DISS. ETH NO. 25154

Combining Ehrenfest Molecular Dynamics with Linear Scaling and Subsystem DFT: Implementation and Application in Device Simulation

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

Samuel Thomas ANDERMATT

M.Sc. in Materials Science, ETH ZURICH

born on 21.08.1987

citizen of Baar ZG

accepted on the recommendation of

Prof. Dr. Mathieu LUISIER, examiner

Dr. Joost VANDEVONDELE, co-examiner

Prof. Dr. Jürg HUTTER, co-examiner

2018

Acknowledgements

First I would like to thank my two supervisors, Prof. Joost VandeVondele who supervised me during my first three years and Prof. Mathieu Luisier who took me over after Prof. VandeVondele moved on to a new position. I thank them both for always being kind, approachable, and supporting. My sincere thanks also go to Prof. Jürg Hutter for co-examining my PhD defence.

I would also like to thank all group members of the Nanoscale Simulations and Nano-TCAD group. I would like to especially mention Florian Schiffmann for helping me in the initial stages, when I started to work with the code he developed and which was largely the starting point of my PhD, Ole Schütt for all the helpful advice when I was working next to him, Fabian Ducry for the fruitful exchanges in the projects on which we collaborated and most of all I would like to thank Mohammad Hossein Bani-Hashemian, who came with me from the Nanoscale Simulations to the Nano-TCAD group for all the professional and personal discussions during the time we worked together. Furthermore, I would like to thank Martin Häufel for his work on the subsystem DFT graph colouring, Sascha Brücks for providing NEGF calculations, and Jinwong Cha for his contributions to the TiO_2 nanoparticle simulations.

I am grateful to my parents, Renate and Severin Andermatt, for their continuous support and also to my brother, who was always there to help me with graphics, whenever I needed him.

Finally, I am also thankful to the Swiss National Supercomputing Centre (CSCS) for providing the computational resources, I needed during my PhD.

3

Abstract

The speed of computers has been steadily increasing over the past decades due to hardware and software progresses. This has enabled the use of numerical simulations to support experiments by either providing additional insight or guidance through predictions. Through the increasing computer power, ever larger and more accurate simulations have been made possible. Still, the increase in model complexity and structure size does not only demand better computers, but also more efficient algorithms. In one of the most successful quantum mechanical theories, which is density functional theory (DFT), a lot of progress has been made in the last two decades to simulate large systems more rapidly. However, most of the progress has been concerned with ground-state applications, where the electronic dynamics are not explicitly simulated. The goal of this work is to build upon the success in ground-state calculations and generalise the existing approaches to include the electron dynamics.

For large systems, the computational performance can be enhanced by exploiting the near-sightedness of electronic matter, which captures the intuitive fact that local properties should only depend on the local environment. The electron dynamics are then included through a method called Ehrenfest molecular dynamics (EMD), which was implemented in the DFT software package CP2K. Furthermore, another approach that is commonly used to speed up ground-state calculations, namely subsystem DFT, was implemented more efficiently in CP2K and combined with EMD. Simulations were carried out to both validate the method and its implementation as well as to gain insight into the behaviour of nanoscale components.

Zusammenfassung

Aufgrund von technologischen Durchbrüchen sowohl auf Seiten der Hardware wie auch der Software, hat die Rechnerleistung von Computern hat über die Jahrzehnte stetig zugenommen. Diese Entwicklung erlaubt die Verwendung numerischer Simulationen zur Unterstützung von Experimenten, entweder indem sie ein tieferes Verständniss der experimentellen Daten ermöglichen oder als Richtschnur für die Planung von Experimenten verwendet werden. Durch die gestiegene Rechnerleistung werden immer grössere und genauere Simulationen möglich. Durch diese Entwicklung werden aber neben besseren Computern aber auch bessere Algorithmen benötigt. In den letzten beiden Jahrzehnten wurde dabei in einer der bedeutendsten Quantenmechanischer Theorien, genannt Dichtefunktionaltheorie (DFT), viel Fortschritt erzielt. Das meiste dieses Fortschritts beschränkt sich aber auf Grundzustandssimulationen, in denen die Elektronendynamik nicht explizit simuliert wird. Das Ziel dieses Werks ist es auf den Erfolgen für Grundzustandssimulationen aufzubauen und die bestehenden Ansätze zu verallgemeinern um die Dynamik der Elektronen mit einzubeziehen.

Dies kann durch die Ausnützung eines Effekts, der die Nahsichtigkeit elektronischer Masse genannt wird, erreicht werden. Dieser Effekt beschreibt die intuitive Vorstellung, dass lokale Eigenschaften durch die lokale Umgebung bestimmt werden. Die Elektronendynamik wird dann mittels einer Methode namens Ehrenfest Molekulardynamik (EMD) simuliert, welche in dem Softwarepaket CP2K implementiert wurde. Ausserdem wird ein weiterer Ansatz der ebenfalls verwendet wurde um Grundzustandssimulationen zu beschleunigen, genannt Subsystem DFT, effizienter in CP2K implementiert und kom-

biniert mit EMD. Simulationen wurden durchgeführt um sowohl die Korrektheit der Methode wie auch der Implementierung zu demonstrieren. Des weiteren wurden auch Simulationen durchgeführt zur Untersuchung des Verhaltens von Komponenten im Nanometerbereich.

Contents

Chapter 1

Introduction

In 1941 Konrad Zuse developed the first universal computer.[1] Since then the speed of computers has grown tremendously. For decades, the computer development has followed Moore's scaling law, which predicts that the number of transistors per area exponentially increases,[2] which has subsequently led to an exponential increase in the performance of computers. Driven by their always improving capabilities, computers have transformed the society on every level. In science, they have enabled a new tool for scientific research, namely numerical simulations. As more powerful computers have been made available, such simulations have become faster, more accurate, and cheaper. However, to optimally exploit these hardware improvements, it is necessary to develop new algorithms that can take advantage of them. The methods which were used for the simulations of small systems in the past are usually inadequate for present-day simulations.

In the field of quantum chemistry, a theory has proven to be very powerful for the numerical simulation of atomic systems and the extraction of their material properties, namely the density functional theory (DFT).[3, 4] DFT is considered to have a good ratio between the computer time needed to perform a simulation and the accuracy it provides. One important measure to characterise the performance of a method is the so-called computational complexity. It describes the relation between the required computer power for a given simulation and the size of the considered system. Traditionally, DFT has

cubic complexity, meaning that doubling the size of a simulated system induces an eight-fold increase in required computational power. The complexity of DFT is better than many other quantum chemistry theories such as the Hartree–Fock [5] or the coupled-cluster [6, 7] method. Nonetheless, for large systems the computational costs become prohibitive so that better algorithms are needed.

In 1996 Walter Kohn, one of the original inventors of DFT who was later awarded the Nobel Prize in Chemistry for his work, introduced a principle, which he called the nearsightedness of electronic matter. [8] This principle dictates that the behaviour of electrons is determined by their local environment. For large systems, the cost of doubling the system size should therefore be dominated by the cost of two separate systems with little overhead. Only few of the electrons can "see" beyond the boundary of the two systems, resulting in a small cost to simulate the interaction between them. In other words, this principle paved the way for the development of linear scaling DFT simulations, where the required computational effort is proportional to the size of the simulated system. Since Kohn introduced the nearsightedness principle, linear scaling DFT has been implemented in a number of quantum chemistry codes, [9–14] among them CP2K, [15, 16] the code used for the majority of this work. This breakthrough has enabled the simulation of millions of atoms, [9, 10] instead of hundreds or thousands as previously possible.

However, all this progress has been focused on ground-state calculations, which assume that electrons are always in the configuration with the lowest possible energy. There are many processes, in which this approximation is violated, for example electron transfer processes or photo absorption. In these cases, the electron movement has to be simulated explicitly. This can be done through real-time propagation (RTP) or Ehrenfest molecular dynamics (EMD). Both types of method allow the use of DFT without making the assumption of an electronic ground-state. Like traditional DFT, RTP and EMD have so far exhibited a cubic scaling of their computation time with respect to the system size and have thus been restricted to the systems of at most a few hundred atoms. The goal of this work was therefore to enable linear scaling for these methods as well by employing the same general approaches that were applied to ground-state simulations in the past.

Figure 1.1: Recent advancements allow simulations at the DFT level of accuracy with linear time complexity. However, while it is possible to use DFT to explicitly simulate the electron dynamics, the linear scaling simulations are so far limited to the electronic ground-state. The goal of this thesis is to remove this shortcoming and thus enabling large scale simulation of the electronic processes.

The thesis is organised as follows: Chapter 2 will provide the theoretical background to understand DFT, RTP, EMD, and the algorithms used to make these simulations faster. Chapter 3 will then give information about the issues encountered during the implementation inside CP2K. In Chapter 4 the simulations that were performed using this code will be presented. Conclusions will be drawn in Chapter 5.

Bibliography

[1] Raúl Rojas. How to make Zuse's Z3 a universal computer. *Ann. Hist. Comput.*, 20(3):51–54, 1998.

[2] Gordon E. Moore. Cramming more components onto integrated circuits. *Electronics Magazine*, 1965.

[3] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864, 1964.

[4] Wallter Kohn and Lu J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[5] Vladimir Fock. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys. A–Hadron Nucl.*, 61(1):126–148, 1930.

[6] Fritz Coester and Hermann Kümmel. Short-range correlations in nuclear wave functions. *Nucl. Phys.*, 17:477–485, 1960.

[7] Jiří Čížek. On the correlation problem in atomic and molecular systems. calculation of wavefunction components in ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys.*, 45(11):4256–4266, 1966.

[8] Walter Kohn. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.*, 76:3168–3171, Apr 1996.

[9] Joost VandeVondele, Urban Borstnik, and Jürg Hutter. Linear scaling self-consistent field calculations with millions of atoms in the condensed phase. *J. Chem. Theory Comput.*, 8(10):3565–3573, 2012.

[10] David R. Bowler and T. Miyazaki. Calculations for millions of atoms with density functional theory: linear scaling shows its potential. *J. Phys. Condens. Mat.*, 22(7):074207, 2010.

[11] Chris-Kriton Skylaris, Peter D. Haynes, Arash A. Mostofi, and Mike C. Payne. Introducing onetep: Linear-scaling density functional simulations on parallel computers. *J. Chem. Phys.*, 122(8):084119, 2005.

[12] Emilio Artacho, Eduardo Anglada, Oswaldo Diéguez, Julian D. Gale, Alberto García, Javier Junquera, Richard M. Martin, Pablo Ordejón, José Miguel Pruneda, Daniel Sánchez-Portal, et al. The SIESTA method; developments and applicability. *J. Phys. Condens. Mat.*, 20(6):064208, 2008.

[13] Xinming Qin, Honghui Shang, Hongjun Xiang, Zhenyu Li, and Jinlong Yang. HONPAS: A linear scaling open-source solution for large system simulations. *Int. J. Quant. Chem.*, 115(10):647–655, 2015.

[14] Stephan Mohr, Laura E Ratcliff, Luigi Genovese, Damien Caliste, Paul Boulanger, Stefan Goedecker, and Thierry Deutsch. Accurate and efficient linear scaling dft calculations with universal applicability. *Phys. Chem. Chem. Phys.*, 17(47):31360–31370, 2015.

[15] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103 – 128, 2005.

[16] Jürg Hutter, Marcella Iannuzzi, Florian Schiffmann, and Joost VandeVondele. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 4(1):15–25, 2014.

Chapter 2

Theory

2.1 Introduction

This chapter gives an overview of the necessary theoretical background to understand the work presented in this thesis. We shall start with an introduction about density functional theory (DFT), [1, 2] the underlying theory behind all simulations in this work. DFT is in principle an exact theory. It has been widely and successfully applied in chemistry, materials science, and condensed matter physics, because it offers a high accuracy compared to the required computer time. Another advantage comes from the ease of its application to new structures without the need for prior empirical investigations.

After the basics of DFT are introduced, two sections will describe possible methods to increase the efficiency of DFT calculations. The first is linear scaling DFT, a way to make the computational costs of DFT calculations scale linearly with respect to the system size, thus enabling the treatment of large systems. The second method is subsystem DFT, where the computational performance is enhanced through a division of the system into a set of smaller subsystems. The benefit of the latter approach is twofold: First, the prior knowledge about the molecular structure of the system can be included. Secondly, the accuracy can be tailored for specific regions of the system, concentrating most of the computer time to critical areas, while

treating the rest at a small fraction of the usual costs.

Finally, an overview of the different ways to move from the static picture to dynamic simulations will be presented. A range of methods exist to simulate the system dynamically (Fig. 2.1), they differ from each other by the particles that are dynamically simulated.

The algorithms presented here have been implemented in the DFT software package CP2K. [3, 4] Since all simulations were performed for systems containing an even number of electrons, spin-polarisation effects will not be considered in this work, although DFT can naturally include them. To simplify the equations, all formulas given in this thesis will be expressed in atomic units.

2.2 Density Functional Theory

2.2.1 Quantum Mechanical Background

Schrödinger Equation

On a fundamental level, atomistic systems are described by the Schrödinger equation. There are two formulations, the more general time-dependent form: [5, 6]

$$i\frac{\partial}{\partial t}\psi = \hat{H}\psi, \tag{2.1}$$

and the simpler time-independent form, which applies to stationary states, where the density of electrons does not change over time:

$$\hat{H}\psi = E\psi. \tag{2.2}$$

Here, $\hat{H}$ is the Hamiltonian operator that describes the time evolution and the total energy of the system, E, denotes the energies associated with the eigenfunctions of the Hamiltonian, and ψ the many-particle wave-function. These quantities are related to the probability distribution of the electronic and atomic positions in the system.

Analytically, the Schrödinger equation can only be solved for the smallest systems, such as a single hydrogen atom. The difficulty lies in the high dimensionality of ψ, which is a $(3N_{el} + 3N_{cores})$ dimensional object. In effect it describes the density of every electron and atomic

Electrons

		Ground-State	Free
Atomic Cores	**Fixed**	Ground-State Energy (Energy)	Real Time Propagation (RTP)
	Free	Ground-State Molecular Dynamics (MD)	Ehrenfest Molecular Dynamics (EMD)

Figure 2.1: Overview of the available methods to simulate the evolution of an atomistic system. They are defined through the choice of the particles to be simulated dynamically. The atomic cores can either be fixed in space (Energy, RTP), or freely move according to their inertia and forces (MD, EMD). The electrons can either be assumed to be in the electronic ground-state of lowest possible energy (Energy, MD), or explicitly moved over time (RTP, EMD).

core at every point in space. For practical calculations, a range of approximations have to be used to make the Eq. (2.1) or (2.2) solvable, which will be covered from here on.

Born-Oppenheimer Approximation

The first approximation is the separation of the electronic and the nuclear part of the wave-function. The validity of this separation is justified based on the large difference between the mass of the atomic cores and of the electrons. [7] The nuclear part of the wave-function can be further simplified by treating the cores as classical particles, thus moving from a probability distribution to a list of coordinates.

Practically, the nucleus and the inner electrons are often grouped together as a single object. The interactions with the rest of the system are then simulated with so called pseudo-potentials. The reasoning is that most of the chemistry is only related to the outer valence electrons and the inner electrons remain closely bound to the nucleus.

After these simplifications, a $3N_{el}$ dimensional problem is obtained, which represents a significant improvement, but still intractable for most calculations.

2.2.2 Single Particle Wave-Functions

The next step consists in approximating the many-particle wave-function through one-electron functions. Ideally, we would like to represent ψ as a simple product of independent one-electron wave-functions. However, such a representation violates the antisymmetry property of the electronic wave-function, [8, 9] which is related to the Pauli exclusion principle. [10] The antisymmetry principle states that the sign of ψ changes when two electrons are exchanged:

$$\psi(x_1, ..., x_i, ..., x_j, ..., x_n) = -\psi(x_1, ..., x_j, ..., x_i, ..., x_n). \qquad (2.3)$$

Instead of a product, the wave-function can be approximated as the sum of product given by a determinant. [11] Such an approach intrin-

sically includes the antisymmetry requirement:

$$\psi(x_1, ..., x_n) \approx \frac{1}{\sqrt{N!}} \det \begin{vmatrix} \psi^1(x_1) & \cdots & \psi^n(x_1) \\ \vdots & \ddots & \vdots \\ \psi^n(x_1) & \cdots & \psi^n(x_n) \end{vmatrix}. \qquad (2.4)$$

This representation is called the Slater-determinant. With the determinant representation it also becomes obvious that no two one-electron orbitals can be identical, as the determinant of a matrix with two identical columns is zero. The one-electron wave-functions will be referred to as molecular orbitals in the remaining of this thesis.

2.2.3 Linear Combination of Atomic Orbitals

An analytical calculation of the molecular orbitals, as introduced above, is typically not feasible. Therefore, we use a linear combination of atomic orbitals (LCAO), [12] which represents the molecular orbitals through a finite set of atom-centred basis functions:

$$\psi^j(r, t) = \sum_\alpha v_\alpha^j \phi_\alpha(r - R_{A_\alpha}), \qquad (2.5)$$

with ψ^j being the orbital of electron j, ϕ_α an atomic basis functions centred around atom A_α, and v_α^j a numerical expansion coefficient. Similarly, the one-electron wave-functions could also be represented by atom-independent basis functions, such as plane-waves or wavelets. [13] Plane-waves have the advantage that the accuracy of the basis set can be systematically enhanced by increasing the energy limit up to which basis functions are generated. However, atom centred basis functions typically require a smaller number of elements to achieve a good representation of the molecular orbitals, which has performance benefits. The CP2K software generally uses an atom-centred basis set, but performs part of the calculations in an auxiliary plane-wave basis, which is called the Gaussian and plane waves (GPW) method. [3]

Ideally, atomic basis functions would have an exponential dependence on the distance from their centres, such functions are named Slater-type orbitals. [14] However, in practice, it is computationally favourable to use Gaussian type orbitals (GTO), because the overlap of Gaussian functions can be analytically calculated. Exponential

functions can be approximated as a linear combination of multiple Gaussians, which form a contracted Gaussian basis function: [15]

$$\phi_\alpha(r) = \sum_k N_k c_k (x - x_{A_\alpha})^l (y - y_{A_\alpha})^m (z - z_{A_\alpha})^n e^{-\alpha_k (r - R_{A_\alpha})^2},$$

(2.6)

where N_k is a normalisation constant and c_k the contraction coefficients. Contracted GTOs differ from each other by the number of primitive Gaussian basis functions, contraction coefficients and exponents. The parameters are typically optimised with regards to specific pseudo-potentials to provide accurate results. [16]

Transformation to an Orthonormal Basis

For reasons that will be explained later, molecular orbitals should be orthonormal. However, this orthonormality is not reflected in the vectors v in Eq. 2.5, as the underlying basis functions ϕ are a non-orthonormal basis themselves. In many situations, it is therefore convenient to transform the basis into an orthonormal basis:

$$\langle \psi^i | \psi^j \rangle = \langle \phi^T v^i | \phi^T v^j \rangle = (v^i)^T S v^j$$
$$= (S^{\frac{1}{2}} v^i)^T (S^{\frac{1}{2}} v^j) = (v_{ortho}^i)^T v_{ortho}^j = \delta_{ij}, \qquad (2.7)$$

The notation above uses the bra-ket notation, [17, 18] which denotes an inner product, defined as:

$$\langle a | b \rangle = \int a(x)^* b(x) dx. \qquad (2.8)$$

The matrix S is a symmetric matrix, which is referred to as the overlap matrix ($S_{\alpha\beta} = \langle \phi_\alpha | \phi_\beta \rangle$). When the basis is changed for molecular orbitals, the matrix representation of the Hamiltonian H has to be modified accordingly.

$$E = v^T H v = v_{ortho}^T H_{ortho} v_{ortho}$$
$$= (S^{\frac{1}{2}} v)^T H_{ortho} S^{\frac{1}{2}} v = v^T S^{\frac{1}{2}} H_{ortho} S^{\frac{1}{2}} v \qquad (2.9)$$
$$\Rightarrow H_{ortho} = S^{-\frac{1}{2}} H S^{-\frac{1}{2}}.$$

The new basis functions are no longer local in space. However, in many situations, it is more suitable to perform intermediate steps of a calculation in the orthonormal basis before transforming the results into the non-orthonormal basis to calculate the real-space density.

