader and more insightful — and learning to interact with people without bias or preconceived notions. These experiences have broadened my worldview and enriched my personal and professional development.

I am sincerely grateful to Monique J. Bruining for her warm welcome to the ICMS organization. Her concise yet impactful introduction to academic life as a PhD student left a lasting impression on me — a journey I have come to summarize as one of collaborative and innovative research. I deeply appreciate her

genuine concern for my career path and her thoughtful guidance along the way. Reflecting on my research journey, I realize that collaboration is an area I still need to strengthen, as evidenced by my publications where, apart from my supervisors, I worked largely independently. This acknowledgment fuels my motivation to seek more collaborative opportunities in the future.

I would also like to extend my thanks to Prof. Michiel Hochstenbach. Serving as a tutor for his Mathematics 1 course was a truly inspiring experience. His unique approach to teaching and learning mathematics profoundly influenced me, reinforcing the idea that mathematics is best learned through experience. The step-by-step process of mastering concepts, gradually improving skills, and reflecting on one's progress has made learning math not only more effective but also more enjoyable.

Beyond the classroom, I am thankful for Prof. Michiel's warm hospitality. I fondly remember the times he invited me and my colleagues to his home, especially the memorable Christmas dinner. The engaging memory games, probability dice games, and English word games created a lively and stimulating atmosphere, where everyone participated enthusiastically. These brain-stimulating games often left me inspired, reminding me of the joy of learning through play and interaction.

I would also like to extend my heartfelt thanks to Pascal den Boef for always being there to practice and play tennis with me. It has been a truly refreshing experience to step outside, feel the warmth of the sun and the fresh air, and engage in physical activity after long hours of intensive brain work in the office. These moments not only kept my body active but also provided a much-needed mental reset. And you are always there to offer me help, thank you for being supportive.

I am also deeply appreciative of the vibrant CASA community. I gave my thanks to the Chair of CASA: Professor Barry Koren, Professor Mark Peletier, Professor Bjorn Baumeier, for their effort of organizing the CASA activities which has benefited all PhDs in the group. The diverse range of activities and events has broadened my academic horizons, exposing me to fascinating research across different groups worldwide. I still clearly remember the stimulating colloquium

sessions every Wednesday, where insightful topics and inspiring speakers left a lasting impact on me. These opportunities not only enriched my academic experience but also fostered a sense of belonging within the research community.

Special thanks to Enna van Dijk and Gea Deddens for organizing and facilitating the daily activities at CASA. Your dedication and hard work have made my time here all the more enjoyable and productive. I am equally grateful to Jonelleke Kamperman for her invaluable assistance in helping me settle in Eindhoven — from securing accommodation to navigating visa applications and registration processes, your support made my transition smoother and less stressful.

I am truly grateful for all these experiences and the wonderful people I have met along this journey. They have not only shaped my academic path but also enriched my personal growth in ways I will cherish forever.

About the author

Zhongquan Chen was born on 1991 in Taishan, Guangdong Province, China. After completing secondary school in Taishan in 2011, he began studying Chemistry and Applied Mathematics at Sun Yat-sen University, obtaining a Bachelor of Science degree in 2015. Subsequently, he pursued a Master's degree in Chemistry at University of Saskatchewan, graduating in 2017 with a thesis title "Compositional Effects on Local Structure in Glassy Binary Lennard-Jones Nanoparticles".

In 2019, he joined the Bjorn Baumeier' group at the Department of Mathematics and Computer Science of Eindhoven University of Technology as a PhD candidate. The research focused on multiscale model of charge transport, which resulted in the dissertation titled "Multiscle Charge Transport Networks: Quantitative Analysis and Structural Characterization".

During the PhD, Zhongquan Chen contributed to course tutoring (Mathematics 1, Numerical Linear Algebra), scientific conferences and publications in journals. After completing the PhD, he plans to continue a career related to computational science.

Bibliography

[1] M. Barbaro, A. Caboni, P. Cosseddu, G. Mattana, and A. Bonfiglio. Active devices based on organic semiconductors for wearable applications. *IEEE Transactions on Information Technology in Biomedicine*, 14(3):758–766, May 2010.

[2] R. J. Hamers. Flexible electronic futures. *Nature*, 412(6846):489–490, August 2001.

[3] Aleksandr Biberman and Keren Bergman. Optical interconnection networks for high-performance computing systems. *Reports on Progress in Physics*, 75(4):046402, April 2012.

