Simulation of Charge Carriers in Organic Electronic Devices: Methods with their Fundamentals and Applications

Karin Zojer

Dedicated to Professor Karl Leo on the occasion of his 60th birthday

1. Introduction

The modeling of organic electronic devices aims at predicting the device performance, for example, the electric current, depending on the point of operation (i.e., applied, external voltage) and external conditions such as the temperature. Device modeling also plays a crucial role to correctly infer material parameters from experiments, for example, from experimentally obtained current voltage curves.

This article is dedicated to collect and compare prevalent simulation approaches that provide the operation of devices by predicting the supply and transport of charges depending on the point of operation. Organic light emitting diodes (OLEDs), organic photovoltaic (OPV) devices, and organic thin-film transistors (OTFTs) operate with hugely different charge carrier densities and different electric field distributions. Of course, a comprehensive simulation of opto-electronic devices such as OLEDs and OPV devices extends beyond mere charge transport description. Processes such as spin-dependent exciton formation and dissociation, exciton migration, and propagation of optical fields inside the device crucially determine the overall functionality. All electronic devices have in common that the operating conditions ought to control the charge transport in a most optimal fashion. In opto-electronic devices, the choice of operating conditions is combined with an optimization of the excitonic and light propagation properties.

With our intention to highlight state-of-the-art options to account for charge transport, approaches to model energy transport (be that the motion of excitons or heat transport) and associated challenges are beyond the scope of this article. Readers interested in the tight integration of charge and exciton transport turn to refs. [1,2]

With the exclusive focus on modeling charge transport in devices, we will introduce in this article the prevalent modeling techniques and highlight for each, how particular device-relevant aspects are considered, be them the role of externally applied voltages, charge carrier injection and extraction, and the consideration of layer sequences and electrode arrangements. Predominantly used device modeling approaches for organic electronics can be distinguished in three categories: i) mesoscopic modeling, ii) drift-diffusion-based modeling, and iii) circuit and compact modeling. Each of these model categories is inspired by modeling approaches, that were successfully established in conventional electronics, and have been demonstrated to support our understanding of the operation of organic devices. The first two categories aim at predicting the combined impact of properties specific for organic semiconductors (OSC) and device-specific geometries and related constraints. The third class rather seeks to mimic the device behavior by assembling known building blocks (lumped elements) into a “circuit.” Hence, associated approaches are inspired, at least in part, by circuit modeling and compact modeling. It will be the nature of charge transport and the large aspect ratios of organic devices that will add a particular flavor to the modeling approaches.

Figure 1 shall serve us as a rough guide for conceptual differences in the considered model categories. Therein, we assume a simple bilayer diode structure, in which two different organic semiconducting layers are sandwiched between

Device modeling is an established tool to design and optimize organic electronic devices, be them organic light emitting diodes, organic photovoltaic devices, or organic transistors and thin-film transistors. Further, reliable device simulations form the basis for elaborate experimental characterizations of crucial mechanisms encountered in such devices. The present contribution collects and compares contemporary model approaches to describe charge transport in devices. These approaches comprise kinetic Monte Carlo, the master equation, drift-diffusion, and equivalent circuit analysis. This overview particularly aims at highlighting the following three aspects for each method: i) The foundation of a method including inherent assumptions and capabilities, ii) how the nature of organic semiconductors enters the model, and iii) how major tuning handles required to control the device operation are accounted for, namely temperature, external field, and provision of mobile carriers. As these approaches form a hierarchy of models suitable for multiscale modeling, this contribution also points out less established or even missing links between the approaches.
two electrodes. We illustrate for the example of the electron concentration, a key quantity to judge the electric current, how differently device simulation methods describe the evolution of the electron concentration \( n(\vec{r}, t) \) as a function of time \( t \) for each position \( \vec{r} \) in the device. Within mesoscopic modeling, there are the kinetic Monte Carlo and the Master equation approach that are both suited to describe hopping transport. Both approaches conceive the organic materials as a set of spatially distributed sites, between which particles (here: mobile charges) can hop. Here, these sites are arranged in a regular lattice and each site is associated with a certain site energy. The difference in site energies will decide whether a thermally assisted hop takes place or not; the electronic nature of the sites determines whether tunneling occurs. The range of possible site energies depends on the material and is schematically indicated as density of states next to the corresponding organic layer. In this way, the method straightforwardly discriminates between different materials. The actual site energies in the layers are indicated with a color code in accord with the associated density of states. In the course of the simulation, the displacement of these particle-like charges are recorded. Averaging the number of the charges encountered at a certain position in the device after a sufficiently long period of time yields the mean number of charges \( \langle n \rangle \) per reference volume.

Also a Master equation perceives the material as a (lattice) arrangement of distinct sites with their energies. Rather than tracking the particles, the Master equation asks for the likelihood with which each site is occupied. In Figure 1, the occupation probability is represented by circles; larger radii correspond to a higher occupation. The lower the site energy, the
more electrons prefer to occupy this site. The charge carrier density in a reference volume is, thus, now solely defined by the occupation likelihood of the sites in that volume. The drift-diffusion approach readily drops this energy-dependent point of view. Rather than looking at particles capable of occupying sites, the method assumes an electron density \( n \) that may vary continuously in space. Drift diffusion further assumes that the device in steady state is still close to thermal equilibrium, so that one can ensure that the electron density is redistributed only within states located lowest in energy. The driving forces responsible for this redistribution neither acknowledge the energy of these states with respect to the Fermi level nor - at least in its conventional form - possible gradients in the state energies. The nature of the layers enters simulations as effective material parameters that connect driving forces to possible charge rearrangements. In equivalent circuit networks, shown at bottom of Figure 1, the clear distinction of the layers is absent. Rather, large signal circuits are designed to mimic the current-voltage relations. In the example shown, branches containing different basic circuit elements (resistors, capacitors, and ideal diodes) are arranged to correctly reflect the current in different voltage regimes. The choice of necessary circuit elements has neither been guided nor justified by the specific properties of the organic semiconductors. The small signal model shown in this example assigns basic elements (here RC links) to each organic layer to be able to interpret the transient behavior in terms of distinct contributions from the two layers.\(^{[4]}\)

Somewhat counter-intuitively, our short list of model approaches to be considered does not comprise atomistic modeling. The consideration of charge redistributions on the atomistic scale including all incurring quantum-mechanical effects might be manageably for molecular junction devices, but the computational effort to capture an entire organic device, that extend to more than 100 nm is forbiddingly excessive. Rather, atomistic modeling has been established as a source for key input for modeling approaches that act at larger length scales.\(^{[5–9]}\) In fact, the above-mentioned mesoscopic approaches combined with atomistic modeling have shaped our current understanding of hopping transport in organic semiconductors. A comprehensive overview on how our understanding emerged with such simulation approaches and to which extent these insights still depend on the assumptions and capabilities of the approaches can be found well detailed in ref. \([1]\).

This article shall provide an in-depth discussion on the foundation of the simulation models and we will explore to which extent the peculiar features of organic semiconductors will enter transport simulations. To do so, we will first collect the relevant aspects of charge transport in organic semiconductors. Equipped with this basis understanding, we will introduce each model category by showing the underlying model equations and lining out how geometry and materials and external tuning handles are considered.

### 2. Underlying Mechanism of Charge Transport

Candidate device modeling approaches must reflect particular aspects related to the basic physics underlying charge transport in organic semiconductors. For a comprehensive overview on established and emerging aspects, the reader may consider ref. \([1]\) and refs. \([10,11]\). Here we collect the major aspects that are relevant for an appropriate consideration of charge transport.

As in any semiconductor, electrons and holes are the two essential charge carriers. Each of these charge carriers is supposed to transfer amongst a group of single electron states that share a comparable nature and energy (Figure 2). For electrons, the associated group of states is derived from the lowest unoccupied molecular orbital (LUMO) of the constituent molecule or building block. Holes move in states derived from the highest occupied molecular orbital (HOMO). Each group of states gives rise to a (possibly spatially varying) density of states \( \rho \) and \( \rho(E) \), located at reference energies \( E_{\text{LUMO}} \) and \( E_{\text{HOMO}} \), as shown in Figure 2. These reference energies do not correspond to the energy levels of the isolated molecules, as depolarization and inter-molecule interactions in the solid and alignment constraints at interfaces strongly modify the latter. For the purpose of this article, the energies \( E_{\text{LUMO}} \) and \( E_{\text{HOMO}} \) will serve us to locate the states in \( \rho \) involved in electron and hole transport with reference level energy \( E_{\text{LUMO}} \) and \( E_{\text{HOMO}} \) in energy.

The key parameters to rationalize or even predict electric currents (e.g., with the drift-diffusion method) are the drift mobilities of electron and holes, respectively. The mobility \( \mu \) is a material-dependent parameter that connects the electric field strength \( F \) with the resulting mean velocity \( \mathbf{v} \) of all charge carriers of one type, for example, for holes \( \mathbf{v}_\text{h} = \mu \mathbf{F} \). Even if we do not necessarily need the mobility for device simulations, the peculiarities of transport in organic semiconductors is reflected in the temperature, field-dependence, and carrier-density dependence of the mobility.\(^{[12,13]}\) In fact, analytical expressions for the mobility and its dependencies on temperature, field and charge carrier density have been conceived on the basis of mesoscopic simulations (to be discussed in Section 3) and have been tremendously helpful to comprehend the transport in devices. The reader is referred to Ref. \([13]\) for a comprehensive and critical evaluation of the related Gaussian disorder model (GDM) and its further developments.

![Figure 2. Density of states \( \rho(E) \) involved in electron transport with reference level energy \( E_{\text{LUMO}} \) and density of states \( \rho(E) \) involved in hole transport with reference level energy \( E_{\text{HOMO}} \).](image-url)
In contrast to crystalline inorganic semiconductors, the mobility of electrons in organic semiconductors is governed by the nature of the phonon-modified electronic states rather than by the frequency of (often phonon-mediated) collisions. In general, organic semiconductors consist of \( \pi \)-conjugated moieties (sites), be them molecules or segments of polymers, whose assembly into a solid matrix is driven by van der Waals interactions. The transport of charges in such a material is governed by the interplay between the electronic coupling between pairs of moieties and electron-phonon interactions, that is, interactions with intramolecular and intermolecular vibrations. Strong electron-phonon coupling leads to a localization of the electronic states on a one or a just a few moieties. Localization is the key to understand coherent motion in single crystals as well as incoherent motion in non-perfectly crystalline and amorphous materials.

2.1. Localization in Ordered Crystals

Charge transport in organic crystals cannot be exclusively explained in terms of coherent motion in delocalized bands, even if their mobility exhibits a band-like temperature dependence.\[^{[14]}\] Though the nature of transport in organic crystals keeps being intensively discussed, there appears to be an increasing consensus, that localization effects govern charge transport, even in highly pure organic crystals.\[^{[11,15]}\]

In general, the dominant causes for localization are static disorder and the modification of the electronic states due to a coupling between electronic states and intramolecular vibrations. Static disorder originates from a lack of strict periodic order in the arrangement of organic molecules or from impurities. The latter give rise non-negligible local variations i) in the density of molecules or hopping sites or/and ii) in the associated site energies. Such variations are a static manifestation of extrinsic influences, for example, due to processing conditions, and play a key role for transport in non-crystalline organic semiconductors, as will be discussed in Section 2.2. Yet, the concept of static disorder also aids to comprehend the peculiar situation in crystals, because it aids to capture the impact of local and temporal changes in the environment of a charge while it moves.\[^{[10,11,16]}\]

The electronic states available for electronic transport in crystals are so-called polaron bands that form due to a coupling between the purely electronic states and predominantly intramolecular vibrations.\[^{[17,18,20]}\] In the absence of vibrational coupling, all electronic states emerge as bands. These bands are narrow in comparison to covalently linked materials, as the electronic coupling is inherently weak. An additional coupling between electrons and phonons lowers the dispersion of the resulting bands even further. If sufficiently strong, the coupling can lead to the formation of quasi-particles, so-called small polarons,\[^{[19]}\] in which electrons are surrounded by phonon clouds. The mobility of electrons described as small polarons reduces with increasing temperature due to a narrowing of the polaron bandwidths.\[^{[21]}\]

In fact, even more decisive for the transport in crystals is a potentially strong coupling between electrons and certain low-frequency, intermolecular phonons. These phonons are associated to lattice vibrations that rigidly displace molecules in the crystal. This implies that also the \( \pi \)-systems of two adjacent molecules are rigidly displaced with respect to each other. Figure 3 illustrates such a vibration with a frequency as low as 23 cm\(^{-1}\) (24 cm\(^{-1}\)). It translates the \( \pi \)-system of a DNTT (C\(_8\)-DNTT) molecule with respect to the other molecule along the long molecular axis. As such vibrations are easily thermally excited, these vibrations causes large amplitudes and, hence, large displacements of the molecules. These displacements, in particular rigid translations, lead to a marked modulation of the electronic coupling between the molecules in the affected pairs. The position of the molecules, electronic coupling strengths, and the state energies encountered in the crystal at a certain position depend now on the moment in time.

This temporary deviation from perfect crystalline order, that is, dynamic disorder, leads to a localization of the wavefunction of the charge carrier. In turn, this localization prevents a description of charge transport as transitions between electronic states with a well-defined crystal momentum. That is, it is not possible to formulate a Bloch–Boltzmann equation; the corresponding lack of momentum conservation already prevents an adequate formulation of scattering processes.

The consequences of strong dynamic disorder for charge transport are collected and explained in Fratini et al.\[^{[20,22]}\] Starting point is a quantum diffusion model that monitors the mean displacement squared of a moving charge.\[^{[24]}\] The diffusivity of this charge is found to depend on the time span of observation. The most important change in the motion pattern occurs at times comparable to the time scale of lattice vibrations, that is, a characteristic time \( \tau = 2\pi/\omega \) being inversely proportional to the frequency \( \omega \) of a lattice vibration. At times shorter or comparable to \( \tau \), the charge carrier experiences the crystal environment as “an essentially frozen, disordered landscape.”\[^{[24]}\] The associated, marked decrease of the diffusivity with time is reminiscent to carrier transport between...
localized sites in a static landscape.\textsuperscript{[14]} Correspondingly, the charge can be viewed as being transiently localized.\textsuperscript{[13,14]} Later, at times exceeding $\tau$, classical diffusive motion is restored.\textsuperscript{[10,14]} That implies that the time required to restore the long-term diffusion limit depends on the nature of the molecules or moieties involved in the vibration. The larger the mass of these molecules, the slower the vibrational motion and the smaller the energy $\omega$, and the more time the charge carrier needs to recover diffusive motion.\textsuperscript{[13]}

Regardless of the time needed to reach the diffusive limit, the associated diffusivities show the qualitative temperature-dependence that one would expect from semiclassical band transport models. The latter transport models rest on the semiclassical Boltzmann equation that describes the motion of a Bloch wave packet $a$ with well-defined momentum in a periodic potential. The wave packet can be scattered by vibrations or impurities into other states with well-defined momentum. The mobility is proportional to the time between consecutive scattering events and inversely proportional to the effective mass. From that relation it can be readily seen that an increasing temperature boosts the likelihood of scattering with phonons, reduces the scattering time, and decreases the mobility along with the diffusivity. Theoretical estimates based on scattering with all phonons\textsuperscript{[23]} and longitudinal intermolecular phonons\textsuperscript{[24,25]} suggest that low energy acoustic phonons are the most important scatterers.

To explain the temperature-dependence in the framework of dynamic disorder, Fratini et al. relate the diffusivities, hence also the mobilities to the characteristic displacement squared $L^2$ encountered at the time $\tau$ via $\mu = L(t)/\tau$. Thus, $L$ acts as a transient localization length.\textsuperscript{[10,22]} An increasing temperature $T$ enhances thermal disorder that, in turn, localizes the charge more and more. The corresponding lowering in $L$ has been found to follow a power-law dependence $L^2 \propto T^p$; hence, the mobility evolves as $\mu \propto T^{-(p+1)}$ with $p + 1 = 2$.\textsuperscript{[13]} However, any temperature dependence of the form $\mu \propto T^q$ with an exponent $q > 0$ is regarded as a fingerprint for band transport. In particular, $\mu \propto T^2$ is expected for band transport in organic crystals possessing bandwidths smaller than $k_BT$.\textsuperscript{[25,26]} For bandwidths exceeding $k_BT$, the mobility decreases with temperature $\mu \propto T^{-3/2}$.\textsuperscript{[25,27]} Hence transport governed by strong dynamic disorder can exhibit the same “fingerprint” temperature dependence as band transport.

The theoretical considerations summarized above provide the fundamental basis to comprehend the motion of charges in organic crystals. This limiting case of coherent transport governed by strong dynamic disorder has been linked into a unified theoretical description that is also capable addressing the case of coherent band transport, that is limited by weak electron–phonon scattering.\textsuperscript{[11]} However, to the best of our knowledge, these considerations have not yet been extended to explore, how the mobility could be affected by local electric fields or the presence of other charges. In terms of device modeling that implies that, at least for the time being, charge transport through materials affected by strong dynamic disorder will be accounted for by an effective, field- and charge carrier-independent mobility. This readily propels the device modeling to the level of drift-diffusion modeling (cf., Section 4). As we will see right below, this is at stark contrast to non-crystalline materials for which the impact of which temperature, local electric fields, and charge carrier density is microscopically explored.

### 2.2. Localization and Hopping Transport

For most optoelectronic applications, single crystals are of subordinate interest. Prevalent are either amorphous or partially crystalline materials. For example, light emitting diodes preferentially rely on amorphous films. As such devices contain large sequences of ultrathin layers,\textsuperscript{[28]} reliable long term operation requires that each interface is well defined and remains smooth within the lifetime of the device. This necessary smoothness is most commonly achieved with amorphous thin films. As another example, state-of-the-art bulk heterojunctions formed of donor-acceptor blends for OPV devices exhibit $\pi$-stacking of donor and acceptor moieties, but lack long-range order.\textsuperscript{[29–34]} In such materials non-negligible static disorder is encountered and decisively influences the charge transport.

In the absence of near-range order, as present in amorphous films, electronic states formed on molecules or polymer chain segments will not spread beyond a few molecules. Charges occupying such states are localized. The motion of a charge in such films is commonly viewed as a hopping motion, that is, as a sequence of incoherent transfers from one localized site to another (Figure 4a), one charge carrier at a time.