2.2.4 Hohenberg-Kohn Theorem

In their seminal paper from 1964,[19] Hohenberg and Kohn proved that all properties of a system in the electronic ground-state can be obtained from the electronic density alone. This finding massively simplifies calculations as the electronic density is only a 3 dimensional object. The total energy of the system is then given by:[2, 19]

$$E = \int v(r)\rho(r)dr + \int\int \frac{\rho(r)\rho(r')}{|r-r'|}drdr' + G[\rho], \qquad (2.10)$$

where

$$\rho(r) = N_{el}\int |\psi(r, r_2, ..., r_n)|^2 dr_2...dr_n \qquad (2.11)$$

is the electron density at position r, $v(r)$ is an external potential, and $G[\rho]$ is a universal functional of the density. The remaining problem is that $G[\rho]$ is unknown. There exists a flavour of DFT, orbital-free DFT, where $G[\rho]$ is approximated with empirical functions.[20] This approach typically requires little computational effort, but suffers from poor accuracy.

2.2.5 Kohn-Sham

In 1965, Kohn and Sham[2] went one step further and calculated the kinetic energy from the Slater-determinant representation of the wave-function:

$$E_{kin} = -\frac{1}{2}\langle\psi|\nabla^2|\psi\rangle$$

$$\approx -\frac{1}{2N!}\langle\det\begin{vmatrix}\psi^1(x_1) & \cdots & \psi^n(x_1)\\ \vdots & \ddots & \vdots \\ \psi^n(x_1) & \cdots & \psi^n(x_n)\end{vmatrix}|\nabla^2|\det\begin{vmatrix}\psi^1(x_1) & \cdots & \psi^n(x_1)\\ \vdots & \ddots & \vdots \\ \psi^n(x_1) & \cdots & \psi^n(x_n)\end{vmatrix}\rangle,$$
$$(2.12)$$

which according to the Slater-Condon rules [11, 21] takes the following simple form for orthonormal orbitals:

$$E_{kin} \approx -\frac{1}{2} \sum_{i=1}^{N} \langle \psi^i | \nabla^2 | \psi^i \rangle . \tag{2.13}$$

The functional $G[n]$ introduced by Hohenberg-Kohn, was then separated into a component representing the kinetic energy, $E_{kin}[\psi]$, and another one labelled, E_{XC}, including the remaining contributions to the energy, namely the exchange and the correlation energies: [2]

$$G[\rho, \psi] = E_{kin}[\psi] + E_{XC}[\rho], \tag{2.14}$$

The exchange energy is connected to the fact that the electronic density does not have the antisymmetry property of the Slater-determinant and must therefore be corrected. On the other hand, the correlation energy accounts for the correlation between the electrons. This term was neglected when the electrons were considered non-interacting. For sufficiently slowly varying densities, the local density approximation [2] (LDA) can be employed, to derive an expression for $E_{XC}[\rho]$ from $\int \rho(r) \epsilon_{XC}(\rho(r)) dr$. The more accurate generalised gradient-approximation (GGA) does not only depend on the electronics density, but also on its gradient ($E_{XC}[\rho] = \int \rho(r) \epsilon_{XC}(\rho(r), \vec{\nabla} \rho(r)) dr$). While the exact form of $E_{XC}[\rho]$ is not known, a vast range of exchange correlation functionals [22–27] have been developed as approximations. Calculations employing the Kohn-Sham matrix representation of the Hamiltonian H_{KS} are in general more accurate than orbital free DFT, this justifying the higher computational cost.

2.2.6 Self-Consistent Field Procedure

To find the ground-state density of a system, it is necessary to perform an self-consistent field (SCF) iteration procedure. The electronic density defines H_{KS}. The ground-state density itself is given by the N eigenvectors of lowest energy of H_{KS}. The electronic density and H_{KS} are considered consistent, when H_{KS} defines the same ground-state density as it was itself constructed from. The SCF procedure is

an iterative process in which such a density is established. The most primitive approach is to calculate the H_{KS} and density in turn from each other. However, such a procedure is not guaranteed to converge and most likely does not. Therefore, typically more sophisticated methods are employed.

One approach to increase the stability is to employ a mixing procedure. The simplest mixing approach is to combine the density of two successive iterations with a constant mixing ratio m:

$$\rho_i = m\rho(H_{KS,i}) + (1-m)\rho(H_{KS,i-1}), \qquad (2.15)$$

where $\rho(H_{KS,i})$ is the density given by the eigenstates of the Kohn-Sham matrix in the ith iteration step. To enhance the speed of convergence, the more elaborate Direct Inversion of the Iterative Subspace (DIIS) method [28] is commonly employed, where the density is a mix of all previous iterations:

$$\rho_{i+1} = \sum_{k=1}^{i} m_k \rho_k, \qquad \sum_{k=1}^{i} m_i = 1 \qquad (2.16)$$

and the mixing coefficients m_t are calculated based on the measured error of previous iterations.

While a mixing procedure does increase the stability, the convergence is still not guaranteed. An alternative approach to obtain the ground-state is to calculate the derivative of the energy with respect to the coefficients c of the molecular orbitals. The issue is that traditional minimisation techniques such as conjugate-gradient (CG) or Broyden-Fletcher-Goldfarb-Shanno (BFGS) [29–33] method, do not intrinsically respect the orthonormality of the molecular orbitals. This issue can be circumvented by the orbital transformation (OT) [34, 35] method. In the OT method the problem is reformulated as finding a rotation matrix which rotates the initial guess into the ground-state density, thus intrinsically enforcing orthonormality. The advantage of this approach is twofold. First, it reduces the dimensionality and second, it transforms the constrained optimisation into an unconstrained optimisation. This allows the minimisation with a method like CG which is guaranteed to converge, although potentially faster but less stable approaches such as BFGS remain a possibility. Furthermore,

even the before-mentioned mixing approach DIIS can be used in this constrained subspace, benefiting from the lack of the need to diagonalise $H_{KS}(t)$, to obtain $\rho(t)$, as $\rho(t)$ be directly calculated from the rotation of the initial guess.

2.2.7 DFT Summary

DFT is in principle an exact theory that makes no approximations beyond the Born-Oppenheimer separation of the electronic and nuclear wave-functions. In practice, only $E_{XC}[\rho]$ must be approximated, whereas the kinetic energy is calculated from the Slater-determinant construction of the wave-function. Nonetheless, DFT is considered an ab-initio theory, which means that it comes from first-principles, or fundamental physical laws. This differentiates it from empirical and semi-empirical methods, which require a range of parameters as input, before they can be applied to a system. Such parameters have to be obtained either from more sophisticated ab-initio computational methods or from experimental observations. The only approximations used in DFT are typically rather transferable. DFT therefore has the advantage over many other approaches that it can be readily employed for new structures without a prior investigation of them. Furthermore, it is more computationally affordable than more rigorous ab-initio methods, e.g. Hartree-Fock, [36] Coupled Cluster, [37, 38] Configuration interaction [11, 39] and Quantum Monte Carlo. [40, 41]

2.3 Linear Scaling DFT

Traditionally, the cost to perform a DFT calculation scales with the third power of the number of atoms.

$$T = cN_{atoms}^3, \tag{2.17}$$

with T being the required computer time, c a constant prefactor that depends on the implementation and N_{atoms} the number of atoms in the system. An algorithm of such computational cost is said to have a computational complexity of $O(N^3)$. With this complexity the cost to investigate larger systems quickly escalates with the system size. It is therefore of interest to investigate algorithms which have a better,

preferably linear $O(N)$, computational complexity. For small systems such algorithms might be slower due to their prefactor but for a sufficiently large system they will always outperform the $O(N^3)$ algorithms.

DFT algorithms that aim for $O(N)$ complexity typically try to exploit the so called nearsightedness of electronic matter. [42, 43] It is a principle that without long range interactions, the local properties can be described by the local neighbourhood. For this reason, the cost to calculate each individual atom should not depend on the system size and be proportional to the number of atoms $(O(N))$. Traditional DFT does not exploit nearsightedness. Instead the eigenstates of the Hamiltonian are calculated, which are typically spread across the whole system. The computational cost is therefore proportional to the number of eigenstates, times the number of eigenstates with which the orthogonality has to be enforced, times the cost of orthogonalisation. All three factors scale individually with the system size, which is why the total complexity is $O(N^3)$. [44] However, the explicit calculation of these eigenstates is not necessary, as the system is completely described by the electronic density [19] or equivalently, within the independent electron approximation, the one particle density matrix P. The elements of the density matrix are known to decay exponentially with the distance for insulators [44] and metals at finite temperatures. [45] A range of algorithms exist to calculate the electronic density with $O(N)$ complexity. An overview can be found in Ref. [44]. The following section shows how linear scaling DFT has been implemented in CP2K [46] and is based on a previously published text in Ref. [47].

2.3.1 Density Matrix Purification

Linear scaling DFT, as implemented in CP2K, [46] is based on density matrix purification. This method uses the fact that in the ground-state the density matrix:

$$P = \sum_{i=1}^{N_{el}} |\psi^i\rangle \langle\psi^i|, \tag{2.18}$$

can be written as a function of the Kohn-Sham matrix

$$P = \frac{1}{2}(I - sign(S^{-1}H - \mu I))S^{-1}, \tag{2.19}$$

where H is the Kohn-Sham matrix, S is the overlap matrix and μ is the chemical potential. This formulation means that the density matrix is given by all eigenvectors of H, which have an eigenvalue lower than the chemical potential. The chemical potential has to be chosen such that the total number of eigenvalues below the chemical potential is equal to the number of electrons.

The sign function $sign(A) = A(A^2)^{-\frac{1}{2}}$ can be computed iteratively through matrix multiplications alone

$$X_{n+1} = \frac{1}{2}X_n(3I - X_n^2). \tag{2.20}$$

This formulation allows the exploitation of the previously mentioned exponential decay of P for insulators in the ground state. [44] If small elements are thus truncated to zero (filtered), P is a sparse matrix for large systems. The same holds for the S and H matrices. If all multiplications are implemented using sparse matrix-matrix multiplication, for example using the DBCSR library, [48] the computation of the density matrix, starting from the H and S matrix can be performed in $O(N)$ cost. However, the sparsity of the matrices is not constant during the multiplication, i.e. if two sparse matrices are multiplied, the sparsity pattern of the result contains more entries. To retain sparsity, small elements, below the filtering threshold, are filtered away from the resulting matrix. Nevertheless, the density matrix is usually significantly less sparse than the Kohn-Sham or overlap matrices. This filling-in of the sparsity pattern contributes to the prefactor of the $O(N)$ algorithm.

We note that the sign matrix based formulation given above is intuitive and useful, but various other linear-scaling techniques are available in CP2K, including the TRS4 method [49] and curvy-steps optimisation. [50] These methods are advantageous for example when the chemical potential is not known, or the self-consistent electronic structure converges slowly.

2.4 Subsystem DFT

There is another approach to achieve linear scaling DFT, or at least significantly speed up DFT, which is called subsystem DFT. It is based on the exploitation of chemical structure that is often present inside the atomic systems. In many cases a system can be further subdivided into molecules. Atoms inside the same molecule have a much stronger interaction with their neighbouring atoms of the same molecule, which are called chemical bonds, than with atoms of other molecules. This allows for a divide and conquer approach, where the computational effort scales linearly with the number of molecules and it furthermore allows for a focus of the computational effort into the regions of main interest, by only applying this scheme in low-interest regions and using standard DFT where more accuracy is required. The following section introduces subsystem DFT and has been previously published in the same paper as the section before. [47]

As described above, in the Subsystem (SS) DFT method, [1, 51, 52] the system is subdivided into (weakly) interacting subsystems, for example molecules. The energy of the full system is obtained from the Hohenberg-Kohn (HK) energy functional [19] for the full system, but corrected by the difference between the Kohn-Sham [2] (KS) and HK energy functionals for each of the subsystems:

$$
\begin{aligned}
E_{SSDFT} =& E_{HK}[\rho] + \sum_i E_{KS}[\rho_i[\{\psi_j\}^i]] - E_{HK}[\rho_i] \\
=& E_s[\rho] + \sum_i (T_s[\{\psi_j\}^i] - E_s[\rho_i]) + \\
& E_{ext}[\rho] + E_{Coul}[\rho] + E_{xc}[\rho],
\end{aligned}
\tag{2.21}
$$

where we have used the definitions of the HK energy functional:

$$
E_{HK}[\rho] = E_s[\rho] + E_{ext}[\rho] + E_{Coul}[\rho] + E_{xc}[\rho],
$$

and KS energy functional:

$$
E_{KS}[\rho[\{\psi_j\}]] = T_s[\{\psi_j\}] + E_{ext}[\rho] + E_{Coul}[\rho] + E_{xc}[\rho].
$$

The index i iterates over all subsystems, while the index j iterates for each subsystem over the electrons of that subsystem. E_{ext}, E_{Coul} and

E_{xc} stand for the external, Coulomb and exchange and correlation energy, respectively. T_s is the KS kinetic energy, which is computed from the single-particle orbitals, while E_s is the HK kinetic energy, which is computed from the electron density. The total density is the sum of the subsystem densities ($\rho = \sum_i \rho_i$), and the subsystem density itself is obtained from the set of molecular orbitals $\{\psi_j\}$ of each of the subsystems as

$$\rho[\{\psi_j\}] = \sum_j ||\psi_j||^2. \qquad (2.22)$$

Note that these molecular orbitals are enforced to be orthogonal only within the set belonging to a single subsystem. In our implementation, these molecular orbitals are expanded in the basis functions of corresponding subsystems only. The lack of orthogonality of the wave functions between subsystems is taken into account by the HK kinetic energy functional. However, since the exact form of this functional is unknown, and must be approximated, the interaction between subsystems is usually described somewhat less accurately in SS than in KS DFT. As in the earlier CP2K implementation [53] this functional is minimised fully self-consistently with respect to the expansion coefficients of the molecular orbitals, which implies that ionic forces are obtained straightforwardly.

In this SS DFT approach, the Kohn-Sham matrix, the density matrix and the overlap matrix of the full system have a clear structure, as they can be made block-diagonal, where each block corresponds to the sub-matrix of a single subsystem. This structure can be exploited to perform the calculations more efficiently. Traditionally, the equations are solved by diagonalising each sub-block, and parallelising the calculation over each of these blocks. While this procedure works well for small systems or homogeneous liquids, it has limitations as soon as the subsystems are very different in size. Indeed, if very large and small subsystems are present at the same time, load balancing the calculation can become problematic, and sometimes calculations would not fit on a single processor further complicating the implementation. Finally, as the very large fragments increase in size, diagonalisation based techniques are not suitable anymore. We therefore prefer to work with the matrix of the full system, and instead

solve the equations with the same techniques as employed in linear-scaling calculations, i.e. with method that directly benefit from the sparsity of the matrices describing the system. In the following, we show that the SS DFT implementation can be combined very easily with linear-scaling DFT, addressing the above mentioned problems, and ultimately allowing for the efficient simulation of very large systems in an environment of molecules.

2.4.1 Combining Subsystem DFT with Linear Scaling DFT

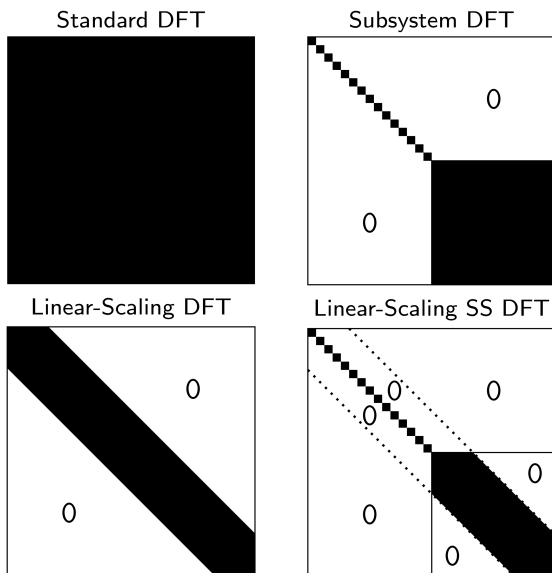Standard DFT — Subsystem DFT — Linear-Scaling DFT — Linear-Scaling SS DFT

Figure 2.2: An illustration of the sparsity pattern of the density matrix of a 1D system, which includes several small molecules and one large subsystem. Linear-scaling DFT and the SS method complement each other to achieve optimal performance.

Within the SS DFT method, using linear-scaling algorithms based on multiplication, P automatically has the same block diagonal struc-

ture of S and H. This is a result of the fact that a multiplication of block diagonal matrices with matching block sizes results again in a block diagonal matrix of equal pattern. Fill-in is thus not present, and the method has a small prefactor. In principle, no changes are needed to the linear-scaling code to be effective in solving the SS DFT equations. We note that subsystems of suitable integer charge are automatically obtained, as a single chemical potential is employed across the system. Furthermore, the combination of the SS DFT with linear-scaling DFT allows for speedups at all length scales as illustrated with the sparsity pattern of P in Figure 2.2. Small subsystems, such as solvent molecules, benefit from the SS DFT scheme, and result in small diagonal blocks. Large subsystems benefit from the exponential decay of P since their blocks are large enough to become sparse and enter the linear-scaling regime. The method is furthermore naturally parallel and GPU accelerated, since the underlying sparse matrix multiplication library is. [48, 54] As an added benefit of this approach, the charge of the individual subsystems is not fixed, but naturally adjusts, in integer steps, to the global chemical potential of system, which is, for example, advantageous if ions are present in solution.

For the same reasons, why it can be combined with linear scaling DFT it can also out of the box be combined with density matrix based EMD or RTP, which will be introduced further down.

