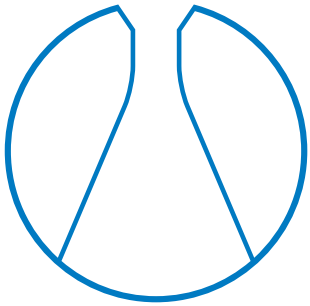
TECHNISCHE UNIVERSITÄT MÜNCHEN

Department Chemie
Lehrstuhl für Theoretische Chemie



Ab Initio Charge Carrier Mobility and Computational Screening of Molecular Crystals for Organic Semiconductors

DISSERTATION

Christoph Otto Schober

TECHNISCHE UNIVERSITÄT MÜNCHEN

Fakultät für Chemie

Lehrstuhl für Theoretische Chemie

Ab Initio Charge Carrier Mobility and Computational Screening of Molecular Crystals for Organic Semiconductors

Christoph Otto Schober

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzende : Prof. Dr. Corinna Hess

Prüfer der Dissertation:

1. Prof. Dr. Karsten Reuter
2. Prof. Dr. Ulrich K. Heiz
3. Prof. Dr. Jochen Blumberger, UC London / UK (schriftliche Beurteilung)
   Prof. Dr. Peter Müller-Buschbaum (mündliche Prüfung)

Die Dissertation wurde am 21.11.2016 bei der Technischen Universität München eingereicht und durch die Fakultät für Chemie
am 13.01.2017 angenommen.

Für Eva.

Abstract

The rapid progress in the field of organic light emitting diodes has raised high expectations for novel organic electronic devices such as organic field effect transistors or organic photovoltaic devices. But before these devices are ready for the market, important device parameters such as the charge carrier mobility need to be improved upon. Theoretical approaches are further complicated by the fact that to date no single model incorporates all crucial effects. Moreover the charge transfer behaviour of such materials is not yet fully understood. In this thesis a systematic analysis of different computational methods to calculate an important charge transfer parameter, the transfer integral H_{ab}, was performed using density functional theory. This was aimed at providing best-practice guidelines to choose between different variants of the popular fragment orbital approach (FO-DFT). A new, hitherto unknown variant of the FO-DFT scheme was developed which yields increased accuracy at negligible additional computational cost. In addition, a robust method to calculate another charge transfer parameter, the internal reorganisation energy λ_{int}, was developed. In this approach, the solid state environment of an individual molecule of a molecular crystal is retained by performing a constrained geometry optimisation for the calculation of λ_{int}. This technique allows for automated calculations without manual monitoring of the individual calculation steps. By applying these methods in a computational high throughput screening study using the two *ab initio*-determined descriptors H_{ab} and λ_{int}, 95 445 molecular crystals from the Cambridge Structural Database were re-evaluated regarding their charge carrier mobility and capability as novel organic semiconductor material. Four promising molecular crystals were selected based on the results of the screening study and presented for further experimental verification. Moving towards full *in silico* computational screening of novel materials the reliable prediction of molecular crystal structures was investigated in a collaborative contribution to the *Sixth Cambridge Crystal Structure Prediction Blind Test*.

Zusammenfassung

Jüngste Fortschritte im Bereich von organischen Leuchtdioden haben hohe Erwartungen an neuartige organische Schaltkreise, wie zum Beispiel organische Feldeffekttransistoren oder organische Solarzellen geschürt. Um diese Materialien zur Marktreife zu bringen, müssen jedoch noch grundlegende Materialeigenschaften verbessert werden. Eine große Herausforderung in diesem Bereich ist die Verbesserung der Ladungsträgermobilität. Zur Berechnung dieser Schlüsselgröße existieren verschiedenste theoretische Modelle. Jedoch gibt es bis heute kein Modell, das alle relevanten Effekte berücksichtigt und zur Berechnung von verlässlichen Ladungsträgermobilitäten genutzt werden könnte. In dieser Arbeit wurde deshalb eine systematische Analyse verschiedener computergestützter Methoden zur Berechnung von Ladungstransferparametern mittels *ab initio* Dichtefunktionaltheorie durchgeführt sowie verbesserte Methoden entwickelt. Dabei wurde die beliebte und weit verbreitete Fragmentorbital-Methode zur Berechnung von Transferintegralen H_{ab} mit Referenzwerten verglichen und eine neuartige, bis dahin unbekannte Variante der Methode mit deutlich erhöhter Genauigkeit entwickelt. Durch die vergleichende Analyse konnten Richtlinien zur Wahl verschiedener Rechenmethoden je nach erforderlicher Genauigkeit oder Rechenaufwand erstellt werden. Zur Berechnung eines weiteren Ladungstransfer-Parameters, der internen Reorganisationsenergie λ_{int}, wurde eine verbesserte, automatisierte Rechenmethode entwickelt. In diesem Ansatz wird die Kristallumgebung eines einzelnen Moleküls in einem molekularen Kristall durch eine einzige Nachbarschale imitiert und dadurch die automatisierte Berechnung der nötigen optimierten Molekülgeometrien ohne manuelle Überwachung ermöglicht. Durch die Anwendung dieser neu entwickelten Methoden in einem Hochdurchsatz-Screening unter Verwendung der Parameter H_{ab} und λ_{int} wurden 95 445 organische Kristalle der Cambridge Structural Database (CSD) unter dem Gesichtspunkt ihrer Ladungsträgermobilität neu untersucht und deren mögliche Eignung als organisches Halbleitermaterial bestimmt. So konnten vier bis dahin unbekannte Molekülkristalle identifiziert werden, die eine hohe intrinsische Ladungsträgermobilität besitzen und nun in weiteren experimentellen Studien untersucht werden können. Um im nächsten Schritt über die in experimentellen Datenbanken wie der CSD vorhandenen Daten hinaus Untersuchungen zu ermöglichen, wurde durch einen in Kooperation entstandenen Beitrag zum *Sixth Cambridge Crystal Structure Prediction Blind Test* die Möglichkeit einer vollständig computergestützten Vorhersage von Kristallstrukturen weiterentwickelt.

Contents

1 Introduction

This thesis is a publication-based dissertation implying that original research was published in international scientific journals. Summaries of the published articles are provided. The focus of this thesis is on providing further background on the theoretical methods employed and on relating the results to relevant literature.

Ever since the rapid development of the semiconductor industry in the 20th century the importance and distribution of electronic devices has increased and plays an irreplaceable role in today's life. Semiconducting materials are for example crucial components in sensors, solar cells, cars, LEDs and many more. Transistors, first proposed in a patent application by Lilienfeld in 1930[1] and successfully built by Shockley, Bardeen and Brattain at AT&T's Bell Labs in 1947, are the most important building blocks of modern electronic devices. In today's applications complex electronic circuits with billions of transistors are combined in integrated circuits. In 1965 Gordon Moore stated in his famous law that *"the number of transistors in an integrated circuit doubles roughly every two years"*[2]. While this prediction still holds true today, the size of individual transistors approaches the fundamental barrier of atomic dimensions, with further miniaturisation impossible for current semiconductor materials. With the substantial progress made since the early 20th century, devices that formerly occupied large rooms such as the first computer can nowadays be carried in a pocket. At the same time, research continued to develop electronic circuits based on different materials, aiming for cheaper production, environmental sustainability or novel functionalities. With the discovery of organic semiconductors by Kallmann and Pope in 1960[4, 5] another class of materials was known, but poor conductivity limited the interest to academic research of this novel phenomena. But partly because of the promising novel properties such as the ability to be manufactured on flexible substrates[6, 7], using for example ink-jet printing [8, 9] or solution processing techniques[10–13], research on these materials continued nevertheless. With increasing theoretical understanding of the fundamentally different charge transfer processes in organic semiconductors and enhanced

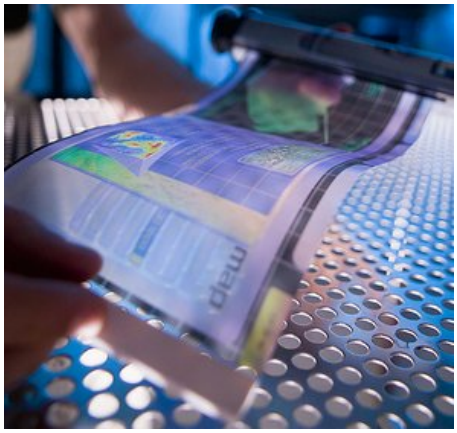


***Figure 1.1**: Example of a flexible display based on organic electronics[3].*

preparation techniques materials with higher conductivity and charge carrier mobility have been found[7, 8, 10, 14–16]. While development for inorganic semiconductors aims at further decreasing the size of individual building blocks, allowing for example for more powerful and resource efficient processors, the main goal for organic semiconductors is to further increase the charge carrier mobility and therefore the range of applications for which such materials can be successfully used. First applications of organic semiconductors already became commercially available, e.g. displays with organic light emitting diodes (OLEDs) and flexible semi-transparent photovoltaic cells (OPVs). According to a recent study by IDTechEx, the organic electronics market will grow from $26.54 billion in 2016 to $69.03 billion in 2026[17]. Within organic electronics OLEDs are the most mature technology, while organic field effect transistors (OFETs) and OPVs still suffer from too low carrier mobilities[18].

As the carrier mobility is the most important device parameter for those applications[19], substantial effort has been spent to enhance existing and to develop new materials in this respect. Different classes of chemical backbones with promising charge transfer properties have been identified over the years, ranging from small molecules[10, 20–22] to polymers[23–26]. Improvements were achieved by modifying molecules based on chemical intuition, optimizing crystal stacking with side chain modifications or devising new molecules based on early rational design guidelines. Unfortunately, the charge carrier mobility is an intrinsic material property and difficult to access experimentally. The measured mobilities are influenced by the employed device architecture[27, 28] or device fabrication process[29–31]. By optimizing these parameters, improvements in carrier mobility on the order of magnitudes have been found[30, 32, 33]. To aid the experimental progress and increase the understanding of the underlying mechanisms of charge transfer in organic semiconductors many theoretical methods have been developed and applied to a broad range of systems[34–43]. A quantitative prediction of charge carrier mobilities even for simple systems is still very demanding[44] and heavily depends on the assumptions built into the different charge transfer models. On the other hand, models such as small polaron hopping[45] can establish relative mobility relationships between different organic semiconductors[19, 46] at only moderate computational cost. By careful examination of different models two important charge transfer parameters can be identified, namely the transfer integral H_{ab} and the reorganisation energy λ. Both serve as an indicator or descriptor for the charge transfer behaviour of a system, allowing rough estimates of the expected intrinsic performance. In a pioneering study Sokolov *et al.*[47] successfully employed the latter descriptor to assess the suitability of seven modifications of the well-known organic semiconductor dinaphtho-thienothiophene (DNTT)[48]. This study can be seen as a proof-of-concept for

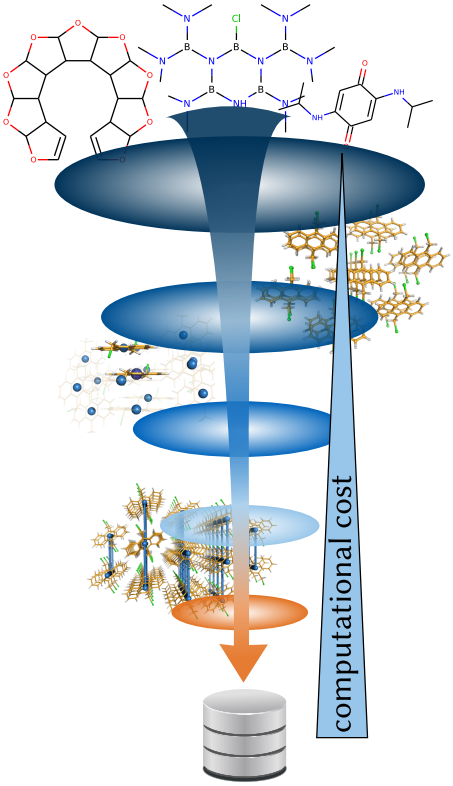


Figure 1.2: *Schematic virtual screening approach with increasing computational cost and accuracy of calculated descriptors.*

a virtual screening approach, in which a huge number of compounds (the *compound library*) is tested automatically for pre-defined quantities using approximate but sufficiently accurate descriptors. Ideally, the systems are tested on different levels of accuracy with increasing computational cost. This approach was initially developed in the context of docking studies in drug design and pharmaceutical research[49–51], but is increasingly adopted in material discovery as well[52–55]. High-throughput virtual screening studies together with a large and systematic compound library, accurate but efficient descriptors and subsequent data analysis allow for computational methods to predict new materials.

With the prediction of novel organic semiconductor materials as ultimate goal, this thesis first addresses the calculation of suitable charge transfer descriptors solely from first principles. While many different methods and models in this area are known (see chapter 3), a systematic comparison and evaluation of the widely used fragment orbital density functional theory (FO-DFT) approach for the calculation of transfer integrals within one computational framework has been done for the first time. In addition, an additional variant of the FO-DFT scheme was developed and shown to perform better than hitherto known flavours when compared with high level *ab initio* reference data (cf. section 8.1), even outperforming more involved constrained DFT (cDFT) calculations for molecular crystals. To complement the descriptor a method to reliably calculate intramolecular reorganisation energies of molecular crystals using an efficient QM/MM approach was developed. This method eliminates the need to carefully verify the obtained geometries[56] and thus permits an automated calculation for a huge number of crystals. With two accurate charge transfer descriptors available for automated calculations, we created a screening workflow for molecular crystals, using increasing levels of computational cost and accuracy. This workflow was used on 95 445 molecular crystals from the Cambridge Structural Database[57], an extensive collection of experimentally measured systems.

The following chapters discuss all important steps towards the *ab initio* high-throughput screening performed in this work, with a special focus on the necessary improvements and new developments in theoretical methodology and computational workflows.

2 Organic semiconductor devices

In this chapter the structure and functionality of the three different organic solid devices commonly referred to as OLED, OPV and OFET and the role the carrier mobility plays for each of those are briefly summarized.

In OLEDs, the primary process is the generation of photons using electric current. To this end, modern OLED designs combine suitable organic hole and electron transfer materials in a multilayer approach (see Fig. 2.1) with an emissive layer and a transparent electrode (usually indium tin oxide, ITO). Charge carriers are injected at the electrodes and form excitons, which then decay under fluorescence emission. The performance, usually measured in terms of the quantum yield of a system depends heavily on the interaction between carrier mobilities, energies and fluorescence[41, 58]. In classical OLEDs, the carrier mobility should be balanced between hole and electrons to allow efficient decay of generated excitons within the boundary of the emissive layer. While an efficient fluorescence quenching is mandatory for high quantum yields, this often prevents a sufficiently high carrier mobility as is mandatory for pure organic devices combining OLED and OFET functionality, so called organic light emitting transistors (OLETs) [58]. Comprehensive multi-scale simulations performed by Kordt and coworkers[41] also confirmed the importance of accurate charge carrier mobilities to obtain reliable device parameters.