This view contains already a “coarse-graining” of the motion description in terms of space and time.\textsuperscript{[15,36]} From the perspective of molecular dynamics, a promotion of the charge from site $i$ to $j$ in Figure 4a involves times scales related to molecular vibrations and rotations, that is, of the order of femtoseconds. Figure 4b shows the sites as local minima in a schematically drawn potential energy landscape. A charge residing in a minimum undergoes a large number of fast motions. Only in a small fraction of cases the charge escapes the local minimum $i$ and lands in local minimum $j$. Correspondingly, the charge spends most of the time in a local minimum. The time a system spends in such a “long-time” state is the retention time. Hence it is convenient to characterize the dynamics of systems in terms of rare, discrete transitions between long-term states. In this way most of the processes are neglected; motions on smaller timescales must be retained only if they change a long-term state.\textsuperscript{[35]} We will see in Section 3.1.2, that this neglect will allow the kinetic Monte Carlo method to reach simulation times of the order of seconds and beyond.

A second implication of hopping is that the coherence of the wavefunction associated to the carrier is lost during a hop, that is, the charge carrier moves ahead without “remembering” the history that brought it to the current site. Note that also descriptions for a coherent motion in a statically disordered energy landscape are emergent.\textsuperscript{[37]}

For a pair of sites, the likelihood of a transfer is given by a transfer rate, $w_{ij}$, that takes different formulations depending on the nature of the charge transfer.\textsuperscript{[38–42]}

Before giving examples for possible rate formulations, let us first introduce the following notation to record the position of all $N_{sc}$ considered charge carriers: A so-called configuration $x$ stores all hopping sites that are occupied with a charge carrier $x = (s_1, s_2,..., s_{N_{sc}})$, that is, charge $s$ occupies site $s_i$, charge 2 occupies site $s_j$, etc. Whenever we perform a transition of a
charge carrier $i$, formally a transition from a configuration $x$ with the occupied site $s_i$ to a configuration $x'$ with site $s_i'$ takes place (Figure 4c). The notion of configurations readily helps to understand that the likelihood of a hop, now rather described with a hopping rate $w_{xx'}$, is not only governed by the electronic coupling between the associated sites $s_i$ and $s_i'$, but also by interaction of the charge with all other charges in the system.

To illustrate typical choices for transition rates, let us turn to the prominently used Miller Abrahams rate and Marcus rate. These rates originate from applying time-dependent perturbation theory to describe the charge transfer process under the assumption of weak electronic coupling. The Miller Abrahams rate expression reads as:

$$w_{xx'}^\text{MA} = V_0 e^{-2\alpha \left| \Delta E_{xx'} \right|}$$

(1)

The spatial decay term $e^{-2\alpha \left| \Delta E_{xx'} \right|}$ contains a charge delocalization constant $\alpha$ and the distance between the two sites $s$ and $s'$ involved in the transition from configuration $x$ to $x'$. This factors adopts an exponentially decaying function of the distance between the two sites to account for tunneling transfer.

The prefactor $V_0$ is the attempt frequency and represents the largest possible rate. Device simulations prefer normalized rates $\tilde{w} = w / V_0$ (complying with Equation (5), see below) and determine with these rates the relative importance of different local hops. After the simulation, values of transport-related quantities are reconstructed by re-scaling the time using the attempt frequency. As illustrated in Figure 4d, the Miller–Abrahams rate expression (Equation (1)) treats all hops downward in energy as equiprobable while hops upward in energy are damped by a Boltzmann factor $e^{-\Delta E_{xx'}/kT}$. Note that Miller Abrahams rates rely on weak electron–phonon interactions and low temperatures. The derivation of the rate assumes that only one phonon can be absorbed or emitted during a hop (reminiscent to the deformation potential theory for electron-phonon scattering). Hence, the energy difference to be overcome during a hop should not exceed the maximum energy of the acoustical phonons (the Debye energy) and the energy of the optical phonons.

In contrast, Marcus rates assume a strong electron-phonon coupling and directly consider the effect of local geometry distortions in the vicinity of the charge. In the high temperature limit, the effect of local distortions can be entirely captured by a single parameter, the reorganization energy

---

**Figure 4.**

(a) Schematic view on a disordered organic semiconductor. Each molecule serves as a site that can be occupied with a charge. (b) Schematic sketch of the potential energy landscape containing the sites $i$ and $j$ as local minima. A charge located in one local minimum undergoes a large number of fast motions, before it escapes the minimum. Coarse graining defines long-time states associated to the energy minima, between which discrete, infrequent transitions occur. Adapted under the terms and conditions of the Creative Commons Attribution (CC-BY) license. Copyright 2018, The Authors, published by MDPI. (c) Transition between a configuration $x$ and a configuration $x'$ in a disordered energy landscape. In this lattice site representation, the site energies are given color coded. The transition removes a charge from a site $i$ and restores it on a site $j$. (d) Evolution of the energy-dependent factors in the Miller Abraham (left panel) and Marcus rate formulation as a function of the energy difference between the two involved configurations $\Delta E_{xx'}$. 

---
The expression for Marcus rates correspondingly reads:

$$w^{M}_{x,s} = v_0 \exp \left( - \frac{(\Delta E_{x,s} + E_n)^2}{4E_k T} \right)$$

(2)

The third term in the rate expression adopts a Gaussian, whose peak is centered at the negative reorganization energy $-E_r$. That implies, as illustrated in Figure 4d, that such hops downward in energy, whose energy difference $\Delta E$ exactly matches the reorganization energy $\Delta E = E_r$, are the fastest; hops for any other energy differences are slower. For small positive or negative energy differences in the order of the reorganization energy, the quantitative behavior of the Marcus and the Miller–Abrahams rates is quite similar. On the other hand, for high negative energy differences the rates obviously disagree.

The above-discussed Miller–Abrahams rate and the Marcus rate are two expressions out of many possible choices. In general, eligible expressions for normalized rates $\tilde{w}_{x,s}$ adopt a common form consisting of two factors:

$$\tilde{w}_{x,s} = p(\bar{r}_s - \bar{r}) b(\bar{r}, E_n; \bar{r}_s, E_x)$$

(3)

Therein, $\bar{r}$ and $\bar{r}_s$ are the positions of site $s$ and $s'$, respectively. The first factor $p$ depends solely on the distance $|\bar{r}_s - \bar{r}|$ of the involved sites and accounts for microscopic, purely spatial relations between the sites due to electronic coupling. The second factor $b$ addresses a possible dependence on total energy of the system before, $E_x$, and after the hop, $E_{x,s}$. The factor indicates that transitions between configurations of different total energies $E_x$ can be considered and accounts for thermally activated hops. It also permits to incorporate a possible influence of a macroscopic, smoothly varying electrostatic potential $\phi(\bar{r})$, that does not depend on the local disorder. To ensure the existence and uniqueness of a steady-state distribution of probabilities of finding the system in configurations $x$, the factor $b$ must comply with the detailed balance condition:

$$\frac{E_{x,s} + \phi(\bar{r}_s)}{4E_k T} b(\bar{r}, E_n; \bar{r}_s, E_x) = \frac{E_{x,s} + \phi(\bar{r})}{4E_k T} b(\bar{r}, E_n; \bar{r}_s, E_x)$$

(4)

with $k_0$ being the Boltzmann constant and $T$ the temperature.

The distance-dependent factor $p$ is subject to two constraints:

$$\int_{\mathbb{R}^3} p(|\bar{r}|) d^3r = 1, \quad \text{and}$$

(5)

$$\int_{\mathbb{R}^3} |\bar{r}|^2 p(|\bar{r}|) d^3r < \infty$$

(6)

The latter ensures finite diffusion currents in the system.

In the formulations used above, these rates are applied to all pairs of sites regardless their position in the site ensemble or the relative alignment of the molecules. An alternative is the use of highly customized, truly site-dependent values of charge transfer rates. These rates are extracted from atomistic simulations that predict the alignment of the molecules in a disordered volume and the electronic coupling between each pair of molecular sites.

In either way, charge carriers must be made to explore a sizable volume of the material if one wants to receive a reliable estimate of the displacement charges due to hopping motion. Hence, the regard of microscopic detail that governs the hopping motion in a disordered organic material incurs substantial costs in device simulations.

2.3. Level Alignment at Interfaces

Inevitably, interfaces crucially control the function of the device in multiple ways. Interfaces enable, maintain, or hamper the current flow, immobilize or release charges. Essentially all organic electronic devices exploit or rely on the properties of interfaces formed with and between organic semiconductors.

It is the energy level alignment, that is, the relative position of the energy levels of occupied and unoccupied states, that determines the charge carrier flux between two materials. Figure 5 schematically shows a multilayer diode, in which each of the semiconducting layers 1–4 may assume a different function. For example, in organic light emitting diodes these layers may serve electron or hole transport, as electron or hole blocking layers, or as emission layers. Once discriminates between the LUMO level offset and HOMO level offset, denoted as $\Delta_H$ and $\Delta_L$ in Figure 5. Further relevant is the offset between the energy levels in a semiconductor and the Fermi level $\mu_\text{F}$ of adjacent electrodes. Whichever offset to a given Fermi level $\mu_\text{F}$ is smaller takes the role of an injection barrier. High workfunction conductors usually serve as hole injecting contacts (anodes) with a barrier $\Delta_p$ in Figure 5. Vice versa, cathode materials are low workfunction conductors that may form an electron injection barrier $\Delta_e$ at the interface.

![Figure 5. Schematic alignment of LUMO and HOMO energies in a fictitious four-layer diode device under flat-band conditions. The offsets between LUMO and HOMO energies at the interfaces between organic semiconductors are denoted as $\Delta_H$ and $\Delta_L$, respectively. The offset between the Fermi level of an electrode (cathode) to the LUMO level in the adjacent semiconductor, $\Delta_H$, defines the injection barrier for electrons. Vice versa, the injection barrier $\Delta_e$ for holes is defined by the offset between Fermi level of an electrode (anode) to the HOMO level in the adjacent semiconductor.](https://www.advancedsciencenews.com/doi/fig/10.1002/adom.202100219)
2.3.1. Organic–Electrode Interfaces

The interfaces formed between the electrode materials and organic semiconductors determine how well charge carriers can be provided or extracted from the device. It even may decisively influence the level alignment between organic layers.\[52\] Device simulations require the information, whether an injection barrier is formed and which shape this injection barrier is going to adopt. Injection of charges is hindered, when a Schottky contact is formed.

The explanation that follows below blends the argumentation collected in the recent review articles of Zojer et al.\[53\] and Xu et al.\[54\]. The former compiles and explains effects associated to OSC interfaces involving conductive substrates; its insights feed the arguments developed in the latter article from a device point of view.

OSC differ from inorganic semiconductors in the following three aspects: OSC are often utilized in their intrinsic state without intentional doping. As their band gap is large in comparison to inorganic crystalline semiconductors, the intrinsic electron and hole densities are inherently very low. The large band gap is also the reason, why usually only one type of charge carriers is involved in establishing the contact. For non-doped organic semiconductors, the alignment of the Fermi level in the electrode with respect to the energy of the transport levels provokes a situation that is fundamentally different from the textbook illustrations for inorganic semiconductors.\[53,54\] To rationalize the shape of injection barriers into OSC, that is, their height and width, it is useful to first comprehend the possible occurrence of band bending and, in a second step, to closely inspect the situation in the vicinity of the contact interface.\[53\]

While the level alignment at the interface determines the offset between Fermi level and OSC right at the interface, band bending determines the extent to which this offset depends on the distance from the interface.

A contact between a metal surface and a pristine OSC is schematically shown in Figure 6a. The DOS for electrons and holes of the pristine OSC are indicated as dark grey Gaussians and deliberately positioned to prevent any overlap with the Fermi level of the metal, that is, to prevent Fermi level pinning.

To see how band bending could affect the alignment of the transport levels in the OSC with \(\mu_F\), it is instructive to revisit the well-known picture for inorganic semiconductors. Schottky contacts between metals and inorganic semiconductors exhibit

![Figure 6](image-url)

**Figure 6.** Schematic illustration of the alignment of the transport levels \(E_{\text{LUMO}}\) and \(E_{\text{HOMO}}\) with the Fermi level \(\mu_F\) of the contact electrode in the absence of doping. A Gaussian-shaped density of states in the organic semiconductor is indicated in dark grey. a) Position of the Fermi level in the metal and the levels \(E_{\text{LUMO}}\) and \(E_{\text{HOMO}}\) in thermodynamic equilibrium \((F = 0)\) before (left) and after (right) the contact is formed. The shown alignment refers to a large offset between Fermi level \(\mu_F\) in the metal and \(E_{\text{LUMO}}^0\). b) Close-up on the situation shown in (a) in the vicinity of the contact as a function of the distance \(r\) from the interface. The offset \(\Delta = E_{\text{LUMO}}^0 - \mu_F\) is the nominal injection barrier for electrons. c) Contact interface shown in (b) with a uniform external field \(|F| > 0\) in reverse bias. The contour \(E_{\text{LUMO}}^0\) is the potential energy of a single injected electron. It accounts for the Coulomb attraction to the image charge in the polarized conductor (red) and the potential energy due to the external field (blue). All energies \(E_{\text{LUMO}}^0\) located above \(\mu_F\) form the injection barrier with an effectively lowered height \(\Delta^*\).
a band bending that is due to a density $N_{\text{imp}}$ of uncompensated charges of dopant ions in the depletion region. The width of the depletion regions relates to this space-charge density $N$ as $1/\sqrt{N_{\text{imp}}}$. Accordingly, very small space-charge densities cause a bending across substantial distances from the interface. For OSC, we may readily expect distances that may exceed the thickness of the organic layer adjacent to the metal contact.

Band bending has been observed for various metal-OSC interfaces\cite{55-65} and has been explained as an electrostatic energy contribution from a space charge density formed due to an accumulation of additional charge carriers.\cite{58} Prime candidates to exhibit band bending are either intentionally or unintentionally doped OSC or OSC that possess extrinsically-induced gap states due to defects or disorder.\cite{55-58,65-65} However, even nominally pristine OSC exhibit band bending.\cite{59-65} The occurrence of space charges for such cases is typically explained by a filling of electronic states of molecules that are located at some distance from the interface.\cite{60,64,66} As rationalized by Oehzelt et al.,\cite{67} the DOS in the organic semiconductor can be populated with charges from the metal when the DOS overlaps with the Fermi level in the electrode, even to a very small extent. The offset at which this overlap is established depends on the shape of the DOS. These states are found at energies in which the DOS tails into the gap and are thought to originate from defects or structural imperfections.\cite{60,64,65,66} Right at the interfaces with metal electrodes, a hybridization of metal and semiconductor states tends to widen the DOS of organic semiconductors tends to widen even more. The formally associated built-in voltage $V_{\text{bi}}$ is the space-charge-induced difference in the energy of a reference state at the interface and far away from the interface.Expressed using the LUMO as reference, the built-in voltage reads as $qV_{\text{bi}} = E_{\text{LUMO}}^0 - E_{\text{LUMO}}$. As, however, the concentration of space charges is usually very low, the space charge region adopts an enormous width, that is, at a very large distance from the interface the energy levels become eventually position-independent. Right at the interface, with a close-up shown in Figure 6b, the “bending” of each level is practically imperceptible. Hence, the nominal injection barrier that electrons or holes need to overcome is typically not affected by the space charge, except for cases with a marked space charge, for example due to chemical reactions in the contact region.\cite{72} For an idealized contact between a non-interacting metal and OSC, the injection barrier height $\Delta_E$ for electrons corresponds to the offset $E_{\text{LUMO}}^0 - \mu_E$ between the LUMO level and the Fermi level of the electrode (Figure 6b). In reality, charge and geometry rearrangements between metal and OSC confined to the interface further shift OSC levels with respect to $\mu_E$.\cite{51,73} Interfacial charge transfer causes a dipole layer (bond dipole) that, in turn, forms a step in electrostatic energy across the OSC layer. This step BD corrects the barrier height $\Delta_E^\text{BD}$ for electrons to $E_{\text{LUMO}}^0 - \mu_E + \Delta_E^\text{BD}$. Analogously, a hole injection barrier $\Delta_p$ is defined as $\Delta_p = \mu_h - E_{\text{HOMO}}^0 + \text{BD}$. Figure 6c shows the situation at the interface region, once an additional electric field is superimposed, for example, due to reverse biasing the interface. For the sake of simplicity, our illustration assumes a uniform electric field. A charge carrier experiences a modified barrier “landscape” upon entering the organic semiconductor from the electrode. The practically rectangular shaped barrier known from thermal equilibrium is superimposed with two further contributions. First one needs to consider the potential energy due to the electric field, $qFr$, where $r$ is the distance of the charge carrier from the contact interface. The second contribution is the potential energy arising from the Coulomb attraction between the injected charge and the polarization in the electrode caused by this charge “hovering” at a distance $r$ outside the electrode. This polarization is accounted for by placing an image charge of opposite sign at the opposite side of the interface at the same separation $r$.\cite{74}

$$U_{\text{BD}}(r) = -\frac{q^2}{16\pi\varepsilon_0\varepsilon_r r}$$

Therein, $q$ is the charge of the carrier, $\varepsilon_0$ the dielectric permittivity, and $\varepsilon_r$ the relative permittivity of the semiconductor. The resulting barrier for the injected charge carrier adopts the shape

$$U(r) = \Delta - \frac{q^2}{16\pi\varepsilon_0\varepsilon_r r} - qFr$$

This barrier shape poses an effective height $\Delta^\text{e}$ that is smaller than $\Delta$. The larger the superimposed field strength, the more pronounced is the lowering of the effective barrier height.

It is possible to alter the Schottky barrier by inserting between semiconductor and electrode ultrathin molecular layers in which uniformly oriented, tightly packed molecular dipoles are arranged in a plane parallel to the contact surface.\cite{53} These interface dipole layers cause an additional offset to $\Delta$ that such injection appears to originate from an effective electrode with a different Fermi level position.\cite{73}

Based on this conception of a Schottky barrier, our device simulation methods need to address two foreseeable challenges: First, one has to address fact that the states in the semiconductor may appreciably spread in energy due to disorder.\cite{74} The most direct route to do so will be provided by kinetic Monte Carlo.\cite{76} Second, we will not necessarily encounter homogeneous fields at the interface. Moreover, presently injected charges modify local fields. Hence, changes in the local fields need to be considered in a self-consistent manner.