2.5   Molecular Dynamics

Electrons

	ground-state	free
fixed	Energy	RTP
free	MD	EMD

Atomic Cores

Molecular dynamics describes the movement of atoms over time. As established before, the cores are treated as classical particles. Therefore, their movement is described by Newtonian mechanics and the forces on the cores can simply be obtained from the derivative of the potential energy towards the positions of the atoms:

$$M_A \ddot{R}_A = -\frac{\partial U}{\partial R_A} - \langle \psi | \frac{\partial H_{el}}{\partial R_A} | \psi \rangle + F_{Pulay}, \tag{2.23}$$

where $\langle \psi | H_{el} | \psi \rangle$ contains the electron dependent contributions to the total energy ,U the remaining electron independent terms and F_{Pulay} the Pulay forces, [55] which originate from the position dependence of the basis set. M_A and R_A are the mass and position of atom A, respectively.

2.5.1 Velocity-Verlet Integration

The simplest way to calculate $R(t + \Delta t)$ and $\dot{R}(t + \Delta t)$ would be to deduce them directly from the position, velocity, and force at time t, though straightforward Taylor expansion:

$$R(t + \Delta t) = R(t) + \Delta t \dot{R}(t) + \frac{\Delta t^2}{2} \ddot{R}(t), \tag{2.24}$$

$$\dot{R}(t + \Delta t) = \dot{R}(t) + \Delta t \ddot{R}(t). \tag{2.25}$$

However, such a propagation scheme would not be stable and the energy would not be conserved over time. The reason for this discrepancy is that the result of Eq. 2.24 and 2.25 is not time-reversible. When we invert the sign of the time and attempt to move back in time, we find:

$$
\begin{aligned}
R_{bw}((t + \Delta t) - \Delta t) =& R(t + \Delta t) - \Delta t \dot{R}(t + \Delta t) + \frac{\Delta t^2}{2} \ddot{R}(t + \Delta t) \\
=& R(t) + \Delta t \dot{R}(t) + \frac{\Delta t^2}{2} \ddot{R}(t) - \\
& \Delta t (\dot{R}(t) + \Delta t \ddot{R}(t)) + \frac{\Delta t^2}{2} \ddot{R}(t + \Delta t) \\
=& R(t) + \frac{\Delta t^2}{2} (\ddot{R}(t + \Delta t) - \ddot{R}(t)) \neq R(t) \quad (2.26)
\end{aligned}
$$

In Eq. (2.26) R_{bw} refers to the backward computed original position, which is not equal to $R(t)$. For this reason, a number of time-reversible integration schemes have been developed. The approach implemented in CP2K and used for this thesis is the velocity-Verlet algorithm:[56, 57]

$$R(t + \Delta t) = R(t) + \Delta t \dot{R}(t) + \frac{\Delta t^2}{2} \ddot{R}(t), \qquad (2.27)$$

$$\dot{R}(t + \Delta t) = \dot{R}(t) + \frac{\Delta t}{2}(\ddot{R}(t) + \ddot{R}(t + \Delta t)). \qquad (2.28)$$

2.6 Real-Time Propagation

Electrons

	ground-state	free
fixed	Energy	RTP
free	MD	EMD

Atomic Cores

Real-time Propagation (RTP), also referred to as RT-TDDFT if the Hamiltonian is modelled through DFT, describes the evolution of the electronic wave-function over time. Contrary to ground-state calculations where ψ is given by the eigenstates of the Hamiltonian, H changes as a function of the time. As a consequence $\psi(t)$ becomes a function of the initial wave-function $\psi(0)$ and all states of the Hamiltonian between the time 0 and t. The propagation of $\psi(t)$ in RTP is based on the time-dependant Schrödinger equation:[5, 6]

$$i\frac{\partial}{\partial t}\psi = \hat{H}\psi. \qquad (2.29)$$

In the LCAO approximation, the derivative of the molecular orbital coefficients takes the following form: [58]

$$i\frac{\partial}{\partial t}v_{ortho} = H_{ortho}v_{ortho}$$

$$iS^{\frac{1}{2}}\frac{\partial}{\partial t}v = S^{-\frac{1}{2}}Hv \qquad (2.30)$$

$$\dot{v} = -iS^{-1}Hv,$$

where H is the Kohn-Sham matrix if DFT is employed. It can also be calculated based on different theories such as Hartree-Fock [36, 59, 60] or the tight-binding (TB) [61, 62] method. For a time-independent Hamiltonian the wave-function can be directly propagated by analytically solving Eq. (2.30):

$$v(t + \Delta t) = e^{-i\Delta t S^{-1} H} v(t). \qquad (2.31)$$

However, the exact propagation of the coefficients over time requires the solution of a time-ordered exponential: [63]

$$v(t + \Delta t) = \hat{U}(t, t + \Delta t)v(t) = \mathcal{T}exp^{-i\int_t^{t+\Delta t} S^{-1}H(\tau)d\tau}v(t). \qquad (2.32)$$

While the direct calculation of the time-ordered exponential is not feasible, it can be approximated by a range of propagators that depend on both $H(t)$ and $H(t+\Delta t)$. These propagators should be time reversible: [63] $U(t, t + \Delta t) = U^{-1}(t + \Delta t, t)$. Examples for such propagators are listed below. [63]
-Crank-Nicholson (CN):

$$\hat{U}_{CN}(t + \Delta t, t) = \frac{1 - \frac{i}{2}\Delta t\hat{H}(t + \frac{\Delta t}{2})}{1 + \frac{i}{2}\Delta t\hat{H}(t + \frac{\Delta t}{2})}; \qquad (2.33)$$

-Exponential midpoint (EM) rule:

$$\hat{U}_{EM}(t, t + \Delta t) = e^{-i\Delta t\frac{1}{2}S^{-1}\hat{H}(t + \frac{\Delta t}{2})}; \qquad (2.34)$$

-Enforces time-reversal symmetry (ETRS):

$$\hat{U}_{ETRS}(t, t + \Delta t) = e^{-i\frac{\Delta t}{2}S^{-1}H(t+\Delta t)}e^{-i\frac{\Delta t}{2}S^{-1}H(t)}; \qquad (2.35)$$

-forth order Magnus expansion (M(4)):

$$\hat{U}_{M(4)}(t, t + \Delta t) = -i\frac{\Delta t}{2}(\hat{H}(t_1) + \hat{H}(t_2)) - \frac{\sqrt{3}\Delta t^2}{12}[\hat{H}(t_1), \hat{H}(t_2)],$$

(2.36)

$$\text{where } t_{1,2} = t + (\frac{1}{2} \mp \frac{\sqrt{3}}{6})\Delta t$$

The Hamiltonians between t and $t + \Delta t$ are obtained through linear interpolation

$$\hat{H}(t + x\Delta t) = (1 - x)\hat{H}(t) + x\hat{H}(t + \Delta t).$$

(2.37)

All the propagators above maintain the orthonormality of the molecular orbitals $(v_{ortho}^{i\,T}(t + \Delta t)v_{ortho}^{j}(t + \Delta t) = \delta_{ij})$. This happens because they are matrix exponentials of skew-Hermitian matrices, which means matrices with an antisymmetric real and a symmetric imaginary part. Such matrix exponentials are so called unitary matrices. All their eigenvalues have a norm equal to 1, which corresponds to a rotation.

The aforementioned propagators use Hamiltonian after the time step $H(t + \Delta t)$, which is initially unknown. It has to be calculated from $v(t + \Delta t)$. A self-consistent iteration is therefore necessary.[63] Before the propagation, $v(t + \Delta t)$ and $H(t + \Delta t)$ have to be obtained through a self-consistent iteration scheme:[63]
(1) Guess an initial $H(t + \Delta t)$ for example through the always stable predictor-corrector (ASPC)[64] scheme.
(2) Propagate $v(t)$ to obtain $v(t + \Delta t)$.
(3) Calculate $H(t + \Delta t)$ from $v(t + \Delta t)$.
(4) repeat the steps 2 and 3 until self-consistency is reached.

2.7 Ehrenfest Molecular Dynamics

Electrons

	ground-state	free
fixed	Energy	RTP
free	MD	EMD

Atomic Cores

Ehrenfest molecular dynamics (EMD) combines the movement of the atomic cores from ground-state (gs) MD with the propagation of the wave-function from RTP. The terms describing the movements of electrons and cores are largely the same. However, when atom-centred basis sets are used, some additional terms have to be considered as the basis functions are implicitly time-dependant:

$$\phi(R(t), x) \neq \phi(R(t + \Delta t), x). \tag{2.38}$$

The dependence of the basis set on the moving atomic cores means that, the total derivative with respect to time has to be used:[58]

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \sum_{A=1}^{N} \dot{R}_A \frac{\partial}{\partial R_A} \tag{2.39}$$

instead of the partial one previously introduced in Eq. (2.30). When we insert Eq. (2.39) into the time-dependant Schrödinger equation

we find:

$$\frac{d}{dt}(\phi(R(t),x)^T v(t)) = -i\hat{H}(\phi(R(t),x)^T v(t)),$$

$$\frac{d}{dt}(\phi(R(t),t)^T)v(t) + \phi(R(t),t)^T\frac{\partial}{\partial t}v(t) = -i\hat{H}(\phi(R(t),x)^T v(t)),$$

$$\phi(R(t),t)\frac{d}{dt}(\phi(R(t),t)^T)v(t) + \phi(R(t),t)\phi(R(t),t)^T\frac{\partial}{\partial t}v(t)$$
$$= -i\phi(R(t),t)\hat{H}(\phi(R(t),x)^T v(t)),$$

$$Bv(t) + S\frac{\partial}{\partial t}v(t) = -iHv(t),$$

$$\dot{v}(t) = -S^{-1}(iH + B)v(t),$$
$$(2.40)$$

with $B = \langle\phi|\frac{d}{dt}\phi\rangle$. Similarly, an additional term must be introduced for the atomic forces: [58]

$$M_A\ddot{R}_A = -\frac{\partial U(R,t)}{\partial R_A} + \sum_{j=1}^{N_e} v^{j*}(D^A - \frac{\partial H}{\partial R_A})v^j, \qquad (2.41)$$

with D containing all the terms coming from the time dependence of the basis set: [58]

$$D_{\alpha\beta}^A = \sum_{\gamma\delta}(B_{\alpha\gamma}^{A+}S_{\gamma\delta}^{-1}H_{\delta\beta} + H_{\alpha\gamma}S_{\gamma\delta}^{-1}B_{\delta\beta}^{A+})+$$
$$i\Big(C_{\alpha\beta}^{A+} - C_{\alpha\beta}^A + \sum_{\gamma\delta}(B_{\alpha\gamma}^+S_{\gamma\delta}^{-1}B_{\delta\beta}^A + B_{\alpha\gamma}^{A+}S_{\gamma\delta}^{-1}B_{\delta\beta})\Big), \quad (2.42)$$

with

$$B_{\alpha\beta} = \left\langle \phi_\alpha \left| \frac{d}{dt} \phi_\beta \right\rangle \right.,$$

$$B_{\alpha\beta}^+ = B_{\beta\alpha}^* = \left\langle \frac{d}{dt} \phi_\alpha \left| \phi_\beta \right\rangle \right.,$$

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \sum_{A=1}^{N_i} \dot{R}_A \frac{\partial}{\partial R_A},$$

$$B_{\alpha\beta}^A = \left\langle \phi_\alpha \left| \frac{\partial}{\partial R_A} \phi_\beta \right\rangle \right.,$$

$$B_{\alpha\beta}^{A+} = B_{\beta\alpha}^{A*} = \left\langle \frac{\partial}{\partial R_A} \phi_\alpha \left| \phi_\beta \right\rangle \right.,$$

$$C_{\alpha\beta}^A = \left\langle \frac{d}{dt} \phi_\alpha \left| \frac{\partial}{\partial R_A} \phi_\beta \right\rangle \right.,$$

$$C_{\alpha\beta}^{A+} = C_{\beta\alpha}^{A*} = \left\langle \frac{\partial}{\partial R_A} \phi_\alpha \left| \frac{d}{dt} \phi_\beta \right\rangle \right..$$

2.7.1 Density Matrix Based Ehrenfest Molecular Dynamics

The following section has been previously published in Ref. [47]. The propagation of the wave function can be reformulated as a propagation of the density matrix P:

$$P(t + \Delta t) = \psi(t + \Delta t)\psi(t + \Delta t)^*$$
$$= U(t, t + \Delta t)P(t)U(t, t + \Delta t)^*, \qquad (2.43)$$

with $U(t, t + \Delta t) = e^{\frac{-\Delta t}{2}(X(t)+X(t+\Delta t))}$ for EM and $U(t, t + \Delta t) = e^{\frac{-\Delta t}{2}X(t+\Delta t)}e^{\frac{-\Delta t}{2}X(t)}$ for ETRS. This formulation offers the prospect to achieve linear scaling in cases where P is sparse. The sparsity could occur either through the nearsightedness of electrons, [42] or is implicitly enforced by the use of the Subsystem DFT method. Note that in the Subsystem DFT approach, electrons are not transferred between subsystems during the propagation.

The straightforward way to propagate P is to calculate the matrix exponential explicitly, for example through a Taylor series expansion.[65] However $P(t + \Delta t)$ can be calculated more efficiently. The multiplication of a Hermitian matrix with a matrix exponential, e^M, from one side and the complex conjugate of the same exponential on the other side can be written as a series similar to a Baker-Campbell-Hausdorff (BCH) expansion (Proof in Appendix):

$$P(t + \Delta t) = e^M P(t)(e^M)^* = \sum_{n=0}^{\infty} Z_n, \qquad (2.44)$$

$$Z_0 = P(t), Z_{n>0} = \frac{1}{n}(MZ_{n-1} + (MZ_{n-1})^*). \qquad (2.45)$$

This sum converges much faster than the explicit calculation of the matrix exponential, requiring only 5-10 terms depending on the accuracy needed. Indeed, if $H(t)$ and $P(t)$ commute, the series terminates immediately independently of the value of Δt, while for small perturbations away from the ground state, only the small non-commuting part remains. The increased sparsity of these higher order terms furthermore reduces their cost with the sparse matrix multiplication employed.

Bibliography

[1] Roy G. Gordon and Yung S. Kim. Theory for the forces between closedshell atoms and molecules. *J. Chem. Phys.*, 56(6):3122–3133, 1972.

[2] Walter Kohn and Lu J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[3] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103 – 128, 2005.

[4] Jürg Hutter, Marcella Iannuzzi, Florian Schiffmann, and Joost VandeVondele. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 4(1):15–25, 2014.

[5] Erwin Schrödinger. Quantisierung als Eigenwertproblem. *Ann. Phys.*, 385(13):437–490, 1926.

[6] Erwin Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Phys. Rev.*, 28(6):1049, 1926.

[7] Max Born and Robert Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys.*, 389(20):457–484, 1927.

[8] Paul A. M. Dirac. On the theory of quantum mechanics. In *Proc. Royal Soc. A: Mathematical, Physical and Engineering Sciences*, volume 112, pages 661–677. The Royal Society, 1926.

[9] Werner Heisenberg. Mehrkörperproblem und Resonanz in der Quantenmechanik. *Z. Phy. A–Hadron Nucl.*, 38(6):411–426, 1926.

[10] Wolfgang Pauli. Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Z. Phys. A–Hadron Nucl.*, 31(1):765–783, 1925.

[11] John C. Slater. The theory of complex spectra. *Phys. Rev.*, 34(10):1293, 1929.

[12] John E. Lennard-Jones. The electronic structure of some diatomic molecules. *J. Chem. Soc. Faraday*, 25:668–686, 1929.

[13] Luigi Genovese, Alexey Neelov, Stefan Goedecker, Thierry Deutsch, Seyed Alireza Ghasemi, Alexander Willand, Damien Caliste, Oded Zilberberg, Mark Rayson, Anders Bergman, et al. Daubechies wavelets as a basis set for density functional pseudopotential calculations. *J. Chem. Phys.*, 129(1):014109, 2008.

[14] John C. Slater. Atomic shielding constants. *Physical Review*, 36(1):57, 1930.

[15] S. Francis Boys. Electronic wave functions. i. a general method of calculation for the stationary states of any molecular system. In *P. Roy. Soc. A-Math. Phy.*, volume 200, pages 542–554. The Royal Society, 1950.

[16] Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys.*, 127(11), 2007.

[17] Hermann Grassmann. *Die Ausdehnungslehre*. verlag von Th. Chr. Fr. Enslin, 1862.

[18] Paul Adrien Maurice Dirac. A new notation for quantum mechanics. In *Math. Proc. Cambridge*, volume 35, pages 416–418. Cambridge University Press, 1939.

[19] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136(3B):B864, 1964.

[20] Yan A. Wang and Emily A. Carter. Orbital-free kinetic-energy density functional theory. In *Theoretical methods in condensed phase chemistry*, pages 117–184. Springer, 2002.

[21] Edward U. Condon. The theory of complex spectra. *Phys. Rev.*, 36(7):1121, 1930.

[22] Paul A. M. Dirac. Note on exchange phenomena in the thomas atom. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 26, pages 376–385. Cambridge University Press, 1930.

[23] CF v Weizsäcker. Zur Theorie der Kernmassen. *Z. Phys. A– Hadron Nucl.*, 96(7):431–458, 1935.

[24] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[25] Axel D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38(6):3098, 1988.

[26] John P. Perdew, John A Chevary, Sy H. Vosko, Koblar A. Jackson, Mark R. Pederson, Dig J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46(11):6671, 1992.

[27] Chengteh Lee, Weitao Yang, and Robert G. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B*, 37(2):785, 1988.

[28] Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chem. Phys. Lett,*, 73(2):393 – 398, 1980.

[29] Charles George Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *J. Appl. Math.*, 6(1):76–90, 1970.

[30] Charles G Broyden. The convergence of a class of double-rank minimization algorithms: 2. the new algorithm. *J. Appl. Math.*, 6(3):222–231, 1970.

[31] Roger Fletcher. A new approach to variable metric algorithms. *Comput. J.*, 13(3):317–322, 1970.

[32] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Math. Comput.*, 24(109):23–26, 1970.

[33] David F Shanno. Conditioning of quasi-newton methods for function minimization. *Math. Comput.*, 24(111):647–656, 1970.

[34] Joost VandeVondele and Jürg Hutter. An efficient orbital transformation method for electronic structure calculations. *J. Chem. Phys.*, 118(10):4365–4369, 2003.

[35] Valery Weber, Joost VandeVondele, Juerg Hutter, and Anders M. N. Niklasson. Direct energy functional minimization under orthogonality constraints. *J. Chem. Phys.*, 128(8):084113, 2008.

[36] Vladimir Fock. Näherungsmethode zur Lösung des Quantenmechanischen Mehrkörperproblems. *Z. Phys. A–Hadron Nucl.*, 61(1):126–148, 1930.

[37] Fritz Coester and Hermann Kümmel. Short-range correlations in nuclear wave functions. *Nucl. Phys.*, 17:477–485, 1960.