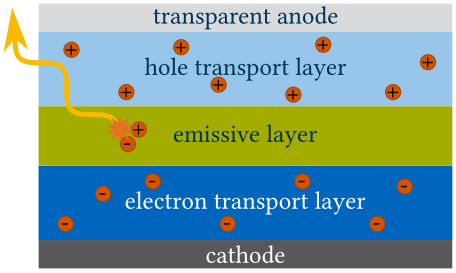


***Figure** 2.1: Schematic depiction of a multi-layer organic light emitting diode (OLED) architecture.*

Organic photovoltaic devices make use of the inverse process, harvesting light in order to generate electric current. Similar to OLEDs different device architectures exist, but the working principle is the same (cf. Fig. 2.2). A donor (hole conductor) and acceptor (electron conductor) are combined between two electrodes. Upon illumination with light, photons are absorbed and excitons form. If an exciton reaches the donor-acceptor (D/A) boundary within its lifetime, the exciton dissociates into free charge carriers. The free carriers then drift and diffuse towards their respective electrodes and can be collected. To improve the carrier yield different device architectures have been developed, with prominent examples being the mixed interlayer or the bulk heterojunction design[59, 60]. Both

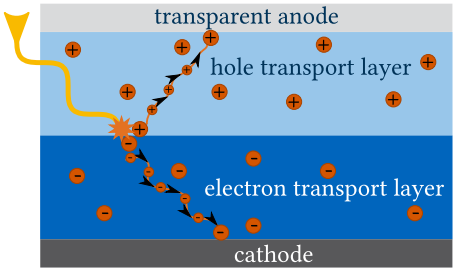


***Figure** 2.2: Schematic depiction of a bilayer organic photovoltaic device (OPV).*

designs aim at improving the number of available free charge carriers by increasing the size of the D/A interface, which is the only region where excitons can dissociate into free charge carriers. These optimizations need to balance the charge carrier mobility, absorption of photons and efficient charge collection at the electrodes, thus rendering them rather complex[61, 62]. Dimitrov and co-workers showed that, in accordance with previous work, the electron mobility plays a crucial role to facilitate the escape of charge carriers from the D/A junction, while a sufficiently high hole mobility is necessary to avoid space charge accumulation [63].

An OFET consists of a semiconducting layer separated from a gate electrode by a thin insulating dielectric. A drain and source electrode are connected to the semiconductor and separated by a channel of certain length and width. OFETs exist in different device architectures (cf. Fig. 2.4), which not only differ in the production process, but also in the measurable device parameters. Depending on the manufacturing process and stability of the semiconducting layer different techniques and materials are used[27, 64]. In contrast to OLEDs and OPVs, where the carrier mobility is only one important parameter among others, for OFETs the carrier mobility is absolutely crucial for the device performance. It affects the on-off current ratio important for OLEDs, with many possible applications enforcing a lower limit on the mobility. Much sought-after are all-organic OLED displays, where the transistors for the active matrix addressing each pixel are OFETs. Such displays can be built on a truly flexible substrate, allowing novel applications. Due to the short time each pixel is addressed, this necessitates OFETs with fast response time, i.e. high carrier mobility. Estimates range from $\mu > 1.5\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$ to $\mu > 10\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$, depending on the architecture of the circuits used to control the pixels[19]. Among the many known organic semiconductors only a small fraction has shown such high mobilities, while most tested materials have mobilities lower than $1 \times 10^{-3}\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$. Although the mobility is the primary parameter, other aspects such as the durability under operating conditions or the manufacturing process also play a major role when commercialising such materials. Much effort goes into the development of high mobility semiconductors which are soluble, allowing cheap production techniques such as spin-coating[30, 65], ink-jet printing[8] or solution-shearing[66].

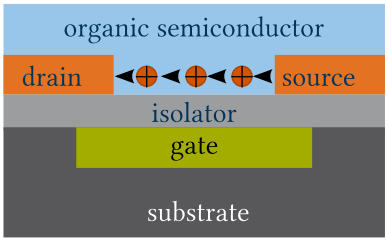


Figure 2.3: *Organic field effect transistor (OFET) in thin film transistor design.*

(a) *bottom contact, bottom gate (BC/BG)*

(b) *bottom contact, top gate (BC/TG)*

(c) *top contact, bottom gate (TC/BG)*

Figure 2.4: *Three examples of common OFET architectures with different positions of source, drain and gate electrodes.*

Correct extraction of the field effect mobility is crucial to assess the suitability of a material for OFET applications. Especially high mobility devices show non-ideal transfer and output charac-

teristics, complicating the determination of the mobility[19]. The mobility of a device with near-ideal FET characteristics can be extracted from two different regions in a I/V-measurement, namely the linear or the saturated region (Fig. 2.5)[7, 19, 67]. All OFET architectures have a

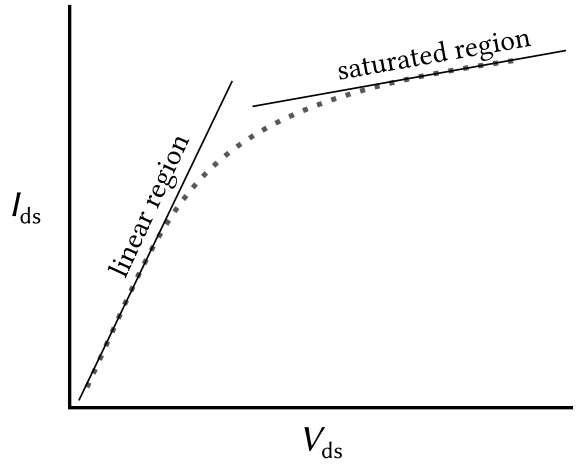


***Figure** 2.5: Schematic curve of an OFET I/V measurement with linear and saturated regions indicated in the graph. With source-drain voltage V_{ds} and current I_{ds}.*

characteristic channel length L, channel width W and the capacitance per unit area of the gate dielectric C_i. The threshold voltage V_T accounts for induced charges that do not contribute to the carrier mobility, for example due to deep traps being filled up before charges can move[64]. The effective gate voltage V_g is then reduced by V_T, $V_\mathrm{g}^\mathrm{eff} = V_\mathrm{g} - V_\mathrm{T}$. For small source–drain voltages V_ds ($V_\mathrm{ds} \ll V_\mathrm{g}^\mathrm{eff}$), a linear charge density gradient between source and drain electrodes is formed and the drain current I_d increases linearly. This is called the linear regime and the mobility can be obtained via

$$I_\mathrm{d}^\mathrm{lin} = \mu \frac{W}{L} C_\mathrm{i} \left(V_\mathrm{g} - V_\mathrm{T}\right) V_\mathrm{ds}, \tag{2.1}$$

With increasing V_ds a depletion region next to the drain electrode is formed. This leads to a space–charge-limited saturation current which will not increase further with increasing V_ds. In this so-called saturation regime ($V_\mathrm{ds} \geq V_\mathrm{g}^\mathrm{eff}$) the mobility can be extracted using a different approach,

$$I_\mathrm{d}^\mathrm{sat} = \mu \frac{W}{L} C_\mathrm{i} \left(V_\mathrm{g} - V_\mathrm{T}\right)^2. \tag{2.2}$$

With both regimes different experimental issues can arise that mask the true mobility and can lead to over- or underestimated mobilities[19]. Although there are many other methods to measure the carrier mobility in experimental devices (e.g. time of flight, carrier extraction by linearly increasing voltage, double injection, or impedance spectroscopy), the OFET technique is widely used due to its comparably simple experimental setup. Potentially superior approaches such as time-of-flight measurements[67], which can give true bulk and field-dependent mobilities only account for a small fraction of published organic semiconductor mobilities. The high demands on the experimental setup for OFET mobilities can obfuscate the intrinsic carrier mobility of molecular crystals, thus making independent and efficient theoretical calculations of carrier mobilities indispensable to reliably compare many different materials.

3 Theoretical description of charge carrier mobility

As has been pointed out in the previous chapter, the charge carrier mobility μ is a crucial parameter to understand and improve organic semiconductor materials. Yet, due to discrepancies that will be explained in greater detail in this chapter, the theoretical description and calculation of the carrier mobility differs from that of the inorganic semiconductors and no single model to date incorporates all effects. The following sections summarize different models with a focus on commonalities. For detailed explanations the following excellent review articles are recommended: [37, 39, 68].

Within organic semiconductors one can differentiate between highly ordered (small molecule) single crystals and highly crystalline thin films on the one hand, and disordered amorphous polymers on the other. The latter can be described by thermally activated hopping processes, where disorder plays a major role in carrier transport. A description of the localized states in terms of a Gaussian distribution was first suggested by Bässler *et al.*[40], with substantial improvements to this model over the last decade[69–71]. These models are successfully employed to explain and understand time-of-flight measurements of disordered organic semiconductors[64].

In highly ordered molecular systems such as single crystals or polycrystalline thin-films different phenomena can be observed, complicating the overall picture. A starting point for the theoretical description is the temperature dependence of μ. In early studies, the carrier mobility in organic semiconductors was shown to increase with increasing temperature, which ruled out classical band transport theories established for inorganic semiconductors. In the last two decades new experimental results challenged this interpretation of the processes in organic semiconductors[37]. With improved preparation techniques ultrapure organic semiconductor single crystals[72, 73] became available and subsequent temperature dependent mobility measurements showed characteristics of band-like transport, i.e. increasing mobility with decreasing temperature[74–76]. In addition, also in polycrystalline thin-film materials band-like mobility characteristics could be found[77]. These findings and the rising industrial interest in organic semiconductors caused many groups to work on different models to correctly describe and understand the processes underlying the carrier transport. Based on ultrapure samples it was concluded that the observed small polaron hopping of most materials is caused by defects[37], which effectively mask the intrinsic mobility of the crystals. Before looking into the different models in more detail it should be pointed out that while an accurate, quantitative description of the charge carrier mobility valid for all organic semiconductors under different conditions is still missing, the existing knowledge already allows to establish qualitative mobility rankings between different materials.

3.1 Carrier mobility μ

In the general definition of μ the velocity response of a charge carrier to an external electric field is defined as

$$\mu_{ij} = \frac{\langle v \rangle_i}{E_j}, \tag{3.1}$$

with $\langle v \rangle_i$ being the time-averaged velocity of the charge and E_j being a component of the electric field. The linearised form of Eq. (3.1),

$$\mu_{ij} = \frac{\partial \langle v \rangle_i}{\partial E_j}, \tag{3.2}$$

is often used when comparing calculations with experimental values. Conceptionally, this is similar to experimental mobility measurements using the time-of-flight technique[67]. In computational studies μ is often calculated using the Einstein–Smoluchowski equation, which relates the mobility to the diffusion coefficient D,

$$\mu_{ij} = \frac{D_{ij}(\mathbf{E})q}{k_\mathrm{B}T}, \tag{3.3}$$

with the charge q. In both formulations, the final mobilities depend on the employed mobility model, either via $\langle v \rangle_i$ or D_{ij}.

3.2 The hopping regime

Supported by the early experimental evidence, the existence of localised charge carriers which move between sites in a crystal in a series of discrete jumps or "hops" quickly became accepted[37]. In fact, hopping models are still widely used[41, 78–83], despite the recent evidence

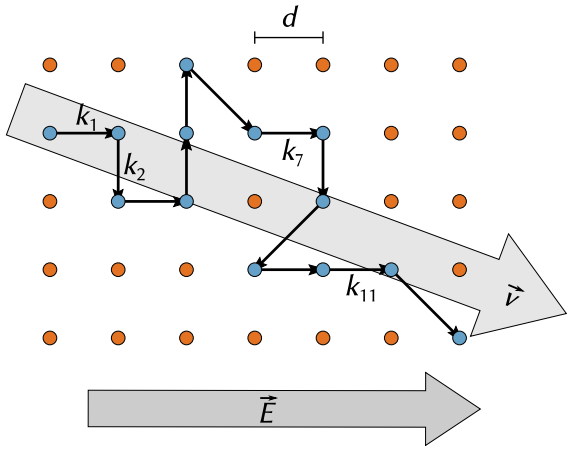


***Figure 3.1**: Schematic depiction of the hopping regime. A charge is localised on a single crystal site, with distance d between sites. The rate k for jumps between distinct sites is proportional to the distance d, the transfer integral H_{ab} and the reorganisation energy λ. The sum of all distinct hops determines the carrier mobility of the system.*

questioning the existence of localised polarons in many organic semiconductors[37, 39, 84]. It is assumed that carriers localise either through their interaction with the surrounding medium or due to thermal fluctuations in the weakly bound molecular crystals, which induce fluctuations in the transfer integrals and annihilate the translational symmetry of the electronic Hamiltonian[84, 85]. Between each site a charge transfer rate k_l exists. The carrier mobility in Eq. (3.1) then becomes

$$\mu_{ij} = \sum_l d_{li} \frac{k_l}{E_j}, \tag{3.4}$$

where d_{li} is the hopping distance (cf. Fig. 3.1). In the most general form based on transition state theory, the charge-transfer-rate is

$$k_{\mathrm{TS}} = \nu_{\mathrm{eff}} \kappa_{\mathrm{el}} \Gamma_{\mathrm{n}} e^{-\beta(\Delta G^{\ddagger} - \Delta)}, \tag{3.5}$$

with the effective vibrational frequency ν_{eff}, the electronic transmission coefficient κ_{el}, the nuclear tunneling factor Γ_n, the diabatic activation energy $\Delta G^{\ddagger}$ and the adiabatic correction factor Δ. By applying semi-classical Landau–Zener theory and looking at the non-adiabatic limit, Eq. (3.5) becomes the famous Marcus rate equation[45]:

$$k_{\mathrm{na}} = \frac{2\pi}{\hbar} \frac{1}{\sqrt{4\pi\lambda k_{\mathrm{B}} T}} |H_{ab}|^2 e^{-\beta \Delta G^{\ddagger}}, \tag{3.6}$$

where

$$\Delta G^{\ddagger} = \frac{(\lambda + \Delta G^0)^2}{4\lambda}, \tag{3.7}$$

with the driving force ΔG^0, the transfer integral H_{ab} and reorganisation energy λ. This is also visualized in Fig. 3.2. The reorganisation energy λ can also be formulated in terms of the local electron–phonon coupling[37],

$$\lambda = 2g^2 \hbar\omega, \tag{3.8}$$

with g representing the strength of the coupling, ω the frequency of a single nuclear mode of the system and $\hbar$ the reduced Planck constant. More details on the parameters H_{ab} and λ, as well as methods to calculate them are given in chapters 4 and 5.

3.3 Band transport regime

On the other side of the spectrum are the band theory models, which are based on fully delocalised charge carriers. In a molecular crystal, the individual orbitals of each molecule form bands that are described by a linear combination thereof. In the case of charge transfer, where for example an electron is removed from the material to form a hole, the hole wavefunction can still be described by the linear combination of HOMO orbitals of the individual molecules. In a simplified (but generalisable) picture[37] of an 1D chain of molecules with a state $|j\rangle$ each, the electronic Hamiltonian can be written as

$$H_{\mathrm{el}} = \sum_j \left(\epsilon |j\rangle\langle j| + H_{ab} |j\rangle\langle j+1| \right), \tag{3.9}$$

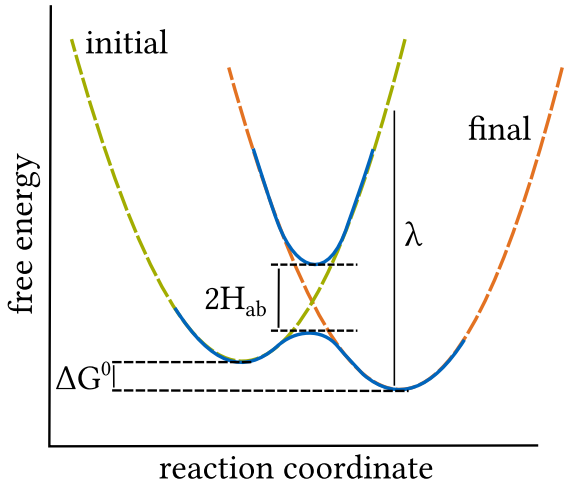


Figure 3.2: *Schematic of two diabatic states (dashed line) compared to the adiabatic picture (continous line). Charge transfer parameters for polaronic hopping H_{ab}, λ and ΔG^0 are indicated in the scheme.*

with the eigenenergy ϵ and again the transfer integral H_{ab}. The mobility μ for such a delocalised charge wavefunction can be described quantitatively by the semiclassical Boltzmann equation[86],

$$\mu = \frac{et_S}{m^*}, \tag{3.10}$$

where m^* is the effective mass of the charge carrier, e is the electron charge and t_S is the relaxation time:

$$m^* = \frac{\hbar^2}{2|H_{ab}|d^2}, \tag{3.11}$$

with the distance between sites d and the transfer integral H_{ab}. The relaxation time t_S is the average time between scattering events in the system which disturb the delocalised wavefunction, therefore limiting the mobility. Scattering events or collisions occur with phononic lattice vibrations as well as with defects and impurities in the material.