### 2.3.2. Organic Heterojunctions

The current understanding of mechanisms that govern the level alignment at interfaces between organic semiconductors is compiled in several reviews, for example in refs.\cite{52,73,77,78}. Device simulations require two sets of information: The extent of the layer offsets $\Delta_E$ and $\Delta_p$ and the possible presence of a built-in field due to the joining of layers. The latter gives rise to a contribution to the electrostatic potential and will, hence, interfere with the external field and local fields due to charges elsewhere in the device.

A reliable determination of $\Delta_E$ and $\Delta_p$ is by no means a trivial task; estimates solely relying on level positions in the individual materials, for example, estimates based on the ionization potential, optical gap, or redox potentials from cyclometrety, might be misleading. Rather, the alignment may sensibly depend on the order of layer deposition and on possible...
chemical and structural rearrangements at interfaces. Ideally, the relative positions of the levels in non-doped organic layers are governed by vacuum level alignment. The position of the vacuum level \( E_{\text{vac}} \) (indicated in Figure 5) in relation to the Fermi level determines the workfunction and, for the case of the semiconductor, in relation to the highest occupied levels (\( E_{\text{HOMO}} \)), the ionization potential. Upon getting in contact, the vacuum levels in each organic layer align and the relative positions of the levels with respect to the vacuum levels are kept. Bound charges, that possibly rearrange upon joining layers, may give rise to interface dipole layers. The potential energy associated to a dipole layer features an abrupt step perpendicular to the interface; superimposed with energy landscape a step in the vacuum level arises. Whether the step pushes the vacuum level up or down in energy depends on the direction of the dipoles. In particularly involved scenarios even built-in fields between the substrate electrode and a top layer are observed.\[52\]

### 3. Mesoscopic Modeling Approaches and Their Foundation

We established above that charge transfer events between sites tend to be predominantly viewed as transitions between localized states. Correspondingly, there are a number of methods that predict how likely and how often these localized sites are occupied for known intersite couplings, a given external field, or a given temperature. Remarkably, only methods considering hopping transport have been deliberately extended to account for device elements other than the organic semiconducting layer.\[36,82\] These methods, namely the kinetic Monte Carlo (kMC) method and the Pauli Master equation, do not regard the full atomistic composition of the sites, but rather focus on the effective coupling between sites. As typical site separations in organic semiconductors set the considered length scale from one to a few nanometers, both methods are considered mesoscopic approaches. These two methods offer a direct and highly illustrative approach to describe the motion of an ensemble of charges as a function of time.

Below, we will dedicate an extended section to introduce the kinetic Monte Carlo method. This is, because the kinetic Monte Carlo method shares a practically deterministic input with the Master equation approach; often both methods are similarly suited for many scientific questions. In addition, kinetic Monte Carlo is the only non-deterministic simulation method that will be discussed in this article. Hence, it is worthwhile to point out the aspects that set this statistical approach apart from all other simulation methods along with assumptions and details that are well scattered in literature. Note that particularly technical details will be offered in dedicated sections, marked with a *, that can be skipped without losing the overarching context.

#### 3.1. Kinetic Monte Carlo

The possibly most overwhelming demonstration of the capability of the kMC approach is the successful prediction the current voltage characteristics of a white organic light emitting diode, considering the motion of electrons, holes, and excitons across multiple layers.\[80\] Though the level of the kMC implementation required for such a prediction already demands substantial computational effort, the very heart of the kMC approach remains intuitive.

The kinetic Monte Carlo approach predicts the progression of a system based on the systems current state and a set of rules. A major advantage of kMC over other Monte Carlo methods is that it is possible to associate the progression of events with a physically meaningful time. The use of kMC to describe charge transport is justified if the motion of the charges in the device can be safely viewed as a sequence of completed individual events, for example, hopping and charge-recombination events. The following overview follows the line of argumentation put forward in ref. [81].

##### 3.1.1. Aim of the Method and Basic Considerations

Bearing in mind the discussion of hopping rates between pairs of sites in Section 2.2, it may appear intuitive to use these rates to propel forward each charge from site to site, that is, to literally trace the fate of each individual charge within the ensemble of charges. This tempting thought is neither useful nor practically feasible. Rather, kMC uses configurations, that is, snapshots of the system in which particles assume certain positions (at sites) and energies, so that we can clearly define an event that propagates the system from one configuration to another. This is particularly advantageous for systems that possess multiple charges that interact with each other. Moreover, the likelihood of an event can be determined without the need to know how charges arrived at their positions prior to a hop. The latter is a prerequisite of kMC, since kMC is a particular form of Markov chain Monte Carlo. kMC explores the time evolution of the system by deciding on the events to happen and performing event after event, traversing through a large number of configurations this way. In doing so, kMC simulations provide a probability \( \pi_x \) to find the system in a given configuration \( x \), that is, a probability \( \pi_x(t) \) for each configuration and for the current simulation time \( t \). All quantities that are of interest will be determined based on the associated time-dependent probability distributions.

The ultimate goal of kMC simulations is either to sample the probability distribution \( \pi_x \) that describes the steady state or to model the temporal evolution of a non-equilibrium system. In the latter case, the simulation starts out from a state that, once left, can never be returned to. Therefore, the simulation rather seeks the sequence of retention times, that is, the configurations that system prefers to visit and the times the system stays in these configurations.\[16,82\]

To establish a kMC simulation, we need to define the configurations, the transitions between them, and the construction of a Markov Chain. The next section contains these definitions and introduces the associated terminology; this section can be skipped by readers without enhanced interest in mathematical formulations.

##### 3.1.2. From Transition Rates to Markov Chain Monte Carlo*

**Configurations and Transitions:** A first and crucial step defines which configurations \( x \) the system can adopt. It is important
that the system possesses only a finite (or at least countable infinite) set of configurations. As seen in Section 2.2, possible transitions between different configurations occur if one charge carrier hops from one site to another (see Figure 4c) and are associated to rates \( w_{x,x'} \). The probability that a particular transition will happen in an infinitesimal time interval \( dt \) is the rate \( w \) multiplied by \( dt \). The total energies \( E_x \) and \( E_{x'} \) of the two configurations involved in the transition critically influence the rate. In the context of continuous-time Markov chains, one commonly collects the rates \( w \) of all allowed transitions between all possible pairs of configurations into a so-called q-matrix. For an allowed transition from a configuration \( x \) to a configuration \( x' \) with associated rate \( w_{x,x'} \), the corresponding matrix entry in the q-matrix would be \( q_{x,x'} = w_{x,x'}. \) The diagonal entries of the q-matrix for our case of kMC simulations are \( q_{x,x'} = - \sum_{x'' \neq x} w_{x,x''} \), that is, is the negative value of the sum of the rates of all transitions going out of configuration \( x \).

Calculation of Observables: In principle, all measurable quantities including the current and all properties of the system are determined by the q-matrix. However, the q-matrix alone is of limited use, as the actual calculation of these quantities is based on the corresponding probability distribution \( \pi_x \). When seeking for steady state quantities, this probability distribution \( \pi_x \) gives the probability that a configuration \( x \) is visited in steady state.\(^{[81]} \) Once the probability distribution \( \pi_x \) is known (existence and uniqueness assumed), the expectation value \( \left< O \right> \) of any quantity \( O \) which takes a value \( O_x \) in configuration \( x \), can be straightforwardly calculated with

\[
\left< O \right> = \sum_{\text{all config } x} O_x \pi_x \tag{9}
\]

The key challenge in using Equation (9) is that the probability \( \pi_x \) for all configurations needs to be known; the calculation of expectation values via Equation (9) becomes increasingly inefficient as the number of configurations becomes large. If we consider a system of \( N_{\text{all}} = 1000 \) sites and \( N_{\text{el}} = 20 \) charge carriers, that is, a size which is still too small to capture a realistic 3D charge transport description, we end up with \( \frac{N_{\text{all}}!}{N_{\text{el}}!(N_{\text{all}} - N_{\text{el}})!} \) configurations, that is, more than \( 10^{44} \) possible configurations.

Markov Chain Monte Carlo: Most of those configurations have a (nearly) vanishing steady state probability \( \pi_x \). Markov Chain Monte Carlo is a powerful tool to simulate steady state probability distributions in a space consisting of a huge number of configurations, in which only a small fraction of the configurations \( x \) is dominantly populated. Kinetic Monte Carlo is a particular form of continuous-time Markov Chain Monte Carlo, i.e., the time evolves continuously as transitions are performed. The transition probability \( P(x_t \rightarrow x_{t+\Delta t}) \), that the system changes from configuration \( x_t \) to configuration \( x_{t+\Delta t} \) within the infinitesimal time interval \( \Delta t \), is determined by a transition probability density \( p(x_{t+\Delta t} | x_t) = q_{x_{t},x_{t+\Delta t}} \) via \( P(x_t \rightarrow x_{t+\Delta t}) = \int p(x_t \rightarrow x_{t+\Delta t}) dx_t dx_{t+\Delta t}. \) One starts out from our known transition probability density \( q_{x_{t},x_{t+\Delta t}} \) to seek the underlying, presently unknown steady state probability distribution \( \pi_x \). During a simulation, a so-called Markov chain is created, that is, a randomly chosen sequence of configurations \( \{ x_t \} \) for a Markov time \( t \in [t_0, t_1] \) within a certain time interval between \( t_0 \) and \( t_1 \). An important property of such Markov chains is that they forget their past and only the currently visited configuration influences the future time evolution of the Markov chains. A configuration \( x_{t+\Delta t} \), that is visited at a time \( t_{\Delta t} \) after \( t_0 \), is solely determined by the configuration \( x_t \), which was visited latest in time, rather than by the complete sequence of previously visited configurations \( \{ x_t \} \) for all times \( t \leq t_0 \).

Construction of a Continuous-Time Markov Chain: Let us suppose that rate expressions and the associated q-matrix of our system leads to steady state probability distribution \( \pi_x \) that is guaranteed to exist and to be unique. Then, one needs to find the sequence of transition events leading toward steady state, that is, lay rules on how to construct the Markov chain. In essence, this sequence can be obtained with one of two mathematically equivalent approaches.\(^{[81]} \) In a first approach,\(^{[84,85]} \) the rates \( q_{x_{t},x'_{t}} \) from a starting configuration \( x \) to all allowed configurations \( x' \) are calculated and the total rate \( \dot q \) determined from the sum of all those rates \( \dot q = \sum q_{x_{t},x'_{t}}. \) Each transition from configuration \( x \) to configuration \( x' \) provides a transition probability \( P(x' | x) = \frac{\dot q_{x_{t},x'_{t}}}{\dot q} \), corresponding to its rate. According to those probabilities, a transition is randomly chosen and performed, so that the system adopts the next configuration. In addition, we need to determine the retention time \( \Delta t \), that is, the time during which the system remains in the present configuration \( x \) before the next hop. Since \( \Delta t \) is exponentially distributed according to the total rate \( \dot q \), a second uniformly distributed random number \( \xi \in (0, 1) \) is needed to calculate \( \Delta t = - \frac{\ln \xi}{\dot q}. \) A second, alternative approach assigns each transition to a retention time \( \Delta t_{x_{t}} = - \frac{\ln \xi}{\dot q_{x_{t},x'_{t}}}. \) Then, the fastest of all these retention times is chosen and the corresponding transition is performed after this fastest retention time elapsed.\(^{[86–88]} \) This requires as many random numbers as allowed transitions are needed. The latter approach is only feasible if most of the rates remain constant, that is, for a transition \( x_0 \rightarrow x_1 \) the condition \( q_{x_0,x_1} = q_{x_1,x_0} \) is fulfilled for most of the configurations \( x'. \) Only then the previously calculated retention times can be reused by setting the new retention time to \( \Delta t_{x_{t}} = \Delta t_{x_{t-1}} - \Delta t_{x_{t-2}} \) with \( \Delta t_{x_{t-1}} \) being the retention time of the performed transition \( x_0 \rightarrow x_1 \). In general, the first approach is preferable due to its superior numerical stability. Nevertheless, in some special cases, the second alternative is beneficial.

### 3.1.3. Simulation Setup

A system ready for performing Kinetic Monte Carlo simulations is defined by i) the set of configurations formed by placing the charge carriers at the spatially localized sites and, ii) the set of hopping rates from site to site that are collected in the q-matrix (cf. Section 3.1.2).\(^{[46,80]} \) This section seeks to collect all necessary steps to setup a simulation.

Formulation of Configurations and Transitions: In principle, the site positions can be distributed within the system volume without constraints.\(^{[50]} \) However, commonly a cubic lattice of sites with a lattice constant \( a \) in the range of \( a = 1 \) nm is employed.\(^{[75,86–88,90–94]} \) The total simulation volume is a lattice determined by the number of sites considered into the three spatial
directions \(N_x \times N_y \times N_z\) with \(N_x\) being the number of sites into spatial direction \(r_i\). Each site is labeled \((i, j, k)\), that places it in the \(i\)th layer in \(r_1\) direction, the \(j\)th layer in \(r_2\) direction, and the \(k\)th layer in \(r_3\) direction. In case that our system contains more than one charge carrier, it is important to stress that each site can be occupied by, at maximum, one charge carrier. Note that this rule readily excludes bipolarons and prevents strong on-site Coulomb interactions. Depending on the purpose of the simulations, appropriate conditions for the volume boundaries must be selected. For the simplest case of a homogeneous bulk system, periodic boundary conditions are an appropriate choice. At any given time, only one charge is allowed to hop from the currently occupied site to another, distinct site with a pre-assigned rate. In practice, it is numerically convenient to limit the hopping to target sites in the vicinity of the starting site, that is, to choose the site within a hopping radius of the initial site.

An important quantity determining the hopping rates is the energy difference \(\Delta E_{x \rightarrow x'}\) between a configuration \(x\) and a configuration \(x'\). This energy difference consists of three parts:

\[
\Delta E_{x \rightarrow x'} = \Delta E_{\text{site}} + \Delta E_{\text{field}} + \Delta E_{\text{Clb}}.
\]

(10)

that is, i) the site energy difference \(\Delta E_{\text{site}}^x = E_{\text{site}}^x - E_{\text{site}}^{x'}\) with the energy \(E_{\text{site}}\) that the charge carriers assume by occupying specific sites in the corresponding configuration, ii) the energy difference \(\Delta E_{\text{field}}\) due to the external electric field \(F\), and iii) the difference in Coulomb interaction energy \(\Delta E_{\text{Clb}}\) between the charge carriers. For an electric field with pointing in \(z\)-direction (along the \(r_3\) component of \(\vec{r}\)), the total energy difference reads:

\[
\Delta E_{x \rightarrow x'} = E_s - E_s' + qF(r_i' - r_i)
\]

\[
+ \frac{e^2}{4\pi \varepsilon_0 \varepsilon_r} \sum_{j=1}^{N_j} \frac{1}{|\vec{r}_i' - \vec{r}_j|} \frac{1}{|\vec{r}_i - \vec{r}_j|}.
\]

(11)

with the vacuum permittivity \(\varepsilon_0\) and the relative permittivity \(\varepsilon_r\) of the organic semiconductor. Therein, \(E_s\) and \(E_s'\) refer to the static energies of sites \(s_i\) and \(s_i'\). Note that only the last, that is, the interaction term, contains the entire configuration of the charge carriers \(x \equiv (s_1, s_2, \ldots, s_N)\). When neglecting the Coulomb interaction, we would need to account solely for the two sites involved in the hop.

In the section below, interested readers may find choices, constraints, and approaches that are required to formulate the total energy difference (Equation (10)) and give rise to Equation (11). Amongst others, the section will clarify how to choose the energy landscape and how Coulomb-interactions are determined in practice.

**Formulation of the Energy Contributions:**

Starting out from Equation (10), we now explain i) the site energy difference \(\Delta E_{\text{site}} = E_{\text{site}}^x - E_{\text{site}}^{x'}\), ii) the energy difference \(\Delta E_{\text{field}}\) due to the external electric field \(F\), and iii) the difference in Coulomb interaction energy \(\Delta E_{\text{Clb}}\) between the charge carriers.

**Site Energy:** The energy of a charge carrier occupying a certain site (without field and interactions being present) is randomly chosen from an underlying density of states prior to the transport simulation. Conceptually, there are no constraints for the shape of density of states (DOS) as long as the DOS contains all potentially accessible states. In the chosen DOS, it is not necessary yet possible to distinguish between nominal transport and trap states. Of course, the DOS may spatially vary, either across an interface of distinct materials (heterojunction) or due morphological reasons (e.g., due a change in molecular packing). Commonly used densities of states reflect static disorder and follow, thus, either a Gaussian distribution, an exponential distribution, or combinations of the two (e.g., for modeling dopants or traps); likewise, the density of states can be also directly suggested from experiments.

For a Gaussian density of states, we draw the static site energy \(E_s\) of site \(s\) from the probability distribution

\[
p(E_s) = \frac{1}{\sqrt{2 \pi \sigma}} e^{-\frac{E_s^2}{2\sigma^2}}
\]

(12)

with the energetic disorder \(\sigma\) determining the width of the Gaussian distribution. It is also possible to impose a spatial correlation between site energy as described in refs. [92,93], in particular if the molecules or sites possess a marked dipole moment.[94] Collective responses of the molecules on local or externally applied electric fields translate into further modifications of the energy landscape of charge carriers.[95] Such responses cause marked shifts between adjacent energy levels, in particular when bulk regions are disordered or “disturbed” by grain boundaries,[96–98] interfaces between organic semiconductors and metals,[99–102] or doping.[103,104] Microelectrostatic models as described, for example, by D’Avino et al.[95] provide such corrections for transport simulations.[96]

With all sites being assigned to a randomly chosen site energy, the site energy of the configuration is the sum of the site energies of all occupied sites:

\[
E_{\text{site}} = \sum_{i=1}^{N_j} E_i.
\]

(13)

Then, the site energy difference between two configurations \(x\) and \(x'\) is

\[
\Delta E_{x \rightarrow x'}^\text{site} = E_{\text{site}}^x - E_{\text{site}}^{x'} = E_s - E_s'.
\]

(14)

The contributions from all other charge carriers that are not hopping during the transition from configuration \(x\) to \(x'\) cancel out.