[4] K. W. Böer and U. W. Pohl. *Semiconductor Physics*. Springer, Cham, 2018.

[5] H. Sirringhaus. 25th anniversary article: organic field-effect transistors: the path beyond amorphous silicon. *Advanced Materials*, 26(9):1319–1335, March 2014.

[6] C. J. Brabec. Organic photovoltaics: technology and market. *Solar Energy Materials and Solar Cells*, 83(2-3):273–292, June 2004.

[7] W. Brütting, editor. *Physics of Organic Semiconductors*. Wiley, 1 edition, May 2005.

[8] A. Dey, A. Singh, D. Das, and P. K. Iyer. Organic Semiconductors: A New Future of Nanodevices and Applications. In *Thin Film Structures in Energy Applications*, pages 97–128. Springer International Publishing, Cham, 2015.

[9] M. Shibata, Y. Sakai, and D. Yokoyama. Advantages and disadvantages

of vacuum-deposited and spin-coated amorphous organic semiconductor films for organic light-emitting diodes. *Journal of Materials Chemistry C*, 3(42):11178–11191, 2015.

[10] J. Lim, W. Lee, D. Kwak, and K. Cho. Evaporation-induced self-organization of inkjet-printed organic semiconductors on surface-modified dielectrics for high-Performance Organic Transistors. *Langmuir*, 25(9):5404–5410, May 2009.

[11] S. Duan, X. Gao, Y. Wang, F. Yang, M. Chen, X. Zhang, X. Ren, and W. Hu. Scalable fabrication of highly crystalline organic semiconductor thin film by channel-restricted screen printing toward the low-cost fabrication of high-performance transistor arrays. *Advanced Materials*, 31(16):1807975, April 2019.

[12] W. Cao and J. Xue. Recent progress in organic photovoltaics: device architecture and optical design. *Energy and Environmental Science*, 7(7):2123, 2014.

[13] M.H. Yoon, A. Facchetti, C. E. Stern, and T. J. Marks. Fluorocarbon-modified organic semiconductors: molecular architecture, electronic, and crystal structure tuning of arene- versus fluoroarene-thiophene oligomer thin-film properties. *Journal of the American Chemical Society*, 128(17):5792–5801, May 2006.

[14] S. Aramaki, Y. Sakai, and N. Ono. Solution-processible organic semiconductor for transistor applications: Tetrabenzoporphyrin. *Applied Physics Letters*, 84(12):2085–2087, March 2004.

[15] J. Mei, Y. Diao, A. L. Appleton, L. Fang, and Z. Bao. Integrated materials design of organic semiconductors for field-effect transistors. *Journal of the American Chemical Society*, 135(18):6724–6746, May 2013.

[16] A. C. Arias, J. Devin MacKenzie, I. McCulloch, J. Rivnay, and A. Salleo. Materials and applications for large area electronics: solution-based approaches. *Chemical Reviews*, 110(1):3–24, January 2010.

[17] J. Liu, H. Zhang, H. Dong, L. Meng, L. Jiang, L. Jiang, Y. Wang, J. Yu, Y. Sun,

W. Hu, and Alan J. Heeger. High mobility emissive organic semiconductor. *Nature Communications*, 6(1):10032, December 2015.

[18] L. Li, J. Liang, S.Y. Chou, X. Zhu, X. Niu, B. Zhi, and Q. Pei. A solution processed flexible nanocomposite electrode with efficient light extraction for organic light emitting diodes. *Scientific Reports*, 4(1):4307, March 2014.

[19] A. W. Hains, Z. Liang, M. A. Woodhouse, and B. A. Gregg. Molecular semiconductors in organic photovoltaic cells. *Chemical Reviews*, 110(11):6689–6735, November 2010.

[20] Y. Lin, Y. Li, and X. Zhan. Small molecule semiconductors for high-efficiency organic photovoltaics. *Chemical Society Reviews*, 41(11):4245, 2012.

[21] K. A. Mazzio and C. K. Luscombe. The future of organic photovoltaics. *Chemical Society Reviews*, 44(1):78–90, 2015.

[22] D. Yang and D. Ma. Development of organic semiconductor photodetectors: from mechanism to applications. *Advanced Optical Materials*, 7(1):1800522, January 2019.