3.4 Polaronic band theory

Band theory assumes fully delocalised charge carriers, therefore neglecting any polarisation effects of the charges on the nuclei. For weakly bound molecular solids, this approximation does not necessarily hold. A more realistic approach for charge carriers in organic solids was developed by Holstein in 1959[87], describing an intermediate region between fully delocalized bands and localised small polarons (cf. Fig. 3.3). Again, following [37], a single harmonic nuclear mode ω per molecule, can be described by the Hamiltonian,

$$H_{\text{nucl}} = \sum_j \frac{\hbar\omega}{2}(q_j^2 + p_j^2), \tag{3.12}$$

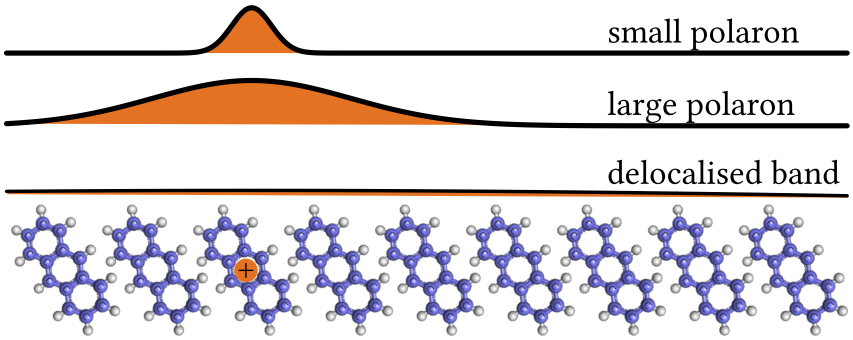


***Figure 3.3**: Schematic representation of various degrees of charge carrier localisation. Top, small polaron with charge localised on a single molecule; middle, large polaron model with charge delocalised over a finite number of molecules and bottom, fully delocalised band.*

with q_j and p_j being the adimensional displacement and momentum on the molecule j. While there is no interaction between electronic and nuclear degrees of freedom, the nuclear positions of a molecule change when charged. This leads to the *local electron–phonon* coupling term,

$$H_{\text{el–nucl}} = g\hbar\omega \sum_j q_j |j\rangle\langle j|, \tag{3.13}$$

with the coupling strength g. It is possible to reformulate the polaronic Hamiltonian similar to the ansatz in Eq. (3.9),

$$H_{\text{polaron}} = \sum_j (\epsilon' |j\rangle\langle j| + H'_{ab} |j\rangle\langle j+1|). \tag{3.14}$$

The on-site energy ϵ' and transfer integral H'_{ab} now differ from those above, with H'_{ab} having an additional dependency on the temperature, via

$$H'_{ab} = H_{ab} e^{-\frac{1}{2}g^2(N_\omega + \frac{1}{2})}, \tag{3.15}$$

with $N_\omega = (e^{\hbar\omega/k_\text{B}T} - 1)^{-1}$ and the Boltzmann constant k_B. The carrier mobility in the polaronic band theory is then calculated in the same manner as in the pure band theory.

3.5 Summary of the presented mobility models

The three mobility models presented in the previous sections represent the boundaries within which many other variants and improved models are developed. In addition, various *direct propagation* schemes exist, which avoid the intricate question of employed approximations altogether by directly propagating the charge in a dynamical simulation[88], albeit at the cost of additional approximations to make them computationally tractable. In the context of this thesis it is critical to note that all of the presented models and schemes share certain physical parameters.

While different models may apply for certain regimes, temperature regions and materials, the transfer integral H_{ab} can be found in all equations. A high transfer integral, or electronic

coupling between molecules in the molecular crystal, is a fundamental requirement for high mobility organic semiconductors.

The second important parameter is the reorganisation energy, which is closely related to the electron–phonon coupling of the crystal. Its significance is directly tangible for small polaron hopping models, such as the famous and widely used Marcus rate equation. Via the electron–phonon coupling strength g (Eq. (3.13)) there is also a connection between the calculated reorganisation energy λ (see chapter 5 for details) and the local electron–phonon coupling term in the polaronic band model.

While these two parameters alone do not allow to calculate quantitative mobilities, they provide valuable insight into the relative ranking of charge carrier mobilities for different materials. Put into the greater context of this thesis, knowledge of both H_{ab} and λ allows to decide whether a certain molecular crystal is expected to show high intrinsic charge carrier mobility or not. For this reason, we developed accurate and efficient methods to calculate both parameters.

4 The transfer integral H_{ab}

In understanding and describing charge transfer reactions, diabatic states are an important theoretical concept used in assessing charge-transfer parameters. Unfortunately, the generally accessible states in most electronic structure methods are the adiabatic ones[45, 89–91]. The latter are also called Born–Openheimer states and are always eigenstates of the electronic Hamiltonian of the system. The so-called non-adiabatic coupling, $\tau_{ab}(\mathbf{R})$ between two adiabatic states $|\Phi_{a,b}\rangle$,

$$\tau_{ab}(\mathbf{R}) = \langle \Phi_a(\mathbf{R}) | \frac{\partial}{\partial \mathbf{R}} \Phi_b \rangle, \tag{4.1}$$

determines the rate of transition between electronic states. While this equation allows the calculation of charge transfer rates, the appearing wave-function derivatives are difficult to handle. A basis representation constructed in a way that the non-adiabatic couplings vanish is called diabatic representation. In a diabatic basis, the transfer integral H_{ab} determines this transition rate. Between two diabatic states, $|\Psi_{a,b}\rangle$, the transfer integral is given by

$$H_{ab} = \langle \Psi_a | \hat{\mathcal{H}} | \Psi_b \rangle. \tag{4.2}$$

Unfortunately, one can show that creating a strictly diabatic basis from an adiabatic one is not possible[89]. Therefore, different strategies have been developed to construct approximate diabatic states. According to van Voorhis, one can distinguish between deductive strategies such as the Generalized Mulliken–Hush (GMH)[92–95] or the Block-Diagonalization (BD)[96, 97] method on the one hand and constructive strategies such as Constrained Density Functional Theory (CDFT)[98–100] or the Fragment Molecular Orbital family of methods[83, 101, 102] on the other. In charge-transfer calculations using DFT the construction of the diabatic states is further complicated by the electron-delocalisation error of DFT[103], which contradicts the target of constructing charge-localised states.

Together with very accurate wave-function based *ab initio* methods such as complete active space self-consistent field (CASSCF) and others, the GMH approach is often used to provide accurate benchmark-data for small systems. Using this technique, Kubas *et al.* calculated the transfer integrals for two sets of small molecules (cf. Fig. 4.1) for both hole- and electron transfer (the HAB11 and HAB7 testsets, respectively)[104, 105]. These were then used to benchmark and compare different computational methods to calculate transfer integrals.

Fragment Orbital DFT (FO-DFT)

This benchmark data was used to investigate the accuracy of the popular fragment orbital DFT scheme, among other methods. Due to its simplicity and ease of implementation in modern electronic structure codes (not limited to DFT) this method is one of the most popular to calculate

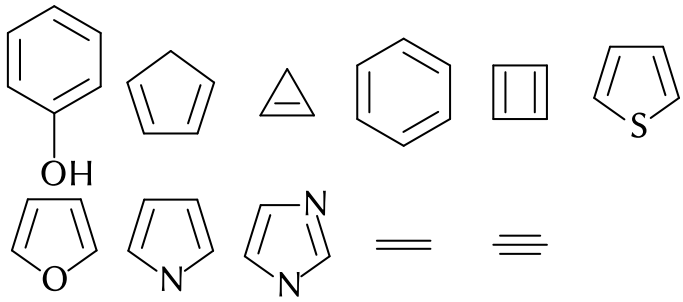


Figure 4.1: *Composition of the hole transfer test-set HAB11. High-level ab initio reference data is available for all systems[104].*

transfer integrals in a variety of different studies[41, 47, 79, 83, 101, 106–108]. In the following overview of the method (for a detailed discussion see section 8.1 and our related paper[102]) a hole transfer between a donor D^+ and an acceptor A is assumed for didactic reasons, with $n-1$ and n electrons, respectively. All three FO-DFT variants are based upon the approximation that only the frontier orbitals directly involved in the charge transfer process between initial and final diabatic state change. The charge-localised states are constructed by separating the calculation of the donor and acceptor molecule, effectively performing two independent calculations. This

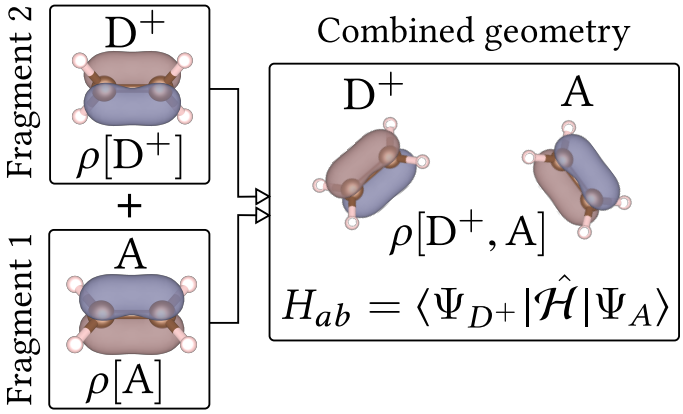


Figure 4.2: *Schematic illustration of the fragment orbital DFT scheme for hole transfer in ethylene. Both isolated ground-state densities ρ_{A,D^+} are calculated separately and then used to construct the approximate diabatic state $\Psi_{a,b}$. The FO-DFT scheme shown here is the $\mathcal{H}^{2n-1}$@D^+A variant[102].*

yields the Kohn–Sham densities ($\rho_{\mathrm{D}^+}, \rho_{\mathrm{A}}$) for each fragment, which are then combined to form the target diabatic state ($\rho_{\mathrm{D}^+\mathrm{A}}$). For this density the Kohn–Sham-Hamiltonian is calculated in a non-selfconsistent approach, preserving the constructed diabaticity. As a direct consequence of the approximations, the calculation of the transfer integral simplifies from Eq. (4.2) to

$$H_{ab} \approx \langle \phi_A^n | \hat{h}^{\mathrm{KS}} | \phi_D^n \rangle, \tag{4.3}$$

with the single-particle Kohn–Sham Hamiltonian $\hat{h}$ and ϕ_{A,D^+} being the Kohn–Sham-orbitals of the individual donor and acceptor fragments. In Fig. 4.2 the whole process is illustrated for an ethylene dimer. In the literature, two different flavours of FO-DFT were already known[83, 101], and a third variant with improved accuracy was developed in this thesis.

$\mathcal{H}^{2n}@DA$: The original approach by Senthilkumar *et al.*[101] uses neutral fragments for both donor and acceptor, no matter what the actual charge state is. This results in a wrong number of electrons in the dimer Hamiltonian, as compared to the true diabatic description of the system.

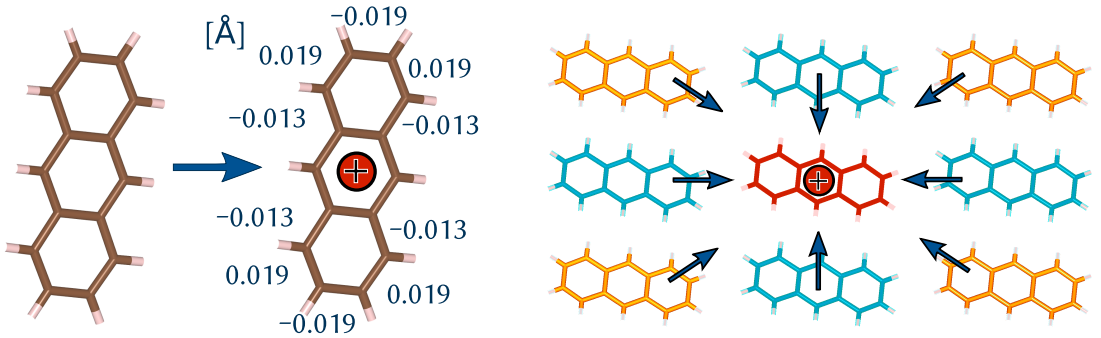
$\mathcal{H}^{2n-1}@DA$: A second version by Oberhofer and Blumberger [83] resolves this issue by resetting the occupation of the ϕ_{D}^{n} orbital in the second step of the FO-DFT calculation. This restores the correct number of electrons in the Hamiltonian.

$\mathcal{H}^{2n-1}@D^{+}A$: By taking a different route to approximate the diabatic state within FO-DFT we developed a third variant. Here, we explicitly calculate charged fragments and therefore construct the dimer with the correct number of electrons in the Hamiltonian. The greatly improved accuracy of this method comes at a small additional computational cost for calculating the charged fragment separately.

We implemented all three variants in the FHIaims all-electron DFT program[109], which allowed us to perform a systematic study of the accuracy of all different flavours, ruling out any influence of different implementations or technical settings. In addition to comparing the different flavours, we also investigated the effect of the different approximations on the calculated transfer integrals, such as neglecting interactions between the fragments or the effect of exact exchange on the Hamiltonian of the dimer state. Our comprehensive and systematic approach enabled us to provide decision guidelines on the usage of the different approximations and flavours of FO-DFT, also taking into account the computational efficiency of different methods. More details can be found in the corresponding publication[102] and the summary thereof in section 8.1. This was then directly utilised to compute a large number of transfer integrals in our computational high-throughput screening for high mobility organic semiconductor materials, as published in [110] (summary in section 8.2).