[38] Jiří Čížek. On the correlation problem in atomic and molecular systems. calculation of wavefunction components in ursell-type expansion using quantum-field theoretical methods. *J. Chem. Phys.*, 45(11):4256–4266, 1966.

[39] Isaiah Shavit. The history and evolution of configuration interaction. *Mol. Phys.*, 94(1):3–17, 1998.

[40] William L. McMillan. Ground state of liquid he 4. *Phys. Rev.*, 138(2A):A442, 1965.

[41] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal. Quantum monte carlo simulations of solids. *Rev. Mod. Phys.*, 73(1):33, 2001.

[42] Walter Kohn. Density functional and density matrix method scaling linearly with the number of atoms. *Phys. Rev. Lett.*, 76:3168–3171, Apr 1996.

[43] Emil Prodan and Walter Kohn. Nearsightedness of electronic matter. *Proceedings of the National Academy of Sciences of the United States of America*, 102(33):11635–11638, 2005.

[44] Stefan Goedecker. Linear scaling electronic structure methods. *Rev. Mod. Phys.*, 71(4):1085, 1999.

[45] Stefan Goedecker. Decay properties of the finite-temperature density matrix in metals. *Phys. Rev. B.*, 58(7):3501, 1998.

[46] Joost VandeVondele, Urban Borštnik, and Jürg Hutter. Linear scaling self-consistent field calculations with millions of atoms in the condensed phase. *J. Chem. Theory Comput.*, 8(10):3565–3573, 2012.

[47] Samuel Andermatt, Jinwoong Cha, Florian Schiffmann, and Joost VandeVondele. Combining linear-scaling DFT with subsystem DFT in Born–Oppenheimer and Ehrenfest molecular dynamics simulations: From molecules to a virus in solution. *J. Chem. Theory Comput.*, 12:3214–3227, 2016.

[48] Urban Borštnik, Joost VandeVondele, Valéry Weber, and Jürg Hutter. Sparse matrix multiplication: The distributed block-compressed sparse row library. *Parallel Comput.*, 40(5):47–58, 2014.

[49] Anders M. N. Niklasson, C. J. Tymczak, and Matt Challacombe. Trace resetting density matrix purification in o (n) self-consistent-field theory. *J. Chem. Phys.*, 118(19):8611–8620, 2003.

[50] Yihan Shao, Chandra Saravanan, Martin Head-Gordon, and Christopher A. White. Curvy steps for density matrix-based energy minimization: Application to large-scale self-consistent-field calculations. *J. Chem. Phys.*, 118(14):6144–6151, 2003.

[51] Christoph R. Jacob and Johannes Neugebauer. Subsystem density-functional theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 4(4):325–362, 2014.

[52] Pietro Cortona. Self-consistently determined properties of solids without band-structure calculations. *Phys. Rev. B*, 44(16):8454, 1991.

[53] Marcella Iannuzzi, Barbara Kirchner, and Jürg Hutter. Density functional embedding for molecular systems. *Chem. Phys. Lett.*, 421(13):16–20, 2006.

[54] Ole Schütt, Peter Messmer, Jürg Hutter, and Joost VandeVondele. *GPU-Accelerated Sparse Matrix–Matrix Multiplication for Linear Scaling Density Functional Theory*, pages 173–190. John Wiley & Sons, Ltd, 2016.

[55] Peter Pulay. Ab initio calculation of force constants and equilibrium geometries in polyatomic molecules: I. theory. *Mol. Phys.*, 17(2):197–204, 1969.

[56] Loup Verlet. Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98, 1967.

[57] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.*, 76(1):637–649, 1982.

[58] T. Kunert and R. Schmidt. Non-adiabatic quantum molecular dynamics: General formalism and case study H_2^+ in strong laser fields. *Eur. Phys. J. D*, 25(1):15–24, 2003.

[59] Douglas R. Hartree. The wave mechanics of an atom with a noncoulomb central field. part i. theory and methods. In *Math. Proc. Cambridge*, volume 24, pages 89–110. Cambridge University Press, 1928.

[60] Douglas R. Hartree. The wave mechanics of an atom with a noncoulomb central field. part II. some results and discussion. In *Math. Proc. Cambridge*, volume 24, pages 111–132. Cambridge University Press, 1928.

[61] Felix Bloch. Über die Quantenmechanik der Elektronen in Kristallgittern. *Z. Phys. A–Hadron Nucl.*, 52(7):555–600, 1929.

[62] John C. Slater and George F. Koster. Simplified LCAO method for the periodic potential problem. *Phys. Rev.*, 94(6):1498, 1954.

[63] Alberto Castro, Miguel A. L. Marques, and Angel Rubio. Propagators for the time-dependent Kohn–Sham equations. *J. Chem. Phys.*, 121(8):3425–3433, 2004.

[64] Jiří Kolafa. Time-reversible always stable predictor–corrector method for molecular dynamics of polarizable molecules. *J. Comput. Chem.*, 25(3):335–342, 2004.

[65] Conn O'Rourke and David R. Bowler. Linear scaling density matrix real time TDDFT: Propagator unitarity and matrix truncation. *J. Chem. Phys.*, 143(10):–, 2015.

Chapter 3

Implementation

3.1 Introduction

This chapter covers topics related to the implementation of RTP/EMD and subsystem (SS) DFT in CP2K.[1, 2]

In the first section a scheme is developed to improve the performance of SS DFT. The next three sections deal with RTP/EMD. Firstly, we will cover the implementation of complex matrix multiplications for EMD. We will then demonstrate that our implementation remains stable with longer time steps than typically feasible for such simulations. Finally, the feasibility of linear scaling ($O(N)$) RTP/EMD will be discussed. The first and the last two sections are based on a previous publication.[3] All simulations mentioned in this chapter are performed with a plane-wave (PW) cutoff of 300 Rydberg or higher, the PBE exchange-correlation functional,[4] GTH pseudopotentials[5] and a DZVP-MOLOPT-GTH/DZVP-MOLOPT-SR-GTH[6] basis set.

3.2 Subsystem Graph Colouring

The implementation of SS DFT in CP2K is unique in its interpretation of the system division as a graph colouring problem. The basics of the graph colouring approach were developed and implemented prior

to the start of this project. However, the performance of the colouring procedure was improved by two orders of magnitude and the final result enhanced through the addition of a pair-switching algorithm. The following section will introduce the reasoning for the graph colouring interpretation and cover some details of the implementation.

The calculation of the energy in SS DFT, according to Eq. (2.21), requires the calculation of the HK energy of all individual subsystems. Each subsystem can be computed fully independently, and the calculation could be parallelised over the various subsystems. However, this can be challenging if large subsystems are present. Furthermore, in CP2K and various other codes that employ regular grids for the electron density, the computation of small subsystems is relatively inefficient, as a sufficiently large auxiliary simulation cells, centred around each subsystem, are needed. This implies an overhead to deal with essentially empty space and complicates the implementation, which was one reason why the older CP2K implementation [7] was replaced by a new scheme. In this equivalent alternative, the sum of the HK energies is computed considering subsets of subsystems that have nonoverlapping densities. The HK energy of such a subset equals the sum of the HK energies of each of the subsystems in the subset. To compute this energy, this subset of nonoverlapping subsystems is placed in the simulation cell that equals the original simulation cell, and the HK energy of the subset is computed. This calculation is repeated for all subsets, each being essentially a relatively sparse system. By packing the subsystems as efficiently as possible in subsets, the amount of empty space is minimised, and the number of iterations over subsets is reduced, speeding up the calculation. As an additional advantage, only few code changes relative to a standard DFT code are needed to perform such a computation; it just implies a standard calculation for each subset of particles. Furthermore, in this approach, large subsystems are not problematic, as the standard parallelism can efficiently be employed.

The computational effort to compute the HK energy is proportional to the number of subsets needed. Finding these subsets is a problem that can be formulated as a graph colouring problem. Graph colouring is a standard problem in computer science. To obtain the graph, each subsystem is considered a node (vertex) in a graph. Two nodes are connected by an edge of the graph if the corresponding

subsystems overlap. This information can be obtained from the overlap matrix of the full system. An allowed partitioning of the graph in subsets is obtained colouring the graph in such a way that two connected nodes have different colours. After colouring, each colour represents a different subset, and all nodes of the same colour belong to the same subset. The colouring of a graph is not unique. The optimal solution with the fewest possible colours is computationally intractable to find, but two established heuristic algorithms to solve the colouring problem were tested. A third algorithm improves on the obtained solutions.

The first algorithm to colour a graph is the greedy algorithm. In the greedy algorithm, all nodes are coloured in an order that is independent of the topology of the graph, for example using the node's index. Each node is assigned the colour with the lowest index that is not yet present in a neighbouring node. The second algorithm is DSATUR. [8] In DSATUR, two criteria are used to decide which node of the graph is coloured next. The first criterion is the degree of saturation. The degree of saturation is the number of unique colours present in the neighbourhood of the node. The algorithm is to choose the node with the highest degree of saturation and give it the lowest colour not yet present in the neighbourhood. In case of multiple nodes with the same degree of saturation, the second criterion takes the node which has the most neighbours.

After an initial solution is obtained, the number of colours is further reduced by switching neighbouring pairs and decreasing their colours where applicable. The pair-switching algorithm can be found in the next subsection. The colouring was tested on an STMV virus in water. [9, 10] The system contains around one million atoms. Most of the subsystems are water molecules, but there are also proteins that are much larger. The greedy algorithm results in 36 colours and the DSATUR algorithm in 28 colours. The pair-switching algorithm reduces the number of colours to 30 and 26, respectively. The colouring of the graph is fast and does not significantly contribute to the total execution time.

3.2.1 Pair Switching Algorithm

for $iter = 1$ **to** max_iter **do**
    for $depth = 1$ **to** max_depth **do**
        for $i = 1$ **to** num_nodes **do**
            if $node(i)\%colour\_index >$
            $max\_colour\_index - depth$ **then**
                lowest_colour=node(i)%colour_index+1
                found=false
                for *each node(i)%neighbor* **do**
                    if *switchable(node(i),neighbor)* **then**
                        switch_colours(node(i),neighbor)
                        x=lowest_possible_colour_index(neighbor)
                        if $x < lowest\_colour$ **then**
                            found=true
                            lowest_colour=x
                            best_partner=neighbor
                        end
                        switch_colours(node(i),neighbor)
                    end
                end
                if *found* **then**
                    switch_colours(node(i),best_partner)
                    x=lowest_possible_colour_index(best_partner)
                    neighbor%colour_index=x
                end
            end
        end
    end
end

Algorithm 1: The pair switching algorithm to decrease the number of colours in the graph.

A pair switching algorithm was used to improve the coloured graph. The goal is to further reduce the number of colours. The algorithm investigates the nodes with high colour indices in the coloured graph and searches the neighbourhood for nodes with which an exchange of colour does not violate the colouring of the graph. All through switching reachable neighbour position are then investigated

for viable reductions in colour index in the switched state. At the end, the colour is switched with the position offering the lowest possible colour index. The algorithm is illustrated in Alg. 1.

The algorithm could be extended further by checking after each successful switch if the nodes neighbouring the switched nodes can be reduced in colour. In case of successful reductions, the neighbours of the neighbours would then need to be checked. However, we did not see this extension to significantly improve the results.

3.3   Complex Matrix Multiplication

EMD requires complex linear algebra for the propagation of the density matrix according to Eqs. (2.44) and (2.45). The straightforward approach is to describe a complex matrix as a sum of two real valued matrices. A matrix multiplication between two complex matrices thus becomes:

$$D = (A + iB)(C + iD) = AC + i(BC + AD) - BD, \qquad (3.1)$$

requiring four real matrix multiplications.

There is an alternative approach which exploits the fact that complex matrix multiplications can be defined through only three real matrix multiplications and that matrix additions are significantly cheaper than multiplications:

$$D = (A + iB)(C + iD) = AC - BD + i((A + B)(C + D) - AC - BD). \qquad (3.2)$$

In this approach, only the computation of AC, BD and $(A+B)(C+D)$ are required. Therefore, only three quarters of the computational time are needed as compared to the straightforward implementation.

Another alternative would be to either directly use complex matrices, or to store complex numbers as two real numbers on a matrix

element level:

$$C(i,j) = \sum_k A(i,k)B(k,j)$$
$$= \sum_k \mathrm{Re}[A(i,k)]\mathrm{Re}[B(k,j)] + i\mathrm{Re}[A(i,k)]\mathrm{Im}[B(k,j)] \quad (3.3)$$
$$+ i\mathrm{Im}[A(i,k)]\mathrm{Re}[B(k,j)] - \mathrm{Im}[A(i,k)]\mathrm{Im}[B(k,j)].$$

CP2K uses blocked sparse matrices, where each element itself is a small matrix. While this approach requires the same amount of operations as the straightforward method, it is twice as efficient in terms of memory bandwidth. This is due to the fact that complex blocks are only twice the size of real blocks, but four times more operations must be executed per loaded block than in the straightforward case. In situations where the speed of the multiplication is limited by the memory bandwidth, this approach could theoretically be faster than the three-multiplications scheme from above. Not that, for applications which are computation limited, the proposed approach is equivalent to the straightforward one in terms of performance. It is thus outperformed by the three-multiplication approach.

In the DBCSR library that CP2K uses for matrix multiplications, calculations are typically limited by the memory bandwidth, except for large basis sets. Hence, it could be computationally favourable to use the complex matrix approach. However, complex multiplications are currently not supported by the DBCSR library so that the three-multiplication approach was implemented as part of this work.

3.4 Optimal Time Step

For the computational cost of an EMD simulation the time step is of crucial importance. Results by O'Rourke and Bowler suggests a time step of 0.06 a.u (1.44 as) for real-time propagation (benzene system) as obtained by the exponential midpoint propagator, only applied self-consistently in the initial 100 steps. [11] For subsystem DFT Krishtal et al. suggested a time step 2 as. [12] The present approach should be stable at longer time steps thanks to the reduction of numerical noise described in detail below and the use of an implicit propagation scheme.
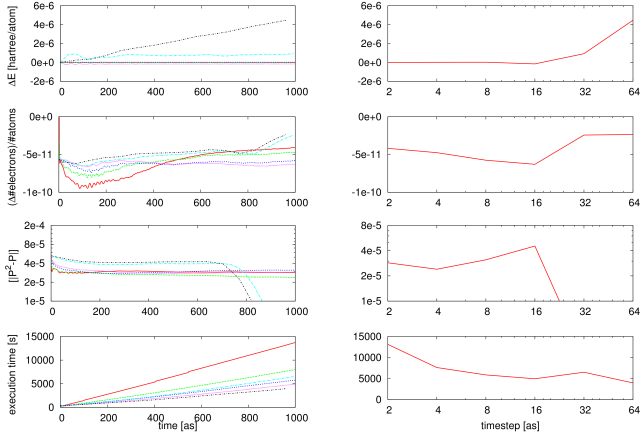
Figure 3.1: An evaluation of the time step dependence of the accuracy and performance of an EMD simulation. The investigated system was a quasi-1D system of 1h ice (2304 atoms). On the left side the time evolution of the energy, charge, idempotency and computational cost are displayed. The execution time refers to the cumulative time consumed over all steps. On the right side the cross-section of the graph at the left side, taken at 960 as, is displayed. The results on the left hand side have been coloured according to the time step: (red) 2 as, (green) 4 as, (blue) 8 as, (violet) 16 as, (turquoise) 32 as, (grey) 64 as.

The stability with respect to the time step has been investigated using a simple condensed phase system with a large band gap that was created replicating in one dimension a cubic simulation cell of ice 1h containing 288 atoms. This quasi-1D system will be used for several tests. For the current testing, a 10^{-7} filtering threshold is employed. As shown in Figure 3.1, charge, energy and the idempotency of P are retained up to large time steps. However, at a time step of 64 as a constant drift ($\sim$4 micro-Hartree/fs/atom) in the energy appears. This drift limits the long-term stability of the simulation. With a small time step, energy conservation is excellent, which suggests that also ionic forces are implemented correctly. We have observed that

55

for large time steps, as well as very small time steps with aggressive filtering, the numerical noise leads to an increase of the occupation of the density matrix during propagation.

Naively, the computational effort is directly proportional to the number of time steps needed to reach a given total simulation time. In practice, the computational effort per step becomes larger with longer time steps. The increase is caused by a slower convergence of the matrix exponential (eq. 2.45) and the higher number of iterations needed to reach self-consistency in the implicit propagation algorithm. For the investigated system this becomes evident as the total execution time remaining roughly constant for time steps of 16 as and more. Given the observed accuracy and efficiency, we thus recommend a time step of approximately 16 as for this and similar systems, which is roughly one order of magnitude larger than the value employed in Ref. [11]

3.5 Can Ehrenfest Molecular Dynamics be $O(N)$?

3.5.1 'Linear-Scaling' Implementation

To study the scalability of the density matrix based EMD code, systems with increasing size have been simulated. This study is based on replicating in one dimension a cubic simulation cell of ice 1h containing 288 atoms up to required size. The one-dimensional nature of the test system facilitates reaching the linear scaling regime. Note that this test does not employ SS DFT, for which the enforced matrix structure would trivially lead to linear scaling, but treats the system as a whole. The wave function and nuclei were propagated for a short time, starting from a ground-state wave function. As shown in Figure 3.2, the computational cost increases linearly with system size. This is consistent with the results reported for RTP in Ref. [11], and supports the correctness of the implementation.

However, the full picture is more complicated, and as described in detail in the following section, at longer timescales the computational cost of EMD gradually increases.
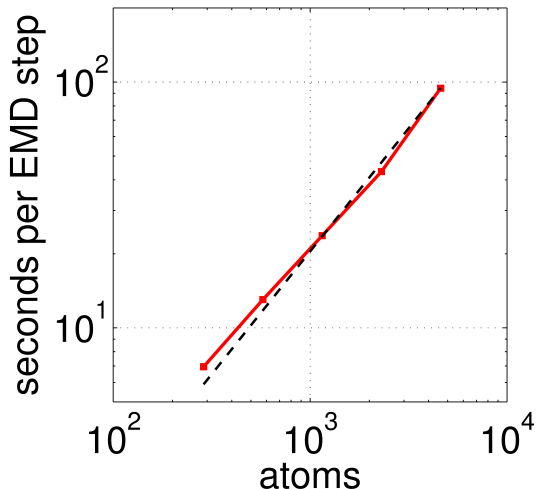
Figure 3.2: Shown are the timings for an EMD simulation of a large-gap quasi-1D system. The red data points represent the average time per time step during the first five EMD steps. The dashed line is a linear fit. The calculations were performed on 25 nodes on the former Cray XC30 (Piz Daint) computer at CSCS.