[23] P. C. Y. Chow and T. Someya. Organic photodetectors for next-generation wearable electronics. *Advanced Materials*, 32(15):1902045, April 2020.

[24] L. Janasz, M. Borkowski, Paul W. M. Blom, T. Marszalek, and W. Pisula. Organic semiconductor/insulator blends for elastic field-effect transistors and sensors. *Advanced Functional Materials*, 32(7):2105456, February 2022.

[25] Z. Wang, L. Huang, X. Zhu, X. Zhou, and L. Chi. An ultrasensitive organic semiconductor NO_2 sensor based on crystalline TIPS-pentacene films. *Advanced Materials*, 29(38):1703192, October 2017.

[26] Z. Wang, L. Huang, X. Zhu, X. Zhou, and L. Chi. Organic gas sensors: an ultrasensitive organic semiconductor NO_2 sensor based on crystalline TIPS-pentacene films. *Advanced Materials*, 29(38):adma.201770277, October 2017.

[27] S. Allard, M. Forster, B. Souharce, H. Thiem, and U. Scherf. Organic semi-

conductors for solution-processable field-effect transistors (OFETs). *Angewandte Chemie International Edition*, 47(22):4070–4098, May 2008.

[28] G. Horowitz. Organic field-effect transistors. *Advance Material*, 10(5):365–377, March 1998.

[29] R. G. Kepler. Charge carrier production and mobility in anthracene crystals. *Physical Review*, 119(4):1226–1229, August 1960.

[30] N. Karl. Charge carrier transport in organic semiconductors. *Synthetic Metals*, 133-134:649–657, March 2003.

[31] V. Coropceanu, J. Cornil, D. A. Da Silva Filho, Y. Olivier, R. Silbey, and J. L. Brédas. Charge transport in organic semiconductors. *Chemical Reviews*, 107(4):926–952, April 2007.

[32] H. Lee, C. Park, D. H. Sin, J. H. Park, and K. Cho. Recent advances in morphology optimization for organic photovoltaics. *Advanced Materials*, 30(34):1800453, August 2018.

[33] V. Bhat, C. P. Callaway, and C. Risko. Computational approaches for organic semiconductors: from chemical and physical understanding to predicting new materials. *Chemical Reviews*, 123(12):7498–7547, June 2023.

[34] S. Hutsch and F. Ortmann. Impact of heteroatoms and chemical functionalisation on crystal structure and carrier mobility of organic semiconductors. *npj Computational Materials*, 10(1):206, September 2024.

[35] M. Muccini. A bright future for organic field-effect transistors. *Nature Mater*, 5(8):605–613, August 2006.

[36] D. C. Young. *Computational Chemistry: A Practical Guide for Applying Techniques to Real-World Problems*. J. Wiley & sons, New York, 2001.

[37] H. Bronstein, C. B. Nielsen, B. C. Schroeder, and I. McCulloch. The role of chemical design in the performance of organic semiconductors. *Nature Reviews Chemistry*, 4(2):66–77, January 2020.

[38] N. M. O'Boyle, C. M. Campbell, and G. R. Hutchison. Computational design

and selection of optimal organic photovoltaic materials. *The Journal of Physical Chemistry C*, 115(32):16200–16210, August 2011.

[39] V. Ruhle, A. Lukyanov, F. May, M. Schrader, T. Vehoff, J. Kirkpatrick, B. Baumeier, and D. Andrienko. Microscopic simulations of charge transport in disordered organic semiconductors. *Journal of Chemical Theory and Computation*, 7(10):3335–3345, 2011. PMID: 22076120.

[40] A. N. Sokolov, S. Atahan-Evrenk, R. Mondal, H. B. Akkerman, R. S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S. C.B. Mannsfeld, A. P. Zoombelt, Z. Bao, and A. Aspuru-Guzik. From computational discovery to experimental characterization of a high hole mobility organic crystal. *Nature Communications*, 2(1):437, August 2011.

[41] H. Bässler. Charge transport in disordered organic photoconductors a monte carlo simulation study. *Physica Status Solidi (b)*, 175(1):15–56, January 1993.

[42] P. W. Anderson. Absence of diffusion in certain random lattices. *Physical Review*, 109(5):1492–1505, March 1958.

[43] S. D. Baranovskii. Theoretical description of charge transport in disordered organic semiconductors. *Physica Status Solidi (b)*, 251(3):487–525, March 2014.