**Field Energy:** An externally imposed electric field \(\vec{F}\), i.e., a field not being caused by other mobile charges in the material but rather due to an applied voltage or by static charges inside the material, is entering the simulation solely via a modification of the transition rates. A charge carrier moving from position \(\vec{r} = (r_1, r_2, r_3)\) to position \(\vec{r}'\) changes its energy whose amount is given by the curve integral

\[
\Delta E_{x \rightarrow x'}^\text{field} = \int_\sigma d^2 \vec{r}' E_{\text{field}}(\vec{r})
\]

(15)

Therein, \(q\) is the charge of the charge carrier and \(E_{\text{field}}(\vec{r})\) the electric field at position \(\vec{r}\). For a constant electric field strength \(\vec{F} = (0,0,F)\) pointing in the \(r_3\) direction, the latter integration leads to

\[
\Delta E_{x \rightarrow x'}^\text{field} = qF(r_i' - r_i).
\]

(16)
when charge carrier \(i\) hops from site \(s_i\) to site \(s'_i\). The spatial position of site \(s\) with respect to the third spatial coordinate is here indicated by \(r_i\).

At this point it is important to note, that the presence of external fields formally prevents the applicability of periodic boundary conditions. In our case that means that when a charge carrier is hopping from layer \(N_s\) to layer \(1\) in positive \(r_3\) direction across the periodic boundary, the field energy difference associated with this hop must be \(qF_a\), corresponding to a hop one lattice constant into field direction, rather than \(-qF_a(N_s-1)\), that is, as if a hop \((N_s-1)\) lattice constants against the field direction would have been performed.

**Coulomb Interactions:** One of key assets of kinetic Monte Carlo is the opportunity to directly consider interactions between charges. The Coulomb repulsion gives rise to a correlated motion of charges. A charge carrier placed at site \(s_i\) feels a Coulomb repulsion of another mobile charge carrier of the same polarity at site \(s_j\). This repulsion can be described by the Coulomb interaction energy

\[
E_{\text{Ch}}^{\text{Clb}}(s_1, s_2) = \frac{q^2}{4\pi \varepsilon_0 \varepsilon_r |r_i - r_j|} \quad (17)
\]

The energy required to place an additional charge carrier at site \(s\) in a given configuration \(x=(s_1, s_2, \ldots, s_{N_c})\) of \(N_c\) charge carriers is the sum of all individual interactions:

\[
E_{\text{Ch}}(s, x) = \frac{q^2}{4\pi \varepsilon_0 \varepsilon_r} \sum_{j=1}^{N_c} \frac{1}{|r_i - r_j|} \quad (18)
\]

Then, the difference in Coulomb interaction energy \(\Delta E_{x,x'}^{\text{Clb}}\) between configuration \(x\) and \(x'\) is the same as the energy needed to remove one charge carrier from site \(s_i\) and add afterward one charge carrier at site \(s'_i\) considering the positions of all other charge carriers:

\[
\Delta E_{x,x'}^{\text{Clb}} = \frac{q^2}{4\pi \varepsilon_0 \varepsilon_r} \sum_{j=1, j \neq i}^{N_c} \left( \frac{1}{|r_i - r_j|} - \frac{1}{|r_i - r'_j|} \right) \quad (19)
\]

**Determination of Coulomb Interactions:** The evaluation of this relation (19) requires significant computational effort.\(^{[305]}\) The required summation in Equation (19) runs over all other mobile charges and their periodic replica if periodic boundary conditions in at least one direction are assumed. For doped systems, the summation must further incorporate the dopant ions that remain as static point charges.\(^{[106,107]}\) For problems involving electrodes (cf. Section 3.1.4), the sum further comprises image charges (and their periodic replica), because image charges allow us to consider the attraction of the charge carriers to the polarized electrodes. For low charge carrier densities, it is advisable to calculate Equation (19) on-the-fly. For high charge carrier densities, an alternative option is to precalculate the monopole potential energy for all charges for all lattice sites before enrolling the chain of hopping events. This has the benefit, that interaction energies, that are required for a charge at a certain site during the simulation, can than conveniently obtained by subtracting the monopole potential energy from the overall stored potential energy. After a hop, this potential energy has to be updated by removing the monopole contribution from the emptied site and adding the monopole contribution at the now occupied site.

Correspondingly, it is desirable to calculate Coulomb interactions efficiently at reasonable accuracy, keeping in mind that particularly short-range Coulomb interactions should be accurately determined.\(^{[106]}\) One possible strategy to save computations aims at truncating the considered interaction distance. Initial kMC implementations consider interactions only within a cut-off radius; there either the Coulomb capture radius\(^{[86]}\) or the Debye length\(^{[109,110]}\) served as cut-offs. In more accurate manner, all additional interactions between charges separated more than the cut-off radius were considered as layer-averaged quantities.\(^{[111,112]}\) If all interactions ought to be considered, i.e., the interaction distance is not truncated, it is helpful to turn to the aforementioned determination of monopole potential energies that is described, for example, in refs. [81,105]. The Ewald method offers an efficient way to determine the required monopole potential energies of each charge carrier depending on its separation from the considered position.\(^{[113,114]}\) Moreover, the Ewald method incorporates also long-range Coulomb interactions, that have been shown to be as relevant as molecular architecture, orientation, and packing.\(^{[115]}\) A truly feasible method to compute long-range interactions became recently available; this fast algorithm evaluates the electrostatic energy based on a multipole expansion.\(^{[116]}\) The algorithm dramatically lowers the computational effort for each KMC step, as it scales linearly in the number of charges.

### 3.1.4. Injection, Generation, and Recombination

Charge Generation and Recombination: The framework of kMC allows to seamlessly incorporate charge generation or recombination events. The latter is crucial to understand the impact of nanostructured heterojunctions on loss mechanisms in organic solar cells\(^{[107]}\) or the efficiency of radiative recombination taking place in markedly disordered layers in OLEDs.\(^{[80,118]}\)

Let us turn to recombination first. Alike hop events, also recombination events are associated to a rate expression. This rate expression may correspond to constants or to model rate expressions. For example, Miller–Abraham rates (Equation (1)) are a convenient choice.\(^{[118–120]}\) As an recombination event is expected to lower the total energy,\(^{[118]}\) the recombination rate in Equation (1) is solely determined by the electronic coupling and the attempt frequency, whose value is adjusted.\(^{[119]}\)

Two circumstances lead to a recombination event: i) The time associated to a recombination event is smaller or comparable to times necessary to perform an alternative hopping event. ii) The charge must be located in the vicinity of an eligible recombination partner, be that a static site of certain properties (e.g., a dopant) or a site currently occupied by a counter charge.\(^{[46,118,119,121,122]}\) For the latter it is particularly desirable to implement the motion of electrons and holes.

Note that implementations, that do not regard other means to resplendish charges, keep electrons and holes lost to recombination in the simulation by reintroducing these charges at other positions.\(^{[118,122]}\)

An illustration for transient simulations studying the impact of disorder on bimolecular recombination is provided...
in Figure 7. Shown is the decay of the charge carrier density $\rho$ as a function of time for four different bulk heterojunctions that differ in the width of an exponential distribution of states. These heterojunctions can be viewed as a blend of small domains so that charges do not have to travel far before they find a domain interface, and, hence, a counter charge. Thus, it is the width of the DOS in such a system that determines the long-time recombination kinetics and the initial delay time after which charges start to recombine.

The generation of electron–hole pairs is also incorporated as discrete events. Implementations in the context of solar cells handle generation as a result of an exciton dissociation event. In such a case also the dynamics of excitons must be considered, that is, their photogeneration, transport, relaxation, and dissociation into a weakly bound or free pair of electron and holes. KMC permits to monitor the motion of excitons in the same spirit as done for charges, as it is possible to describe exciton migration as a sequence of transfers between distinct sites with an adequately chosen rate expression. For methodological details specific for exciton motion, we refer interested readers to, for example, refs. 1,109,123. In such a case also the dynamics of excitons must be considered, that is, their photogeneration, transport, relaxation, and dissociation into a weakly bound or free pair of electron and holes. KMC permits to monitor the motion of excitons in the same spirit as done for charges, as it is possible to describe exciton migration as a sequence of transfers between distinct sites with an adequately chosen rate expression. For methodological details specific for exciton motion, we refer interested readers to, for example, refs. 1,109,123. Excitons dissociate into a free pair of electrons and holes when the excitons arrive at a location that favors exciton dissociation, in most cases a heterojunction. In this way, any charge generation event is coupled to the dynamics of the exciton arrival at heterojunctions and occurs independent from the events associated to hops or recombination of charges. Note that the definition of the generation and recombination rates do not explicitly require the band gap, that is, the separation of electron and hole transport states.

Charge Transfer to Electrodes: A simulation of charge transport through a device requires the consideration of an exchange of charges with electrodes. In the above-described framework, the process of injection can be viewed as a thermally assisted hop that moves the charge carrier from an electrode site to an adjacent site in the OSC. Charges are extracted from the device via hops from the semiconductor into the metal. Alike the semiconducting material, also the conducting electrode is perceived as an arrangement of sites, in which each site is ready to accept or provide a charge; possible hops between electrode sites are not considered. In practice, each electrodes is inserted as an additional layer of sites that is placed at the appropriate surface of the simulated semiconductor volume. In the example of Figure 1a, the two layers that sandwich the semiconductors form the cathode and anode of the organic diode in the simulation.

Hops between electrode and OSC sites are described with a hopping rate $w$, whose form usually follows the expression chosen to describe hops in the semiconductor. Eventhough each electrode site enters $w$ with the same site energy $\mu_F$, the hopping rates may profoundly vary across the contact due to distributed site energies and local changes in the electrostatic potential. Due to the causes of these variations, each hopping rate has to be reassessed after the completion of a hop.

The impact of such spatially varying hopping rates across the contact was investigated by Zacher and Armstrong. After assigning various effective activation energies for charge hopping on quadratically arranged contact sites in a solar cell, a fraction of sites was found not to participate in charge extraction. As illustrated by selected trajectories in Figure 8, the charges are not necessarily extracted when they hit the contact the first time. If the charge hits a contact site with an unfavorably low rate, the charge rather hops further in the OSC near the contact until it finds a site with a favorable extraction rate (red trajectory).

Another electrode-related challenge is that each charge induces an image charge in each electrode (cf. Section 2.3), regardless whether mobile charge carriers or static charges due to doping are concerned. As these image charges attractively interact will all charges in the system, electrodes can be expected to profoundly interfere with remote charge transport in the bulk material. The consideration of image charges increases the effort described in Section 3.1.3 to evaluate local electric fields.
3.1.5. Performing Transitions

Performing a simulation step in kMC comprises two stages:
The selection of a hop and the updates of transition rates. As
a hop is completed, the time is advanced by the corresponding
retention time of the hop.

For the system in a given configuration, the target configura-
tion and the retention time of the next hop is selected using
random numbers.

Determination of Hopping Event and Retention Time:
As briefly mentioned in Section 3.1.2, this can be done in two ways.
Be \( N_{\text{hop}} \) the number of allowed hops, \( w_i \) the calculated rate of an
allowed hop, and \( w = \sum w_i \) sum of all those rates \( w = \sum w_i \).
In a first approach, a uniformly distributed random number
\( \xi \) drawn from the interval \((0,1)\) (0 excluded and 1 included)
determines the next configuration: We choose the configura-
tion to hop to whose rate \( w \) fulfills the ordering condition
\[
\sum_{i=1}^{N_{\text{hop}}} w_i < \xi w \leq \sum_{i=1}^{N_{\text{hop}}} w_i
\]
with respect to the cumulative sum over the hopping rates. A second uniformly distributed random number \( \xi \) in the interval \((0,1)\)
(excluding both 0 and 1) determines the retention time \( \Delta t \) for the transition. In a second way, the retention time \( \Delta t \) is directly assigned to each individual
allowed transition. One uniformly distributed random number
\( \xi \) in the interval \((0,1)\) (excluding 0 and 1) is needed for each
transition, so that the retention times are \( \Delta t = -\frac{\ln \xi}{R_i} \). The trans-
mission with the shortest retention time is performed. As random
numbers should be used and, hence, generated as rarely as pos-
sible, the latter scheme is not recommended except for situa-
tions in which the first reaction method can be used as update
mechanism.\[82\]

Once a hop is completed, the rates of i) the just dis-
placed charge carrier to its nearby sites and ii) the rates of all
charge carriers that did not hop need to be calculated anew.
Displacing a charge carrier with a hop alters the Coulomb in-
teraction energy of all sites. There are several strategies to assess
energy changes due Coulomb interactions. Of course, it is most
accurate to recalculate the energies and related rates for pos-
sible hops of all charge carriers to their nearby sites. Amongst
these new rates, the rate associated to the next hop is selected
with the selection procedure that needs only two random num-
bers. For systems with very low charge carrier densities, that
is, in systems in which the Coulomb interaction is nearly neg-
ligible,\[88\] the first reaction method offers an alternative update
mechanism.\[82\] Therein, the Coulomb interaction considered in an approxi-
mate way. If Coulomb interaction would be neglected completely,
only the rates of the just displaced charge carrier need to be recalculated, because the rates of all other charge
carriers did not change during the previous hop. With recal-
culated rates of the previously hopped charge carrier at hand,
a transition is randomly chosen from these few rates and the
retention time picked according to the selection scheme
relying on two random numbers. Then, also all former reten-
tion times of the charge carriers, that have not been moved,
are updated by subtracting the retention time of the just per-
formed transition. That bears the convenience, that per charge
carrier only the chosen hop and the corresponding retention
time is stored. In comparison to the accurate update mecha-
nism, the first reaction method provides a system update at
strongly reduced computational costs, because the most time
consuming part within performing a transition is the evalua-
tion of the Coulomb interaction. However, the incomplete
consideration of the Coulomb interaction leads to erroneous
and untrustworthy values for high charge carrier densities.
And even though the first reaction method is an adequate
choice for low charge carrier densities, it might lead to wrong
results and may complicate the interpretation of the ongoing
effects. In attempt to harvest the benefits of the first reac-
tion method and the accurate update mechanism, one might
limit the number of charge carriers to be updated: The hop-
ping rate is recalculated only for charge carriers that are
encountered within a distance of an update-hopping radius
from the site, to which the currently hopped charge carrier
moved;\[127\] the other rates remain unchanged. On the basis of
the combined old and new rates, a transition and a retention
time can be selected by using the sum of all rates and two
random numbers.

Choice of the Starting Configuration: The routine to per-
form transitions discussed in the previous section opens
an important question: How a starting configuration has to be
chosen? This choice strongly depends on the purpose of the
simulation.

a) For time-dependent problems, the starting configuration is
must be well-defined and provided in advance.\[82\] For example,
electron and holes are randomly placed at sites selected
from the entire volume\[120\] or related to the donor:acceptor
interface.\[126\]

b) If one aims at describing steady state, the particular choice
of the starting configuration ought to be irrelevant. An appro-
priate starting configuration is to randomly place the desired
amount of charge carriers at sites. As this arbitrary pick of
the configuration introduces an unintentional bias, one needs
to thermalize the system before any observable can be deter-
mined. This implies that the system has to reach steady state,
before steady state quantities can be extracted. This thermal-
ization is done by performing a certain number of hops until
the system becomes independent of the starting configuration.
As a rule of thumb, thermalization requires a number of hops
that corresponds to a tenth of the total number of \( N_{\text{main}} \) hops
needed after thermalization to determine steady-state quanti-
ties. This number \( N_{\text{main}} \) is not significantly increasing the com-
putational effort. Moreover, if thermalization is not successful
within \( N_{\text{main}} \) hops, the total simulation is correlated and the
measured values are meaningless anyway.

3.1.6. Determination of Observables and Uncertainties

As kMC is a statistical method, it is important to comment on
the determination of observables and the associated uncertain-
ties. In practice, this determination is a “measurement” within
the sequences of configurations amassed in a simulation. A
measurement ought to start at a time \( t_{\text{start}} \). For steady-state sim-
ulations, \( t_{\text{start}} \) must be a time after thermalization is achieved.
During the time span between \( t_{\text{start}} \) and the final \( t_{\text{end}} \), at which
the measurement ends, a certain number of \( N_{\text{main}} \) transitions
will be performed.
Static quantities such as local charge carrier densities only depend on the configurations that are visited and the corresponding time that the system stays in those configurations. It is not necessary to know any details regarding the transitions between those configurations. We exemplarily show the calculation for the charge carrier density $n_{cc}(V)$, which gives the average charge carrier density in a certain volume $V$ (e.g., a certain site). Once steady state is reached, the total time $\Delta t_{av}$ that the system spends in a certain configuration $x$ during the whole simulation, corresponds to the steady state probability distribution of this configuration $\pi_x = \frac{\Delta t_x}{\sum \Delta t_x}$ Combined with Equation 9, we can calculate the charge carrier density

$$n_{cc}(V) = \frac{1}{V} \sum_{i=1}^{N_{cc}} N(i, V) \frac{\Delta t_i}{t_{end} - t_{start}}$$  \hspace{1cm} (20)

The retention time of transition $i$ is given by $\Delta t_i$ and $N(i, V)$ is the number of charge carriers, that is occupying volume $V$ in the configuration prior to the $i^{th}$ transition. If we, for example, want to calculate the charge carrier density of a certain site, one has to sum all time spans during which this site is occupied; after the simulation finished, this occupation duration is divided by the volume that this site is related to and the total simulation time.

When calculating dynamic quantities, such as velocities, we can start with Equation (9) and formulate

$$\langle O \rangle = \sum_{i=1}^{N_{acc}} O_i \frac{\Delta t_i}{t_{end} - t_{start}}$$  \hspace{1cm} (21)

Instead of expensively calculating $O_i$ from the configuration visited prior to hop $i$, we can use an estimator for $O_i$ extracted from the transitions. This alternative is stochastically equivalent to evaluating Equation 9. For the case of a velocity, this estimator is simply $\langle v \rangle = \frac{\Delta r}{\Delta t}$ containing the hopping distance $\Delta r$ of the transition. Then, the average velocity of one charge carrier results from i) averaging this estimate of the velocity with respect to all hops performed and ii) dividing this average by the total number of charge carriers:

$$\langle v \rangle = \frac{1}{N_{cc}} \sum_{i=1}^{N_{acc}} \frac{\Delta r_i}{\Delta t_i} = \frac{1}{N_{cc}} \frac{\Delta t_i}{(t_{end} - t_{start})} \sum_{i=1}^{N_{acc}} \Delta r_i$$  \hspace{1cm} (22)

that is, we simply have to sum up the hopping distances of all charge carriers and, after the simulation, divide it by the elapsed time and the total number of charge carriers.

If we are interested in the mobility, we have to calculate the average velocity of one charge carrier in field direction $\langle v_{field} \rangle$ in the same way as $\langle v \rangle$ and get the mobility by $\mu = \frac{\langle v_{field} \rangle}{F}$.