5 The reorganisation energy λ

In polaron hopping theories the reorganisation energy describes the electrostatic response of the system to changes in the charge state. In the context of charge transfer theories this usually is a charge transfer between two sites in the crystal. Due to the long range nature of the electrostatic interaction ($F_{\mathrm{el}} \propto r^{-2}$), *ab initio* calculations of the full reorganisation energy λ require appropriately sized periodic unit cells. While todays electronic structure codes and the available computing power in principle allow such calculations for small to medium sized molecules, the practical costs are too high. To solve this problem, λ is usually split into two separate contributions, the short-range *inner-sphere* (or *internal*) and the long range *outer-sphere* (or *external*) contribution, $\lambda = \lambda_{\mathrm{int}} + \lambda_{\mathrm{ext}}$. This is also shown in Fig. 5.1 for the example of a molecular crystal. The external contribution λ_{ext} can then be calculated using for example the Marcus formula[45],



(a) λ_{int} ***(b)*** λ_{ext}

Figure 5.1: *Exemplary representation of the geometry changes in an antracen crystal in the presence of a charge. (a) Changes in bond length shown by an individual antracen molecule when calculating with and without charge (intramolecular part, λ_{int}) (b) Response of the whole crystal environment to a charge (intermolecular part, λ_{ext}).*

$$\lambda_{\mathrm{ext}} = (\Delta q)^2 \left[\frac{1}{2r_{\mathrm{D}}} + \frac{1}{2r_{\mathrm{A}}} - \frac{1}{R}\right]\left[\frac{1}{\epsilon_{\mathrm{op}}} - \frac{1}{\epsilon_{\mathrm{s}}}\right], \tag{5.1}$$

with the transferred charge q, the distance between the donor and acceptor R, the effective radii of donor and acceptor $r_{\mathrm{D,A}}$ and the optical (ϵ_{op}), respectively static (ϵ_{s}) dielectric constant of the medium. In most applications of polaron hopping theories, the external contribution is neglected[111]. This approximation is motivated by the size of λ_{ext}, which has been shown to be vanishingly small compared to λ_{int}[80, 111, 112].

In molecular crystals the internal contribution λ_{int} is calculated as the energy difference between the initial (a) and final (b) diabatic state of a single molecule, $\lambda_{\mathrm{int}} = E_b(\mathbf{R}_b) - E_a(\mathbf{R}_a)$. The

standard procedure to approximate this contribution is called *4-point-scheme*, which separates the contributions of the donor and acceptor (shown here for hole transfer):

$$\lambda_{\text{int}} = \left[E_D(\mathbf{R}_D^+) + E_A^+(\mathbf{R}_A)\right] - \left[E_D^+(\mathbf{R}_D^+) + E_A(\mathbf{R}_A)\right]. \tag{5.2}$$

This equation can also be interpreted in terms of the polaron binding energy, E_{pol}, stabilizing the charge localized on a single lattice site and creating a connection to band theories via the local electron–phonon-coupling[113]. In order to calculate a single λ_{int}, four different energies and two different geometries need to be obtained. The ground state energy on the optimal geometry of the uncharged molecule ($E[\mathbf{R}]$), the energy of the charged system with the optimal charged geometry ($E^+[\mathbf{R}^+]$) and two energies representing the vertical transitions in the transfer reactions ($E^+[\mathbf{R}]$ and $E[\mathbf{R}^+]$). While this technique is routinely used in studies to calculate λ_{int}[41, 56, 78–83], it has been shown that having the correct solid-state molecular geometry is important to reproduce the relative carrier mobilities in molecular crystals[56]. It has been known for a long time that the optimal gasphase geometry of organic molecules can differ from their optimal geometry in the crystalline phase, with additional constraints by the packing requirements in the solid state (cf. Fig. 5.2). Mas-Torrent *et al.* used manual constraints in their study[56] to retain

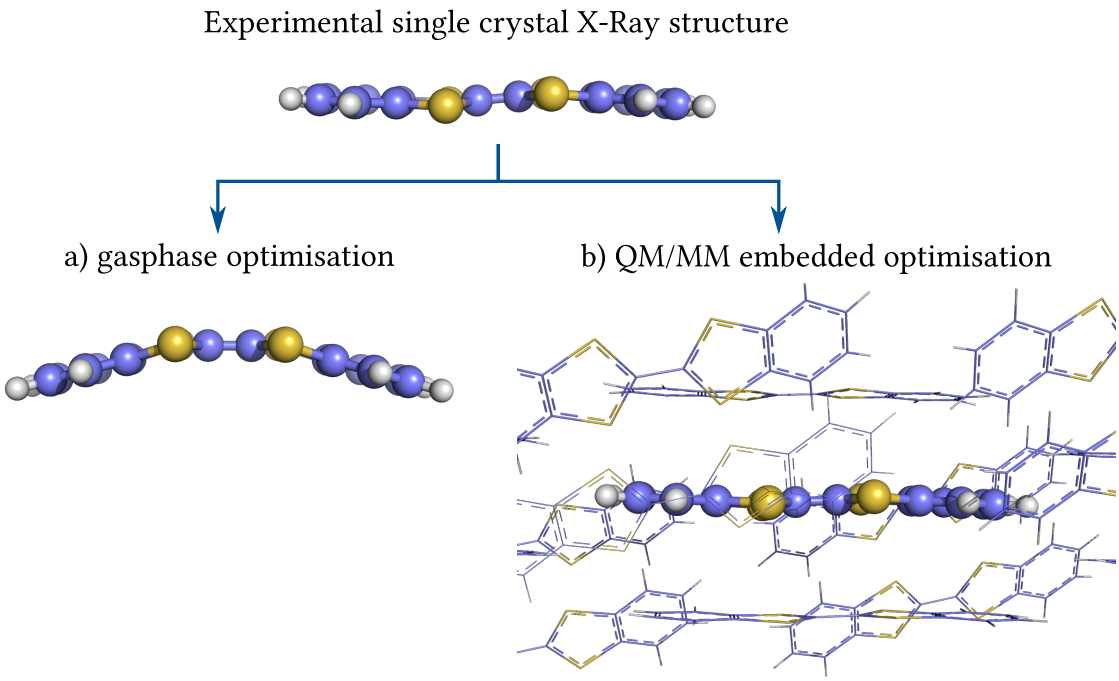


Figure 5.2: *Different final lowest-energy conformations for an identical molecule optimised in (a) standard isolated gasphase approach, and (b) the QM/MM vdW-embedding scheme for a tetrathiofulvalene derivative using the PBE GGA-functional. Gasphase optimisation yields a bent structure which differs from the solid-state structure.*

the planarity of the solid state environment. For a high-throughput screening (HTS) study with thousands of molecules as carried out in this thesis such manual techniques cannot be used.

To this end we developed a variant of the efficient QM/MM scheme[114, 115] specifically tailored for molecular crystals[110]. We use a single shell of neighbouring molecules to mimic the solid state environment during the geometry optimisation of a single (central) molecule, with the neighbouring molecules (the MM region) being constraint to their experimental coordinates during the whole optimisation. As our target is to re-optimise the experimental structure on the chosen DFT level, changes to the geometry are small and neglecting these changes in the neighbouring molecules does not hamper their ability to provide the correct constraints for the geometry optimisation. The developed methodology played a critical role in the high-throughput screening study[110] (section 8.2).

6 High-throughput (computational) screening

The concept of high-throughput screening (HTS) was mainly established in drug design and pharmacological research[116]. To discover new drug molecules, chemists synthesize and test as many different molecules as possible. Due to the large and diverse chemical space it is hard to decide which molecule to test next, and testing all of them is impossible. To aid this dilemma, automated synthesis and testing capabilities were developed. These are based on fluorescence measurements and microtitre plates with up to 384 wells per plate[116], increasing the number of tested compounds per chemist per year from about 100 to more than 1×10^6[49, 118]. In a staged approach, compound libraries are tested for hits, which undergo various filters and tests before being selected as a lead structure and being considered for detailed investigation. The amount of hits an empirical HTS produced is relatively low, with many inactive compounds being tested and discarded in each run[119]. Nevertheless, empirical HTS screening is one of the cornerstones of modern drug discovery. Substantial work is also invested in the design of compound libraries, with many companies offering curated libraries with selected classes of molecules.

Figure 6.1: *Experimental High-Throughput Screening setup using multiple robotic screening platforms[117].*

With the substantial progress in computational chemistry, making calculations of relevant pharmacological molecules feasible, computer aided drug design (CADD) was established in 1981[119]. Initially, great hopes had been placed into CADD, but with the rapid progress in empirical HTS virtual screening methods had been pushed into the background. Only in the last decade computational methods re-emerged as potential tool to reduce the number of compounds to be screened in HTS approaches and focus on more active compounds as indicated by the computational assessment[49, 119]. The basic working principle is the same as in experimental HTS, but laboratory experiments are replaced by computational experiments and calculations, respectively. To assess the eligibility of a tested compound, a theoretical model for the target property is necessary. To allow fast testing and high throughput, model parameters should be computationally inexpensive, without being meaningless. This is achieved with similarity based searches, predicting biological activity using statistical models (QSAR), pharmacophore mapping or virtual docking studies.

In recent years, these virtual screening techniques have been increasingly adapted for the search of novel materials in other fields. This process was somewhat slowed by the complexity of the underlying theoretical descriptions of material properties in computational material science as compared to drug design, but recent successes showed the prospects of overcoming these difficulties[53, 54, 120, 121]. With the *Harvard Clean Energy Project (CEP)* a large-scale project was initiated in 2008, aiming at the discovery of novel organic photovoltaic materials using computational methods[52]. An extensive molecular compound library was generated for this project using combinatorial approaches, yielding about 1×10^7 distinct structures. Molecular descriptors were chosen, allowing calculations using the generated molecular motifs in the compound library. The disadvantage of such artificial compound libraries is the lack of information on the stability and synthesis of generated compounds. In the context of the CEP, Sokolov *et al.*[47] evaded this problem in a small computational screening study aiming for high mobility organic semiconductors by modifying a single, well-known parent backbone dinaphtho-thienothiophene (DNTT)[48]. They chose the reorganisation energy λ_{int} as their primary descriptor and calculated 7 different derivatives of their parent structure. For selected structures, they calculated the crystal structure under symmetry considerations related to the parent molecule and used these for the determination of transfer integrals and relative charge carrier mobilities. This proof-of-concept study successfully identified a compound with high expected mobility, which was also experimentally verified.

In this thesis we choose to develop a general, unbiased screening workflow to identify high mobility semiconductors not only based on prior knowledge of suitable backbones, but by scanning a diverse subset of the full chemical space. According to the illustration in the previous chapters our initial target descriptors are the transfer integral H_{ab} and the intramolecular reorganisation energy λ_{int}. While the latter is – in principle – a molecular quantity, we developed a QM/MM embedding scheme (cf. chapter 5) to reproduce the solid state environment and ensure the comparability of all calculated reorganisation energies. Therefore, both descriptors are dependent on the knowledge of the solid state structure, and our compound library needs to reflect this. While crystal structure prediction is making huge progress (see chapter 7 for details and future opportunities), we decided to first make use of already existing knowledge. Synthetic organic chemists manufactured and characterised a plethora of organic structures over the last century, with a majority not being considered as organic semiconductor, but for completely different purposes. These structures are collected in the *Cambridge Structural Database*[57], with more than 811 138 structures available today[1]. All of them are experimentally known structures, and for more than 94 % of the structures single-crystal X-Ray data is available. In addition to providing us with the information we need for the calculation of our *ab initio*-based descriptors, this approach also gives us additional, high quality meta data such as the synthesis route, the stability or existing applications. By combining this data with the availability reports from the freely accessible ZINC15-database[122], we can easily identify commercially available compounds. We extracted our initial selection from the CSD Database using the ConQuest interface[123] (version 1.17). To reduce the number of poor-quality structures we chose a refinement value of $r \leq 0.05$. In addition, we also excluded structures with metallo-organic components, more than one different molecule per crystal, missing 3D coordinates, known structural disorder, structures containing ions or polymeric components. The resulting collection of 95 445 crystals constitutes our initial compound library.

[1] CSD Summary Statistic, 6. January 2016

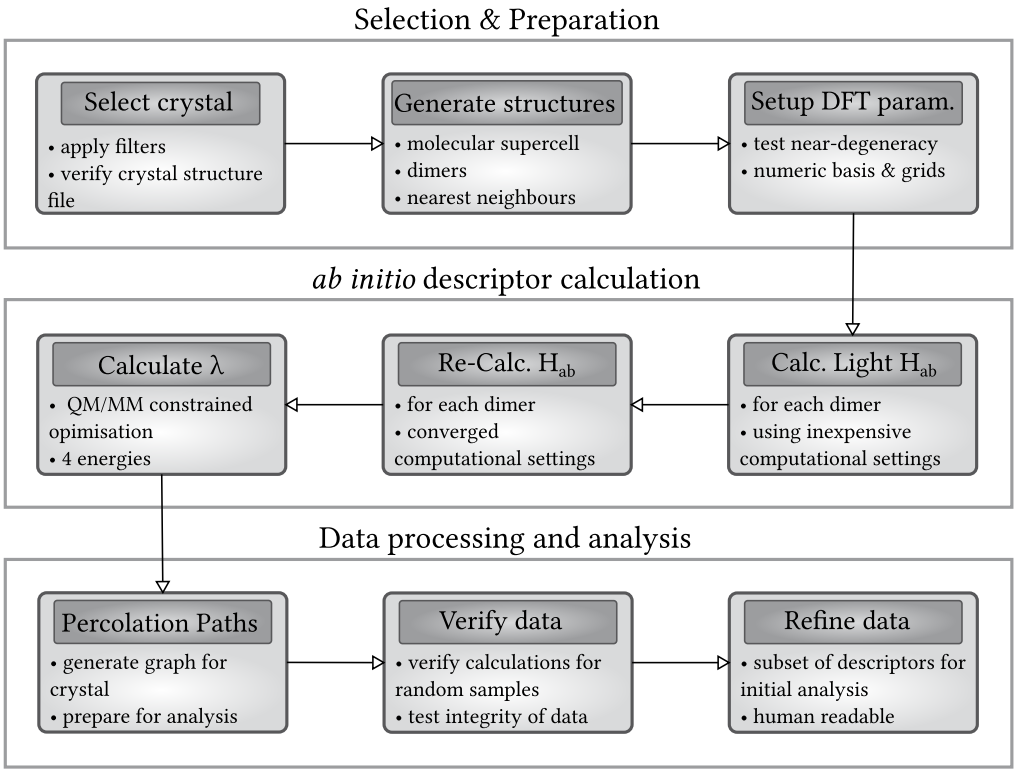


Figure 6.2: *Simplified representation of the developed dynamic screening workflow for organic semiconductor materials based on the two ab initio descriptors H_{ab} and λ. Each task (Firework) consists of multiple atomic steps (FireTasks), efficiently implementing the preparation or calculation of a single quantity. If a step fails, the workflow automatically adjusts and the system in question is either removed or the failed task repeated with a different method.*

Suitable descriptors and the compound library are not the only prerequisites for a successfull high-throughput screening study. For each system, a number of preparation steps, verifications and calculations have to be performed in an automated fashion, without disrupting the overall screening procedure if individual steps fail. Ideally, error identification and handling are included without user interaction. In our setup this was achieved by using the *FireWorks* workflow management framework[124]. This framework allows the definition of *workflows*, which consist of different so-called *Fireworks*, each of which again can be composed of *FireTasks*. The latter are individual, atomic calculation steps. The high-level workflow developed for this HTS study is shown in Fig. 6.2, with the individual *FireTasks* for an exemplary calculation step detailed in Fig. 6.3. The large number of molecular crystals and the necessary *ab initio* calculations for our descriptors result in more than a million individual steps, with each step producing miscellaneous data. In order to cope with the huge amount of generated data, a suitable data collection and post-processing scheme has been developed. The *FireWorks* framework uses MongoDB[125] as data storage engine. We adopted this engine for our primary data, storing information on each individual calculation step, resources and results. To facilitate statistical data analysis we also created a SQL-based secondary database[126] with processed and accumulated information on a per-crystal basis. This rich collection of data is the primary result of our high-throughput screening study. *Ab initio* transfer integrals are available for each dimer