3.5.2 EMD Density Matrix Filling

In CP2K the density matrix is not filtered based on an a priori fixed sparsity pattern, for example as determined from a spatial cutoff [11] or from the fixed sparsity pattern of the overlap matrix. Instead, the elements are simply filtered based on their magnitude, as computed during the matrix multiplications. The advantage of this scheme is that the accuracy is determined by one parameter, the filtering threshold, independently of the chemistry of the system. Arbitrary accuracy can be obtained reducing the filtering threshold. However, as a result of this filtering method, the occupation of P depends on the decay of the physical density matrix, and also on the quality of the numerical scheme. To achieve linear scaling, the occupation of P needs to stay significantly below 1.

To investigate the evolution of the density matrix sparsity during

EMD, a simple test system has been prepared and let to evolve close to the ground state using EMD started from the electronic ground state, but with finite ionic temperature. The test system is a quasi-1D system, with a large band gap, namely the ice 1h system described above, replicated in one direction to contain approximately 2000 atoms. As such, the ionic evolution is very similar to that of a ground-state Born-Oppenheimer (gs-BO) MD. Within gs-BO MD, the density matrix is sparse at all times.

As shown in Figure 3.3, a straightforward implementation of density matrix based EMD, which directly implements the formulae given with filtering sparse matrix multiplication, results in a density matrix that quickly fills up over the course of a simulation. This leads to a significant slowdown of the calculation, ultimately scaling $O(N^3)$ with system size. A closer investigation shows that the filled-in elements are typically small and close to the filtering threshold. Even though the energy and total number of electrons remain largely constant, the idempotency of the density matrix is not well conserved.

We considered two possible explanations for P becoming non-sparse. The first is a building up of numerical noise. Numerical noise introduced by the filtering process has more severe effects in EMD than it has in gs-BO MD. In gs-BO MD the recalculation of P from H, which has a fixed sparsity pattern, resets the noise at each time step, while in EMD all noise in P adds up over the course of a simulation. A second possible explanation is that P could be intrinsically nonsparse, i.e. would become non-sparse also if computed with essentially infinite precision. EMD simulates nonadiabatic dynamics, and the proofs for the exponential decay of P however were given for the ground state at zero temperature for insulators and finite temperature for metals. [13] To control the numerical noise caused by the filtering of P two techniques were combined. The first is the restoration of the idempotency of P. This can be achieved through McWeeny purification: [14]

$$P = 3P^2 - 2P^3. \tag{3.4}$$

An idempotent P, such as the true density matrix, remains invariant, while a noisy P moves back in the direction of idempotency. In our experience, one or two purification steps after each implicit

propagation are sufficient to maintain good idempotency. McWeeny purification has therefore little impact on the short term performance, but improves the long-term stability. The second technique is to use two different filtering thresholds during the simulation. Steps in the simulation that are critical for numerical errors are performed with a tighter filtering threshold. Once the simulation leaves the critical region the matrices are filtered at the normal filtering threshold. The critical operations are the McWeeny purification and the application of the propagator to P. In our calculations, reducing the filtering threshold by two orders of magnitude for these steps yielded stable results at low computational cost. The reduction of the numerical noise through these control mechanisms is displayed in Figure 3.3, showing that this significantly outperforms the straightforward implementation.

As illustrated in Figure 3.4 for 2000 time steps (8000 as), the combination of both techniques allows for an effective control of the numerical noise. For the large band gap system employed, P remains sparse (0.2 occupation) at a filtering threshold of 10^{-5}. However, this does not hold at arbitrary precision. Using a tighter threshold (10^{-7}), a dense matrix is obtained after 6000 as. As illustrated in Figure 3.5 for the simulation of a boron-nitride nanotube (a smaller gap system), this threshold depends on the system studied. Even at a filtering threshold of 10^{-5} the density matrix becomes essentially dense after 1000 as. Furthermore, the unfiltered EMD simulation in figure 3.5 shows that the filtered simulations slightly underestimate the true occupation of the density matrix, thus demonstrating that the fill-in is not a result of the inaccuracy introduced through the filtering process. The loss of the sparsity can thus be reproduced with nonfiltered EMD or MO based EMD, and suggests that this is indeed the intrinsic behaviour for a system described by EMD.

3.5.3 Hückel Theory

Finding an apparently intrinsic density matrix filling in EMD, the results are compared to a simple model, a Hückel chain.[15] The Hückel chain can be solved free of numerical uncertainty in a few lines of MATLAB code, and thus complements the numerical experiments performed using the more complicated implementation of DFT based EMD in CP2K. It is based on a time-independent, non-self-consistent

Hamiltonian and has no moving ions. The Hamiltonian is a symmetric tridiagonal matrix. The off-diagonal elements are alternating between 1 and c, where c is a parameter that can be tuned to have a small (c=1) or a large (c=0) gap at half filling. The diagonal was chosen to be zero. The Hamiltonian does not depend on time. The initial density matrix is chosen to be the ground state for a given value of the parameter c in the Hamiltonian. The band gap is then slightly decreased by decreasing the value of c. At this point, P no longer corresponds to the ground state of the Hamiltonian and can be propagated over time. For small changes, this mimics the behaviour of a system where ions move and electrons follow adiabatically.

At $t = 0$, P shows the well-known exponential decay. However, during the course of the EMD simulation the exponential decay of P is preserved only at short distance, and for large values of the matrix elements. At longer distances the decay, if any, becomes less than exponential. The results are illustrated in Figure 3.6. This decay is consistent with our previous observations that P only remained sparse at large filtering thresholds. The short-ranged exponential decay is enough to achieve sparsity for large filtering thresholds.

To conclude, non-exponential decaying density matrices were found in both the realistic simulations and in a simple Hückel chain model. This establishes the non-exponential decay of P within EMD, and makes linear-scaling EMD hard to reach at arbitrary accuracy. We emphasise again that this problem does not occur in the framework of the SS method, where the block diagonal structure of the Hamiltonian gives an upper limit for the occupation of P. It remains an open question how in general nearsightedness can be exploited within a non-Born-Oppenheimer linear-scaling setup. The observation made here for linear scaling EMD is likely related to a known short-coming of EMD, namely the electronic overcoherence, and we speculate that approaches developed in that context [16–18] might provide a viable approach to obtain a truly linear-scaling non-adiabatic MD.
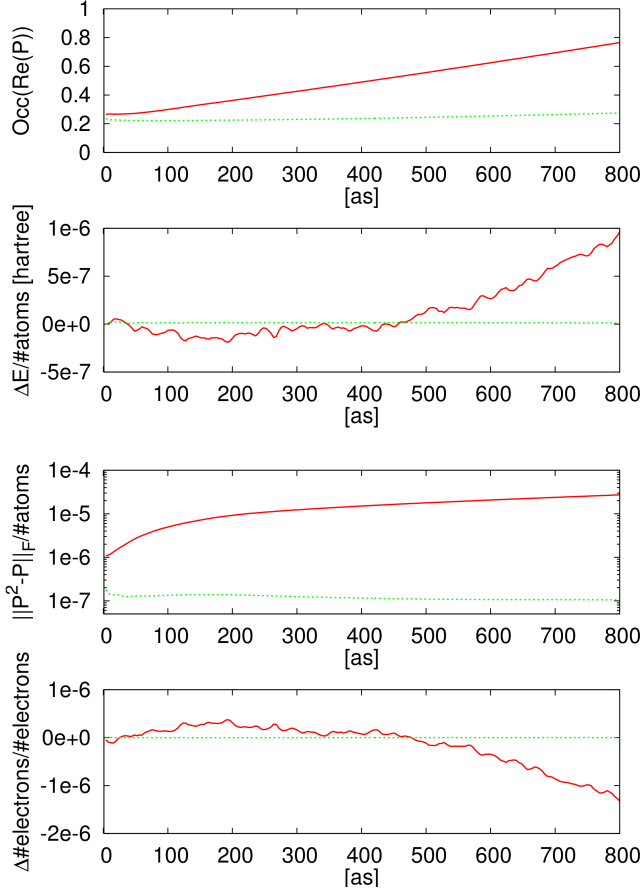
Figure 3.3: Shown are results for an EMD simulation of a large-gap quasi-1D system. The reported quantities are the occupation of P, the total energy drift, the drift in the number of electrons and the Frobenius norm of the deviation of P from idempotency. The aspect ratio of the system is 8:1 and contains the system contains 2304 atoms. The time step is 4 as, the filtering threshold is 10^{-5}. The advanced implementation with McWeeny purification and a two level filtering threshold (green, dotted) shows distinctly better properties than the straightforward implementation (red, solid).
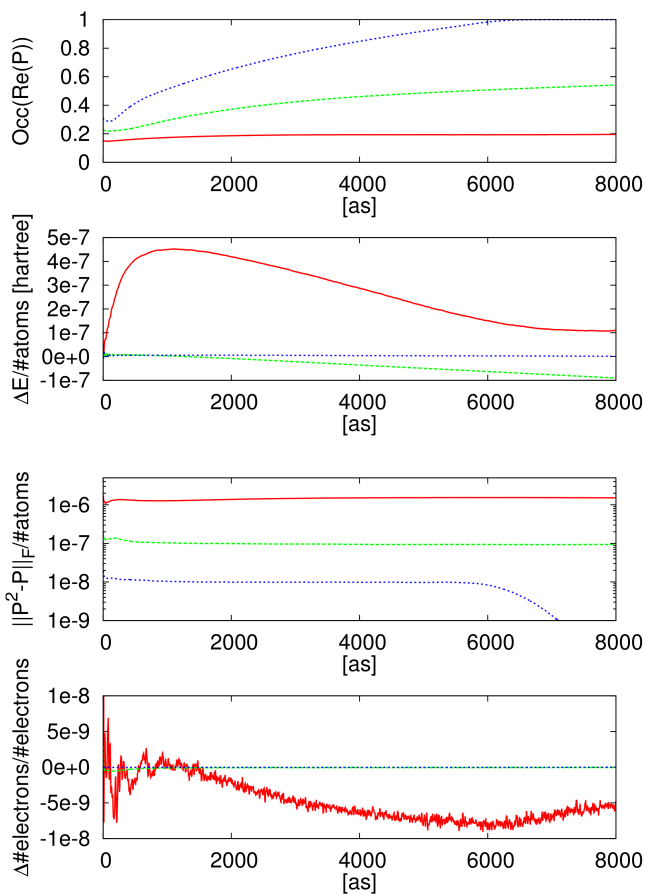
Figure 3.4: A comparison of noise-filtered EMD at different filtering thresholds. The system and the measured quantities are the same as in Figure 3.3 and the time step is 4 as. (red, solid) 10^{-5} filter threshold. (green, long dash) 10^{-6} filter threshold. (blue, short dash) 10^{-7} filter threshold

62
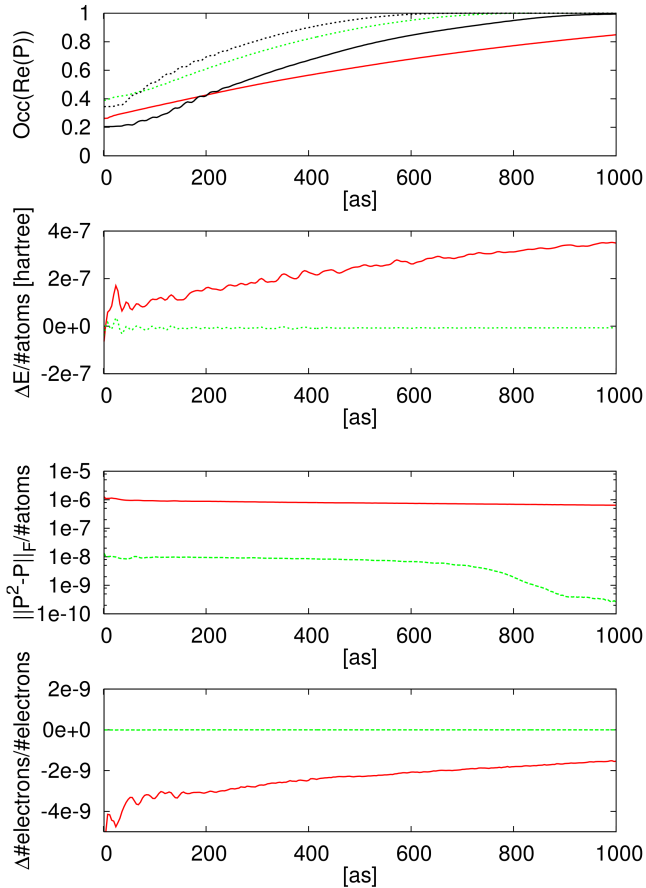
Figure 3.5: As in Figure 3.4 noise-filtered EMD is compared at different filtering thresholds. This system is a boron-nitride nanotube. The nanotube is about 10 nm long, hydrogen terminated and contains 2880 atoms. (red, solid) 10^{-5} filter threshold. (green, dotted) 10^{-7} filter threshold, (black, solid) The fraction of elements of size 10^{-5} or higher in an unfiltered EMD simulation, (black, dashed) The fraction of elements of size 10^{-7} or higher in an unfiltered EMD simulation.
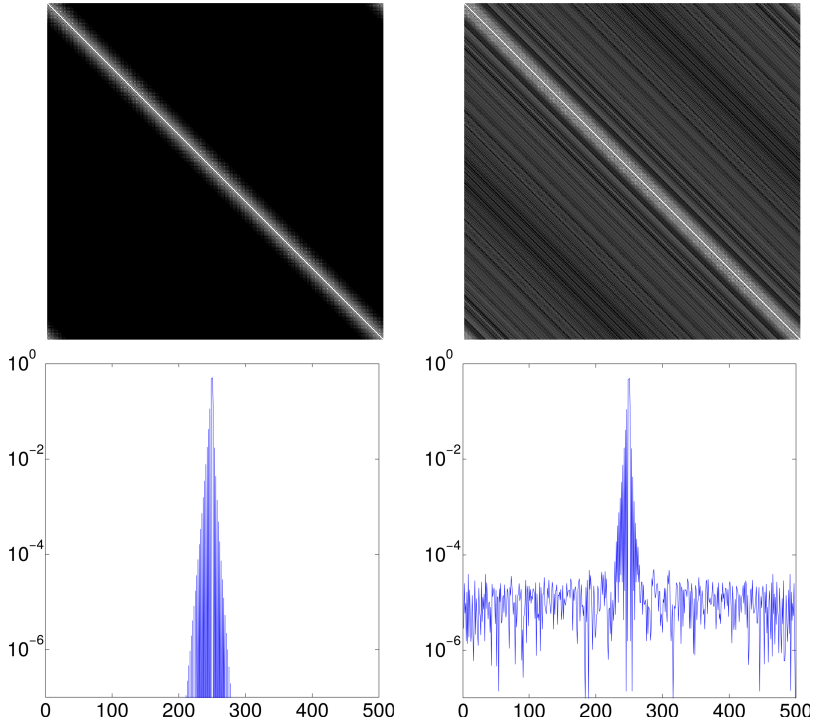
Figure 3.6: The panels on the top show a logarithmic representation of the absolute values of the elements of the density matrix. P(t=0) is shown on the left and P(t=500) at the right. The bottom panels show the absolute values of the elements along a single row of the density matrix at the same times. The system is a cyclic Hückel chain of length 500. The initial density matrix is obtained for $c = 0.5$, while c was shifted by 0.001 to propagate away from the ground state.

Bibliography

[1] Jürg Hutter, Marcella Iannuzzi, Florian Schiffmann, and Joost VandeVondele. cp2k: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 4(1):15–25, 2014.

[2] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103 – 128, 2005.

[3] Samuel Andermatt, Jinwoong Cha, Florian Schiffmann, and Joost VandeVondele. Combining linear-scaling dft with subsystem dft in bornoppenheimer and ehrenfest molecular dynamics simulations: From molecules to a virus in solution. *J. Chem. Theory Comput.*, 12:3214–3227, 2016.

[4] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[5] S. Goedecker, M. Teter, and J. Hutter. Separable dual-space gaussian pseudopotentials. *Phys. Rev. B*, 54:1703–1710, Jul 1996.

[6] Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys*, 127(11), 2007.

[7] Marcella Iannuzzi, Barbara Kirchner, and Jürg Hutter. Density functional embedding for molecular systems. *Chem. Phys. Lett.*, 421(13):16–20, 2006.

[8] Daniel Brélaz. New methods to color the vertices of a graph. *Commun. ACM*, 22(4):251–256, April 1979.

[9] Steven B. Larson, John Day, Aaron Greenwood, and Alexander McPherson. Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution1. *J. Mol. Biol.*, 277(1):37 – 59, 1998.

[10] Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson, and Klaus Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14(3):437 – 449, 2006.

[11] Conn O'Rourke and David R. Bowler. Linear scaling density matrix real time tddft: Propagator unitarity and matrix truncation. *J. Chem. Phys.*, 143(10):–, 2015.

[12] Alisa Krishtal, Davide Ceresoli, and Michele Pavanello. Subsystem real-time time dependent density functional theory. *J. Chem. Phys.*, 142(15), 2015.

[13] Stefan Goedecker. Linear scaling electronic structure methods. *Rev. Mod. Phys.*, 71(4):1085, 1999.

[14] Roy McWeeny. Some recent advances in density matrix theory. *Rev. Mod. Phys.*, 32:335–369, Apr 1960.

[15] Erich Hückel. Quantentheoretische beiträge zum benzolproblem. *Z. Phys.*, 70(3-4):204–286, 1931.

[16] Michael J. Bedard-Hearn, Ross E. Larsen, and Benjamin J. Schwartz. Mean-field dynamics with stochastic decoherence (mf-sd): A new algorithm for nonadiabatic mixed quantum/classical molecular-dynamics simulations with nuclear-induced decoherence. *J. Chem. Phys.*, 123(23), 2005.

[17] Neil Shenvi, Joseph E. Subotnik, and Weitao Yang. Simultaneous-trajectory surface hopping: A parameter-free algorithm for implementing decoherence in nonadiabatic dynamics. *J. Chem. Phys.*, 134(14), 2011.

[18] Alexey V. Akimov, Run Long, and Oleg V. Prezhdo. Coherence penalty functional: A simple method for adding decoherence in ehrenfest dynamics. *J. Chem. Phys.*, 140(19), 2014.

Chapter 4

Applications

4.1 Introduction

This chapter deals with numerical simulations conducted to validate the implemented methods, to demonstrate applications of RTP/EMD, as well as to support the design of experimental devices. An overview of the chapter can be found in Table 4.1.

All simulations proposed in this chapter have been carried out with the DFT software package CP2K.[1, 2] Unless specified otherwise, DZVP-MOLOPT-GTH or DZVP-MOLOPT-SR-GTH contracted Gaussian basis sets[3] have been used, together with an auxiliary plane-wave basis with a cut-off of 300 Rydberg or higher. For the exchange-correlation energy a Perdew-Burke-Ernzerhof (PBE)[4] functional has been employed, with the adiabatic local density approximation (ALDA). To treat the interactions of the atomic cores Goedecker-Teter-Hutter[5] pseudopotentials were employed. For all simulations the electrons were propagated using the enforced time-reversible symmetry (ETRS) propagator.[6] Parts of this chapter are taken from previous publications,[7–9] because the material originates from different publications, RTP and RT-TDDFT are used interchangeably.