[44] V. Rühle, A. Lukyanov, F. May, M. Schrader, T. Vehoff, J. Kirkpatrick, B. Baumeier, and D. Andrienko. Microscopic simulations of charge transport in disordered organic semiconductors. *Journal of Chemical Theory and Computation*, 7(10):3335–3345, October 2011.

[45] S. Olthof, S. Mehraeen, S. K. Mohapatra, S. Barlow, V. Coropceanu, J. L. Brédas, S. R. Marder, and A. Kahn. Ultralow doping in organic semiconductors: evidence of trap filling. *Physical Review Letters*, 109(17):176601, October 2012.

[46] D. Abbaszadeh, A. Kunz, G. A. H. Wetzelaer, J. J. Michels, N. I. Craciun, K. Koynov, I. Lieberwirth, and P. W. M. Blom. Elimination of charge carrier trapping in diluted semiconductors. *Nature Materials*, 15(6):628–633,

June 2016.

[47] V. Podzorov, E. Menard, A. Borissov, V. Kiryukhin, J. A. Rogers, and M. E. Gershenson. Intrinsic charge transport on the surface of organic semiconductors. *Physical Review Letters*, 93(8):086602, August 2004.

[48] M. Kuik, L. J. A. Koster, G. A. H. Wetzelaer, and P. W. M. Blom. Trap-sssisted recombination in disordered organic semiconductors. *Physical Review Letters*, 107(25):256805, December 2011.

[49] A. Rose. Space-charge-limited currents in solids. *Physical Review Journals*, 97(6):1538–1544, March 1955.

[50] Z. A. Lamport, H. F. Haneef, S. Anand, M. Waldrip, and O. D. Jurchescu. Tutorial: organic field-effect transistors: materials, structure and operation. *Journal of Applied Physics*, 124(7):071101, August 2018.

[51] A. Laudari and S. Guha. Bandlike transport in ferroelectric-based organic field-effect transistors. *Physical Review Applied*, 6(4):044007, October 2016.

[52] E.M. Muller and J.A. Marohn. Microscopic evidence for spatially inhomogeneous charge trapping in Pentacene. *Advanced Materials*, 17(11):1410–1414, June 2005.

[53] T. He, Y. Wu, G. Davino, E. Schmidt, M. Stolte, J. Cornil, D. Beljonne, P. Paul Ruden, F. Würthner, and C. D. Frisbie. Crystal step edges can trap electrons on the surfaces of n-type organic semiconductors. *Nature Communications*, 9(1):2141, May 2018.

[54] E. Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28(6):1049–1070, December 1926.

[55] M. Born and R. Oppenheimer. Zur quantentheorie der molekeln. *Annalen der Physik*, 389(20):457–484, January 1927.

[56] V. May and O. Kühn. *Charge and Energy Transfer Dynamics in Molecular Systems*. Wiley, 1 edition, February 2011.

[57] Mark E. Tuckerman. *Statistical mechanics: Theory and molecular simulation*. Oxford University Press, October 2023. Publisher Copyright: © Mark E. Tuckerman 2023. All rights reserved.

[58] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136(3B):B864–B871, November 1964.

[59] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140(4A):A1133–A1138, November 1965.

[60] K. Burke. Perspective on density functional theory. *The Journal of Chemical Physics*, 136(15):150901, April 2012.

[61] F. Malet, A. Mirtschink, Klaas J. H. Giesbertz, L. O. Wagner, and P. Gori-Giorgi. Exchange–correlation functionals from the strong interaction limit of DFT: applications to model chemical systems. *Physical Chemistry Chemical Physics*, 16(28):14551–14558, 2014.

[62] J. P. Perdew. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Physical Review B*, 33(12):8822–8824, June 1986.

[63] R. A. Marcus. On the theory of oxidation-reduction reactions involving electron transfer. I. *The Journal of Chemical Physics*, 24(5):966–978, May 1956.

[64] R. A. Marcus. Chemical and electrochemical electron transfer theory. *Annual Review of Physical Chemistry*, 15(1):155–196, October 1964.

[65] R. A. Marcus. Electron transfer reactions in chemistry. Theory and experiment. *Reviews of Modern Physics*, 65(3):599–610, July 1993.