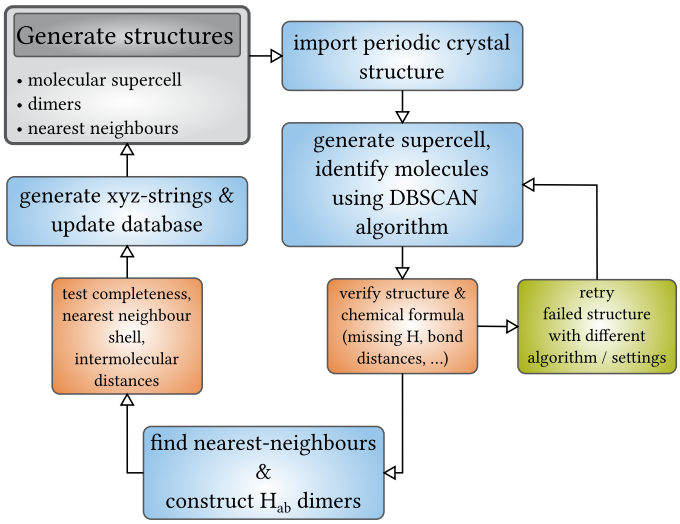
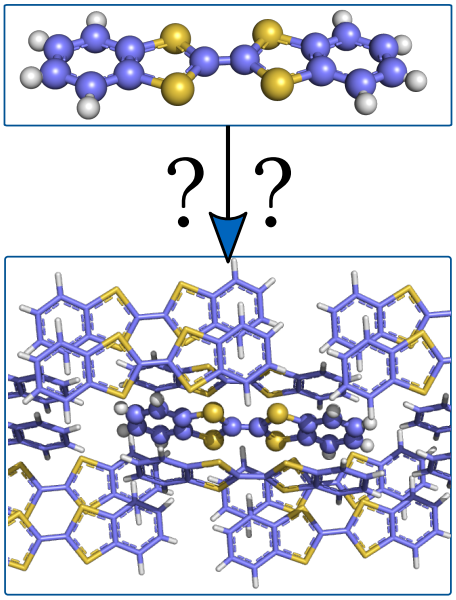


***Figure 6.3**: Worktasks (FireTasks) assigned to the GENERATE STRUCTURES step of the overview workflow shown in Fig. 6.2. The different tasks depicted in this scheme are carried out dynamically, with various tests and checks ensuring only correct structures are processed further. Problematic structures are first reevaluated with different settings/algorithms and ultimately discarded from the database, if they still fail.*

in a nearest neighbour shell of 64 729 molecular crystals, with 13 180 systems having at least one transfer integral above 50 meV. For 10 215 of these structures, the intramolecular reorganisation energy λ_{int} has been calculated using our robust QM/MM scheme. The DFT GGA energies of molecular HOMO and LUMO orbitals are also available. In addition, for all systems above the threshold of 50 meV, the commercial availability of the components was evaluated. Of those, 2815 molecules are listed as *in-stock* or *for-sale*, making them ideal candidates for experimental testing of the charge carrier mobility. Our collected data also allows us to determine the charge percolation pathways in each crystal, therefore testing another important requirement for real high mobility materials[43]. We analysed our screening results and selected 4 novel organic semiconductors for which we expect a high carrier mobility according to our descriptors. These systems are presented in [110] (section 8.2).

7 Crystal structure prediction for molecular solids

True *in silico* materials design is only possible when all necessary information can be obtained using computational methods. As we pointed out before, one remaining challenge in the context of molecular solids is the knowledge of the crystal structure of the material. Already in 1988 John Maddox wrote in an article for *Nature*, not being able to predict the crystal structure only from the knowledge of the molecular composition is "one of the continuing scandals in physical science [...]"[127]. Since then much has happened, but the correct prediction of crystal structures is still an open and active field of research[128, 129]. The Cambridge Crystallographic Data Centre initiated its first *Crystal Structure Prediction Blind Test* (CSP) in 1999[130], an international and collobarative effort to test and present the state-of-the-art of crystal structure prediction. In this blind test hithertho unpublished crystal structures are targeted by different research groups, each group trying to calculate the correct structure solely based on the knowledge of the chemical composition of each molecule. In 2014 the sixth CSP blind test was launched, in which we participated[131]. The reliable prediction of crystal structures is part of our efforts to predict new high mobility organic semiconductors by designing suitable molecular crystals. A challenge in predicting the correct crystal structure for a molecular solid is the existence of polymorphs. Depending on the symmetry of the unit cell, the same molecule can form different crystal structures, with small energy differences between each structure. The thermodynamically most stable structure is then considered to be the correct one[132]. This has two implications: First, the employed computational method to establish the relative energy ranking between different polymorphs must be able to exactly reproduce even tiny energy differences between similar systems and second, systems crystallizing in a metastable polymorph defy this approach. For this reason, usally not only the most stable, but also other systems with energies close to the minimum are considered. For pharmacological applications the knowledge of such *hidden* polymorphs is essential[133].



To explore the configurational space of a given compound different structures need to be generated and ranked. The computational cost of the energy ranking method also determines the requirements for the structure generation method. Random structure search is only feasible for computationally inexpensive ranking methods such as empirical models or force fields[134], while for *ab initio* ranking methods structure generation needs to ensure to find enough low-energy structures without wasting resources on unstable polymorphs. In our contribution we used a genetic algorithm (GA) approach[135, 136] to efficiently sample the configurational space. In contrast to many other approaches (cf. Table 2 in [131]) we employed *ab initio* energy evaluations throughout our prediction workflow. This was made possible by our efficient implementation of the Harris approximation[137] within the FHIaims framework[109]. All initially generated structures were evaluated and ranked with the un-biased Harris approach. The work towards reliable prediction of crystal structures for acceptable computational cost carried out in this study[131] (summary in section 8.3) goes beyond the scope of our HTS study[110] and will enable entirely *in silico* studies in the future.

8 Publications

8.1 Critical analysis of fragment-orbital DFT schemes for the calculation of electronic coupling values

Summary of *Critical analysis of fragment-orbital DFT schemes for the calculation of electronic coupling values* by C. Schober, K. Reuter and H. Oberhofer, J. Chem. Phys. **144**, 054103 (2015).

In this work we investigate the popular fragment-orbital scheme for the calculation of electronic coupling values and critically assess possible variants or 'flavours' which arise from different approximations and implementations of the method. In the fragment-orbital schemes, the diabatic or charge-localized state of a system is constructed from the non-interacting fragment densities obtained via isolated calculations. Subsequent combination of such densities to the full system and calculation of the non-selfconsistent Hamiltonian allow the extraction of reliable electronic coupling elements for a broad range of systems. Approximations in both steps and combinations thereof result in three possible flavours of FO-DFT, $\mathcal{H}^{2n}@DA$ (F1), $\mathcal{H}^{2n-1}@DA/\mathcal{H}^{2n+1}@D^{-}A^{-}$ (F2) and $\mathcal{H}^{2n-1}@D^{+}A/\mathcal{H}^{2n+1}@D^{-}A$ (F3). The systematic description, including a consistent nomenclature for the first time, of these flavours together with methodical testing of the accuracy within one DFT framework is the main result of this work. In addition, common assumptions about the errors of this method such as neglecting interactions between fragments or the effect of hybrid functionals were thoroughly tested. Two datasets with high level *ab initio* reference data published by Kubas *et al.*, HAB11 (hole transfer) and HAB7 (electron transfer), were used to reliably assess the accuracy of our data. Using the previously known variants F1 and F2, we show that calculated H_{ab} are fully comparable between three different DFT frameworks, namely ADF, CPMD and FHIaims, despite their diverse technical foundations (Slater-type orbitals, plane waves and numeric atomic orbitals). After establishing the consistency with existing implementations, we focus on the effect of the different possible approximations and compare all three flavours within the same technical framework, therefore ruling out any influences of different implementations and technical settings such as the choice of the basis set. Our results show differences in the accuracy for hole and electron transfer. For hole transfer, our new method F3 performs best with a mean relative signed error (MRSE) of only −22.4%, as compared to −24.6% for F1 and −37.7% for F2. In the case of electron transfer, the original method F1 performs best (MRSE −22.4 %), closely followed by our new implementation F3 (−22.9 %) and F2 (−27.1 %). By using a potential embedding approach, we ruled out the neglect of polarization between the fragments as possible source of the error in H_{ab}. Using a novel approach entitled hybrid-GGA crossover FO-DFT we investigate the effect of exact exchange on the calculated H_{ab} by separating the effect of exact exchange on the self-consistent electron

density and the full systems' Hamiltonian. This allows us to show that the expected decrease in H_{ab} due to increased localization of the electron density is indeed observable, but overcompensated by an increase of H_{ab} due to the improved description of the systems' Hamiltonian in the second FO-DFT step. This improves the MRSE for hole (electron) transfer from −22.4 % to −7.3 % (−22.9 % to +11.8 %).

Individual contributions:

I implemented all possible variants of FO-DFT within the all-electron DFT-framework FHIaims and created a convenient and customizable Python package implementing a workflow to allow easy and semi-automatic calculation of electronic coupling elements, greatly reducing the necessary user interaction as compared to existing solutions. All calculations with the FHIaims code and subsequent data analysis were performed by myself. Dr. Harald Oberhofer edited the manuscript and Prof. Dr. Karsten Reuter proofread the final draft.

8.2 Virtual screening for high carrier mobility in organic semiconductors

Summary of *Virtual screening for high carrier mobility in organic semiconductors* by C. Schober, K. Reuter and H. Oberhofer, J. Phys. Chem. Lett. **19**, 3973 (2016).

In this publication we present our *in silico* high-throughput screening study to identify novel organic semiconductor materials. Our approach is based on two primary descriptors, the transfer integral H_{ab} and the internal reorganisation energy λ_{int}. Both descriptors are calculated using *ab initio* DFT, a novel approach which, to our best knowledge, was used for the first time for such a large number of molecular crystals. This was made possible by our previous work on method development for efficient calculation of H_{ab} and λ_{int}. To increase the value of our predicted materials we chose to use the *Cambridge Structural Database* (CSD) as our primary compound library. This is motivated by the large number of high quality experimental molecular crystal structures in this database, of which many have not yet been tested for their charge carrier mobility, but originally have been synthesised for different purposes. This rich collection of data not only provides us with the necessary structural information for our *ab initio* descriptors, but also allows us make additional statements on the stability or the synthesisability of each compound. The latter is especially important since for many of the structures contained in combinatorial compound libraries this information is not available, making the predicted materials less suited to immediate experimental validation. In our approach, we use our primary descriptors to identify promising candidates and also determine the charge percolation network in each material, which is only possible with the knowledge of H_{ab} for the full crystal. After identifying such promising candidates, we further examine the stability and complexity of synthesising each compound, ultimately presenting four novel materials for which we expect high charge carrier mobility and which are easily accessible for experimental verification.

Individual contributions:

I developed and implemented a workflow to calculate H_{ab} and λ_{int} starting from experimental crystallographic information files (cif). To this end, I implemented the initial generation of molecular supercells, nearest neighbour shells and dimers from experimental structures, without any user interaction, in the *Atomic Simulation Environment* (ASE). For the automated calculation of both descriptors the Python package *aimsutils* was created. To allow to automatic calculation of 100 000 molecular crystals I design a suitable workflow for the *Fireworks*-workflow management framework. For data collection, verification and processing two different database solutions were designed, with a convenient per-system data collection for further statistical analysis. All calculations and subsequent data analysis were performed by myself. Dr. Harald Oberhofer edited the manuscript and Prof. Dr. Karsten Reuter proofread the final draft.

8.3 Report on the sixth blind test of organic crystal-structure prediction methods

Summary of *Report on the sixth blind test of organic crystal-structure prediction methods* by A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio Jr, A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, C. R. Groom, Acta Crystallogr., Sect. B, **72**, 439 (2016)

and

GAtor (individual contribution, Supplementary Information) by F. Curtis, X. Li, C. Schober, K. Cosburn, S. Lohani, F. Vacarro, H. Oberhofer, K. Reuter, S. Bhattacharya, Á. Vázquez-Mayagoitia, L. M. Ghiringhelli, N. Marom, Acta Crystallogr., Sect. B, **72**, 439 (2016).

With this work, we contributed to the "Sixth CSP Blind Test of Organic Crystal Structure Prediction Methods", "an open and international collaborative effort to chart the progress and state of the art in predicting the crystal structures of small organic molecules"[138]. We targeted molecule *XXII* (tricyano-1,4-dithiino[c]-isothiazole, $C_8N_4S_3$), a chiral molecule with a hinged six-membered ring. During the course of the blind test this turned out to be a challenge for many of the force field based approaches, while we employed a purely DFT-based *ab initio* workflow. Initially, four pools with 50000 structures each were generated with different symmetry and number of molecules per unit cell. All initial structures were ranked using the Harris approximation together with the Tkatchenko-Scheffler (TS) pairwise van-der-Waals scheme. In subsequent steps of increasing computational cost the geometries were relaxed until full unit cell relaxation on PBE+TS level was achieved. This was then used as the initial pool of structures for a genetic algorithm. While we did not find the experimental structure in our initial submission, we were able to show that the energy ranking we employed is able to identify the most stable structure and an optimized version of the structure generation (GAtor) part would have been able to predict the correct structure.

This work was carried out in collaboration with Prof. Dr. Noa Marom from Carnegie Mellon University, Pittsburgh, Dr. Luca M. Ghiringhelli from the Fritz-Haber-Institut der Max-Planck-Gesellschaft, Berlin, Álvaro Vázquez-Mayagoitia from Argonne National Laboratory, Lemont and us.

Dr. Luca M. Ghiringhelli contributed the genetic algorithm (GA) which was further modified by Prof. Dr. Noa Marom and included into the GAtor framework. This framework implements a multistage approach to rank the energy of the predicted crystal structures from the GA. We contributed an efficient implementation of the Harris functional within FHIaims, which was used as the initial high-throughput energy ranking step for all systems in the four initial pools obtained from the genetic algorithm. With this fast and efficient technique, the initial energy ranking could be performed on an *ab initio* level, avoiding complications in force field parametrisation. This was shown to work very well for molecular crystals, since the superposition of molecular densities employed here, together with the van-der-Waals correction, is able to reproduce the majority of the total energy of the system.

Individual contributions:

I created a Python package which implements all necessary steps to allow efficient and accurate calculations of Harris energies with FHIaims. By making use of the symmetry of the molecular entities in each crystal the number of calculations necessary for any crystal is reduced to the number of inequivalent individual molecules per cell. The self-consistent wave-function of such a molecule is then transformed using modified Wigner-D matrices and superimposed to yield the appropriate wave-function for the full crystal structure.

9 Conclusions & Outlook

The interest of researchers in both academia and industry has multiplied with the ongoing improvements in the field of organic semiconductors. With a growing OLED market, technologies to completely replace inorganic semiconductor materials also in the electronic circuits of devices will allow novel shapes and properties. First prototypes of truly flexible displays have already been presented and many start-up and spin-off companies compete in their race to find the best / the cheapest / the most durable materials for organic electronics. While our understanding of the fundamental charge transfer processes in organic semiconductors lags behind these developments, research is being conducted to understand charge transfer and charge carrier mobility in real materials. In carefully designed experimental studies more and more extrinsic factors are eliminated to reveal the intrinsic carrier mobility, while theoreticians strive to improve their models and employ approximations closer matching the experimental conditions. The development of methods to improve the accuracy and efficiency of calculated parameters allows researchers to investigate more and larger systems with less cost, increasing the benefit of theoretical modelling. In parallel, the discovery of new materials is also an active field of research. Here, approaches range from deriving design guidelines from existing materials to experimental screening of potential candidates.