To determine the uncertainties of measured quantities, essentially all methods established for Monte Carlo simulations can be used. For example, we can recommend jackknife method as an easy and reliable approach.[128] To estimate uncertainties, the measured data set is divided into $N_{bk}$ blocks of (approximately) equal size. Often 100 blocks are chosen, that is, if we perform one million hops, each of the blocks contains $10^4$ hops. To estimate the uncertainty of a desired observable $O$ from a set of data $D$, the latter observable is calculated $N_{bk} + 1$ times, that is, (i) once with the whole data set $D$ leading to the estimator of the observable $O^D$ and (ii) $N_{bk}$ times with the data set $D_i$ in which block $i$ is excluded from $D$, resulting in $O^{D_i}$. An estimator for the error $\Delta O$ is given by

$$\Delta O = \sqrt{\frac{N_{bk} - 1}{N_{bk}} \sum_{i=1}^{N_{bk}} (O^{D_i} - O^D)^2}$$  \hspace{1cm} (23)

with the average $O^D = \frac{1}{N_{bk}} \sum_{i=1}^{N_{bk}} O^{D_i}$ of the jackknife data sets $D_i$. To justify the application of Jackknife, the blocks must be shown to be uncorrelated; correlated blocks yield unreliable error estimators. Prior to applying jackknife one has to check whether the condition of uncorrelated blocks holds, for example, by measuring autocorrelation times (cf.[128]) or, at least, by inspecting the time series of the corresponding quantity.

For the calculation of transient properties, the simulation starting out from the same initial configuration has to be run multiple times.[82,126,129] Each trajectory is simulated until a certain number of steps passed or until the charge (or exciton) leaves the simulation, be that via escape through the contact or via recombination (dissociation for excitons). The trajectories are averaged to obtain a mean value for an observable value at a specific time.

### 3.1.7. Inherent Weaknesses

The kinetic Monte Carlo method is computationally demanding and calls for parallelization.[130] A possible reason that drives up costs is the slow if not even absent convergence in the case of very small fields. As in kMC it is not possible to refuse a hop, the anticipated small or negligible displacements are realized by an unduly lengthy sequence of hops back and forth. Similarly inefficiently invested hops occur when regions with deep traps need to be bypassed; to avoid such situations, a modified workflow has been suggested.[131] A last potential issue is that charges need to explore a region of a certain volume to reach steady state. The carrier displacement achieved in such volumes may exceed the layer thickness in realistic devices. In turn, that means that carriers traversing a layer did not necessarily reach the energetically most favorable pathways yet.

### 3.2. Master Equation

The Master equation approach as the second, less demanding mesoscopic approach lead to important stepping stones toward modeling device characteristics, be that for OLEDs[132–134] or OPV devices.[135] These examples already illustrate that also the Master equation is suited to describe charge as well as exciton transport. For a comprehensive discussion of the differences between the Master equation (ME) and the kinetic Monte Carlo approach, the reader is referred to Casalegno et al.[123]

It is also worthwhile to highlight here, that our understanding of the dependence of the mobility on our major tuning handles (temperature, field, charge carrier density) [12.136–139] and
the impact of injection on 3D transport[108] has been decisively shaped with simulations based on the Master equation. Such simulations have validated emerging analytical and semianalytical expressions for the mobility that used to analyze transport without the need of simulations. The most prominent analytical expressions were derived the Gaussian disorder model that branched into many different extensions.[113]

In general, ME is suitable to describe all particle types, whose motion can be viewed as a sequence of discrete hopping events, that is, electrons, holes, or excitons. The description of transport requires the definition of distinct sites and hopping rates between pairs of these sites. Correspondingly, ME is able to address quantities that may vary at length scales comparable to the extension of a hopping site. The definition of the spatial site arrangement and the chosen hopping rates controls the extent to which disorder in site energies and site interactions or a morphology, that may include distributed and sharp interfaces, are considered in the simulation. To date, isotropic rates such as the Miller–Abrahams or the Marcus type sites that operate in regular lattices are still prevalent.

What sets ME most apart from kMC is that discrete charges are replaced by a more continuous charge density. This density is represented by a local site occupation probability \( n_i \), that is allowed to adopt continuous values between zero and one, as indicated in Figure 1b). Then, hopping causes continuous changes in site occupation probabilities, rather than kMC-inherent discrete changes of the site occupations related to a transition between configurations.

Each particle type requires the formulation and solution of a dedicated Master equation. Exemplarily shown for electrons, the occupation probability \( n_i \) of a site \( i \) evolves according to the Pauli Master equation[124,123,140]

\[
\frac{dn_i}{dt} = \sum_{j, p=\pm} \left( w_{ij, p}(1-n_j) - \sum_j w_{ji, p}(1-n_j) \right) - R_i + G_i
\]

Therein \( w_{ij} \) denotes the hopping rate between a site \( i \) and a site \( j \). The factors \( (1-n_j) \) prevent hops to a site \( i \) if the site is occupied.[140] The terms \( R_i \) and \( G_i \) are rates that address events that annihilate (due to recombination) and generate charges at site \( i \), respectively. The steady-state solution requires \( \frac{dn_i}{dt} = 0 \) and results in a large system of non-linear equations that has to be solved self-consistently for a given number of particles in a system.[142] Solving this system of equations (24) simultaneously addresses all possible hops in the system and yields the values \( n_i \) for each site. Note that this solving is an order of magnitude faster than performing kMC[123] and essentially follows refs. [41,140], a procedure that is more detailed in ref. [123,141]. With these values \( n_i \) the steady-state current flowing through a device and a concomitant the charge carrier mobility can be calculated.

The rates \( w_{ij} \) either correspond to model expressions, for example, Equations (1, 2), or are extracted from atomistic modeling as we briefly discussed in Section 2. As seen in kMC before, such rates depend on the difference \( \Delta E_{ij} \) in total energy before (particle at site \( i \)) and after a hop (particle at site \( i \)). Contributions \( \Delta E_{ij}^w \) and \( \Delta E_{ij}^r \) are evaluated as in kMC (Equations (14) and (15)).

However, the remaining contribution \( \Delta E_f \), that accounts for the Coulomb interaction including the image charges in the electrodes,[108,123] is more difficult to include, because the individual charges are replaced with an averaged site-occupation probability (mean field approximation). This mean-field approximation affects the rates and hence the relative importance given to certain processes. It is prone to underestimate attractive[142,108,123] and repulsive forces[142] in regions where charges tend to accumulate. Charge accumulations may occur, for example, near heterojunctions[142] or when considering high charge density (i.e., approaching 1 charge per 100 sites).[142] A method to partially consider Coulomb interactions in a self-consistent manner has been described in Ref. [123].

Simulations based on the Master equation are initialized alike kinetic Monte Carlo simulations. Required are an energy landscape, the number of considered charges, and the hopping rates. Heterojunctions are introduced via a rigid offset of the energies at the interface. Charge carrier injection and abstraction at the contact is treated as in kMC.

4. Drift-Diffusion Modeling

4.1. Merits and Weaknesses of the Method

The drift-diffusion model (DDM) is the possibly most mature and wide-spread model approach in organic electronics. It offers several appealing features for device modeling.

A first appeal arises from the ability to consider the full complexity of device architectures and geometries, while being able to account for material-specific transport-relevant properties of the employed materials. Such complex geometries comprise, for example, a large number of diverse layers or electrodes that are arranged at different positions and orientations with respect to the electrically active layers. Correspondingly, the method holds the promise to faithfully describe inhomogeneous field distributions and to explore their impact on the device operation.[141–145] In contrast to the mesoscale model discussed in Section 3, it is also possible to combine the consideration of transport across very large distances, due to electrode spacings exceeding several micrometers for example, with a description of transport across very short distances, for example, across thin interlayers or accumulation at interfaces.

The DDM classically describes charge transport from a macroscopic, continuum-inspired perspective. It hinges to two central equations that ensure charge conservation (continuity equation) and quantify the driving forces (current density equation). A second asset is that the driving forces contain the material properties by means of effective parameters, such as charge carrier mobility and diffusion constant. This fact allows to address transport phenomena that are as different as hopping, transient localization-limited transport, or band transport (as most traditional application)—provided the effective parameters reflect the signature behavior if these transport mechanisms correctly.

The transport of electrons and holes are just two possible transport processes out of many others that can be conveniently described with a dedicated set of continuity and current density equations, for example for the charge exchange with trap states or for the transport of ions, heat, or excitons. So,
as a third appealing asset, drift–diffusion-based models can be rather easily extended to address other transport processes and their coupling to the motion of charges, at least at the formal level of formulating additional model equations. As the dissipation of heat and the motion of excitons are serious topics in the development of organic optoelectronic devices, drift–diffusion based solvers are clearly an excellent starting point for device models that aim at coupling charge transport to additional transport and propagation phenomena. The need to build up a sophisticated simulation framework in a step-by-step manner has an unintended yet enormously useful consequence. For discriminating the role of one quantity or mechanism, one just increases the complexity of a simulation accordingly or switches off a mechanism without necessarily influencing other processes. As a last appeal, DDMs offer a reasonable comprise between swiftness and accuracy in predicting device characteristics.

Despite their many virtues, drift-diffusion-based simulations are not universally suitable. There are a number of aspects that limit the use of a drift-diffusion model. A first important aspect is that the underlying model is only valid for devices, in which the steady state situation does not strongly differ from thermal equilibrium. For example, the DDM is not suited to describe and maintain strong local changes in the occupation of the DOS. A second reason for caution is that the filament-nature of transport pathways in disordered materials cannot be accounted for. Rather, DDMs assume effective material properties that emerge from a spatially averaging of local processes. Letting extrinsic modifications such as doping or biasing aside, a material is considered homogeneous so that for any position in space slowly changing, continuous quantities such as a carrier density and electrostatic potential can be defined. Hence, DDMs are not well suited to investigate nanostructured devices.[146] Though it is possible to impose nanostructured domains, DDMs can only qualitatively help to understand the formation of current pathways.[147]

A final concern relates to the commonly used shape of the DDM equations. The traditional model equations were conceived for band transport and will not fully account for disorder, whenever phenomena related to disorder cannot be mapped on effective material parameters. For instance, DDM does not acknowledge the actual occupation of a DOS, because the model originally assumes transport to occur exclusively near a band (energy) extremum. Below we will line out two possible routes to modify DDMs to account for the hopping nature of transport in a disordered semiconductor (cf. Sections 4.2.1 and 4.5)

### 4.2. Underlying Model Equations

In general, a continuity equation is the differential form of a conservation law. Each conserved physical quantity, such as particle number, charge, and energy satisfy a continuity equation. This equation links an appropriately chosen density, a flux associated with this density \( j \), and a source/sink term accounting for a spatially dependent, additional introduction or removal of this quantity.

In practice, a continuity equation is formulated for each of the mobile species to be considered. The continuity equation for electrons (with \( q \) the elementary charge) links the electron density \( n \) and the electrical current:

\[
\frac{\partial n(r,t)}{\partial t} = \frac{1}{q} \nabla \cdot \mathbf{j}_e(r,t) + G(r,t) - R(n,p,t)
\]

(25)

Therein, the net source term consists of \( G \), the generation rate, and \( R \) the recombination rate. For holes of the density \( p \) and the charge \( q = e \) (e being the elementary charge), the corresponding equation reads

\[
\frac{\partial p(r,t)}{\partial t} = \frac{1}{q} \nabla \cdot \mathbf{j}_h(r,t) + G(r,t) - R(n,p,t)
\]

(26)

with \( \mathbf{j}_e \) being the hole current density.

Each continuity equation is paired with a current density equation that establishes, how the flux of charges depends on its spatially and temporarily varying density. At this instance the actual transport mechanism enters the model. A rigorous derivation of the current density equation, starting from the Boltzmann equation, reveals two contributions: one due to a possibly present local electric field \( \mathbf{F} \) (drift) and a second one due to charge density gradients (diffusion). Exemplarily formulated for electrons, the drift term is

\[
\mathbf{j}_{e,\text{drift}}(n) = q n \langle \mathbf{v}_e \rangle
\]

(27)

and accounts for an average drift velocity \( \langle \mathbf{v}_e \rangle \) that the electrons adopt under the influence of a constant electric field strength \( F \). In first approximation, this velocity can be linearly related to the electric field, \( \langle \mathbf{v}_e \rangle = \mu_e F \) with \( \mu_e \) the drift mobility of electrons. As the second considered mechanism, the diffusion current \( \mathbf{j}_{e,\text{diff}} \) seeks to reduce concentration gradients in the semiconductor. Its magnitude is material dependent, described by a diffusion constant \( D_e \):

\[
\mathbf{j}_{e,\text{diff}}(n) = q D_e n \nabla n(r)
\]

(28)

The net current density for electrons is then

\[
\mathbf{j}_e(n) = \mathbf{j}_{e,\text{drift}}(n) + \mathbf{j}_{e,\text{diff}}(n)
\]

(29)

For holes, the current density \( \mathbf{j}_h \) is

\[
\mathbf{j}_h(p) = q p \mathbf{v}_h + q D_h \nabla p(r)
\]

(30)

The formulation of Equations (29) and (26) has two important implications. The mobilities \( \mu_n, \mu_p \) (short-hand \( \mu_{n/p} \)) as well as the diffusion constants \( D_n, D_p \) consider in an effective manner, how the relevant transport mechanism and the properties of the material quantitatively impact the swiftness of motion. To ensure that the values of \( \mu_{n/p} \) and \( D_n, D_p \) are well-defined, we must consider sufficiently large volumes for averaging the local transport properties. Though it is possible to consider spatially resolved mobility \( \mu_\sigma(r) \) even within one layer of a material, possible changes in mobilities should not be resolved finer than ca. 25 nm.

The overall current density is

\[
\mathbf{j} = \mathbf{j}_e + \mathbf{j}_p + \mathbf{j}_{AC}
\]

(31)
The current density contribution $j_{AC}$ is a displacement current that only occurs for a time-dependent, usually harmonically driven device:

$$j_{AC} = \varepsilon_0 \varepsilon r \frac{\partial \hat{E}_{AC}(t)}{\partial t}$$  \hspace{1cm} (32)

To fully express the current density, the local electric field strength $\hat{F}$ is needed and usually determined from the electrostatic potential $\phi$ with the Poisson equation:

$$\nabla (\varepsilon(\vec{r}) \nabla \phi(\vec{r})) = \frac{\rho(\vec{r})}{\varepsilon_0}$$  \hspace{1cm} (33)

Therein, the $\varepsilon(\vec{r})$ is the dielectric permittivity. The charge density $\rho(\vec{r})$ contains all local contributions, be them static or mobile:

$$\rho(\vec{r}) = q \left( p(\vec{r}) - n(\vec{r}) + N_d(\vec{r}) - N_a(\vec{r}) + N_{trap}(\vec{r}, t) \right)$$  \hspace{1cm} (34)

Contributions other than the local charge carrier concentrations $n$, $p$ stem from charged species with fixed positions with their density $N$, for example, donor or acceptor ions ($N_d$, $N_a$) or trap states that temporarily may capture or release charges ($N_{trap}$).

The Equations (25),(26),(29),(30), and (33) form a system of equations the solution of which yields the current densities $j_n$, $j_p$ and the charge carrier densities $n$, $p$. At this stage, the material-specific properties of our device have already entered the model via the effective material parameters $\varepsilon_r$, $\mu_n/p$, and $D_{n/p}$ in Equations (29, 30, 33). Due to the spatial dependence of the material parameters and (intentionally introduced) doping densities, also layer sequences and thicknesses are passed to the model.

Note that current density-voltage relations obtained with Equations (29) and (30) can well reproduce the current densities obtained by elaborate mesoscopic simulations (Figure 9), provided that the mobility and the treatment of the contacts are adequately chosen. However, Equations (29) and (30) do not provide information on the variance of the current density. For example, the DDM is not capable of reproducing the formation of current filaments (right panel of Figure 9), in which the current density strongly exceeds the volume-averaged current density.$^{[108]}$

To fully describe the device in the anticipated mode of operation and to solve the model equations, we need to pass three further sets of information to the partial differential equations: Boundary conditions, initial conditions, and specified source terms (generation and recombination). One boundary condition of the Poisson equation describes the externally applied voltages, $V$. At each contact or conductive layer, the potential $\phi$ is considered constant. The external voltages $V$ are the difference in potentials between each pair of electrodes, for example, $V = \phi_A - \phi_C$ between anode and cathode. For complete devices it is necessary to block a flow of mobile charge through boundaries into non-conducting materials or into the surrounding of the device. For instance, we block the flow of holes from an organic semiconductor to air or to an insulating layer, be that an encapsulation layer or the insulator in an OTFT. This can be done in

\[\ \]

Figure 9. (Left) current density–voltage characteristic of a single layer, single carrier diode as a function of temperature simulated with a 3D Master equation (dots) and a 1D continuum model. b) Current density distribution in the 22 nm thick diode at an applied voltage of 2 V at $T = 300$ K obtained from the Master equation. The formation of current filaments can be clearly discerned. Reproduced and adapted with permission.$^{[108]}$ Copyright 2009, American Physical Society.
a numerical manner by forcing the current density vector to zero in the insulator and letting the electric field vanish beyond an external boundary interface. At interfaces between organic semiconductors adjacent to electrodes or conductive layers, the boundary conditions control the charge carrier injection. The related choices will be discussed in more detail in Section 4.4.

Thanks to the ease of imposing several externally applied electric fields to the sample volume it is possible to account for a complex superposition of local and external electric fields. Figure 10 shows the hole density distribution within a planar organic transistor with five electrodes. Besides source, gate, and drain electrodes, this transistor also possess a control source and control drain contact. The current and the direction of the current (shown by white arrows) can be controlled with four independently applied voltages. The 2D DDM-based simulations reveal that the channel region (5) is supplied with charges through a low resistivity path. Without the control electrodes, the resistivity of the regions leading to and from the channel is determined by the conductivity of the OSC and, thus, usually high.