[66] B. Baumeier, O. Stenzel, C. Poelking, D. Andrienko, and V. Schmidt. Stochastic modeling of molecular charge transport networks. *Physical Review B*, 86(18):184202, November 2012.

[67] N. E. Jackson, Brett M. Savoie, L. X. Chen, and MaM.rk A. Ratner. A simple index for characterizing charge transport in molecular materials. *The*

Journal of Physical Chemistry Letters, 6(6):1018–1021, March 2015.

[68] N. E. Jackson, L. X. Chen, and M. A. Ratner. Charge transport network dynamics in molecular aggregates. *Proceedings of the National Academy of Sciences of the United States of America*, 113(31):8595–8600, August 2016.

[69] H.J.C. Berendsen, D. Van Der Spoel, and R. Van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, September 1995.

[70] Martin Stroet, Bertrand Caron, Koen M. Visscher, Daan P. Geerke, Alpeshkumar K. Malde, and Alan E. Mark. Automated Topology Builder Version 3.0: Prediction of Solvation Free Enthalpies in Water and Hexane. *Journal of Chemical Theory and Computation*, 14(11):5834–5845, November 2018.

[71] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, January 2007.

[72] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, December 1981.

[73] W. F. Van Gunsteren and H. J. C. Berendsen. A Leap-frog Algorithm for Stochastic Dynamics. *Molecular Simulation*, 1(3):173–185, March 1988.

[74] C. M. Breneman and K. B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. the need for high sampling density in formamide conformational analysis. *Journal of Computational Chemistry*, 11(3):361–373, 1990.

[75] B.T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chemical Physics*, 59(3):341–350, August 1981.

[76] J. Wehner, L. Brombacher, J. Brown, C. Junghans, O. Caylak, Y. Khalak, P. Madhikar, G. Tirimbo, and B. Baumeier. Electronic excitations in complex molecular environments: Many-body green's functions theory in votca-xtp.

Journal of Chemical Theory and Computation, 14(12):6253–6268, 2018. PMID: 30404449.

[77] G. Tirimbò, V. Sundaram, O. Çaylak, W. Scharpach, J. Sijen, C. Junghans, J. Brown, F. Z. Ruiz, N. Renaud, J. Wehner, and B. Baumeier. Excited-state electronic structure of molecules using many-body green's functions: Quasiparticles and electron–hole excitations with votca-xtp. *The Journal of Chemical Physics*, 152(11):114103, 03 2020.

[78] B. Baumeier, J. Wehner, N. Renaud, F. Zapata Ruiz, R. Halver, P. Madhikar, R. Gerritsen, G. Tirimbo, J. Sijen, D. Rosenberger, J. S. Brown, V. Sundaram, J. Krajniak, M. Bernhardt, and C. Junghans. Votca: multiscale frameworks for quantum and classical simulations in soft matter. *Journal of Open Source Software*, 9(99):6864, 2024.

[79] P. O. Löwdin. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *The Journal of Chemical Physics*, 18(3):365–375, March 1950.

[80] P. O. Löwdin. On the Nonorthogonality Problem. In *Advances in Quantum Chemistry*, volume 5, pages 185–199. Elsevier, 1970.

[81] K. Suzuki, S. Maruyama, and K. Araki. Generalized method for calculating Löwdin orbitals. *Bulletin of the Chemical Society of Japan*, 46(2):355–358, February 1973.

[82] B. Baumeier, J. Kirkpatrick, and D. Andrienko. Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies. *Physical Chemistry Chemical Physics*, 12(36):11103, 2010.

[83] B. D. Hughes. *Random Walks and Random Environments: Volume 2: Random Environments*. Oxford University Press, Oxford, New York, June 1996.

[84] J. R. Norris. *Markov Chains*. Cambridge series on statistical and probabilistic mathematics. Cambridge University Press, Cambridge, UK ; New York, 1st pbk. ed edition, 1998.

[85] Tom Fischer. On simple representations of stopping times and stopping

time sigma-algebras. *Statistics & Probability Letters*, 83(1):345–349, January 2013.

[86] J. Ding, J. Lee, and Y. Peres. Cover times, blanket times, and majorizing measures. *Annals of Mathematics*, 175(3):1409–1471, May 2012.

[87] J. L. Snell. Finite Markov chains and their applications. *The American Mathematical Monthly*, 66(2):99–104, February 1959.

[88] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Undergraduate texts in mathematics. Springer, New York, NY Heidelberg Berlin, reprint edition, 1983.