This thesis has attempted to contribute to the advancement of the field by improving the availability, accuracy and understanding of *ab initio* methods to calculate charge carrier mobilities, and by using such methods to demonstrate the power of computational modelling in high-throughput screening. This was done by first investigating one of the most common methods to calculate the transfer integral H_{ab}, the fragment orbital approximation. While this method has been known since 2003 and is widely used especially in the field of molecular semiconductors, systematic approaches to gauge its accuracy compared to other methods have been missing. In 2014 Kubas and co-workers laid the foundation for these studies by calculating accurate benchmark data for two sets of small organic molecules and comparing different methods to calculate transfer integrals. Based upon their reference data a comprehensive study of the fragment orbital family of methods was conducted in this thesis. To rule out any influences of implementation details and technical settings, all different flavours were implemented in the same DFT framework. The variant developed in the course of this work showed much greater accuracy at small additional computational cost. The best practice rules derived from this work allow others to choose the most suited variant of FO-DFT, depending on their requirements for accuracy and efficiency. Aiming at automated calculations a new method to calculate the internal reorganisation energies for molecular crystals was devised. To get rid of the need to control the final structures after the gasphase geometry optimisation, the de-facto standard approach up to now, a QM/MM embedding scheme is used to mimic the correct solid state environment at negligible additional computation cost.

To contribute to the discovery of new organic semiconductors and demonstrate the viability of the developed methods a virtual high-throughput screening study was performed. In contrast to other, small-scale approaches, the computationally most expensive task of calculating crystal structures was postponed, and the freed resources were used to screen the largest collection of molecular crystals known, the Cambridge Structural Database. Most of the structures in this database have never been considered as organic semiconductors before, making them ideal candidates for a computational screening. Built on the developed methods an automated *ab initio* workflow to calculate two charge transfer descriptors was designed. To render the calculation of 100 000 molecular crystals possible this was integrated in a workflow management framework, combining different supercomputing centers and local resources, as well as sophisticated data management tools. As a first result of this project, four novel molecular crystals with high expected intrinsic charge carrier mobility were selected from the final database. The software developed for this project was written with reusability in mind and hopefully will be a big step towards routine calculations of charge transfer parameters, eventually allowing researchers from different fields to make use of these techniques.

Naturally, the work done in this thesis is all but complete, and especially the huge amount of data generated and collected during the high-throughput screening will be of great value for further research towards structure-property relationships, design guidelines and novel organic semiconductor materials. As this was anticipated, the data was carefully processed and conditioned, without loosing the ability to work on the primary data if this turns out to be necessary, for example for big data mining approaches using machine learning techniques. Further enhancements will involve the accessibility of the workflows and methods to encourage routine calculations of charge transfer parameters similar to the way (DFT) structure optimisations are used in many experimental research groups today. Improving the (re)usability of computational methods is one of the great challenges in our quest to speed up computational material discovery and foster the wide-spread use of computational methods.

A The OrgEl environment and databases

In this technical appendix the detailed structure of both databases containing the screening data is described. In addition, an overview of the different tools and software written to generate, process and analyse the data is given. The purpose of this chapter is to allow other people to immediately work with the data, and be able to understand the generation process after working through the provided code.

This chapter is organized as follows: First, general information on the data, location and format is summarized. Then, the final result database with the processed per-system data is explained, followed by the database containing the raw data. Next, the tools to work with both databases are explained.

A.1 General Information

All data is available in the archive `PhD_Christoph_Schober_2013-2016.tpxz`. This is the persistent form of the data generated during this thesis, for access please ask the admin (`admin@theo.ch.tum.de`).

A.2 Functionality published elsewhere

The crystal structure preparation and neighbour generation routines have been included into the Atomic Simulation Environment (ASE)[1]-package. The relevant methods are `ase.utils.clustering.cluster` and `ase.utils.clustering.get_nearest_neighbours`, both are used in the screening workflow during the crystal preparation tasks. The code is also available in the TUM.TheoChemCodeArchive[2].

A.3 The OrgEl result database

The per-system result database is stored in a SQLite[3] database file (`screening.db`) and is best used together with the Pandas Python Libary for Data Analysis and Statistics[4]. In this database are 4 tables:

[1] `https://wiki.fysik.dtu.dk/ase/`
[2] `https://gitlab.lrz.de/theochem_archive/2016_orgel_ase`
[3] `https://www.sqlite.org/`
[4] `https://http://pandas.pydata.org/`

- SCREENING_CALCS: The final result database with 64 725 rows, each describing a calculated molecular crystal system. For each system, 164 different properties are saved. All fields are listed and explained in Tabs. A.1 to A.3.
- SITE_ENERGIES: Per-system site energies of neutral molecules in the crystal. There are 172 436 site-energies for 10 212 different molecular crystals.
- FOR_SALE: Subset of the full SCREENING_CALCS-table. Only systems which are available for sale by a commercial vendor according to the ZINC15 database are listed here. The latest update was done at 2016-07-28 and showed 2815 systems available.
- MONGO_QUERIES: [Technical table] All dates at which the running screening calculations were parsed for finished calculations.

The most important table to work with the HTS data is SCREENING_CALCS, where all the results for the calculated systems are collected. This includes the primary screening descriptors $H_{ab}^{\max}$ and λ_{int}. In addition, data from the experimental crystal structure determination such as symmetry, space group or publication reference are summarized. Only details such as the full H_{ab}s for all dimers or the optimised molecular geometries for the calculation of λ_{int} are kept elsewhere, in the OrgEl raw database (see appendix A.4).

Usage example

The following code example demonstrates the steps to load one table from the database file and extract all systems with certain parameters.

```
import pandas as pd
from sqlalchemy import create_engine

# define the path to the database file
db_path = 'sqlite:////path/to/screening.db'
# connect to the database using sqlalchemy
db = create_engine(db_path)

# load the target table from the database as
# pandas.DataFrame in variable df
df = pd.read_sql_table('screening_calcs', db)
# workaround - pandas sometimes confuses 'lambda'
# with the lambda-function in Python
df['lamda'] = df['lambda']
select = df.query("(lamda < 120) and (50 < hab_refined < 300)")
```

***Table A.1**: This table summarises all fields generated during the screening approach. They can be classified in calculated (descriptor) values and meta-information related to the systems.*

field	explanation
Technical	
level_0	technical SQLite field
identifier	MD5-based hash of cif-name for identification
index	unique integer index of system
System information	
file_name	Filename of downloaded .cif-file
name_cif	CSD identifier
paper_ref	Paper reference if included in .cif-file
inchi_key	INCHI-key of molecule
zinc_id	ZINC15 database ID of molecule
xyz_string	XYZ geometry string for single molecule (experimental geometry)
rdkit_smiles	SMILES identifier of molecule calculated by rdkit
availability	Commercial availability of molecule according to ZINC15 database
chemical_formula	Chemical formula after clustering
max_atom_number	Heaviest atom in molecule
Calculated	
last_task	Last successfully completed task in screening workflow
status	Status of screening workflow
n_atoms	number of atoms per molecule
min_dist	minimum distance between molecules in nearest neighbour cell
com_dist	Center-of-mass distance between molecules with highest H_{ab}
hab	Approximate (light) H_{ab}
hab_refined	Re-calculated (tight) H_{ab}
state_c	KS-state of charged fragment
state_n	KS-state of neutral fragment
lambda	Calculated reorganisation energy
homolumogap	HOMO-LUMO gap in neutral system (eV)
homo_en	HOMO-energy of neutral system (eV)
lumo_en	LUMO-energy of neutral system (eV)
stacking	type of crystal stacking (0: not assigned, 1: layered, 2: flattened h-bone, 3: herringbone, 4: sandwhich herringbone)

Table A.2: *These fields are all related to the crystallographic information file (cif). For details, please see the IUCr specification (`http://www.iucr.org/__data/iucr/cifdic_html/1/cif_core.dic/index.html`).*

field	explanation
c_audit_date	Date added to CSD
c_cell_angle_alpha	cell angle α
c_cell_angle_beta	cell angle β
c_cell_angle_gamma	cell angle γ
c_cell_length_a	cell vector a
c_cell_length_b	cell vector b
c_cell_length_c	cell vector c
c_cell_volume	cell volume
c_cell_formula_units_Z	number of molecules per unit cell
c_chemical_formula_moiety	chemical formula
c_chemical_name_common	common chemical name (if available)
c_chemical_name_systematic	systematic (IUPAC) name
c_crystal_color	crystal color
c_exptl_special_details	additional details on measurement or system
c_journal_year	year of publication
c_max_atomic_num	Heaviest atom in molecule
c_n_atoms	number of atoms per molecule
c_n_non_hydrogen	number of non-hydrogen atoms per molecule
c_num_structures	number of different structures in cif-file
c_refine_ls_R_factor_gt	Residual factor for the reflections, quality criteria
c_symmetry_cell_setting	cell for space-group symmetry
c_symmetry_space_group_name_HM	name of symmetry space group
c_temperature	temperature of measurement

Table A.3: *In this list all calculated cheminformatics descriptors are collected. All have been calculated/generated using rdkit (`http://www.rdkit.org`). For a explanation of each descriptor, please see rdkit.Chem.Descriptors (`http://www.rdkit.org/Python_Docs/rdkit.Chem.Descriptors-module.html`).*

NumAromaticRings	fr_nitrile	fr_phenol
HeavyAtomCount	fr_nitro	fr_phenol_noOrthoHbond
HeavyAtomMolWt	fr_nitro_arom	fr_phos_acid
MolMR	fr_nitro_arom_nonortho	fr_phos_ester
MolWt	fr_nitroso	fr_piperdine
NHOHCount	fr_oxazole	fr_term_acetylene
NOCount	fr_tetrazole	fr_thiazole
NumAliphaticCarbocycles	fr_thiophene	fr_thiocyan
NumAliphaticHeterocycles	fr_unbrch_alkane	fr_oxime
NumAliphaticRings	fr_Nhpyrrole	fr_para_hydroxylation
NumAromaticAtoms	fr_C_O_noCOO	fr_diazo
NumAromaticBonds	fr_Al_COO	fr_dihydropyridine
NumAromaticCarbocycles	fr_Al_OH	fr_epoxide
NumAromaticHeterocycles	fr_Al_OH_noTert	fr_ester
NumBonds	fr_aldehyde	fr_ether
NumConjugatedBonds	fr_alkyl_carbamate	fr_furan
NumDoubleBonds	fr_alkyl_halide	fr_guanido
NumHAcceptors	fr_allylic_oxid	fr_halogen
NumHDonors	fr_amide	fr_hdrzine
NumHeteroatoms	fr_amidine	fr_hdrzone
NumRadicalElectrons	fr_aniline	fr_HOCCN
NumRotatableBonds	fr_Ar_COO	fr_imidazole
NumSaturatedCarbocycles	fr_Ar_N	fr_imide
NumSaturatedHeterocycles	fr_Ar_NH	fr_Imine
NumSaturatedRings	fr_Ar_OH	fr_isocyan
NumValenceElectrons	fr_ArN	fr_isothiocyan
RingCount	fr_aryl_methyl	fr_ketone
SSSR	fr_azide	fr_ketone_Topliss
fr_piperzine	fr_azo	fr_lactam
fr_priamide	fr_barbitur	fr_lactone
fr_prisulfonamd	fr_benzene	fr_methoxy
fr_pyridine	fr_benzodiazepine	fr_morpholine
fr_quatN	fr_bicyclic	fr_N_O
fr_SH	fr_C_O	fr_Ndealkylation1
fr_sulfide	fr_C_S	fr_Ndealkylation2
fr_sulfonamd	fr_COO	fr_NH0
fr_sulfone	fr_COO2	fr_NH1
fr_urea	fr_NH2	

A.4 The OrgEl raw database

A full data backup of the production database is provided in the subfolder `mongo_dump`. This data backup can be restored using the tools provided by MongoDB[5]. In NoSQL-databases such as MongoDB *documents* are stored in *collections*, and a *collection* is part of a *database*. The MongoDB instance consists of two important sub-databases: `fireworks` and `processed`. Example methods to work with the data collected in the OrgEl raw database can be found in the modules `orgel_screening.fw_admin` and `orgel_screening.analyse`.

`fireworks`

This database includes all information of the FireWorks workflow management framework, such as the individual workflows per system, "launches", paths and stored data. The database layout is described in the documentation of the package (FireWorks docs[6]).

`processed`

Data for all calculated system from the FireWorks-database is accumulated in this separate database in the collection `data` to allow easier access to the calculated properties. Depending on the final state of the workflow, different data is stored. The following two examples represent data stored for a failed and a successful workflow.

Failed at initial clustering step If a workflow failed, the entry in the database will show the final worktask in which it failed, together with the error message provided by the workflow.

```
{u'_id': ObjectId('568b9ba60df61366062d4ca7'),
 u'failed_at': u'Clustering',
 u'failed_trace': u'Traceback (most recent call last): ...',
 u'filename': u'ZOTKAU.n_atoms_100_to_149.cif',
 u'fireworks': [95313],
 u'total_runtime': [],
 u'wflow_state': u'FIZZLED'}
```

Successfully completed full workflow In a successful workflow information on the full crystal environment comprises the main data in each document. Beside the descriptor values for H_{ab} and λ_{int} there also are the final optimised geometries for charged and neutral QM/MM optimisations or the distances between molecules in the dimers.

```
{u'_id': ObjectId('568bd27b0df61335a7dbd9ec'),
 u'c_lumo': 48,
 u'chem_formula': u'C8H9NS2',
 u'com_dists': [4.983..., ..., 11.857...],
 u'com_hab_max': 6.3929278362504585,
```

[5] https://docs.mongodb.com/manual/reference/program/mongodump/
[6] https://pythonhosted.org/FireWorks/

```
 u'com_vecs': [[-2.457..., 3.820..., -2.049...], [...]],
 u'degenerated': False,
 u'filename': u'YOKROF.n_atoms_20_to_29.cif',
 u'fireworks': [22331, 239991, 263409, 263432, 263510],
 u'hab_max': 191.8,
 u'hab_max_refined': 190.17,
 u'habs_refined': [7.73, ..., 0.05],
 u'lambda': 168.7170915029128,
 u'last_task': u'Lambda QMMM',
 u'max_atom_num': 16,
 u'min_dists': [2.574400749875095, ..., 3.45446670962288],
 u'min_hab_max': 2.530719011742661,
 u'n_atoms': 20,
 u'n_homo': 48,
 u'total_runtime': [2.68, 188.44, 121.04, 393.34, 1493.91],
 u'wflow_state': u'COMPLETED',
 u'xyz_central': u'...',
 u'xyz_neighbours': [u'...', u'...', u'...'],
 u'xyz_opt_charged': u'...',
 u'xyz_opt_neutral': u'...'}
```

Usage example for a local MongoDB instance

In this example a connection to the MongoDB database using PyMongo[7] is created and some data extracted.

```
from pymongo import MongoClient
from fireworks import LaunchPad

# create connection to local mongodb instance
client = MongoClient('mongodb://admin_user:PASSWORD@127.0.0.1')

# extract data for system 'YOKROF.n_atoms_20_to_29.cif'
processed_db = client.get_database('processed')
data = processed_db.get_collection('data')
d_yokrof = data.find_one({'filename':
 'YOKROF.n_atoms_20_to_29.cif'})

# or: find all failed at clustering step
d_clu = data.find({'failed_at': 'Clustering'})
```

A.5 Calculation archive files

The raw data of the performed DFT and MM calculations is stored in a number of pixz[8]-archives. They are available in the subfolder *archive_rawdata*. Individual folders can be extracted using

[7] https://api.mongodb.com/python/current/
[8] http://www.pixz.com

the functionality provided by *pixz*, the relevant path is stored in the *OrgEl Raw Database* (for example code see the OrgEl package appendix A.6, `orgel_screening/scripts/stage_2_analysis/PP6_parse_site_energies.py`).