From the current density \( \mathbf{j} \) (Equation (31)) resulting for an externally applied voltage, or for a set of voltages in the case of transistors, the total current \( I \) in the device for the given applied voltage is obtained by integrating the current density \( \mathbf{j} \) with respect to an area of a plane parallel to the contact:

\[
I = \int_A \mathbf{j} \cdot d\mathbf{A}
\]

4.2.1. Generalized Einstein Relation

Experiments and Monte Carlo-based simulations on transport in organic semiconductors reveal deviations from the conventional current density equations (29) and (30). Particularly the ratio between diffusion coefficient \( D \) and mobility \( \mu \) would deviate from the standard Einstein relation,[155-153]

\[
\frac{D_n}{\mu_n} = \frac{k_B T}{q}
\]

that strictly holds for transport in band extrema occupied according to the Boltzmann distribution.[154]

The generalized Einstein relation offers a route toward a more consistent description of the current density. For an in-depth discussion including the interpretation for a large selection of model scenarios in disordered semiconductors the reader is referred to ref. [155]. The generalized relation between diffusion coefficient and mobility has the form (written here for electrons):[156,157]

\[
\frac{D_n}{\mu_n} = \frac{1}{q n \frac{dn}{d\eta}}
\]

in which \( \eta_{F,n} \) is the quasi-Fermi level for electrons. An analogous expression holds for holes and requires the quasi-Fermi level \( \eta_{F,p} \) for holes. For each charged species (in full analogy to the electrochemical potential[157]), the quasi-Fermi level connects the density of states with the occupied density of states assuming that the occupation of the DOS is described with the Fermi–Dirac distribution \( f_{FD} \):

\[
f_{FD}(E, \mu_F) = \frac{1}{1 + e^{\frac{E - \mu_F}{k_B T}}}
\]

With \( g_L(E) \) being the density of states related to the LUMO, the quasi-Fermi level \( \eta_{F,n} \) fulfills the condition:

\[
n(T) = \int g_L(E) f_{FD}(E, \mu_{F,n}) dE
\]

The use of quasi-Fermi levels is only valid for situations near thermodynamic equilibrium. In principle, a quasi-Fermi level can also be defined for other distribution functions \( f(\eta_{F,n}) \) with the argument \( \eta_{n} = \frac{E - \mu_{F,n}}{k_B T} \) or \( \eta_{p} = -\frac{E - \mu_{F,p}}{k_B T} \), for example for the Maxwell–Boltzmann distribution \( f(\eta) = e^{-\eta} \) or the Gauss–Fermi integral.[158,159]

Inspired by Mensfoort and Coehoorn the generalized Einstein relation can be cast into this form:[157]

\[
\frac{D_n}{\mu_n} = \frac{k_B T}{q} \mathcal{G} (\eta_n)
\]

in which \( \mathcal{G}(\eta_n) \) is the diffusion enhancement factor with \( \eta_n = \frac{E - \mu_{F,n}}{k_B T} \).
relation (40) offers the advantage that the mobilities $\mu_{n/p}$ can be expressed via the diffusion coefficients $D_{n/p}$ and vice versa, that is, one of the two material constants can be eliminated from the current density Equations (29) and (30).

This factor $G$ contains the information on the distribution function $f$ and the density of states \[ G(\eta_n) = \frac{\int dE_g(E) f(\eta_n)}{\int dE_g(E) df/d\eta_n} \] \[ (41) \]

The enhancement factor is exactly unity when $f(\eta)$ is the Maxwell distribution. An enhancement factor different from unity indicates a deviation from the standard Einstein relation (Equation (36)). Evidently, possible deviations stem from the actually employed distribution function $f(\eta)$ and the shape of the density of states $g(E)$. For example, large charge carrier concentrations described with $f_{\text{trap}}$, in particular in connection with doping, provoke a non-constant $G(\eta_n)$ with marked maxima. \[ \text{If, on the other hand, the occupation of a Gaussian-shaped DOS in a disordered semiconductor remains weak due to small charge carrier densities, the standard Einstein relation is approximately valid.} \[ \text{[162]} \]

### 4.2.2. Generation and Recombination Terms

The generation and recombination rates $G$ and $R$ are key to describe OLEDs and OPVs as only these mechanisms allow us to introduce and remove mobile charges. \[ \text{[163–165]} \] Such rates give the number of events per unit time and enter the DD simulation via analytic expressions. The term $G - R$ in Equations (25, 26) is a compact notation (“source and sink” term) for all processes that affect the number of available mobile charge carriers apart from motion. Hence, both $G$ and $R$ ought to be perceived as a composite, in which all perceivable processes can be stored, for example like:

\[ G = G_{\text{ex}} + G_{\text{trap}} \text{, and} \]
\[ R = R_{\text{BTM}} + R_{\text{trap,ext}} + R_{\text{trap}} \] \[ (42) \]
\[ (43) \]

Below we will give examples for the contributions chosen here.

Generation, being particularly vital in OPV, is connected to the ability of the material to absorb light in a given position in the device. As discussed in Section 3.1.4, light absorption primarily generates excitons, that migrate and, eventually dissociate or recombine. Thus, also DD model implementations prefer to resort to associate the generation rate $G_{\text{ex}}$ to the dissociation likelihood of simultaneously monitored excitons.

Recombination processes aim at removing excess electrons and holes to restore thermal equilibrium. Among the bimolecular processes feeding $R_{\text{BTM}}$, recombination events involving electron and holes stemming from different excitons or from injection at contacts are most relevant. For OPV applications, it is particularly important to capture recombination at heterojunctions as a major loss mechanism, as the concentration of prospective partners for recombination is particularly high. Recombination events are expected to occur diffusion-limited and, hence, to follow the Langevin theory. \[ \text{[166]} \]

\[ R_{\text{ass}} = k(np - n^0 p^0) \] \[ (44) \]

in which $k$ is the Langevin rate coefficient, and $n^0$ and $p^0$ are the (negligibly small) charge carrier densities in thermal equilibrium. Though being a constant, note that $k$ describes a two-step process: The likelihood that electron and hole meet and the likelihood that such electrons and holes entering a bound relation ultimately recombine. Based on the formalisms laid out in Section 3.1.4, it was found that the Langevin rate coefficient does not reproduce actually found recombination rates. In fact, the apparent rate coefficients $k$ were found to depend on charge density (including $n$ and $p$), disorder, morphology, electric fields, and the presence of heterojunctions. \[ \text{[118,119,121,122,167]} \]

Though there are analytic expressions\[168–172\] that emerge from the theory of encounter, essentially based on Onsager’s work\[173\] and on recent findings of Burke et al.,\[174\] their use is, strictly speaking, limited to participating materials that are isotropic and homogeneous. Possible alternatives are empirical rate expressions, provided that they are justified for material and morphology in question.\[179\]

Besides interactions amongst mobile charge carriers, also traps assist ultimate recombination\[175\] and have to be regarded as $R_{\text{trap,ass}}$ in a Shockley Read Hall process.\[176\] The notion of incorporating traps into recombination rates can even be lifted on a more general level within DDM. As we saw above, generation and recombination rates establish the connection between electron and holes whose transport is described with formally independent equations. Hence, it is conceivable to introduce rates $G_{\text{trap}}, R_{\text{trap}}$ to also establish an exchange of charges between nominal transport states and trap states.\[164,177,178\] In this way, it is not only possible to introduce any kind of trap type with its spatial distribution, but also describe (hopping) transport within a trap type, provided its density $N_t$ is large enough to permit transport; then, an independent set of continuity and continuity equation has to be provided.

The second option is that rates allow us to account for strongly localized processes even within the framework of a DDM. Rate-described processes can be placed anywhere in the device, in the bulk of layers or at interfaces. Hence, the apparently low spatial resolution dictated by the need to define effective materials can be substantially refined where needed.

**Figure 11** gives an example that DDM can consider trap distributions of different nature and that these traps are essential to reproduce current voltage relations. The shown current–voltage curves relate to a pentacene thin-film transistor, that contains a thin interlayer between pentacene and insulating layer.\[178\] The interlayer releases protons that form a charge density right at the interface and cause a shift of the onset voltage by tens of volts. While DDM simulations with constant mobilities reproduce this behavior semi-quantitatively (Figure 11a), an additional distribution of interface-trap states is required to correctly reproduce the subthreshold region above 20 V (Figure 11b). The hysteresis was explained by bulk traps in the pentacene layers. The concentration of filled deep traps was monitored with an extra continuity equation. To quantitatively explain the difference between up and down sweeps, trapping and detrapping rates in the order of seconds were required (Figure 11c,d).
4.2.3. Remarks on Solving Strategies*

The drift-diffusion model gives rise to systems of time-dependent partial differential equations that contain the applied voltage as boundary condition. As these equations contain variables that continuously vary in space and time (in contrast to the lattice-based formulation of the Master equation), their solution requires a discretization of the spatial differential operators. Established methods are finite differences (FD),[154,163,164] finite element methods (FEM),[154] and finite volumes.[159,180] Depending on this choice, the device geometry is discretized in meshes containing either a discrete value (in FD) or a simple function (in FEM) representing each spatially dependent quantity in a mesh cell. Along with the geometry, also spatially dependent quantities such as $\phi$, $n$, $p$ are now discretely defined in each mesh cell. State-of-the-art discretization schemes rely on the Scharfetter–Gummel discretization and account for the generalized Einstein relation.[159,180,181] The Scharfetter–Gummel algorithm (or Gummel algorithm for steady-state simulations)[182] replaces the native interpolation of the densities $n$ and $p$ between adjacent mesh points in the FD and FEM methods by an exponential interpolation.[183] Exponential changes between adjacent grid point are consequence of the particular form of the current density equations (29, 30) and are difficult to address in finite difference and finite element methods, because they interpolate either linearly (FD, FEM) or with polynomials of the order less than four (FEM). This rather technical twist enhances the stability of the solver.

The spatial discretization of the equations yields a system of equations with the unknown discrete quantities associated to each mesh cell. For a detailed shape of these equations, the reader may consult the references[163,164] for FD, [148] for FEM, or[180] for FV. A large toolset is available to solve these mesh-related equations self-consistently. Possible methods range from Gauss-Seidel schemes,[164] the Newton method[184] to a strongly implicit method.[168] Implementations that require time derivatives are usually approximated with the FD method. Stepping forward in time implies an update of all locations in the simulation volume. This is realized implicitly (Euler backward),[150] explicitly (Euler forward),[144,185] or, less commonly, with a mixed implicit a schemes, such as the Crank Nicholson method.[186]

4.3. Consideration of Heterointerfaces

Heterojunctions, that is, interfaces that are formed by two semiconductors, may enter a drift-diffusion model in two distinct ways. If the two semiconductors are intimately mixed, for example in a bulk heterojunction, the model cannot resolve domains with diameters lower than $\approx 25$ nm in terms of mobility, diffusion constant, and dielectric permittivity. Correspondingly, it is not useful to distinguish changes in the reference energies $E_{\text{LUMO}}$ and $E_{\text{HOMO}}$. Rather, the entire blended domain is considered as an effective medium.[163,187,188] Abrupt or gradual heterojunctions are difficult to consider. The drift-diffusion model in its common form (Equations (29)
and (30)) assumes that charge motion occurs in a homogeneous medium. The energy at which transport takes place is assumed to be constant; spatial changes in energy that could give rise to a driving force are not considered. Letting generation and recombination processes aside, the description of motion does not formally require the energy at which transport takes place.

At heterojunctions, the electron and hole-related densities of states of the materials are set off by an energy \( \Delta L_{1-2} = E_{LUMO,2} - E_{LUMO,1} \) and \( \Delta H_{1-2} = E_{HOMO,2} - E_{HOMO,1} \). For the case of electron transport across the heterojunction displayed in Figure 12a, this offset poses an energy barrier. For sharply defined energy levels, that is, the transport of each charge type takes place in a very narrow range of energies, it is possible to employ a description that has been explored for inorganic heterojunctions (Figure 12b).[189] Therein, the current across an interface has the form of Equation (29) in which the electric field \( \mathbf{F} \) is extended by an additional driving force associated with the band offset.[189] The current density \( j_{n, \text{int}} \) acts as boundary condition to either of the bulk regions.[191,192]
is a generalized electric field $\vec{F}_{n(p)}$ derived from a generalized potential $\tilde{\phi}_{n(p)}(r)$:

$$\vec{F}_{n(p)} = -\nabla \tilde{\phi}_{n(p)}(r)$$  \hspace{1cm} (45)

The generalized potentials $\tilde{\phi}_{n(p)}(r)$ for electrons and holes is a superposition of the actual electrostatic potential $\phi$ and a band parameter $\Theta_{n(p)}(r)$, that accounts for the energetic alignment of the transport levels:

$$\tilde{\phi}_{n(p)}(r) = \phi(r) - \Theta_{n(p)}(r) / q$$ \hspace{1cm} (46)

The band parameters reflect the spatial offset in band locations when going across the interface.[189] In the case of organic-organic interfaces, a convenient choice for $\Theta_{n(p)}$ would be $\Theta_{n,\text{left}} = \Theta_{p,\text{left}} = 0$, $\Theta_{n, \text{right}} = E_{\text{LUMO}, 2} - E_{\text{LUMO}, 1}$, and $\Theta_{p, \text{right}} = E_{\text{HOMO}, 2} - E_{\text{HOMO}, 1}$.[146] The derivative of the band parameters imposes an artificial electric field on the location of the interface that aids current flow depending on the orientation and height of the offset. This approach is particularly suited for gradual heterojunctions, because the band parameter evolves smoothly across the interface and has a well-defined derivative. For abrupt heterojunctions, the value of the generalized electric field $\vec{F}$ has to be determined with great care. Due to the rapid change of the band parameters, the necessary derivation of the generalized potential might be dependent on the cell size of the spatially discretized mesh.

The formalism as introduced above poses two disadvantages. It excludes the possibility to consider thermal activated currents across an energy barrier. Furthermore, we cannot discriminate the impact of the shape of the DOS present at the interface. Wider DOS allow an exchange of charges in a wide range of energies at the interface; in case of a profound overlap of the densities of states even a practically resonant transfer is conceivable. Hence, the energy barriers estimated from the LUMO and HOMO level offsets are larger than the barriers that are effectively overcome.

An important refinement suggested by Cottaar et al. retains the idea of effective electric fields at the interface that modify the current density in Equation (29).[190] In the formulation using generalized Einstein relation, this effective field $\vec{F}$ contains the actual electric field and a possible difference between the Fermi levels $\mu_{\text{F}, 1}$ and $\mu_{\text{F}, 2}$ (Figure 12c):

$$\vec{F} = F(r) + \left( \mu_{\text{F}, 1}(n_i) - \mu_{\text{F}, 2}(n_f) \right) / (q a)$$ \hspace{1cm} (47)

The constant $a$ is the separation of two mesh planes at the interface. This choice abandons the need for band parameters, as the Fermi level difference across the interface readily accounts for the energy offset, possible modifications of the shape of the density of states, and different charge densities.[190] Figure 13 illustrates that DD simulations tend to overestimate the current across the heterojunction. Only dropping the assumption of thermal equilibrium at the interface and the inclusion of short-range Coulomb interactions via effective fields can reproduce the currents obtained from a 3D reference simulation based on the Master equation.

A second idea to address these disadvantages is indicated in Figure 12d. The heterojunction interfaces are removed from the drift–diffusion simulated domain and replaced by boundary conditions that mediate the current exchange between the now separated bulk regions that are still modeled with drift diffusion. The boundary conditions permit an interface current consistent with the hopping rates across the interface.[191,193] Then, the net current density across the interface ($j_{\text{int}}$ as in Figure 12a) readily enters two boundary conditions:[192]

$$j_{n, 1}(n_1, n_2) = j_{n, 1} \text{ and } j_{n, 2}(n_1, n_2) = j_{n, \text{int}}$$ \hspace{1cm} (48)

Therein, $j_{n, 1}$ and $j_{n, 2}$ are the bulk current densities left and right of the interface. These are not only dependent on the bulk charge densities $n_1$ and $n_2$, but particularly on the charge densities found on the left ($n_1$) and right boundary ($n_2$) of the interface (Figure 12d). To satisfy detailed balance, the net current density across the interface is viewed as a superposition of forward $j_{n, 1}$ and backward fluxes $j_{n, \text{int}}$:[192]

$$j_{n, 1} = j_{n, 1 - 2} + j_{n, \text{int}}$$ \hspace{1cm} (49)

as indicated in Figure 12d. The contributions $j_{n, 1 - 2}$ and $j_{n, 2 - 1}$ are now associated with transmission factors $T_{1 - 2}$ and $T_{2 - 1}$:

$$j_{n, 1 - 2} = a_i n_i T_{1 - 2} \text{ and } j_{n, 2 - 1} = a_i n_i T_{2 - 1}$$ \hspace{1cm} (50)

These transmission factors contain the hopping rates $h(\Delta E_{x, x'})$ across the interface multiplied with the density of available target sites beyond the interface $a(N_0 - n)$, averaged within the plain of the interface:

$$T_{n, x - 1} = \int h(a(N_0 - n)) w(\Delta E_{x, x'}) \, d\varphi \, dN$$ \hspace{1cm} (51)

Therein, the factors $a$ and $N_0$ are the intersite distance and the total density of available sites on the target side of the
interface. Note that the barrier offset \( \Delta_{l,12} \) is defined as the difference between two sharp energy levels \( E_{LUMO,1} \) and \( E_{LUMO,2} \). A possibly present electric field component \( \vec{F} \) perpendicular to the interface is accounted for by the term containing \( \phi(l_i^{(e)} - \phi(l_i^{(h)}) \) in Equation (51). Hops are restricted to sites located directly at the interface, that is, from the last layer left from the interface to the first layer right on the interface and vice versa (Figure 12d).

The actual form of the transmission factors in Equation (51) can be obtained by choosing an appropriate model for the rates \( \omega(\Delta E_{i,j}) \) expressions for \( T_{n,k-l} \) for Miller Abraham rates are provided in ref. [192]. Note that the transmission factors \( T_{p,k-l} \) can be analogously formulated for holes surmounting a barrier \( \Delta_{H,12} \).

When solving the drift-diffusion model, the current continuity condition in Equation (48), establishes a connection between the carrier densities \( n_{i,1} \) and \( n_{i,2} \) consistent with the available densities \( n_1 \) and \( n_2 \) in the bulk and the electric field present at the interface. This connection readily permits temperature-activated currents according to the underlying hopping rate \( w \). In order to also incorporate the shape of the DOS as a parameter to control the interface current density, it is conceivable to modify the present formulation of the transmission \( T_{n,k-l} \) between two sharp energy levels (Equation (51)). While the initial energy could be the most likely occupied level, \( T_{n,k-l} \) could be extended by an integration with respect to multiple target energies. [192] Another tempting by-product of the above-described approach is that recombination currents across the interface, that is, electrons recombine with holes present on the opposite side of the interface and vice versa, can be formulated on the same footing. [192]

### 4.4. Consideration of Contact Interfaces

The second type of relevant interface is the contact formed between the electrode and the semiconductor. The electrodes enter drift-diffusion models as boundary conditions. These conditions reflect the two major purposes the contacts serve during operation. The first purpose is to apply an external voltage \( V_i \) between pairs of electrodes. This voltage \( V_i \) corresponds to the difference \( V_{ij} = \phi_C,i - \phi_C,j \) in the electrostatic potentials \( \phi_C \) adopted in the two electrodes \( i \) and \( j \). Each \( \phi_C \) enters the Poisson Equation (33) as boundary conditions for the electrostatic potential. [134] The second purpose of the contact is to exchange charges between the semiconductor and the electrode.