[89] Hans U. Gerber and Shuo-Yen Robert Li. The occurrence of sequence patterns in repeated experiments and hitting times in a Markov chain. *Stochastic Processes and their Applications*, 11(1):101–108, March 1981.

[90] Haiyan Chen and Fuji Zhang. The expected hitting times for finite Markov chains. *Linear Algebra and its Applications*, 428(11-12):2730–2749, June 2008.

[91] P. G. Doyle and J. L. Snell. Random walks and electric networks. *arXiv*, 2000. Publisher: arXiv Version Number: 1.

[92] N. G. van Kampen. *Stochastic processes in physics and chemistry*. North-Holland personal library. Elsevier, Amsterdam ; Boston, 3rd ed edition, 2007. OCLC: ocm81453662.

[93] D. C. Lay, S. R. Lay, and J. McDonald. *Linear Algebra and Its Applications*. Pearson, Boston, fifth edition edition, 2016. OCLC: 875056178.

[94] K Binder. Applications of Monte Carlo methods to statistical physics. *Reports on Progress in Physics*, 60(5):487–559, May 1997.

[95] K. Binder. *Monte Carlo Methods in Statistical Physics*. Number 7 in Topics in Current Physics. Springer, Berlin Heidelberg, second edition edition, 1986.

[96] K. A. Fichthorn and W. H. Weinberg. Theoretical foundations of dynamical Monte Carlo simulations. *The Journal of Chemical Physics*, 95(2):1090–

1096, July 1991.

[97] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, 2008.

[98] Endre Süli and David F. Mayers. *An Introduction to Numerical Analysis*. Cambridge University Press, 1 edition, August 2003.

[99] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

[100] R. W. Freund and N. M. Nachtigal. QMR: a quasi-minimal residual method for non-Hermitian linear systems. *Numerische Mathematik*, 60(1):315–339, December 1991.

[101] H. A. Van Der Vorst. Bi-CGSTAB: A fast and smoothly converging variant of Bi-CG for the solution of nonsymmetric linear systems. *SIAM Journal on Scientific and Statistical Computing*, 13(2):631–644, March 1992.

[102] David L. Cheung and Alessandro Troisi. Modelling charge transport in organic semiconductors: from quantum dynamics to soft matter. *Physical Chemistry Chemical Physics*, 10(39):5941, 2008.

[103] Masahiro Sato, Akiko Kumada, and Kunihiko Hidaka. Multiscale modeling of charge transfer in polymers with flexible backbones. *Physical Chemistry Chemical Physics*, 21(4):1812–1819, 2019.

[104] T. Brereton, O. Stenzel, B. Baumeier, D. Andrienko, V. Schmidt, and D. Kroese. Efficient simulation of Markov chains using segmentation. *Methodology and Computing in Applied Probability*, 16(2):465–484, June 2014.

[105] W. F. Pasveer, J. Cottaar, C. Tanase, R. Coehoorn, P. A. Bobbert, P. W. M. Blom, D. M. deLeeuw, and M. A. J. Michels. Unified description of charge-carrier mobilities in disordered semiconducting polymers. *Physical Review Letters*, 94(20):206601, May 2005.

[106] W. F. Pasveer, P. A. Bobbert, H. P. Huinink, and M. A. J. Michels. Scaling of current distributions in variable-range hopping transport on two- and

three-dimensional lattices. *Physical Review B*, 72(17):174204, November 2005.

[107] M. Casalegno, A. Bernardi, and G. Raos. Numerical simulation of photocurrent generation in bilayer organic solar cells: Comparison of master equation and kinetic Monte Carlo approaches. *The Journal of Chemical Physics*, 139(2):024706, July 2013.

[108] L. J. A. Koster. Charge carrier mobility in disordered organic blends for photovoltaics. *Physical Review B*, 81(20):205318, May 2010.

[109] J. Cottaar, R. Coehoorn, and P.A. Bobbert. Modeling of charge transport across disordered organic heterojunctions. *Organic Electronics*, 13(4):667–672, April 2012.

[110] R. A. Marsh, C. Groves, and N. C. Greenham. A microscopic model for the behavior of nanostructured organic photovoltaic devices. *Journal of Applied Physics*, 101(8):083509, April 2007.