A.6 The orgel Python package

In this package most of the HTS related functionality for data generation and analysis is provided. A version of the package is provided with the archived data, but the git-repository is available at the TUM.TheoChemCodeArchive[9]. The overall structure of the package is as follows:

```
orgel_screening
├── docs/
├── LICENCE
├── orgel_screening/
├── orgel_workflows/
├── README
├── scripts/
├── setup.py
└── tests/
```

Initial setup The code to setup the initial databases with the information for all systems and add the workflows to the FireWorks-database is found in files `scripts/setup_screening.py`, `scripts/convert_h5_sql.py` and `scripts/feed_the_mongo.py`.

FireWorks workflows The full workflow described in this thesis and used for the calculations is defined in the module `orgel_workflows.tasks` in `orgel_workflows/tasks.py`.

Data processing After the FireWorks-workflow finished the data is aggregated using the scripts `scripts/DB_PP0_add_columns_once.py` and `scripts/DB_PP1_extract_data.py`.

Data analysis This is by far the most extensive part of the package, with all the scripts and tools used to process the initial data, add new information based on the data and external resources and populate the SQL-database. The functionality for the analysis of the data is mainly aggregated in the folder `orgel_screening/scripts` and `orgel_screening/scripts/stage_2_analysis`. In the former folder, the initial analysis is done (`PP1_add_smiles.py`, `PP2_add_descriptors.py`), while the latter contains different plotting functions (e.g. `plot_1d_hists.py`) and additional analysis and data processing (`PP1_parse_homolumogap.py` to `PP9_query_ZINC15.py`). Functionality to calculate the percolation pathways based on the data can be found in `scripts/stage_2_analysis/PP5_find_percolation.py` and plotting the pathways in `scripts/stage_2_analysis/percolation_plots/nice_plot.py`.

[9] `https://gitlab.lrz.de/theochem_archive/2016_orgel_screening`

Bibliography

[1] (J. E. Lilienfeld), US1745175 (A), CIB: H01L29/00; H01L29/786; H03F3/04, **1930** (cit. on p. 1).

[2] G. Moore, Moore's Law | Computer Science | Britannica.com, **1965**, `https://www.britannica.com/topic/Moores-law` (visited on 2016-08-30) (cit. on p. 1).

[3] Army to Display Flexible Technology (Public Domain), `https://www.flickr.com/photos/rdecom/4146880795/` (visited on 2016-08-30) (cit. on p. 1).

[4] H. Kallmann, M. Pope, *Nature* **1960**, *186*, 31–33 (cit. on p. 1).

[5] H. Kallmann, M. Pope, *J. Chem. Phys.* **1960**, *32*, 300–301 (cit. on p. 1).

[6] X. Cai, D. Ji, L. Jiang, G. Zhao, J. Tan, G. Tian, J. Li, W. Hu, *Appl. Phys. Lett.* **2014**, *104*, 063305 (cit. on p. 1).

[7] S. R. Forrest, *Nature* **2004**, *428*, 911–918 (cit. on pp. 1, 2, 7).

[8] H. Minemawari, T. Yamada, H. Matsui, J. Tsutsumi, S. Haas, R. Chiba, R. Kumai, T. Hasegawa, *Nature* **2011**, *475*, 364–367 (cit. on pp. 1, 2, 6).

[9] K.-J. Baeg, M. Caironi, Y.-Y. Noh, *Adv. Mater.* **2013**, *25*, 4210–4244 (cit. on p. 1).

[10] Y. Sun, G. C. Welch, W. L. Leong, C. J. Takacs, G. C. Bazan, A. J. Heeger, *Nat Mater* **2012**, *11*, 44–48 (cit. on pp. 1, 2).

[11] D. S. Chung, J. W. Park, W. M. Yun, H. Cha, Y.-H. Kim, S.-K. Kwon, C. E. Park, *ChemSusChem* **2010**, *3*, 742–748 (cit. on p. 1).

[12] F. Silvestri, M. D. Irwin, L. Beverina, A. Facchetti, G. A. Pagani, T. J. Marks, *J. Am. Chem. Soc.* **2008**, *130*, 17640–17641 (cit. on p. 1).

[13] A. Facchetti, *Nat Mater* **2013**, *12*, 598–600 (cit. on p. 1).

[14] D. J. Gundlach, *Nat Mater* **2007**, *6*, 173–174 (cit. on p. 2).

[15] P. Anikeeva, R. A. Koppes, *Science* **2015**, *350*, 274–275 (cit. on p. 2).

[16] E. Stavrinidou, R. Gabrielsson, E. Gomez, X. Crispin, O. Nilsson, D. T. Simon, M. Berggren, *Sci. Adv.* **2015**, *1*, e1501136 (cit. on p. 2).

[17] IDTechEX, Printed, Organic & Flexible Electronics Forecasts, Players & Opportunities 2016-2026: IDTechEx, **2015**, `http://www.idtechex.com/research/reports/printed-organic-and-flexible-electronics-forecasts-players-and-opportunities-2016-2026-000457.asp` (visited on 2016-08-30) (cit. on p. 2).

[18] Editorial, *Nat Mater* **2013**, *12*, 591–591 (cit. on p. 2).

[19] H. Sirringhaus, *Adv. Mater.* **2014**, *26*, 1319–1335 (cit. on pp. 2, 6, 7).

[20] Y. Lin, Y. Li, X. Zhan, *Chem. Soc. Rev.* **2012**, *41*, 4245–4272 (cit. on p. 2).

[21] J. E. Anthony, *Chem. Rev.* **2006**, *106*, 5028–5048 (cit. on p. 2).

[22] Y. M. Sun, Y. Q. Ma, Y. Q. Liu, Y. Y. Lin, Z. Y. Wang, Y. Wang, C. A. Di, K. Xiao, X. M. Chen, W. F. Qiu, B. Zhang, G. Yu, W. P. Hu, D. B. Zhu, *Adv. Funct. Mater.* **2006**, *16*, 426–432 (cit. on p. 2).

[23] L. Biniek, B. C. Schroeder, C. B. Nielsen, I. McCulloch, *J. Mater. Chem.* **2012**, *22*, 14803–14813 (cit. on p. 2).

[24] J.-M. Jiang, M.-C. Yuan, K. Dinakaran, A. Hariharan, K.-H. Wei, *J. Mater. Chem. A* **2013**, *1*, 4415–4422 (cit. on p. 2).

[25] J. W. Jung, F. Liu, T. P. Russell, W. H. Jo, *Energy Environ. Sci.* **2012**, *5*, 6857–6861 (cit. on p. 2).

[26] J. R. Tumbleston, A. C. Stuart, E. Gann, W. You, H. Ade, *Adv. Funct. Mater.* **2013**, n/a–n/a (cit. on p. 2).

[27] M. Uno, T. Uemura, Y. Kanaoka, Z. Chen, A. Facchetti, J. Takeya, *Organic Electronics* **2013**, *14*, 1656–1662 (cit. on pp. 2, 6).

[28] E. Orgiu, J. George, J. A. Hutchison, E. Devaux, J. F. Dayen, B. Doudin, F. Stellacci, C. Genet, J. Schachenmayer, C. Genes, G. Pupillo, P. Samorì, T. W. Ebbesen, *Nat Mater* **2015**, *14*, 1123–1129 (cit. on p. 2).

[29] C. Reese, Z. Bao, *Materials Today* **2007**, *10*, 20–27 (cit. on p. 2).

[30] Y. Yuan, G. Giri, A. L. Ayzner, A. P. Zoombelt, S. C. B. Mannsfeld, J. Chen, D. Nordlund, M. F. Toney, J. Huang, Z. Bao, *Nat Commun* **2014**, *5*, 3005 (cit. on pp. 2, 6).

[31] Y. Zhou, C. Fuentes-Hernandez, J. Shim, J. Meyer, A. J. Giordano, H. Li, P. Winget, T. Papadopoulos, H. Cheun, J. Kim, M. Fenoll, A. Dindar, W. Haske, E. Najafabadi, T. M. Khan, H. Sojoudi, S. Barlow, S. Graham, J.-L. Brédas, S. R. Marder, A. Kahn, B. Kippelen, *Science* **2012**, *336*, 327–332 (cit. on p. 2).

[32] C. Goldmann, S. Haas, C. Krellner, K. P. Pernstich, D. J. Gundlach, B. Batlogg, *J. Appl. Phys.* **2004**, *96*, 2080–2086 (cit. on p. 2).

[33] H. Dong, X. Fu, J. Liu, Z. Wang, W. Hu, *Adv. Mater.* **2013**, *25*, 6158–6183 (cit. on p. 2).

[34] C. Motta, S. Sanvito, *J. Chem. Theory Comput.* **2014**, DOI 10.1021/ct500390a (cit. on p. 2).

[35] G. Raos, M. Casalegno, J. Idé, *J. Chem. Theory Comput.* **2013**, DOI 10.1021/ct400854a (cit. on p. 2).

[36] J. Cornil, S. Verlaak, N. Martinelli, A. Mityashin, Y. Olivier, T. Van Regemorter, G. D'Avino, L. Muccioli, C. Zannoni, F. Castet, D. Beljonne, P. Heremans, *Acc. Chem. Res.* **2013**, *46*, 434–443 (cit. on p. 2).

[37] A. Troisi, *Chem. Soc. Rev.* **2011**, *40*, 2347–2358 (cit. on pp. 2, 9–12).

[38] A. Troisi, *Adv. Mater.* **2007**, *19*, 2000–2004 (cit. on p. 2).

[39] F. Ortmann, F. Bechstedt, K. Hannewald, *Phys. Status Solidi B* **2011**, *248*, 511–525 (cit. on pp. 2, 9, 11).

[40] H. Bässler, *phys. stat. sol. (b)* **1993**, *175*, 15–56 (cit. on pp. 2, 9).

[41] P. Kordt, J. J. M. van der Holst, M. Al Helwi, W. Kowalsky, F. May, A. Badinski, C. Lennartz, D. Andrienko, *Adv. Funct. Mater.* **2015**, *25*, 1955–1971 (cit. on pp. 2, 5, 10, 16, 20).

[42] C. Poelking, E. Cho, A. Malafeev, V. Ivanov, K. Kremer, C. Risko, J.-L. Brédas, D. Andrienko, *J. Phys. Chem. C* **2013**, *117*, 1633–1640 (cit. on p. 2).

[43] T. Vehoff, B. Baumeier, A. Troisi, D. Andrienko, *J. Am. Chem. Soc.* **2010**, *132*, 11702–11708 (cit. on pp. 2, 26).

[44] V. Coropceanu, H. Li, P. Winget, L. Zhu, J.-L. Brédas, *Annu. Rev. Mater. Res.* **2013**, *43*, 63–87 (cit. on p. 2).

[45] R. A. Marcus, *Rev. Mod. Phys.* **1993**, *65*, 599–610 (cit. on pp. 2, 11, 15, 19).

[46] J.-L. Brédas, D. Beljonne, V. Coropceanu, J. Cornil, *Chem. Rev.* **2004**, *104*, 4971–5004 (cit. on p. 2).

[47] A. N. Sokolov, S. Atahan-Evrenk, R. Mondal, H. B. Akkerman, R. S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S. C. B. Mannsfeld, A. P. Zoombelt, Z. Bao, A. Aspuru-Guzik, *Nat Commun* **2011**, *2*, 437 (cit. on pp. 2, 16, 24).

[48] T. Yamamoto, K. Takimiya, *J. Am. Chem. Soc.* **2007**, *129*, 2224–2225 (cit. on pp. 2, 24).

[49] J. Bajorath, *Nat Rev Drug Discov* **2002**, *1*, 882–894 (cit. on pp. 3, 23).

[50] B. K. Shoichet, *Nature* **2004**, *432*, 862–865 (cit. on p. 3).

[51] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat Rev Drug Discov* **2004**, *3*, 935–949 (cit. on p. 3).

[52] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, A. Aspuru-Guzik, *J. Phys. Chem. Lett.* **2011**, *2*, 2241–2251 (cit. on pp. 3, 24).

[53] E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, A. Aspuru-Guzik, *Annu. Rev. Mater. Res.* **2015**, *45*, 195–216 (cit. on pp. 3, 24).

[54] J. J. Low, A. I. Benin, P. Jakubczak, J. F. Abrahamian, S. A. Faheem, R. R. Willis, *J. Am. Chem. Soc.* **2009**, *131*, 15834–15842 (cit. on pp. 3, 24).

[55] I. Y. Kanal, S. G. Owens, J. S. Bechtel, G. R. Hutchison, *J. Phys. Chem. Lett.* **2013**, *4*, 1613–1623 (cit. on p. 3).

[56] M. Mas-Torrent, P. Hadley, S. T. Bromley, X. Ribas, J. Tarrés, M. Mas, E. Molins, J. Veciana, C. Rovira, *J. Am. Chem. Soc.* **2004**, *126*, 8546–8553 (cit. on pp. 3, 20).

[57] F. H. Allen, *Acta Cryst. B* **2002**, *58*, 380–388 (cit. on pp. 3, 24).

[58] J. Liu, H. Zhang, H. Dong, L. Meng, L. Jiang, L. Jiang, Y. Wang, J. Yu, Y. Sun, W. Hu, A. J. Heeger, *Nat Commun* **2015**, *6*, 10032 (cit. on p. 5).

[59] B. Kippelen, J.-L. Brédas, *Energy Environ. Sci.* **2009**, *2*, 251–261 (cit. on p. 5).

[60] E. Wang, W. Mammo, M. R. Andersson, *Adv. Mater.* **2014**, *26*, 1801–1826 (cit. on p. 5).

[61] M. A. Ruderer, P. Müller-Buschbaum, *Soft Matter* **2011**, *7*, 5482–5493 (cit. on p. 6).

[62] H. Fan, H. Shang, Y. Li, X. Zhan, *Appl. Phys. Lett.* **2010**, *97*, 133302 (cit. on p. 6).

[63] S. D. Dimitrov, J. R. Durrant, *Chem. Mater.* **2014**, *26*, 616–630 (cit. on p. 6).

[64] J. Zaumseil, H. Sirringhaus, *Chem. Rev.* **2007**, *107*, 1296–1323 (cit. on pp. 6, 7, 9).

[65] T. Izawa, E. Miyazaki, K. Takimiya, *Chem. Mater.* **2009**, *21*, 903–912 (cit. on p. 6).

[66] G. Giri, E. Verploegen, S. C. B. Mannsfeld, S. Atahan-Evrenk, D. H. Kim, S. Y. Lee, H. A. Becerril, A. Aspuru-Guzik, M. F. Toney, Z. Bao, *Nature* **2011**, *480*, 504–508 (cit. on p. 6).