In the most simple case of ideal contacts, charges demanded or offered by the semiconductor are readily transferred from or to the contact. The current density equations (29) and (30) allow us to account for this behavior by assuming an inexhaustible surplus of charges located in the electrodes. In practice, this is realized by introducing an additional “ghost” layer beyond the boundary, at which a large carrier density \( n_C > n(p) \) is placed. [134,146,164]

However, most contact situations are not ideal (i.e., non-ohmic), in particular when the organic layers are not intentionally doped. In such cases, the actually exchanged current density \( j_C \) needs to be incorporated as boundary condition to the current density equations (29) and (30).

All available formulations of \( j_C \) aim at correctly describing the net-current density across the interface. These current-density expressions are parameterized with respect to the charge densities \( n, p \) and/or electric field \( \vec{F} \) and have to be determined self-consistently in the course of solving the drift-diffusion model. In steady-state, the net current as well as its components become dependent on the current flowing in the device. Charge densities and local fields are dictated by the current demand and determine the relative contribution of individual current contributions across the interface. This is illustrated in Figure 14 for the example of a top-contact, bottom-gate thin-film transistor. Its source-drain current increases exponentially with the charge carrier mobility. For small mobilities, thermally injected carriers (red) are slowly carried away from the injecting interface. This injection current is almost fully compensated by a backflow of charges residing next to the injecting interface into the contact (green). At higher mobilities, the injected carriers are moved away swift enough to reduce the likelihood of backflow. An electric field is established next to the contact that reshapes the injection barrier until enough injected charges supply the large bulk-limited current. [143,194,195] At very high mobilities, this field is large enough to permit tunneling.

Net expressions for \( j_C \) keep one current expression floating, while other formulations keep several contributions floating in the course of the simulation. [194,195,201]

#### 4.4.1. Definition of Net Current Densities Across the Contact

Often, the net current density \( j_C \) is expressed using a surface recombination velocity \( v_s \)[196,202]

\[
\vec{j}_{C,n} \hat{n} = q(n - n^0) \nu_n \tag{52}
\]

\[
\vec{j}_{C,p} \hat{n} = -q(p - p^0) \nu_p \tag{53}
\]

with \( \hat{n} \) being the normal vector on the contact surface. The charge carrier densities \( n, p \) refer to the densities at the interface,
and \( n^0 \) and \( p^0 \) are the densities at the interface encountered in thermal equilibrium. Originally, surface recombination velocities were introduced as “collection velocities” in the context of thermionic diffusion-assisted currents across barriers at the contact.\[202]\] This surface recombination velocity acts as a parameter that controls, how efficiently charges are interchanged with the contact. While \( v = 0 \) indicates no net charge exchange with the contact, \( v \to \infty \) corresponds to an ideal exchange. Hence, the floating current density \( j_{\text{fc}} \) depends indirectly to external conditions via changes in the charge carrier density in the semiconductor adjacent to the contact. The barrier is incorporated as a correction to the locally available electric field. The equilibrium densities at the contact are related to the barrier \( \Delta \) at the contact and might be expressed as

\[
n^0 = N_n e^{-\frac{\Delta}{kT}} \quad \text{and} \quad p^0 = N_p e^{\frac{\Delta}{kT}}
\]

with \( N_n \) and \( N_p \) being the effective densities of electron and hole states of the semiconductors next to the electrode.

Expressions as in Equations (52) with constant valued \( \nu \) are predominantly applied for devices of photovoltaic devices, specifically in forward bias and for small electric fields associated with voltages \( V < V_{\text{OC}} \) smaller than the open circuit voltage \( V_{\text{OC}} \).\[193,199,200,203\] For this operation range, \( \nu \) can be extracted from experiment and was shown to be field-independent.\[200\]

We will see below that it is possible to further parameterize \( \nu \) in terms of temperature, fields, or material parameters such as the charge carrier mobility.\[196,197\]

Other formulations split the net current into independent current density contributions. Davids et al., for example, consider thermionic emission \( j_{\text{th}} \) and tunneling \( j_{\text{tu}} \) as responsible mechanisms for injection charge carriers into the semiconductor, whereas interface recombination governs the backflow into the contact.\[201\]

\[
J_{\text{c,n}} = j_{\text{th}} + j_{\text{tu}} = j_{\text{th}}
\]

Such a formulation is particularly appealing, if the field-dependence of the Schottky barrier with an effective height \( \Delta^\nu(F) \) needs to be accounted for. The thermionic emission current density reads as

\[
j_{\text{th}} = AT^2 e^{-\frac{\Delta^\nu(F)}{kT}}
\]

Here, \( A \) is the Richardson constant and \( F_{\text{int}} \) the field strength near the interface. Note that \( \Delta^\nu \) is set to \( \Delta \) if the direction of the field does not lower the barrier. The interface recombination term \( j_{\text{ir}} \) is set to be proportional to the charge carrier densities at the interface and formulated here for electrons:

\[
j_{\text{ir}} = \frac{AT^2}{N_n} n
\]

where \( N_n \) is the density of chargeable sites. The latter expression recovers the ansatz using surface recombination velocities (Equation (52)). An additional tunneling contribution \( j_{\text{tu}} \) is determined by height and width of the Schottky barrier and occurs only if the field at the interface is oriented such that the nominal force \( qF_{\text{int}} \) would pull charge carriers from metal into the OSC. A convenient expression for \( j_{\text{tu}} \) has been derived using the WKB approximation for tunneling through a potential barrier of the shape Equation (8):

\[
j_{\text{tu}} = j_1 \left( \frac{F_{\text{int}}}{E_{\nu}(y)} \right)^2 \exp \left( -\frac{E_{\nu}(y)}{F_{\text{int}}} \right)
\]

in which

\[
E_\nu = \frac{4\Delta^2}{3qa_0R}
\]

with \( R \) the Rydberg radius and \( a_0 \) the Bohr radius. The quantity \( y \) bears information on the barrier shape

\[
y = 1 - \Delta^\nu
\]

and enters

\[
\nu(y) = \sqrt{1 + y \left[ E_{\nu}(\bar{F}) - yK(\bar{F}) \right]}
\]

\[
t(y) = \frac{1}{\sqrt{1 + y \left[ (1 + y)E_{\nu}(\bar{F}) - yK(\bar{F}) \right]}}
\]

with \( \bar{F} = ((1 - y)/(1 + y))^{1/2} \) and the complete elliptic integrals \( E_{\nu}(y) \) and \( K(y) \) of the first and second kind, respectively. Each contribution to the net current density in Equation (55) is adjusted in the course of the simulation, be that with respect to an updated field strength or charge carrier density.

Considering that \( j_{\text{tu}} \) is the time-reversed process of \( j_{\text{th}} \), Scott and Malliaras combine these current density contributions into a joint expression reminiscent to an effective drift current across the interface:\[197\]

\[
j_{\text{c,n}} = 4q\mu_n F_{\text{int}} \left[ \frac{1}{f} - \frac{1}{\sqrt{f}} - \sqrt{\frac{1 + 2\sqrt{f}}{f^2}} \right] \frac{\Delta^\nu}{e^{\nu(T)}} \epsilon_\nu
\]

in which \( f \) is a reduced field

\[
f = \frac{q^2}{k_B T} \frac{F_{\text{int}}}{4\pi\epsilon_0 \epsilon_r}
\]

Note also here that the field-dependence in Equation (63) applies only if the field direction causes a barrier lowering. This formulation removes a floating expression from the simulation and explicitly considers the dependence of the injected current on the electric field strength.

### 4.5. Linking Mesoscopic and Drift–Diffusion Models

The presumably most intuitive link to connect mesoscopic modeling to a drift-diffusion model is to cast the effective motion of a charge into a material parameter, the charge carrier mobility. Indeed, the incorporation a mobility provided by mesoscale simulations into macroscopic simulations represents the most exploited link. What appears to be intuitive relates, however, to...
a common practice known from conventional microelectronic device modeling. This practice was preceded by a rigorous derivation of the drift–diffusion equations from the fundamental underlying mesoscopic equation, the Boltzmann equation.\cite{149,204}

An important outcome of this derivation is that the drift mobility in connection with the diffusivity coefficient fully captures the effective motion of both electrons and holes in a crystalline semiconductor. If the validity of the Einstein relation can be ensured, the mobility can be conveniently connected to the diffusion coefficient and serves, hence, as the only necessary parameter to capture the motion of charge carriers through a material.

However, for all non-perfectly ordered organic semiconductors, that is, most of the employed organic materials, it is not possible to rigorously assign the effective motion to a drift-mobility. In the absence of crystalline order, that is, of a strict lattice periodicity, it is not possible to associate a crystal momentum to the charge carriers. Hence, a Boltzmann equation as the crucial starting point for the existing argumentation cannot be formulated.

Mendels and Tessler found clear evidence, that the sum of drift and diffusion current densities (27, 28) alone are not always able to fully capture currents observed on Monte Carlo simulations.\cite{153} Rather, an additional current density contribution appeared that depends on the local occupation of the DOS. Using an effective local energy $\bar{E}(x)$ to capture the most probable energy of a mobile charge at a position $x$, this additional current would appear whenever local differences $d\bar{E}/dx$ would occur. This gradient $d\bar{E}/dx$ would act as a driving force to complement the drift-diffusion current density $j$ (Equation (29)) as a third contribution:

$$j = q\mu_N D + n\mu_g \frac{d\bar{E}}{dx}$$ (65)

This implies that besides the previously discussed generalized Einstein relation not only globally, but also locally operating corrections are necessary.

Such additional local contributions inherently occur when hopping transport is considered, even if just the transport of a single charge is considered. This has been demonstrated in the only attempt (to the best of our knowledge) to rigorously derive macroscopic drift-diffusion equations from the governing equations for hopping transport.\cite{48} The essential starting idea is to formulate hopping rates that consider also possible changes in the energy landscape due to external potentials. These hopping rates are reformulated to separate short-term and long-term and short-range and long-range contributions into individual factors. Assuming then a macroscopic perspective, the short-term and short-range contribution will impact transport in an averaged manner, while long-range and long-term factors will govern the spatial and temporal advance of transport on macroscopic length and time scales. For details the reader shall be referred to Ref. [48]. While the continuity equation in the form of Equation (25) is recovered, the resulting current density equation contains besides the established drift and diffusion terms ($j_{n, \text{diff}}$) now also two further current contributions $j_{n, \text{hopp}}$:

$$j_n = j_{n, \text{diff}} + j_{n, \text{hopp}}$$ (66)

The latter $j_{n, \text{hopp}}$ splits in two contributions

$$j_{n, \text{hopp}} = -e n(\bar{f}) \mu_N(\bar{f}) \nabla N(\bar{f}) - e n(\bar{f}) \mu_g(\bar{f}) \nabla g(\bar{f})$$ (67)

that contain additional driving forces due to changes in the local density of states $g$ and changes in the density $N$ of locally available sites. These driving forces contain the response of the material via hitherto unrecognized and spatially dependent mobility-like parameters $\mu_N$ and $\mu_g$. To the best of our knowledge, there are no estimates of $\mu_N$ or $\mu_g$ values available yet, even though parts of the "recipe" are indicated in ref. [48]. Dedicated mesoscopic simulations on well-prepared model systems in the spirit of Mendels and Tessler\cite{153} appear to be necessary to quantify both $\mu_N$ and $\mu_g$ and to explore their potential impact. Intriguing use cases certainly are organic-organic heterojunctions, at which the density of states and the density $N$ of available hopping sites may markedly change on rather short distances. Currently needed approaches to augment DDM to deal with heterojunctions (cf. Section 4.3), such as the effective field approach of Cottaar et al.\cite{190} can be possibly recast into hopping-inherent corrections.

In conclusion, a rigorous adaption of the classical DDM to hopping transport will prospectively enable the consideration of important factors such as local gradients in i) the density of available states and ii) the occupation thereof. So far only two suggestions addressing independently either of these factors are available. Despite the potential improvement of the continuum model, one must be aware of the fact that all effective parameters to be employed in such an adapted model will still represent spatial averages and will neglect possibly present short-distance changes.

4.6. DD versus kMC: Comparison of Accessible Length and Time Scales

After having the discussed underlying ideas, assumptions and peculiarities of mesoscopic and continuum device simulation methods, naturally the questions arises on which length- and time-scales it is safe or recommendable to apply each of the approaches. Figure 15 attempts to sort characteristic times and lengths that are characteristic for transport in OSC and for organic electronic devices. The figure readily shows that mesoscopic models (here represented by kMC) and drift-diffusion models appear to share a wide range of scales rather than operating at distinct, complementary scales. A combination of mathematical and physical considerations together with constraints in affordable computational effort determine the involved scales. In the end, as will become evident from the discussion below, most of times the actual purpose of the simulation decides which model to pick.

In terms of time-scales, both methods operate orders of magnitude beyond times associated with molecular vibrations. Both methods allow to access times ranging from microseconds to seconds and even beyond. Within the kMC method, this wide accessible range can be explained with the chosen rate expression, that determines the observed retention times. Due to the exponential terms containing energy and temperature, already a small variation in energy causes a large variety of rate values.
and, hence, retention times. Moreover, rates obtained from model expressions, such as the Miller–Abrahams rate (Equation (1)) and the Marcus rate (Equation (2)) contain a hopping prefactor that sets the highest attempt frequency and, hence, the time scale. For charge transport, the time scale is, however, determined by the relevant energy scales. The latter must contain the Coulomb interaction energies of electrons at adjacent sites. In practice, the hopping prefactor is set to match observables obtained from experiment, for example mobilities [86,109].

The “dynamic range” in DDM-based methods is remarkably wide, for example illustrated with the dark transient currents simulated in Ref. [150]. This range appears wide enough to capture signal delays in OTFTs, in particular the times as small as $10^{-7}$ s as observed for transistors with operating frequencies in the MHz range. Two independent aspects contribute to the wide range of times scales. The DDM consists of a system of partial differential equations. Simulations solve the model on discretized geometries and discrete time steps. In principle, the spatial and time resolution can be chosen as needed. In this way, highly resolved simulations are conceivable. Yet, a chosen resolution in space usually defines a resolution in time, in particular if explicit time-dependent solvers are used (cf. Section 4.2.3). A mathematical stability condition ensures that the time step is small enough to guarantee, that charges do not entirely pass even the smallest mesh cells. Even though the time resolution might become fine, the response of the material on this time scale is not necessarily correctly described, since material constants such as the mobility and diffusion constant are long-time averages. The second aspect relates to the characteristic times associated with trapping and detrapping times that enter the model via recombination and generation processes. These times can be chosen from a large range; the solvers need to ensure small enough time steps and sufficiently long simulation times to capture these.

A more stringent combination of the two aspects can be achieved by recasting the drift diffusion model to consistently account for the wide range of time constants. The variable-order time fractional drift–diffusion equation inherently considers the motion of mobile charges and their trapping and detrapping [205]. The equation monitors the motion of a charge carrier density consisting of mobile charges and charges immobilized in traps. The relationship between the time rate of change of trap and free charge carrier densities takes the form of a continuity equation (Equation (25)) and is based on an asymptotically power-law distribution function for the random residence time in the localized states [205].

The transient current density obtained from such a model can readily capture the dependence of photo-generated currents on the light intensity in time-of-flight experiments [205] shown in Figure 16. At low intensity (black curve), charges require a long time to leave the device. An initially small amount of generated charges quickly occupy deep traps and move forward by traversing through mostly deeply lying states via trapping and deep trapping. At high intensity (red curve), a large
fraction of generated charges escape the device much faster. Due the large amount of initially generated charges, the lowest lying states become readily populated. This blocking of low-energy states enables all remaining carriers to migrate with lesser and shorter-lasting trapping, that is, with higher mobility. These more mobile charges form a weakly-time dependent “plateau” in the transient current response, while the deeply trapped charges arrive considerably later in a long-time tail with a more pronounced time-dependence.

In terms of length scales, the kMC method acts at lengths related molecular distances, that is, in the order of a nanometer, up to micrometers. Hence, devices such as OLEDs and OPVs with a typical thickness of 50 nm thick[144,145,194,195] as already explained in Section 4.2.2. generation and recombination rates can be offered at selected locations, be that in the bulk or at interfaces. Along similar lines, also static local charge distributions can be incorporated to mimic certain electrostatic effects, for example, by placing dipole charge densities at interfaces[144,208] Note, however, that dynamic processes that occur at the interface must be converted into bulk rate models. This conversion to a bulk model introduces approximations in the recombination and generation rates. Such rate expressions are not necessarily capable of accounting for localized changes in fields and charge densities caused, for example, by a specific morphology at donor:acceptor interface. A consideration of such effects ultimately calls for mesoscopic models.[209,210]

In conclusion, it appears safe to state that both methods are suitable for a wide range of devices. Exceptions are either small devices that call for a mesoscopic treatment and very extended devices (e.g., planar thin film transistors) that are described best with a DDM. For all other devices the method is rather picked depending on the needed detail of the considered local effects.

5. Approaches Inspired by Circuit-Modeling

The actual aim of circuit modeling is to replicate the behavior of an electronic device or of devices connected in a circuit. In strong contrast to the previously discussed methods, each kind of model that reproduces the function of a device is eligible. Candidate models to describe the function of a device may contain and connect RC-networks, compact analytical expressions, and look-up tables generated from measurements. This enormous freedom to propose an equivalent circuit model holds the promise to highly accurately predict the device function at much lesser cost compared to the simulations methods introduced before.