[111] U. Wolf, V. I. Arkhipov, and H. Bässler. Current injection from a metal to a disordered hopping system. I. Monte Carlo simulation. *Physical Review B*, 59(11):7507–7513, March 1999.

[112] M. Casalegno, C. Carbonera, S. Luzzati, and G. Raos. Coarse-grained kinetic modelling of bilayer heterojunction organic solar cells. *Organic Electronics*, 13(5):750–761, May 2012.

[113] M. Bouhassoune, S.L.M. Van Mensfoort, P.A. Bobbert, and R. Coehoorn. Carrier-density and field-dependent charge-carrier mobility in organic semiconductors with correlated Gaussian disorder. *Organic Electronics*, 10(3):437–445, May 2009.

[114] J. Frenkel. On pre-breakdown phenomena in insulators and electronic semi-conductors. *Phys. Rev.*, 54:647–648, Oct 1938.

[115] T. J. Brereton, D. P. Kroese, O. Stenzel, V. Schmidt, and B. Baumeier. Efficient simulation of charge transport in deep-trap media. In *Proceedings Title: Proceedings of the 2012 Winter Simulation Conference (WSC)*, pages

1–12, Berlin, Germany, December 2012. IEEE.

[116] S. Sarkar and Jeroen A. S. Witteveen. *Uncertainty quantification in computational science: theory and application in fluids and structural mechanics*. World Scientific Publishing Co. Pte. Ltd, Singapore, 2017.

[117] J. O'Connor, S. Laizet, A. Wynn, W. Edeling, and P. V. Coveney. Quantifying uncertainties in direct numerical simulations of a turbulent channel flow. *Computers & Fluids*, 268:106108, January 2024.

[118] A. Chernatynskiy, S. R. Phillpot, and R. LeSar. Uncertainty Quantification in Multiscale Simulation of Materials: A Prospective. *Annual Review of Materials Research*, 43(1):157–182, July 2013.

[119] D. Suleimenova, H. Arabnejad, W. N. Edeling, D. Coster, O. O. Luk, J. Lakhlili, V. Jancauskas, M. Kulczewski, L. Veen, D. Ye, P. Zun, V. Krzhizhanovskaya, A. Hoekstra, D. Crommelin, P. V. Coveney, and D. Groen. Tutorial applications for Verification, Validation and Uncertainty Quantification using VECMA toolkit. *Journal of Computational Science*, 53:101402, July 2021.

[120] P. V. Coveney, D. Groen, and A. G. Hoekstra. Reliability and reproducibility in computational science: implementing validation, verification and uncertainty quantification *in silico*. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2197):rsta.2020.0409, 20200409, May 2021.

[121] P. V. Coveney and R. R. Highfield. When we can trust computers (and when we can't). *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2197):rsta.2020.0067, 20200067, May 2021.

[122] C. Poelking, M. Tietze, C. Elschner, S. Olthof, D. Hertel, B. Baumeier, F. Würthner, K. Meerholz, K. Leo, and D. Andrienko. Impact of mesoscale order on open-circuit voltage in organic solar cells. *Nature Materials*, 14(4):434–439, April 2015.

[123] C. Poelking and D. Andrienko. Long-range embedding of molecular ions

and excitations in a polarizable molecular environment. *Journal of Chemical Theory and Computation*, 12(9):4516–4523, September 2016.

[124] B. Baumeier, J. Kirkpatrick, and D. Andrienko. Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies. *Physical Chemistry Chemical Physics*, 12(36):11103, 2010.

[125] J. P. Perdew, M. Ernzerhof, and K. Burke. Rationale for mixing exact exchange with density functional approximations. *The Journal of Chemical Physics*, 105(22):9982–9985, December 1996.

[126] A. A. L. Marques, J. Vidal, M. J. T. Oliveira, L. Reining, and S. Botti. Density-based mixing parameter for hybrid functionals. *Phys. Rev. B*, 83(3):035119, January 2011.

[127] C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *The Journal of Chemical Physics*, 110(13):6158–6170, April 1999.

[128] Ik Jang Ko, Hyuna Lee, Jin Hwan Park, Gyeong Woo Kim, Raju Lampande, Ramchandra Pode, and Jang Hyuk Kwon. An accurate measurement of the dipole orientation in various organic semiconductor films using photoluminescence exciton decay analysis. *Physical Chemistry Chemical Physics*, 21(13):7083–7089, 2019.