[67] A. Kokil, K. Yang, J. Kumar, *J. Polym. Sci. B Polym. Phys.* **2012**, *50*, 1130–1144 (cit. on pp. 7, 10).

[68] H. Oberhofer, K. Reuter, J. Blumberger, *Be Submitt.* **2016** (cit. on p. 9).

[69] S. T. Hoffmann, F. Jaiser, A. Hayer, H. Bässler, T. Unger, S. Athanasopoulos, D. Neher, A. Köhler, *J. Am. Chem. Soc.* **2013**, *135*, 1772–1782 (cit. on p. 9).

[70] I. I. Fishchuk, A. Kadashchuk, S. T. Hoffmann, S. Athanasopoulos, J. Genoe, H. Bässler, A. Köhler in AIP Conference Proceedings, *Vol. 1610*, 15TH INTERNATIONAL CONFERENCE ON TRANSPORT IN INTERACTING DISORDERED SYSTEMS (TIDS15), AIP Publishing, **2014**, pp. 47–52 (cit. on p. 9).

[71] I. I. Fishchuk, A. Kadashchuk, S. T. Hoffmann, S. Athanasopoulos, J. Genoe, H. Bässler, A. Köhler, *Phys. Rev. B* **2013**, *88*, 125202 (cit. on p. 9).

[72] N. Karl, *Journal of Crystal Growth* **1990**, *99*, 1009–1016 (cit. on p. 9).

[73] N. Karl, *Synthetic Metals*, Proceedings of the Yamada Conference LVI. The Fourth International Symposium on Crystalline Organic Metals, Superconductors and Ferromagnets (ISCOM 2001). **2003**, *133–134*, 649–657 (cit. on p. 9).

[74] O. Ostroverkhova, D. G. Cooke, F. A. Hegmann, J. E. Anthony, V. Podzorov, M. E. Gershenson, O. D. Jurchescu, T. T. M. Palstra, *Appl. Phys. Lett.* **2006**, *88*, 162101 (cit. on p. 9).

[75] O. D. Jurchescu, J. Baas, T. T. M. Palstra, *Appl. Phys. Lett.* **2004**, *84*, 3061–3063 (cit. on p. 9).

[76] V. Podzorov, S. E. Sysoev, E. Loginova, V. M. Pudalov, M. E. Gershenson, *Appl. Phys. Lett.* **2003**, *83*, 3504–3506 (cit. on p. 9).

[77] T. Sakanoue, H. Sirringhaus, *Nat Mater* **2010**, *9*, 736–740 (cit. on p. 9).

[78] W.-Q. Deng, L. Sun, J.-D. Huang, S. Chai, S.-H. Wen, K.-L. Han, *Nat. Protocols* **2015**, *10*, 632–642 (cit. on pp. 10, 20).

[79] H. Tamura, I. Hamada, H. Shang, K. Oniwa, M. Akhtaruzzaman, T. Jin, N. Asao, Y. Yamamoto, T. Kanagasekaran, H. Shimotani, S. Ikeda, K. Tanigaki, *J. Phys. Chem. C* **2013**, *117*, 8072–8078 (cit. on pp. 10, 16, 20).

[80] R. S. Sánchez-Carrera, S. Atahan, J. Schrier, A. Aspuru-Guzik, *J. Phys. Chem. C* **2010**, *114*, 2334–2340 (cit. on pp. 10, 19, 20).

[81] E. M. García-Frutos, E. Gutierrez-Puebla, M. A. Monge, R. Ramírez, P. de Andrés, A. de Andrés, R. Ramírez, B. Gómez-Lor, *Organic Electronics* **2009**, *10*, 643–652 (cit. on pp. 10, 20).

[82] D. L. Cheung, D. P. McMahon, A. Troisi, *J. Phys. Chem. B* **2009**, *113*, 9393–9401 (cit. on pp. 10, 20).

[83] H. Oberhofer, J. Blumberger, *Phys. Chem. Chem. Phys.* **2012**, *14*, 13846–13852 (cit. on pp. 10, 15–17, 20).

[84] A. Troisi, G. Orlandi, *Phys. Rev. Lett.* **2006**, *96*, 086601 (cit. on p. 11).

[85] A. Troisi, G. Orlandi, *J. Phys. Chem. A* **2006**, *110*, 4065–4070 (cit. on p. 11).

[86] N. W. Ashcroft, N. D. Mermin, *Solid State Physics*, Saunders College, Philadelphia, **1976** (cit. on p. 12).

[87] T. Holstein, *Annals of Physics* **1959**, *8*, 325–342 (cit. on p. 12).

[88] J. Spencer, F. Gajdos, J. Blumberger, *J. Chem. Phys.* **2016**, *145*, 064102 (cit. on p. 13).

[89] T. Van Voorhis, T. Kowalczyk, B. Kaduk, L.-P. Wang, C.-L. Cheng, Q. Wu, *Annu. Rev. Phys. Chem.* **2010**, *61*, 149–170 (cit. on p. 15).

[90] R. A. Marcus, N. Sutin, *Biochimica et Biophysica Acta (BBA) - Reviews on Bioenergetics* **1985**, *811*, 265–322 (cit. on p. 15).

[91] J. C. Tully, *J. Chem. Phys.* **2012**, *137*, 22A301 (cit. on p. 15).

[92] R. S. Mulliken, *J. Am. Chem. Soc.* **1952**, *74*, 811–824 (cit. on p. 15).

[93] N. S. Hush in *Progress in Inorganic Chemistry*, (Ed.: F. A. Cotton), John Wiley & Sons, Inc., **1967**, pp. 391–444 (cit. on p. 15).

[94] R. J. Cave, M. D. Newton, *Chemical Physics Letters* **1996**, *249*, 15–19 (cit. on p. 15).

[95] R. J. Cave, M. D. Newton, *J. Chem. Phys.* **1997**, *106*, 9213–9226 (cit. on p. 15).

[96] T. Pacher, L. S. Cederbaum, H. Köppel, *J. Chem. Phys.* **1988**, *89*, 7367–7381 (cit. on p. 15).

[97] T. Pacher, L. S. Cederbaum, H. Köppel in *Advances in Chemical Physics*, (Eds.: I. Prigogine, S. A. Rice), John Wiley & Sons, Inc., **1993**, pp. 293–391 (cit. on p. 15).

[98] Q. Wu, T. Van Voorhis, *Phys. Rev. A* **2005**, *72*, 024502 (cit. on p. 15).

[99] Q. Wu, T. Van Voorhis, *J. Chem. Phys.* **2006**, *125*, 164105–164105–9 (cit. on p. 15).

[100] H. Oberhofer, J. Blumberger, *J. Chem. Phys.* **2010**, *133*, 244105–244105–10 (cit. on p. 15).

[101] K. Senthilkumar, F. C. Grozema, F. M. Bickelhaupt, L. D. A. Siebbeles, *J. Chem. Phys.* **2003**, *119*, 9809–9817 (cit. on pp. 15–17).

[102] C. Schober, K. Reuter, H. Oberhofer, *J. Chem. Phys.* **2016**, *144*, 054103 (cit. on pp. 15–17).

[103] A. J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **2008**, *321*, 792–794 (cit. on p. 15).

[104] A. Kubas, F. Hoffmann, A. Heck, H. Oberhofer, M. Elstner, J. Blumberger, *J. Chem. Phys.* **2014**, *140*, 104105 (cit. on pp. 15, 16).

[105] A. Kubas, F. Gajdos, A. Heck, H. Oberhofer, M. Elstner, J. Blumberger, *Phys. Chem. Chem. Phys.* **2015**, *17*, 14342–14354 (cit. on p. 15).

[106] Y. Yi, L. Zhu, J.-L. Brédas, *J. Phys. Chem. C* **2012**, *116*, 5215–5224 (cit. on p. 16).

[107] H.-L. Wei, Y.-F. Liu, *Appl. Phys. A* **2014**, 1–7 (cit. on p. 16).

[108] F. Hernández-Fernández, M. Pavanello, L. Visscher, *Phys. Chem. Chem. Phys.* **2016**, *18*, 21122–21132 (cit. on p. 16).

[109] V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter, M. Scheffler, *Computer Physics Communications* **2009**, *180*, 2175–2196 (cit. on pp. 17, 28).

[110] C. Schober, K. Reuter, H. Oberhofer, *J. Phys. Chem. Lett.* **2016**, *7*, 3973–3977 (cit. on pp. 17, 21, 26, 28).

[111] J. Nelson, J. J. Kwiatkowski, J. Kirkpatrick, J. M. Frost, *Acc. Chem. Res.* **2009**, *42*, 1768–1778 (cit. on p. 19).

[112] J. E. Norton, J.-L. Brédas, *J. Am. Chem. Soc.* **2008**, *130*, 12377–12384 (cit. on p. 19).

[113] V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, J.-L. Brédas, *Chem. Rev.* **2007**, *107*, 926–952 (cit. on p. 20).

[114] H. M. Senn, W. Thiel, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229 (cit. on p. 21).

[115] D. Berger, A. J. Logsdail, H. Oberhofer, M. R. Farrow, C. R. A. Catlow, P. Sherwood, A. A. Sokol, V. Blum, K. Reuter, *J. Chem. Phys.* **2014**, *141*, 024105 (cit. on p. 21).

[116] A. Smith, *Nature* **2002**, *418*, 453–459 (cit. on p. 23).

[117] Screening Robotics for HTS - Wikimedia Commons (Public Domain), `https://commons.wikimedia.org/wiki/File:Screening_robotics_for_HTS-1-.jpg` (visited on 2016-10-10) (cit. on p. 23).

[118] W. P. Walters, M. T. Stahl, M. A. Murcko, *Drug Discovery Today* **1998**, *3*, 160–178 (cit. on p. 23).

[119] G. Sliwoski, S. Kothiwale, J. Meiler, E. W. Lowe, *Pharmacol Rev* **2014**, *66*, 334–395 (cit. on p. 23).

[120] J. Greeley, T. F. Jaramillo, J. Bonde, I. Chorkendorff, J. K. Nørskov, *Nat Mater* **2006**, *5*, 909–913 (cit. on p. 24).

[121] Y. J. Colón, R. Q. Snurr, *Chem. Soc. Rev.* **2014**, *43*, 5735–5749 (cit. on p. 24).

[122] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.* **2005**, *45*, 177–182 (cit. on p. 24).

[123] I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson, R. Taylor, *Acta Crystallogr B* **2002**, *58*, 389–397 (cit. on p. 24).

[124] A. Jain, S. P. Ong, W. Chen, B. Medasani, X. Qu, M. Kocher, M. Brafman, G. Petretto, G.-M. Rignanese, G. Hautier, D. Gunter, K. A. Persson, *Concurr. Comput. Pract. Exp.* **2015**, n/a–n/a (cit. on p. 25).

[125] MongoDB Inc., MongoDB for GIANT Ideas, **2016**, `https://www.mongodb.com` (visited on 2016-09-02) (cit. on p. 25).

[126] W. McKinney in Proceedings of the 9th Python in Science Conference, (Eds.: S. van der Walt, J. Millman), **2010**, pp. 51–56 (cit. on p. 25).

[127] J. Maddox, *Nature* **1988**, *345*, 201–201 (cit. on p. 27).

[128] S. M. Woodley, R. Catlow, *Nat Mater* **2008**, *7*, 937–946 (cit. on p. 27).

[129] E. R. T. Tiekink, *Chem. Commun.* **2014**, *50*, 11079–11082 (cit. on p. 27).

[130] J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen, W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, D. E. Williams, *Acta Crystallogr. B* **2000**, *56*, 697–714 (cit. on p. 27).

[131] A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio Jr, A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, C. R. Groom, *Acta Crystallogr. Sect. B* **2016**, *72*, 439–459 (cit. on pp. 27, 28).

[132] S. L. Price, *Chem. Soc. Rev.* **2014**, *43*, 2098–2111 (cit. on p. 27).

[133] G. A. Stephenson, R. A. Forbes, S. M. Reutzel-Edens, *Advanced Drug Delivery Reviews*, Characterization of the Solid State **2001**, *48*, 67–90 (cit. on p. 27).

[134] M. A. Neumann, *J. Phys. Chem. B* **2008**, *112*, 9810–9829 (cit. on p. 28).

[135] A. R. Oganov, C. W. Glass, *J. Chem. Phys.* **2006**, *124*, 244704 (cit. on p. 28).

[136] S. Bhattacharya, B. H. Sonin, C. J. Jumonville, L. M. Ghiringhelli, N. Marom, *Phys. Rev. B* **2015**, *91*, 241115 (cit. on p. 28).

[137] K. Berland, E. Londero, E. Schröder, P. Hyldgaard, *Phys. Rev. B* **2013**, *88*, 045431 (cit. on p. 28).

[138] CSP Blind Tests - The Cambridge Crystallographic Data Centre (CCDC), `https://www.ccdc.cam.ac.uk/Community/initiatives/cspblindtests/` (visited on 2016-09-06) (cit. on p. 33).

Acknowledgements / Danksagung

This work has been carried out at the Chair of Theoretical Chemistry of Prof. Dr. Karsten Reuter at the Technische Universität München between April 2013 and September 2016. It would not have been possible without the support and input of many people, but foremost I want to thank my supervisors Prof. Dr. Karsten Reuter and Dr. Harald Oberhofer. They initiated this project and gave me the freedom to change and to evolve the project according to my own interests. Prof. Dr. Karsten Reuter made it possible to visit many different international conferences, meet scientists from different groups and fields and was in general very helpful in answering all the open questions, but especially the ones related to strategic decisions during this project. Dr. Harald Oberhofer had to endure many discussions on all levels (albeit the DFT+U project will never be excelled in this prospect), from implementation details in FHIaims to specific wording in paper drafts and many more. It was always informative and helpful to talk to him.

I would like to thank Ruth Mösch for her continuous work "in the background", always making sure contracts were extended, Dienstreiseanträge signed and sent, and the opportunity to ask any questions related to the famous bureaucracy of the university.

A constant source of knowledge during my time in the group was Christoph Scheurer, who not only created and provided the technical and computational environment I used during my work, but also encouraged me to make many interesting discoveries in our system by just giving me the root-password and not mentioning what I should or should not do. Scientifically, he also helped in many ways, and often could point me into the right direction ("*Your problem sounds familiar to something the Quantum Chemists / Computer Scientists/ ...have encountered in the 1970s/1980s/...*").

Work would not have been so much fun without all the people in the group. Therefore, I like to thank all of them for many discussions, events and the good time here in Munich. Especially the *10 o'clock Morning Tea/Coffee-break*-group (H.O., C.S, c.S, S.R, G.'M'.M. and many more), which discussed anything from science, university gossip, world politics or why this break is productive. Another important institution was the *Kaffee-Kombinat* (R.M., S.R., c.S., G.'M'.M.), which entered the fray for better coffee. At some point there will be espresso!

Thanks to Harald Oberhofer, Georg Micheltisch and Simon Rittmeyer who proof-read the thesis and helped to identify many smaller and larger issues to change and improve. Special thanks to Simon Rittmeyer for providing extensive typesetting and layout feedback and finding every single font `mismatch` in this thesis.

Many thanks go to my family and my wife for their constant support and their ability to listen to strange topics and explanations, until usually a solution for the problem in question presented itself.

Finally I should mention the financial support by the Solar Technologies Go Hybrid initiative of the State of Bavaria and computational resources provided by the Leibniz Supercomputing Centre.