The choice of the equivalent circuit model strongly depends on the intended purpose of the simulation. Associated attempts aim to predict i) the time-dependent response of organic devices,[4,211] ii) the time-dependent response of a circuit containing such devices, and the impact of iii) heterogeneous loss regions, and iv) coupled electrical and thermal transport in devices with particularly large aspect ratios.[212–215]

Approaches derived from circuit modeling essentially aim at reproducing the current-response of a device for given applied voltages, be that for steady state or for time-dependent operation. There are two relevant classes of model approaches: Lumped element models and physical models. Lumped element models (LEM) simplify electronic circuits with two major assumptions. Important features of the circuit, such as resistance, capacitance, inductance, and gain are condensed into idealized components. These components are connected by perfectly conducting wires and often have a linear current-voltage relation in common. If the latter is true, one refers to the so-called small signal analysis.
that there is no rigorous protocol on how to partition the underlying electronic circuit into the idealized components. In a more refined approach, so-called diode and transistor models approximately predict the operation of a single real device. In going beyond idealized components such as resistors, inductances, or capacitances, such models permit to consider components with markedly non-linear current voltage relations within lumped element models. However, the complexity of these models (so-called large signal models) renders them less attractive for circuit design. Model formulations can either originate from physical insight into device operation or be empirically conceived from fitting a parametric model to the electrical characteristics. Physical models seek to approximate the underlying phenomena taking place in the devices. Parameters that enter these models are independently quantifiable material and geometry properties such as the charge carrier mobility, oxide thickness, channel length, etc. The presumably most ubiquitous transistor model predicts the source-drain current as a function of the length, etc. The presumably most ubiquitous transistor such as the charge carrier mobility, oxide thickness, channel length, etc. The presumably most ubiquitous transistor model predicts the source-drain current as a function of the drain-source and gate-source voltage in the Gradual channel approximation.[216] Empirical models, often called compact models, essentially rely on determining functions and associated set of parameters that fit the measured data best.

5.1. Main Idea of Equivalent Circuit Analysis

Here we briefly explore the basic idea of equivalent circuit modeling to rationalize i) the form of the required circuit components and ii) the superior modeling speed. The modeling of equivalent circuits is based on equivalent circuit analysis. A multitude of textbooks and web-based tutorials illustrate the basic idea of equivalent circuit analysis, for example ref. [217]. Equivalent circuits are analyzed in terms of the following terminologies. The circuit consists of interconnected pathways that contain components. A pathway is a single line of connected elements and sources. Connections or junctions of pathways are referred to as nodes. Then, a loop is a closed path that contains each node of a circuit strictly once. Note that a loop does not necessarily need to carry a current. The analysis of such circuits essentially exploits Kirchhoff’s laws. Kirchhoff’s current law (KCL) ensures charge conservation and states that the amount charge or current entering a node must equal the amount charge or current leaving a node. Formulated for a node connected with multiple paths $i$ with currents $I_i$, the KCL reads:

$$\sum I_i = \sum I_j^{\text{in}} - \sum I_k^{\text{out}} = 0$$  (68)

in which the sums over $j$ and $k$ run over the incoming and outgoing leads, respectively. Note that for each of the currents $I_i$ there must exist a closed circuit path, be that through wires or other circuit components. The second Kirchhoff law, the Kirchhoff voltage law (KVL), ensures energy conservation by stating that the sum of all voltages must be equal to zero within a closed loop:

$$\sum V_i = 0$$  (69)

These voltages $V_i$ are the voltages that drop across each branch that participate in the loop. The last important term is the branch, that is a single component or a group of components that are terminated by two nodes.

The circuit analysis exploits both Kirchhoffs laws along the following recipe to provide a system of equations:

1. Label voltages $V$ and impedances $Z$ (such as $V_1$, $V_2$, $V_3$, R1, R2, Z1, etc.)
2. Assign a current $I_i$ to each branch, either clockwise or anticlockwise. The latter orients $I_i$ currents so that $I_i$ may enter equations with a positive or negative sign depending on their orientation with respect to the considered node.
3. Formulate KCL (Equation (68)) for each node. Each node produces a linear equation in the current $I_i$.
4. Formulate KVL (Equation (69)) for each independent loop. The voltage drops across the branches are expressed in terms of the known current-voltage relations for the components in each branch.
5. Solve the system of equations provided by KCL and KVL to obtain the branch currents $I_i$.

In the most simple case, the ohmic law is employed to formulate the KVL, that is,

$$V_i(t) = Z_i \times I_{i,\text{branch}}(t)$$  (70)

in which the impedance $Z$ is a complex number $Z = Z e^{i\theta}$ to address also time-dependent responses. Impedances do not only consider resistors, but also capacitances or inductances. For steady state operation, the impedance of a resistor takes the simple form $Z = R$. Then, KVL produces equations linear in the branch currents $I_i$, so that the entire system of equations to be solved is linear as well. In the presented formulation, the number of equations entering the system depends on the choice of loops and nodes. For remove this ambiguity, there are a number of strategies that aim at removing all redundant equations by tying the relations either to nodes or internal loops. For details, the interested reader shall be referred to ref. [217]. Beyond impedances, also current and voltage sources can be considered while preserving the convenient matrix form of the equations.

Non-linear relations between voltage and current, to be entered in the KVL (69), considerably complicate the circuit analysis. While equations may now contain functions in $I_i$, for example, the exponential function in diode models, some equations may even provide just an implicit relation between voltage and current.[218]

As long as one can justify to work in the approximation of Ohms law (Equation (5.1)), the circuit analysis is very fast as one just needs to solve a linear system of equation. Ohms law is particularly well justified in the small signal case, that is, if one is interested in changes due to small changes in current or voltage:

$$V_i(t) = V_0 + Z_i \times I_{i,\text{branch}}(t)$$  (71)

That implies, that circuit components are assigned to impedances in an equivalent circuit, that is, to the idealized components in lumped element models.
Regardless of the level of simplification of the current-voltage model relations, the notion of branches and circuits allows us to readily spot how the external voltage enters the model.

5.2. Application of Equivalent Circuit Models

OLEDs may serve us here as a particular rich illustration, how the intended purpose guides the conception of highly distinct lumped element models for the very same device type. Later, for reasons explained in Section 5.3, we will also incorporate transistor models. In general, it is useful to discriminate between steady-state operation and transient operation.

5.2.1. Steady State Operation of OLEDs

Lumped element models for steady state operation mainly seek to predict the current-voltage characteristics of a device. For this purpose, OLED-related equivalent circuits tend to contain not only resistances and capacitances, but also diode elements with a \( I(V) \) relation reminiscent to Shockleys law for p-n junction diodes.\(^{[219]}\) Note that the actual shape of the \( I(V) \) relation is given by the particularly chosen diode-model, that readily permits to also consider temperature in the device simulation.\(^{[211,220]}\) Such a nonlinear current voltage \( I(V) \) relation complicates calculations of related circuits. In an effort to easier incorporate the diode-model into a conventional equivalent circuit solver while maintaining a reasonable accuracy, it has been suggested to approximate the exponential voltage dependence a Taylor series.\(^{[220]}\)

Resistors and capacitances augment the diode element to capture non-ideal diode behavior\(^{[220,222]}\) and to discriminate operating regimes.\(^{[213,221,223]}\) The latter is illustrated with the network shown in Figure 17. A typical current voltage characteristic of an OLED is divided in three forward voltage regions, illustrated as colored regions in Figure 17a. Each of these regions is accounted for by dedicated branches in the equivalent circuit, that are highlighted with the same color in Figure 17b. For the smallest voltages (green), the current is solely governed by the contact resistance \( R_C \), and a parallel resistance \( R_P \) to account for a leakage current. Capacitance \( C_G \) denotes the geometrical capacitance of the diode layer stack. In voltage region 2 (orange) beyond a built-in voltage \( V_{bi} \), charges start to be injected and flow through the series resistances \( R_C \) and \( R_P \), where the latter is located in the left-most diode branch (orange). The capacitance \( C_D \) stores charges that are accumulated in excess of \( C_G \). Note that \( R_C \) typically exceeds \( R_C \) and \( R_P \), so that the right-most green path is considered an open circuit in voltage regions 2 and 3. In the final region 3, the second diode branch is activated (mauve branch in Figure 17b). The diode element describes an appreciable current that is expected to set in beyond the threshold voltage \( V_0 \). The series resistance \( R_s \) accounts the parasitic series resistance. Each branch has been designed to suggest individual static and time-dependent measurements of each of its components, that is, to populate the equivalent circuit of an OLED without fitting. Note that capacitances in general are not relevant in steady state, because the OLED equivalent capacitance will behave as an open circuit and has no effect on OLED voltage and current.\(^{[223]}\)

5.2.2. Transient Operation and Impedance Spectroscopy

Equivalent circuit models are the backbone of the interpretation of transient responses and associated impedances. Commonly, these models are intentionally kept simple and belong to small signal models. To explore the transient response of the OLED, the OLED is operated with a pulsed signal that is superimposed with a DC voltage, \( V_{DC} \), that defines the point of operation. For impedance spectroscopy, for example, a sine-shaped AC voltage is used:\(^{[4]}\)

\[
V(t) = V_{DC} + V_{AC} \sin(2\pi ft)
\]

(72)

Therein, \( f \) is the frequency and \( V_{AC} \) the amplitude of the AC signal. The response of an organic diode structure, be that the current or the luminous flux, is going to be delayed and characteristically shaped depending on the impedance of the structure.\(^{[211]}\) In other words, the perturbing pulsed signal probes the capacitances and resistances associated with the steady-state operation at \( V_{DC} \).\(^{[224]}\)

The choice of the equivalent circuit depends on which specific features of the device should be accounted for.\(^{[224]}\) Two main classes of choices may be discerned: A first class seeks to determine characteristic time-constants associated with the underlying transport process and enable the analysis of pulse-shapes.\(^{[211]}\) Associated equivalent circuits provide a succession of RC links, each consisting of a resistance \( R \) in parallel to capacitance \( C \). Each RC link has a unique impedance given by its \( R \) and \( C \) value and corresponds to a time constant. ECM of a second class focus on the determination of capacitances and resistances associated with layers and interfaces between, often with the intent to interpret capacitance voltage and capacitance frequency spectra.\(^{[4,224–226]}\) This direction is in line with the notion to conceive equivalent circuit models that relate to specific regions in the device. We will discuss these efforts further in the upcoming section.

![Figure 17. Equivalent circuit model of an OLED proposed by Bender et al.\(^{[227]}\) (b) to capture distinct regions of the \( I-V \) characteristic (a).](image-url)
5.3. Consideration of Spatially Distinct Device Regions

Particular interesting in the context of the device modeling hierarchy are developments that attempt to rigorously assign the elements in the equivalent circuit to spatial regions in the device. We collected selected developments in this separate section, because these types of equivalent circuit models represent the presumably strongest links to device simulations with spatially resolved potentials and current densities, be them performed on the continuum or even on the microscopic scale.

Nowy et al. and Weis et al. provide particularly intuitive examples, in which each electrically active layer is assigned to a parallel RC element; connected in series these RC links describe the full OLED stack.\(^4\) Within this model, the layer-specific \(R\) and \(C\) are permitted to change their values depending on the operating conditions. In such a way, it is possible to distinguish between low-resistance and high-resistance states of the layers for a given operating voltage.

Devices with non-homogeneous lateral distributions of charges such as large-area OLEDs\(^{212,214,215,228}\) and organic transistors\(^{227}\) call for spatially resolved network models. The function of such devices is not only governed by processes within the transversal device cross-section that contains all active layers, but also by laterally formed distribution patterns of charge density, potential, and temperature.\(^{394,395,229,230}\) Here the overarching idea to conceive ECM is illustrated in Figure 18 for an organic transistor.\(^{227}\) The laterally extended device is partitioned into building blocks. Each building block contains the entire device cross-section (i.e., layer sequence) so that it is possible to assign them a qualitative current–voltage behavior. In case of the transistor shown in Figure 18, the device is first partitioned into three distinct regions that are further subdivided into building blocks. Building blocks in the central region assume the \(I-V\) characteristics of an ideal transistor to represent gate-assisted charge accumulation in the channel region. Building blocks associated to the contact regions consist of branches containing not only an ideal transistor, but also RC links to account for the transport in high-resistance regions between channel and electrodes. Note that the proposed branches in these high resistance regions contain voltage-dependent resistances \(R\) to account for a non-ohmic current voltage relation. Assembling all building blocks according to Kirchhoffs laws readily yields a map of potential distribution and resistors across the device. Such networks offer a convenient way to introduce non-ideal effects such as non-ohmic injection and ohmic losses and to identify regions that hamper ideal operation most.

ECM also supported our understanding on the temperature and brightness distribution in large area OLEDs under operating conditions.\(^{229}\) Such OLEDs are partitioned into regular grids of building blocks. These building blocks can be connected to electrical, but also to thermal networks.\(^{212}\) In thermal networks, the electric current is replaced by the heat flux and the voltage by the temperature. The choice of the current–voltage and current-temperature response in each building block provides the coupling between electrical and thermal network. This lends to an appreciable flexibility in parameterizing the building block. With the intent to capture instabilities due to self-heating, thermally activated conductivities were mapped on building blocks that formally act as thermistors.\(^{214,215,229,230}\)

Another vivid field is the conception of compact transistor models.\(^{211–213}\) For transistors, all distinct device regions that can act as capacitances are of central importance. These capacitances inform to which extent and how fast charge carriers are rearranged in the transistors. Correspondingly, knowing the capacitances and the charges stored therein allow one to predict the switching speed of the transistor\(^{234}\) or how fast a sensor can be read out reliably. Transistor models aiming at compact capacitance models inherently discriminate different device regions, as these models build

![Figure 18.](image-url)
on the existing physical understanding of charge rearrangements in different parts of the transistor.\cite{231,233,235} For the latter reason, such models are conceived and bench-marked in a tight feedback loop with drift-diffusion-based simulations.\cite{231,233} Typical applications comprise the determination of voltage- and frequency-dependent capacitances in transistors\cite{231,233} and small circuits,\cite{235} or time-dependence of charge accumulation.\cite{231}

While it appears to be straightforward to inform the latter transistor models with more detailed drift-diffusion-based simulations, it is not established whether it is possible to rigorously extract a network of building blocks (seen above for OLEDs and transistors) from any drift-diffusion-based simulation. So far, the choice of building blocks appears to be rather ambiguous, as the choice is governed by physical insight and numerical considerations regarding the optimal number of building blocks. In an intriguing way to remove such ambiguities, at least in part, is inspired by Liero et al.\cite{214} They suggest a protocol to rigorously translate a set of a partial differential equations describing electrothermal transport in an OLED to an equivalent circuit network.

### 6. Conclusions

State-of-the-art simulations of the electric behavior of organic electronic devices rely on a broad set of models. Organic, in particular optoelectronic devices cannot be comprehensively simulated with an ultimate, all-purpose model. Rather, we are offered distinct model approaches, each of which addresses device features at certain length scales and with specific assumptions. In the present paper, we provided an overview on the prevalent model approaches and an in-depth discussion on their foundations. Each of the models has been inspected in terms of its capability to capture, how temperature, external electric fields, and charge carrier concentrations govern the device function under operating conditions. To this end, we first provided a brief summary of key aspects, that set the charge transport in organic materials apart from transport regimes known in conventional inorganic semiconductors.

With a proper choice of the underlying model it is possible to elucidate a large variety of distinct influences, stretching from peculiarities of molecular packing at an interface to the impact of a parasitic capacitance on the transient behavior of an entire device in a circuit.

For the particular case of hopping transport, the model approaches pose ample leads to tightly integrate them into parameter-free multi-scale simulation frameworks. In fact, once combined, models may partially compensate individual weaknesses and extent the range of their applicability. A highly encouraging example of bridging and integrating multiple length scales and model levels into simulation workflows is the prediction of charge carrier mobilities due to hopping conduction in organic semiconductors.\cite{236}

This opens the question in which direction future developments can be expected. The current state-of-the-art offers ample opportunity to integrate and to conceive tailor-made frameworks for certain devices or research tasks. Still immature or desirable developments are:

1. Rigorous connections of charge transport models or device models are necessary stepping stones. An obvious candidate appears to be the strict derivation of drift–diffusion equations from microscopic charge hopping. A second candidate is a stricter conception of compact and lumped-element models on the basis of elaborate drift–diffusion-based descriptions.

2. On an even more fundamental level, it appears to be desirable to explore dynamic-disorder-limited transport under realistic operating conditions for organic transistors, that is, in devices that will exploit crystals most likely. To the best of our knowledge, there is no device simulation relying on the strong-dynamic-disorder model. As an important intermediate step toward such a model, the strong disorder model was used to perform kMC calculations\cite{227} and to predict a drift velocity.\cite{238} An additionally necessary step would be to explore, how the charge carrier mobility is affected by other mobile or static charges.

3. All above-mentioned efforts lend themselves to a more efficient prediction of the device functionality under accurate consideration of multiple, possibly competing effects. With such “forward” models, high-throughput screenings and machine learning approaches can be lifted from highly local, microscopically defined figures of merit\cite{239} or single criteria such as optical densities\cite{240} toward device models considering transport and optical aspects.\cite{241}

4. Holistic simulations of devices comprise at least the transport of charges and energy. Future developments will seek to find the most efficient model stage for each of the transport processes.\cite{212,240} Such efforts will clearly fuel activities to rigorously couple the model stages. Though being apparently not in the focus of current attention, well-integrated model stages may lead to future simulation frameworks with “accuracy-on-demand.” Such frameworks ought to self-reliantly determine which device region requires an in-depth microscopic description and which device region can be treated on a coarser level—possibly to the extent of compact model.

### Acknowledgements

The author thanks M. Krammer, G. Krikun, M. Gruber, B.A. Stickler, and S. Possanner for fruitful discussions that reflected their engagement in pushing the limits of device simulations. The author acknowledges financial support by the Austrian Science Fund (FWF) through the FWF Elise Richter Fellowship (V317-N20).

### Conflict of Interest

The author declares no conflict of interest.

### Keywords

charge transport, circuit modeling, drift-diffusion, kinetic Monte Carlo, master equation, organic devices

Received: January 31, 2021
Revised: May 17, 2021
Published online: June 23, 2021
Karin Zojer received her Ph.D. in Physics (2003) from the Technical University of Dresden. Afterwards, she joined the group of Jean-Luc Brédas for 4 years at the University of Arizona and at the Georgia Institute of Technology (2003–2007). She moved to Graz University of Technology in 2007 and established her research group. Her research has always involved organic semiconductors and comprised topics such as optical spectroscopy, materials modeling, and modeling of organic devices. In recent years, the focus of her activities has been on understanding the impact of device design on the performance of organic devices.