[129] Yu-Fan Chang, Cheng-Han Yu, Shao-Chun Yang, I-Hong Hong, Sin-Ciang Jiang, Hsin-Fei Meng, Heh-Lung Huang, Hsiao-Wen Zan, and Sheng-Fu Horng. Great improvement of operation-lifetime for all-solution OLEDs with mixed hosts by blade coating. *Organic Electronics*, 42:75–86, March 2017.

[130] Anirban Mondal, Leanne Paterson, Jaeyoung Cho, Kun-Han Lin, Bas van der Zee, Gert-Jan A. H. Wetzelaer, Andrei Stankevych, Alexander Vakhnin, Jang-Joo Kim, Andrey Kadashchuk, Paul W. M. Blom, Falk May, and Denis Andrienko. Molecular library of oled host materials—evaluating the multiscale simulation workflow. *Chemical Physics Reviews*, 2(3):031304, 07 2021.

[131] Andrea Saltelli, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, 181(2):259–270, February 2010.

[132] I.M Sobol. Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, February 2001.

[133] J. Frenkel. On Pre-Breakdown Phenomena in Insulators and Electronic Semi-Conductors. *Phys. Rev.*, 54(8):647–648, October 1938.

[134] K. K. Tsung and S. K. So. Carrier trapping and scattering in amorphous organic hole transporter. *Applied Physics Letters*, 92(10):103315, March 2008.

[135] S. Kaiser, T. Neumann, F. Symalla, T. Schlöder, A. Fediai, P. Friederich, and W. Wenzel. De Novo calculation of the charge carrier mobility in amorphous small molecule organic semiconductors. *Frontiers in Chemistry*, 9:801589, December 2021.

[136] A. Lukyanov and D. Andrienko. Extracting nondispersive charge carrier mobilities of organic semiconductors from simulations of small systems. *Phys. Rev. B*, 82:193202, Nov 2010.

[137] T. P. Schulze. Efficient kinetic Monte Carlo simulation. *Journal of Computational Physics*, 227(4):2455–2462, February 2008.

[138] O. Stenzel, C. Hirsch, T. Brereton, B. Baumeier, D. Andrienko, D. Kroese, and V. Schmidt. A general framework for consistent estimation of charge transport properties via random walks in random environments. *Multiscale Modeling & Simulation*, 12(3):1108–1134, January 2014.

[139] William J. Stewart. *Introduction to the numerical solution of Markov chains*. Princeton university press, Princeton, 1994.

[140] U. Von Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, December 2007.

[141] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982.

[142] C. Fraley and A. E Raftery. Model-based clustering, discriminant analysis, and density estimation. *Journal of the American Statistical Association*, 97(458):611–631, June 2002.

[143] S. Still and W. Bialek. How many clusters? An information-theoretic perspective. *Neural Computation*, 16(12):2483–2506, December 2004.

[144] R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(2):411–423, July 2001.

[145] E. Rudberg, E. H. Rubensson, and P. Sałek. Kohn-Sham density functional theory electronic structure calculations with linearly scaling computational time and memory ssage. *Journal of Chemical Theory and Computation*, 7(2):340–350, February 2011.

[146] Z. Chen, P. van der Hoorn, and B. Baumeier. Graph random walk for time-of-flight charge mobilities, 2024.

[147] O. Çaylak, A. Yaman, and B. Baumeier. Evolutionary approach to constructing a deep feedforward neural network for prediction of electronic coupling elements in molecular materials. *Journal of Chemical Theory and Computation*, 15(3):1777–1784, March 2019.

[148] Yi-Siang Wang, Chun-I Wang, Chou-Hsun Yang, and Chao-Ping Hsu. Machine-learned dynamic disorder of electron transfer coupling. *The Journal of Chemical Physics*, 159(3):034103, July 2023.

[149] Onur Çaylak and Björn Baumeier. Machine learning of quasiparticle energies in molecules and clusters. *Journal of Chemical Theory and Computation*, 17(8):4891–4900, August 2021.

[150] Tianhao Tan and Dong Wang. Machine learning based charge mobility prediction for organic semiconductors. *The Journal of Chemical Physics*, 158(9):094102, March 2023.

[151] Jan Hermann, Zeno Schätzle, and Frank Noé. Deep-neural-network solution of the electronic Schrödinger equation. *Nature Chemistry*, 12(10):891–897, October 2020.