System	Process	RTP	EMD	LS-gs DFT	SS DFT
boron-nitride nanotube	electron injection	✓	✗	✗	✗
metallic CNT between palladium slabs	electron transport	✓	✗	✗	✗
semiconducting CNT between palladium slabs (CNTFET)	electron transport	✓	✗	✗	✗
germanium selenide between TiN contacts (Selector)	electron transport	✓	✗	✗	✗
copper filament in SiO_2 matrix connecting copper slabs (CBRAM)	plasmonics + electron transport	✓	✓	✗	✗
TiO_2 nanoparticles in acetonirile solution (DSSC component)	bandgap determination	✗	✗	✓	✓
satellite tobacco mosaic virus in water	force relaxation	✗	✗	✓	✓
N3 dye in acetonirile solution (DSSC component)	photo-absorption	✗	✓	✓	✓

Table 4.1: Overview of the simulated structures, together with the therein investigated processes and the methods used for that purpose.

4.2 Electron Injection into a Boron-Nitride Nanotube
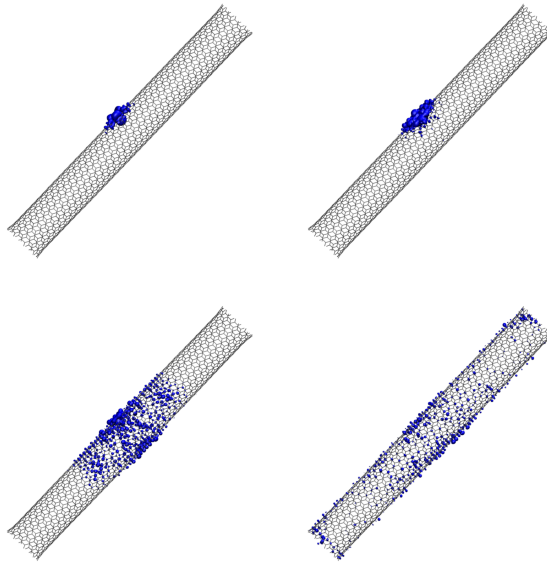
Figure 4.1: Visualising the injection of an excess electron into a boron-nitride nanotube. From top left to bottom right: 25 as, 500 as, 2500 as and 7500 as. Without the injected electron the boron-nitride nanotube is a closed-shell molecule. Therefore, the electronic density of the spin difference is displayed to track the injected electron.

The developments in this work have been triggered by our interest in electron dynamics beyond the linear response. This includes the simulation of electron transfer across interfaces as e.g. in dye sensitised solar cells (DSSCs), [10–16] in the presence of time-dependent external fields as found in various spectroscopies. [17–19]

To test electron dynamics beyond the linear response, the dynamics of an electron on a boron-nitride (BN) nanotube was investigated using density matrix based EMD. The initial configuration was a localised electron in the centre of the tube. The nanotube, generated
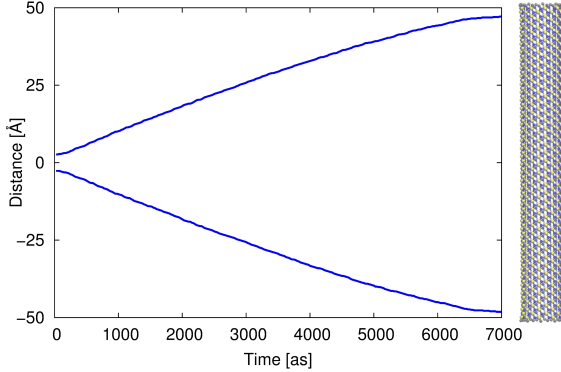
Figure 4.2: Quantifying the spreading of the excess electron over a boron-nitride (BN) nanotube. The spreading is calculated as the width of the tube segment that contains 95 percent of the total spin density with the remaining spin density split evenly between the two sides.

with TubeGen, an online tool for the creation of nanotube geometries,[20] has chirality (13,0) and is hydrogen terminated. The nanotube is 10 nm long and contains 1440 atoms. This nanotube is an insulator and therefore the density matrix is sparse at the ground state. As discussed previously, the occupation of the density matrix quickly reaches 1 during propagation so that the density matrix based EMD code operates in the cubic regime. A visual representation of the spreading of the excess electron is given in Figure 4.1. We note that these simulations are performed with GGA functionals, which suffer from self-interaction error. Whereas simulations with hybrid functionals are possible in this context, the computational cost is significantly higher. A more quantitative illustration of the spreading of the electron can be found in Figure 4.2. The electron was injected quickly into the nanotube and then spread over the whole nanotube. The front of the spreading moved at approximately 7×10^5 m/s. To our knowledge the saturation velocity in boron nitride nanotubes is not known. However, the saturation velocity in semiconducting carbon nanotubes was predicted through semi-classical models to be around
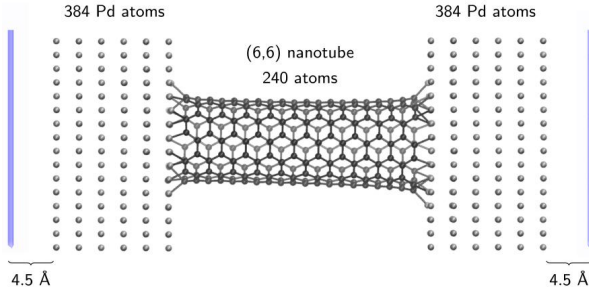
Figure 4.3: Atomistic simulation setup for the calculation of the current through a metallic (6,6) carbon nanotube. The planes on the side of the system are constraints for the Poisson equation solver, where the potential is fixed. The palladium contacts serve as reservoirs from which electrons are taken away from/added to the system, during the simulation. The length of the nanotube is approximately 2.5 nm and the width of the contacts about 1.3 nm each. The structure is composed of a total of 1008 atoms, which corresponds to a Hamiltonian and overlap matrix of size 22320×22320.

$3 - 5 \times 10^5$ m/s [21, 22] and determined from experimental data to be 2×10^5 m/s, [23] which is close to the velocity which we observed. In this system, the electron dynamics is very fast, and completed on a timescale of about 10 fs, which is readily accessible. One EMD step (5 as) takes 120 seconds on 144 computational nodes on the Piz Daint supercomputer (CPU: Xeon E5-2670, GPU: NVIDIA Tesla K20x, available configuration till the end of 2016).

4.3 Simulation of Current through a Metallic Nanotube

4.3.1 Introduction

Two different ab-initio approaches to simulate electronic transport through nanostructures have been compared for a carbon nanotube (CNT). The first one is real-time time-dependant density functional

theory (RT-TDDFT), the second one the Non-Equilibrium Greens Function (NEGF) formalism. Both methods are applied here to the computation of the channel resistance of a CNT placed between two Pd electrodes. The device structure seen in Fig. 4.3. In the RT-TDDFT case, microcanonical/closed boundary conditions are used. The voltage is applied through two planes of fixed electrostatic potential,[24] one on each side of the system. It has been shown that such closed boundary conditions can correctly produce a temporary steady-state.[25, 26]

4.3.2 Computational Setup

The NEGF calculations have been performed at the DFT level via the coupling of CP2K and the quantum transport code OMEN[27–29] and the Pd slabs had to be extended to introduce open-boundary conditions. The time step for the RT-TDDFT simulation was limited to five attoseconds and, while the filtering threshold for the Hamiltonian, the overlap, and the density matrices was 10^{-12}.

4.3.3 Results

At the beginning of the RT-TDDFT simulation the ground-state density is calculated and used as starting point for the electronic wavefunction. During the evolution of the system, following the application of a bias of 0.25 V, three stages are observed (Figs 4.4 and 4.5). Initially, the field is constant across the system, which causes a unidirectional current across both the contacts and the tube. After approximately 100 attoseconds, the field in the metallic contacts vanishes as a result of the contact polarisation. At this point, the current inside the contacts is rather chaotic, but it is uniformly distributed in the nanotube, as the electric field. The current leads to an accumulation of charges at the boundaries of the device. These charges start to shield the applied voltage and therefore the field across the tube slowly decreases, followed by the current intensity. In the final stage, the system reaches an equilibrium. The accumulated charges completely hide the potential difference and no effective current or field remains inside the CNT.
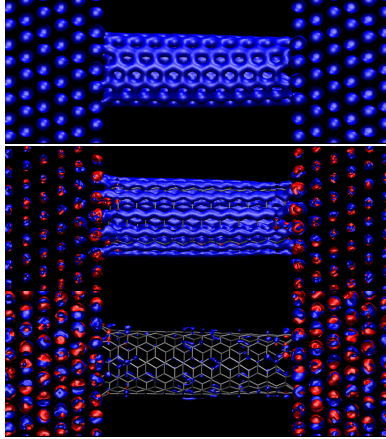
Figure 4.4: Visualisation of the current density through the CNT structure, taken at 25, 500, and 5000 (top to down) attoseconds simulation time. The colour indicates the direction of the local current, blue for right to left propagation, red for left to right.
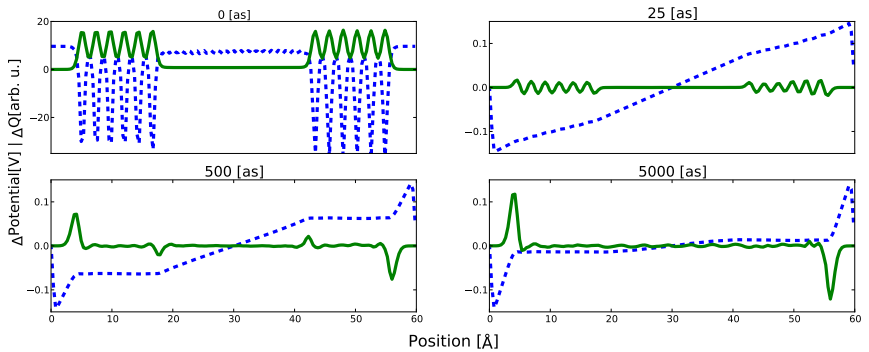

Figure 4.5: Evolution of the electrostatic potential (dashed blue lines) and electronic density (solid green line) over time. Initial state (t=0) (top-left). Changes in electrostatic potential and electronic density between the initial state and the one after 25 (top-right), 500 (bottom-left), and 5000 (bottom-right) attoseconds.

The resistivity of the nanotube can be extracted from the change in charges on the contacts as a function of time. We used two methods (Fig. 4.6) to calculate the resistivity and derived a value of 10.1 and 11.4 $k\Omega$, respectively, which is close to the resistance of 12.9 $k\Omega$, obtained from quantum transport simulations. The first method focuses on the initial phase of the simulation, while the second extracts the resistance from the total simulation, through an equivalent circuit model. The remarkable agreement between both extraction methods is interesting, since the simulation is longer than the time it takes for backscattering to occur inside the contacts. The dependence of the resistivity on the voltage has been carefully analysed by performing simulations at different applied voltages and the system was found to be ohmic, as expected from a metallic nanotube. For the NEGF simulations the contacts must be expanded with more layers of palladium atoms to allow for the introduction of open boundary conditions. The initial wave-function is obtained from a CP2K energy minimisation. The transmission probability through the CNT structure is then calculated semi-self-consistently at zero bias (low-field approximation) and 300 Kelvin in the post-SCF phase. The obtained NEGF resistance is equal to 11.1 $k\Omega$, a result very close to the one derived from RT-TDDFT, thus confirming the validity of the latter approach for transport calculations.

Furthermore, the impact of several defect types (e.g. hydrogen atoms at the metal/CNT interface or Stone-Wales defects[30]) (Fig. 4.7), in the nanotube and the contact regions, was studied. None of the considered configurations was found to significantly alter the resistance.

4.3.4 Conclusion

We have determined the electric resistance of a carbon nanotube connecting two metallic contacts using two different approaches, RT-TDDFT and NEGF. The calculated electric resistances are in good agreement. This supports the usage of RT-TDDFT/EMD as an efficient alternative to NEGF for the simulation of electronic processes from first-principles, RT-TDDFT also has the advantage that it can be applied to non-steady-state cases, such as transient regimes. In the following section, larger and semiconducting instead of metallic systems will be investigated.
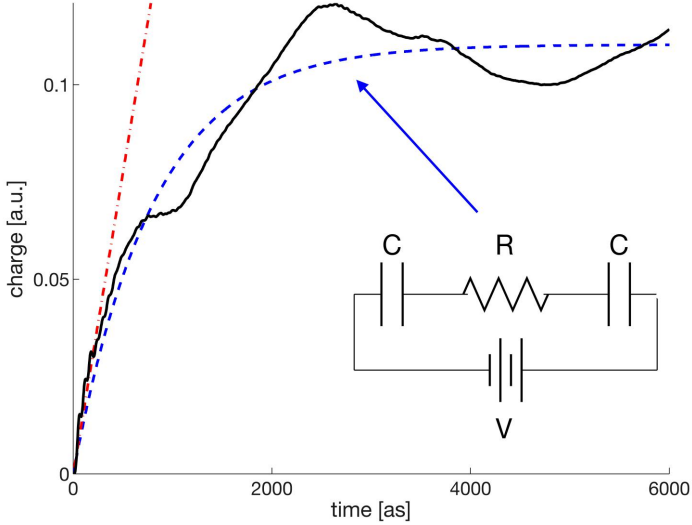
Figure 4.6: Extraction methods for the device resistance based on the electric charge that moves from one metal contact to the other as a function of time. (solid black line) First method: The resistance is calculated from a linear fit of the charge transfer at t=0 (dash-dotted red). Second method: The resistance is estimated from a circuit model, as illustrated in the inset (dashed blue line). The circuit is composed of two capacitors that account for the coupling between the planes with a fixed potential and the Pd contacts and of a resistor that corresponds to the nanotube channel. The analytical solution of the circuit is an exponential function of the form $n(t) = \frac{CV}{2}(1 - exp^{-\frac{2t}{RC}})$.
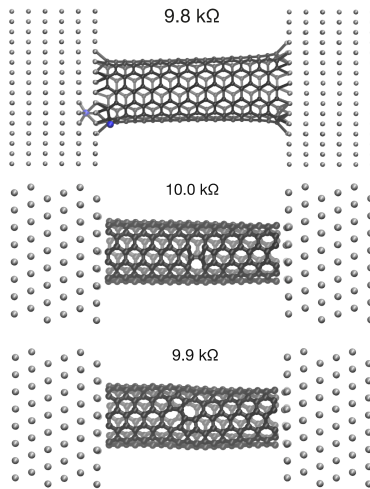
Figure 4.7: Nanotube structure with various defects introduced. The resistances have been calculated with the first method from Fig. 4.6 (time step: 1 as, potential: 1 V). (top) Hydrogen atoms added at the Pd/nanotube interface. (middle, bottom) Stone-Wales defects inside the nanotube in two different orientations.

4.4　Simulation of Current through a Semiconducting Nanotube

4.4.1　Introduction

As a more realistic setup compared to the previous section we now calculate the electron transport through a semiconducting CNT that has larger dimensions as in Section (4.3). The approach used here treats the boundaries as finite capacitors, as in the previous section. It is known to produce an accurate quasi steady-state [26]. We furthermore improve the stability of this method by inserting a dielectric medium inside the capacitor. This technique enables us to simulate systems one order of magnitude larger than previously possible, as RT-TDDFT transport simulations have been so far limited to at most 200 atoms [26, 31–40].

4.4.2　Computational Setup

The dielectric constant between the constraints and the contacts is defined using a spatially, density-dependant function, [24] which confines the dielectric medium to the regions of vanishing electronic density. In the RT-TDDFT simulations the density matrix is filtered with a threshold of 10^{-9} and the time step is set to 10 attoseconds. The NEGF calculations are performed with the same DFT parameters as the RT-TDDFT simulations, which removes the need for the expensive inclusion of Hartree-Fock exchange required to obtain results comparable to experimental data. The contact unit cells interact with up to two or three next nearest-neighbour unit cells.

4.4.3　Results

The current through a carbon nanotube field-effect transistor (CNT-FET) with a floating gate, as in Fig. 4.8, has been calculated. Such logic switches are serious candidates to replace silicon FinFETS at the end of Moore's scaling law. A microprocessor based on CNTFETs has already been demonstrated. [41] The semiconducting nanotube is generated using the TubeGen [20] online tool and the contacts have been prepared with the help of the ASE [42] engine. The length of

left potential — 600 Pd atoms — 9 nm (17,0) CNT 1496 atoms — 600 Pd atoms — right potential

4.5 Å $(\epsilon = 20)$ 4.5 Å $(\epsilon = 20)$

Figure 4.8: Semiconducting CNT connecting two palladium contacts. A voltage is applied through constraints in the Poisson solver, as described in the previous section. The dielectric constant between constraints and the contacts is set to $\epsilon_r = 20$ to allow for a large charge transfer over the course of the simulation. The dielectric constant of the remaining empty space is set to 80 (CP2K's default value). This rather unphysical choice does not change the conclusions of the paper, since the resulting electrostatic potential is used in NEGF as well. The distance of 4.5 Angstrom between the constraints and the contacts ensures the absence of electron density overlap between them.

the nanotube is chosen such as to approximately match experimental CNTFETs. [43] The geometry of the combined system is relaxed with respect to both the atomic positions and the system dimensions in order to establish a stress-free contact region between the CNT and the Pd slabs. As basis functions, DZVP is employed with 25 contracted Gaussians per Pd atom and 13 per carbon one, which results in a Kohn-Sham matrix of 47952×47952 elements for RT-TDDFT and 92952×92952 elements for the NEGF system with extended slabs.

The simulation is initiated from the unbiased ground-state electronic density. From there a bias of 2 V is applied to generate an electric current. In the early stage, the current is irregular due to polarisation effects and initial reflections, but after about 10 fs, a pseudo-steady state establishes. It can be observed in Fig. 4.9. The resistance in this state is calculated from the current, given by the electron drift, and the effective potential drop across the CNT, as reported in Fig. 4.10. This potential difference diminishes over time as electrons accumulate at the far end of the contacts and shield the applied bias, a phenomenon that can be seen in Fig. 4.11. Different approaches were tried to interpret the current as a single number from the time-dependant density (see in appendix). The most reliable one consists in calculating the current based on the charge difference between the contacts as given by their Mulliken charges Q:

$$I(t) = \frac{\partial}{\partial t}\left(\frac{Q^{left} - Q^{right}}{2}\right). \tag{4.1}$$

The numerical derivative is calculated through an interpolation in an interval of $\pm$ 2.5 fs around each time step. The width of this interval serves here as a smoothing parameter.

The resistance is then compared to NEGF calculations, which has been shown to agree well with RT-TDDFT [38, 39] for small systems. A self-consistent NEGF calculation for a system of this size is not yet feasible. The potential of the steady-state has therefore to be approximated. In the central region it is extracted from the RT-TDDFT simulation at different time steps. The contacts are then extended to introduce open boundary conditions. To extract the electrostatic potential from RT-TDDFT at time t, the difference between its value at time t and t_0 is calculated. The result is finally added to the ground-state potential in the central region of the extended system to perform
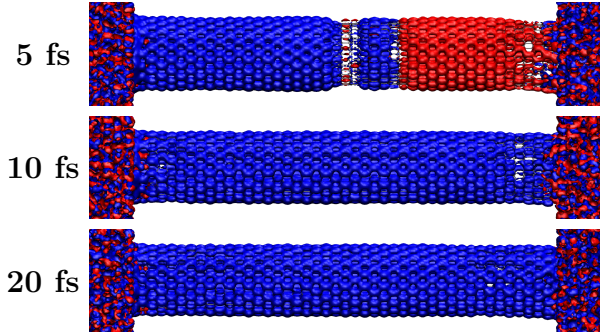
Figure 4.9: Overview of the RT-TDDFT current density at different time steps. The current from right to left is coloured in blue and the current in the opposite direction in red. The currents at $t = 5$, 10, and 20 fs are represented, from top to bottom.

the NEGF calculation:

$$V_{NEGF} = V_{GS} + V_{TDDFT}(t) - V_{TDDFT}(t_0). \qquad (4.2)$$

The potentials in the extended slabs are shifted to be continuous with the central region.

One of the input parameters of NEGF calculations is the temperature of the system, as the occupation of states in the contacts is modelled by a Fermi-Dirac distribution in the Landauer-Büttiker formalism:[44]

$$I(V_b) = \frac{2e}{\hbar} \int_{-\infty}^{\infty} \frac{1}{2\pi} \mathcal{T}(E,V) w_{FD}(E,T) dE,$$
$$w_{FD}(E,T) = f_{FD}^L(E,T) - f_{FD}^R(E,T), \qquad (4.3)$$

where $f_{FD}(E,T)$ is the Fermi-Dirac distribution function at the energy E and temperature T, the factor 2 accounts for the spin degeneracy and $\mathcal{T}(E,V)$ refers to the transmission function at energy E. In RT-TDDFT simulations the initial electronic density is computed at zero Kelvin to obtain an idempotent density matrix. While this

82
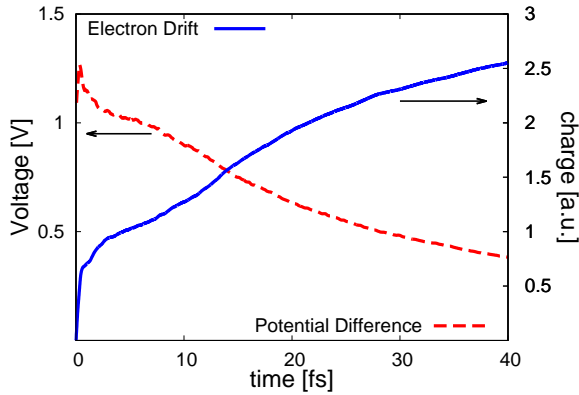
Figure 4.10: Voltage difference between the left and the right contact. With voltages measured as planar averages through the middle layers of the respective slabs (dashed line); total charge moved from one side to the other as determined from the Mulliken charges in the contacts (solid line).
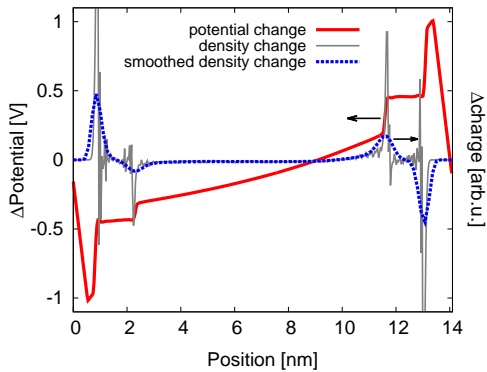

Figure 4.11: Change in the electrostatic potential (thick red line) and charge density (real data: thin grey line, smoothed data: dashed green line) in the CNT system after 10 fs as compared to the initial state.

suggests the use of a zero Kelvin temperature in NEGF as well, it is found that such a choice does not accurately represent the situation after the application of a bias to the RT-TDDFT simulation. In this case, the eigenstates of the density matrix and the Kohn-Sham matrix are no longer identical.

To determine a suitable temperature for the NEGF calculation the density-of-states (DOS) at time t is interpreted as the eigenvalues of the time-dependent Kohn-Sham matrix $H(t)$. The occupation at each energy level can then be determined by projecting the eigenstates of the density matrix $P(t)$ onto the eigenstates of $H(t)$. This has been done here after 20 fs of simulation time, with an atomic resolution. By calculating the DOS in the metallic contacts, as shown in Fig. 4.12, a similar window-function as in the Landauer-Büttiker formalism can be established:

$$w_{RT-TDDFT}(E) = f^L_{RT-TDDFT}(E) - f^R_{RT-TDDFT}(E), \qquad (4.4)$$

where $f^{L/R}_{RT-TDDFT}(E)$ is the relative occupation of the DOS inside the left and right contact respectively. The function computed in this way is not fully equivalent to $w_{FD}(E,T)$, but it can be used to fit the temperature of the NEGF scheme. It turns out that the RT-TDDTFT window function $w_{RT-TDDFT}(E)$ can be matched remarkably well by $w_{FD}(E,T)$ at $T = 300$ K, which is then chosen for the NEGF calculations, see Fig. 4.13. It should be noted that the equivalence to room temperature is a pure coincidence and is not enforced by the simulation setup. Using this temperature, the electrostatic potential is extracted every 10 fs of the RT-TDDFT run and is then transferred to the NEGF solver in order to compute the stationary current of the system. A resistance is finally extracted from the current and compared to the RT-TDDFT results in Fig. 4.14. The potential drop at each selected time step is also calculated and used to convert the time to a voltage. A fairly good agreement between the NEGF and RT-TDDFT methods is found and highlighted in Fig. 4.14.
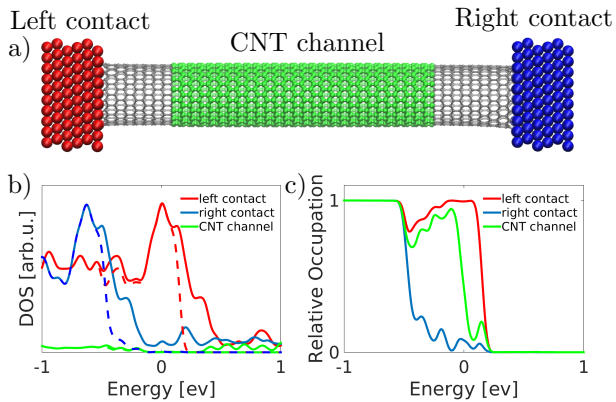
Figure 4.12: a) Atomic structure in which the DOS is calculated after 20 fs of RT-TDDFT simulation. Extractions in the left contact, CNT channel and right contact are performed. b) Total DOS in each region (solid lines) as well as energy resolved density of the occupied states, as obtained by projecting the density matrix onto the eigenstates of the Kohn-Sham matrix (dashed lines). c) Fraction of the occupied states at any given energy for the three different parts of the system (left and right contact, CNT channel).
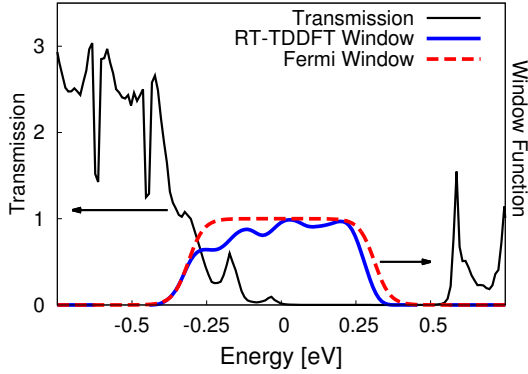
Figure 4.13: Transmission function as obtained from NEGF (black, solid) and difference in the occupation of states between the two contacts $w(E)$, as given by the RT-TDDFT simulation (blue, solid) and by the Fermi-distribution at 300 K at $t = 20 fs$ (red, dashed).
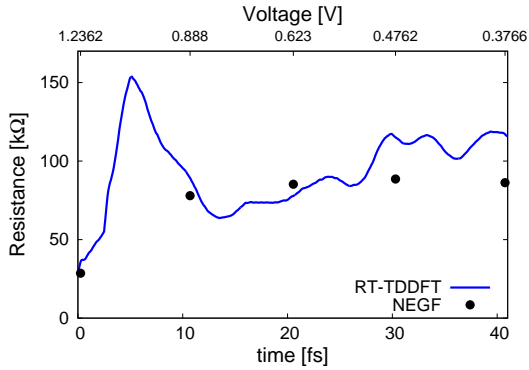
Figure 4.14: Resistance of the CNT FET in Fig. 4.8 as calculated from both RT-TDDFT (solid line) and NEGF (dots). The potential of the central region in the NEGF calculations is taken from the RT-TDDFT simulation at the time indicated by the bottom x-axis. The corresponding voltage difference between the left and right contact is given in the top x-axis.

4.5 Current-Voltage Characteristics of Germanium Selenide Selectors

As another illustration of the RT-TDDFT method, we have simulated a GeSe selector composed of a $Ge_{0.5}Se_{0.5}$ slab surrounded by two TiN electrodes as in Fig. 4.15. Such structures may find applications in crossbar arrays, where they could minimise between-line read and write interferences. [45]

The amorphous model of $Ge_{0.5}Se_{0.5}$ has been generated with CP2K. We have combined GTH pseudo-potentials with a localised single-zeta valence (SZV) basis sets in a melt-and-quench algorithm: The initially randomly decorated atomic GeSe configuration is melted at 1200 K for 6 ps, then quenched at 800 K for 12 ps, and finally at 400 K during 6 ps with a larger basis set (DZVP). The final structure is relaxed.

The considered voltage is too high to employ a low-field approximation, where the transmission function in Eq. (4.3) is assumed to be bias independent. Consequently, the reference NEGF calculations are performed after imposing a non-self-consistent linear potential drop between the metallic contacts. The non-equilibrium Hamiltonian is generated by solving the Poisson equation under a bias potential that is forced to be flat in the contact regions and to behave linearly in the GeSe channel. Such a condition on the potential is justified by the very low charge density in the $Ge_{0.5}Se_{0.5}$ central region. The potential of the left TiN contact is set to 0, while it is equal to -qV in the right contact region. Moreover, homogeneous Neumann boundary conditions are applied in the transport direction, z, and periodic boundary conditions along the x and y directions. Once the Hamiltonian is created, it is passed, together with the overlap matrix, to OMEN that evaluates the transmission through the device. The operation is repeated at different voltages.

The RT-TDDFT simulations are started at different initial biases and evolve over time till the applied potential is completely screened by the electric field coming from the displaced charges. A good agreement with NEGF is found in both cases, as reported in Fig. 4.16. Furthermore, it should be emphasised that RT-TDDFT produces a similar I-V curve, regardless of the initial bias. This indicates that
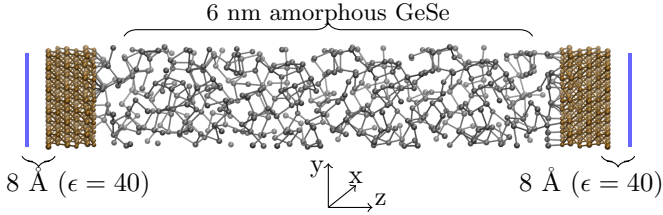
Figure 4.15: Amorphous $Ge_{0.5}Se_{0.5}$ slab between two TiN contacts. The calculation is performed with a DZVP basis with 13 basis functions for the N, Ge, and Se atoms and 26 for the Ti one. The system is $1.3 \times 1.3 \times 7.6$ nm^3 large and contains 744 atoms.

the current is determined by the bias alone, not by the previous state of the system.

4.6 Plasmonic Effects in Conductive Bridging Random Access Memories

4.6.1 Introduction

As part of this thesis conductive bridging random access memory (CBRAM) cells, [46] a promising technology [47] for non-volatile memory and neuromorphic computing, have been investigated. The operation principle of CBRAM is illustrated in Fig. 4.17. The motivation for this study originates from the experimental observation that plasmons can be used to switch CBRAM cells. [48, 49] To gain insight into the atomistic processes occuring during the switching process, the electrostatic forces through the applied field as well as the plasmonic forces caused by the so-called "Optical Tweezer" [50] effect have been computed. The goal was to understand if the switching is a direct result of the tweezer force, or a result of the thermal heating related to the absorption of the plasmon. Furthermore, RTP simulations were conducted to investigate optical rectifications, an effect where an oscillating optical field induces a constant electric field and therefore an unidirectional force acting on atoms. As the systems were already
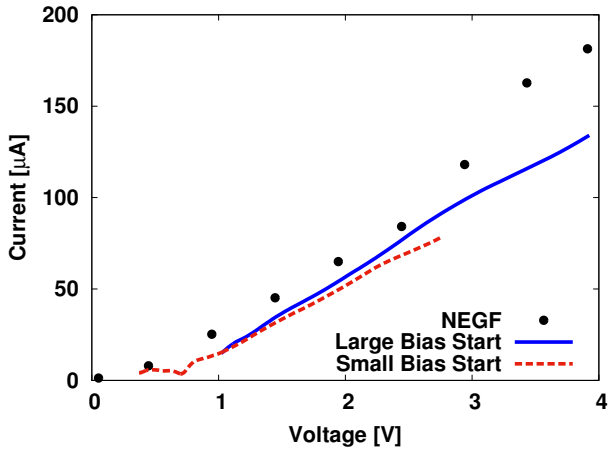
Figure 4.16: Current vs. voltage characteristics of the $Ge_{0.5}Se_{0.5}$ selector in Fig. 4.15 as obtained from two different runs of RT-TDDFT simulations that are started at different initial biases (lines) and from NEGF (dots). The first femtosecond is cut to avoid polarisation effects. The current is extracted with the same method as in the CNT-FET with interpolation in a $\pm$ 2 femtoseconds interval.
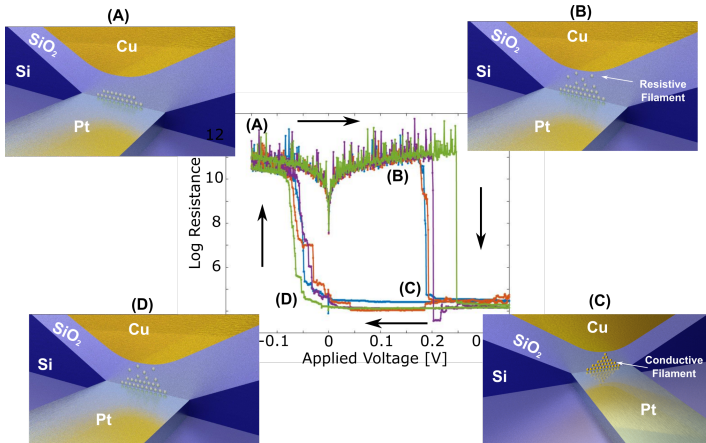
Figure 4.17: Operation principle of a CBRAM cell. (A) The gap between a copper and a platinum electrode is filled with amorphous SiO_2. The resistance between the electrodes is therefore intrinsically high. (B) A voltage is applied so that atoms start to diffuse from the active copper electrode towards the inert platinum electrode, where they accumulate. (C) When enough copper atoms have moved towards the platinum electrode, a conductive filament is formed between both electrodes. The resistance drops as a result. Even after turning out the applied voltage, the filament remains stable and the resistance low. A CBRAM cell is therefore typically non-volatile. (D) To dissolve the filament and switch it back to its high resistance state, a reverse bias must be applied.

created, we decided to use them to further validate the transport simulations based on RT-TDDFT introduced in the previous sections, and compare it to the "current vs. voltage" characteristics obtained from NEGF.

4.6.2 Electrostatic Forces

The electrostatic forces are the main driving force for the formation of metallic filaments between two electrodes in a CBRAM cell. One of the electrodes must be inert, e.g. Pt, the other active, e.g. Cu. The forces can be evaluated through EMD simulations by performing two calculations: One with an applied electric field and another that serves as a reference and that can either be an EMD run at zero bias or a ground-state simulation of each of the atomic configurations in the original EMD trajectory. The electrostatic component of the atomic forces can then be extracted from the force difference between the main EMD and the reference simulation. For the latter, we opted for an unbiased EMD calculation and furthermore used an RTP simulation to obtain the electric field.

Setup

The growth of the nanofilament was not simulated in this work. We directly started from a crystalline Cu filament in an amorphous SiO_2 matrix surrounded by two Cu plates for simplicity, as in Fig. 4.18. The stresses in the simulated structure were relaxed through a ground-state DFT geometry optimisation. The simulated structure contains 3031 atoms. The copper atoms of the filament and metallic contact are modelled with a DZVP basis with 25 basis functions per atom and the surrounding amorphous SiO_2 matrix with a 3SP [51] basis with 12 basis functions per atom. This results in a density matrix of size 55690×55690. The cutoff of the auxiliary PW basis was chosen to be 500 Rydberg and the time step of the simulation 10 attoseconds. The electrostatic forces were extracted after approximately one femtosecond simulation time. Due to the relatively short duration of the simulation, a dielectric constant of 5 was sufficient to represent the vacuum between the contacts and the constraints in the potential.
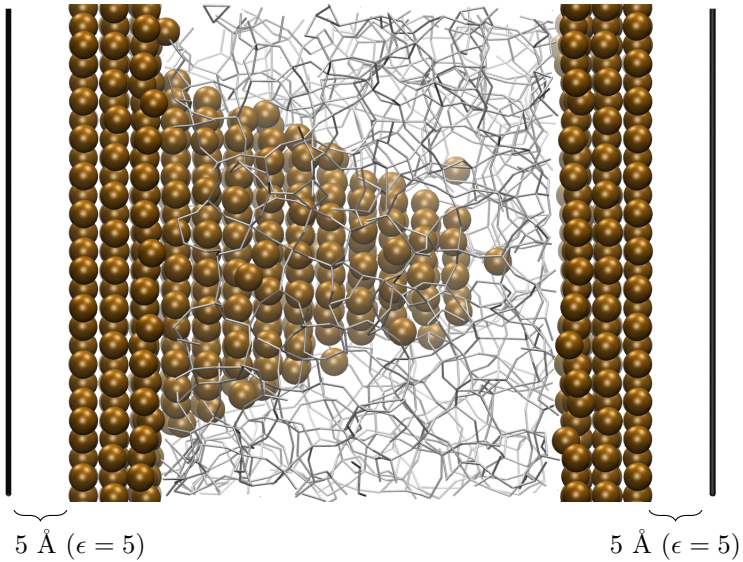
5 Å ($\epsilon = 5$)   5 Å ($\epsilon = 5$)

Figure 4.18: Model system of the considered CBRAM cell. A crystalline copper filament is placed in an amorphous SiO_2 matrix. It extends from one copper electrode to another. The left copper electrode corresponds to the inactive platinum electrode from Fig 4.17 and the right to the active electrode. In black are the planes of constraint electrostatic potential, which are used to apply the bias. Since the filament growth process is not explicitly simulated, the normally inert electrode can be replaced by a Cu one with very little influence on the results.
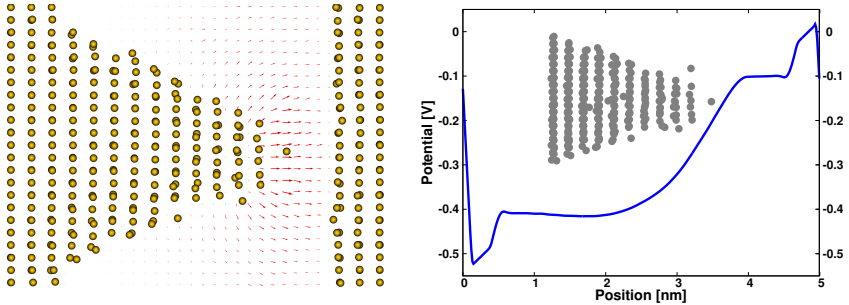
Figure 4.19: Electric field and potential under an applied bias in the CBRAM cell from Fig. (4.18), as obtained from an RTP simulation. (Left) Electric field in a slice taken in the middle of the system. (Right) One-dimensional profile of the electrostatic potential. For the profile, the electrostatic potential is averaged in the two dimensions, corresponding to the dimensions perpendicular to the current direction. As a result, the potential appears to drop off after 2 nm, although the filament extends until 3 nm. This drop only represents potential drop in the surrounding SiO_2 matrix, not the filament itself.

Results

As shown by the RTP simulation in Fig. 4.19, the electric field, as expected, is limited to the SiO_2 matrix and reaches its strongest value in the small gap situated between the filament tip and the contact. The effective potential drop across the gap is approximately 0.3 V and therefore slightly lower than the potential difference of 0.5 V between the constraints.

From Fig. 4.20, we can see that only the atoms at the tip of the filament experience significant electrostatic forces. This is the result of the field being the strongest at the tip, and a consequence of the positive charge of the surface copper atoms. It has to be mentioned that the direction of the DC field corresponds to the process of the dissolution of the filament. Hence, the atoms at the tip are pulled away from the filament with a force of approximately 100 pN.
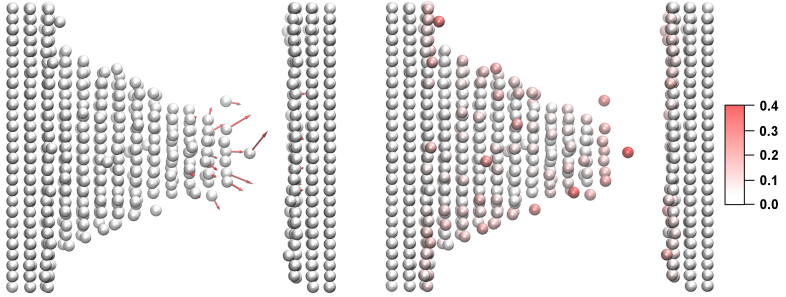
Figure 4.20: (Left) Electrostatic forces acting on a crystalline copper filament under a DC bias of approximately 0.3 V. The arrows indicate the direction and magnitude of the force. (Right) Hirshfeld charges of the copper atoms.

4.6.3 Plasmonic Forces

The optical tweezer effect describes how a polarisable particle is attracted towards the maximal intensity of an optical field. This attraction is caused by the momentum change of photons during refraction, as illustrated in Fig. 4.21. The equation for the force caused by the field gradient on a particle in a dielectric medium is given by Ref.: [52]

$$F_{grad} = 4\pi n_2^2 \epsilon_0 a^3 \left(\frac{m^2 - 1}{m^2 + 2}\right) \frac{1}{2} \nabla E^2(r, t), \qquad (4.5)$$

where n_2 is the dielectric constant of the surrounding medium, $m = \frac{n_1}{n_2}$ the relative dielectric constant of the particle, a the particle radius, and ∇E the gradient of the electric field. For single atoms, however, it is not immediately obvious how the dielectric constant should be chosen. Therefore, we decided to directly simulate the plasmon as a non-propagating AC field and calculate the resulting forces instead if applying Eq. (4.5). This has been done by following the same procedure as in the previous section, i.e. by performing two simulations and extracting the forces between them. As some of the filament atoms have a positive charge, they oscillate in the field. The force is thus averaged over the period of a single plasmon oscillation, here 2.5 fs, corresponding to a frequency of 400 THz and a wavelength of approx-
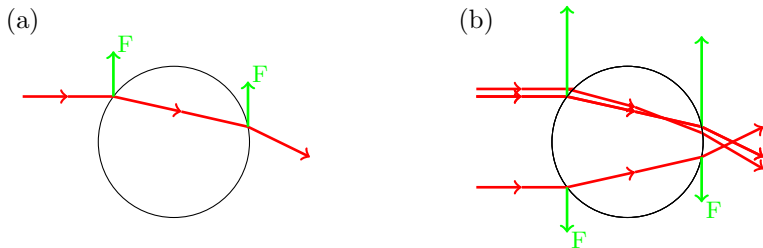
Figure 4.21: "Optical Tweezer" effect: (a) A photon hits the particle and is refracted. The momentum is conserved through a force on the particle, opposite to the momentum change of the photon. (b) If the intensity of light on both sides of the particle are different a net Force in the direction of the higher intensity is the result.

imately 750 nm, cancelling out the effect of the constant charge and isolating the tweezer force.

System

The plasmonic force was calculated in a system with a silver filament connecting silver contacts, as in, as in Fig. 4.22. As in the calculation of the electrostatic forces, the PW cutoff was set to 500 Rydberg and a DZVP basis with 25 basis functions per atom was used for the silver atoms, while a 3SP [51] basis with 12 basis functions per atom described the amorphous SiO_2 matrix, resulting in a total matrix size of 21840×21840. The simulation was run for 7.5 femtoseconds with a timestep of 10 attoseconds. The plasmon was simulated as a one volt AC field with a frequency of 400 THz applied through constraints in the Poisson solver. [24]

Results

The Ag atoms in the filament were found to exhibit forces of up to 20 pN after 5 femtoseconds and up to 15 pN after 7.5 femtoseconds, as reported in Fig. 4.23. However, the direction of the forces does not seem to reach an equilibrium over the time of the simulation. Also,

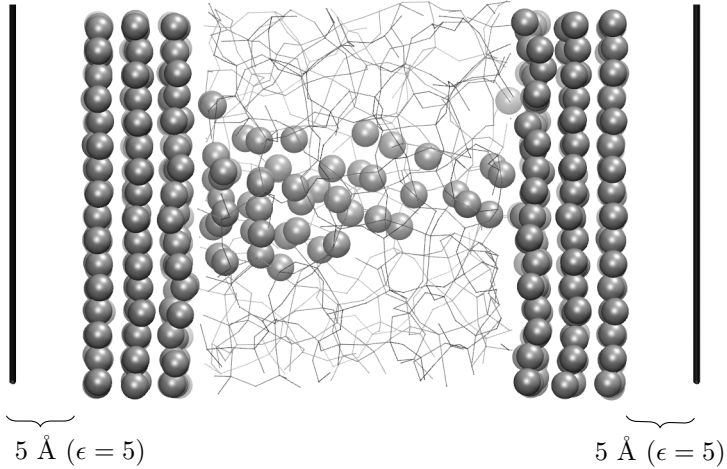5 Å ($\epsilon = 5$)   5 Å ($\epsilon = 5$)

Figure 4.22: Ag filament connecting two Ag slabs. The constraints on the side are used to apply a 1 V AC field with a frequency of 400 THz, corresponding to a wavelength of approximately 750 nm.
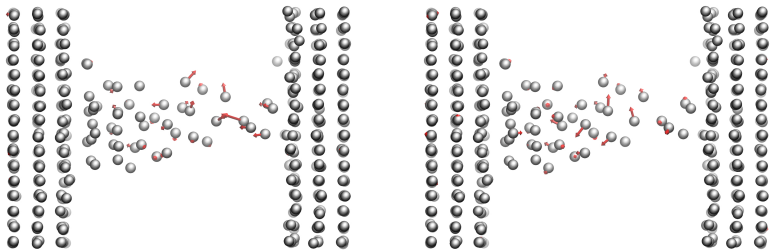

Figure 4.23: (Left) Plasmonic forces extracted after 5 fs or 2 cycles. (Right) Plasmonic forces extracted after 7.5 fs or 3 cycles. The arrows indicate the force strength and direction.

no clear force direction can be identified. Longer simulations are not possible, since molecular dynamics trajectories from identical initial configurations, such as the EMD simulation including the plasmon and the reference simulation, are known to diverge exponentially over time.[53] The difference in the atomic configurations between the two simulations causes an additional force difference between them, which over time dominates the plasmonic component of the forces.

4.6.4 Current Simulations

Introduction

As another validation of electron transport simulations with RT-TDDFT, besides the semiconducting CNT tube and the GeSe selector, the current flowing through a CBRAM cell was also simulated. The idea is to compare the IV curve obtained from an RTP simulation with that created through NEGF. In the NEGF simulations, the low-field approximation is employed, which means that at all bias points, the transmission is calculated from the unbiased Kohn-Sham matrix.

Simulation Domain

The selected device structure is made similar to an experimental CBRAM cell[54] and consists of an amorphous SiO_2 layer surrounded by two Cu electrodes and a metallic filament in between, as depicted in Fig. 4.24. The amorphous silicon oxide is obtained using a melt and quench approach as described in Ref. [55]. Subsequently, copper contacts are attached and the filament inserted by converting all atoms within a cone to copper. The resulting structure is geometry optimised and annealed for more than two picoseconds using ground-state DFT molecular dynamics. For this system an SZV basis set with 9 basis functions per atom is employed for the copper atoms during the NEGF and RT-TDDFT simulation and a DZVP basis with 25 basis functions during the annealing procedure in the systems creation. For the amorphous SiO_2 matrix a 3SP[51] basis set with 12 basis functions per atom is used. It was verified with NEGF that the SZV+3SP combination gives essentially the same result as DZVP, but at much lower computational cost, which allows for the simulation of more realistic structures.
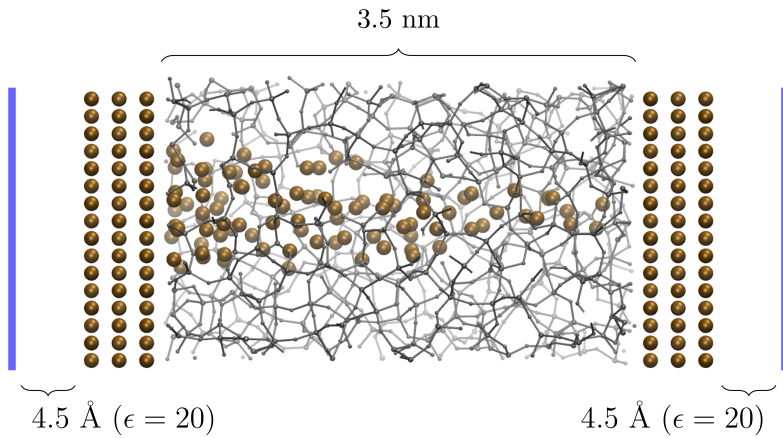
Figure 4.24: CBRAM cell with a copper filament extending from one copper electrode to the other. The filament is surrounded by an amorphous silicon oxide matrix. The device cross section measures 2.06×2.24 nm^2, the total length is set to 4.56 nm for a total of 1569 atoms.

Results

RT-TDDFT is compared to NEGF calculations without direct input from the RT-TDDFT simulations. Instead, the resistance is computed within the low-field approximation, applying the Landauer-Büttiker formalism directly to the unbiased transmission function, i.e. using the same transmission functions at all biases. It has to be mentioned that the Fermi-level of the microcanonical and of the open-boundary structures cannot be accurately aligned.

After the creation of the initial CBRAM structure a ground-state molecular dynamic simulation is performed at 800 K. The resistance is calculated with both NEGF and RT-TDDFT at four different time steps that correspond to atomic relocations and therefore different filament configurations. Results are given in Fig. 4.25.

The MD runs induce a resistance change by one order of magnitude. This is not a true ON-OFF switching process, which would require a longer simulation time and a change in the conductivity by several orders of magnitude. In the low resistance regime, a good agreement between RT-TDDFT and NEGF is found. For the high resistance regime, the use of RT-TDDFT is not straightforward, since the total current is very low, thus leading to a higher susceptibility to noise. Nonetheless, the behaviours qualitatively agree.

4.6.5 Optical Rectification

As a further investigation, an effect from non-linear optics called optical rectification [56] was examined. In optical rectification, an AC field is partially transformed into a DC field, when applied to an asymmetric system. In the present case this phenomenon could appear when the plasmon interacts with the copper filament. We were especially interested in the impact of the filaments conductivity on the rectification.

The system was set up equivalently to the previous section, but instead of a DC field, an AC field with one volt magnitude and 197 THz frequency was applied. We simulated configurations taken from the same MD trajectory mentioned previously and investigated the configurations after 435 fs for the non-conducting case and after 1605 fs for the conducting system.
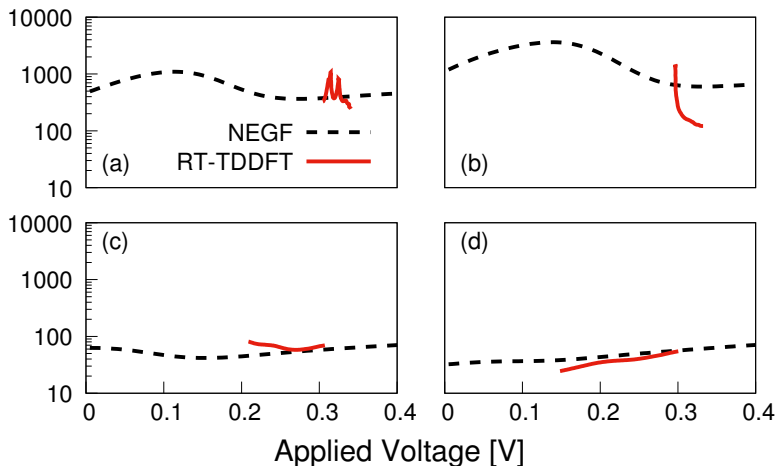
Figure 4.25: Evolution of the electric resistance of the copper filament system in Fig. 4.24 during a molecular dynamics simulation. The frames are taken after 0 fs (a), 657 fs (b), 1456 fs (c), and 1605 fs (d). The resistance calculated by NEGF (black, dashed) and RT-TDDFT (red, solid) are reported. The first 2 fs of the simulation are omitted to eliminate initial polarisation effects. For the calculation of the current, the width of the interval for the interpolation for the numerical derivative of the charge is 1 fs wide.

To extract the DC component of the field, the electrostatic potential along the filament was averaged over the time period of the plasmon oscillation, thus eliminating the AC component.

Results

As seen in Figure 4.26, a rectification effect can be observed in the conducting filament, where a 30 mV strong DC bias is found between the two contacts, which is close to experimental results for similar systems.[57] The associated field appears to be uniform across the filament. In the nonconducting filament, the potential is lowered in the centre of the filament, but the voltage difference between the slabs is negligible.

4.6.6 Conclusions

To establish the main contribution to the plasmonic switching of CBRAM devices, the strength of the "Optical Tweezer" effect was investigated. The atomic forces were then compared to electrostatic forces originating from the DC bias. The electrostatic forces were found to be at least an order of magnitude stronger and thus the tweezer effect is unlikely the cause of the plasmonic switching. This supports an alternative explanation of the switching, which is based on the heating of the filament as a result of the absorption of the plasmon.[54] Furthermore the presence of an optical rectification effect was demonstrated, the implications of this effect remain to be investigated in future research.

4.7 The Band Gap of TiO_2 Nanoparticles in Acetonitrile Solution

The electronic structure of TiO_2 anatase nanoparticles in an acetonitrile solvent was investigated, as a test of the ground-state Born Oppenheimer (gs-BO) subsystem (SS) DFT implementation. Such particles are central for dye-sensitised solar cells (DSSCs) by serving as the conduction path for the photo-excited electrons from dye
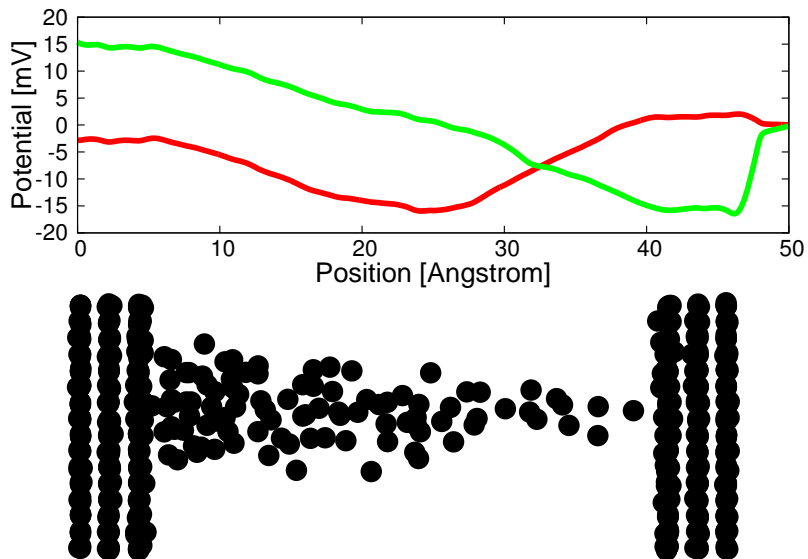
Figure 4.26: (top) Change in potential due to optical rectification in a conducting (green) and insulating (red) copper CBRAM after 9 femtoseconds of simulation time. The potential is averaged over the plasmon frequency. (bottom) Investigated filament structure.
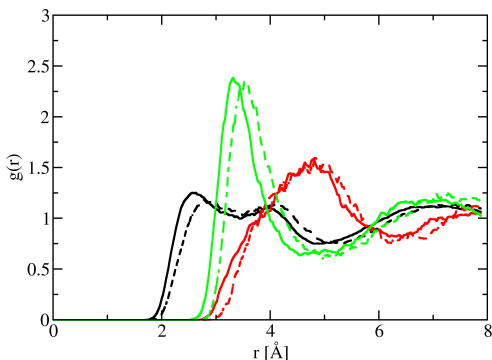
Figure 4.27: Comparison of the pair correlation functions $g(r)$ obtained for liquid acetonitrile with SS DFT (solid line) and KS DFT (dashed line). black: N-H, red: nitrile C-C, green: methyl C-N.

molecules. [58] To validate the performance of the SS DFT implementation, several tests were performed.

First, the structure of the neat liquid has been computed using MD based simulations (10-15ps in length) based on both KS DFT and SS DFT. Pair correlation functions for a cubic unit cell (edge 15.74Å) containing 45 molecules are shown in Figure 4.27. Overall, the structure of the liquid is well described by the SS DFT approach. All peaks are well reproduced and are similar in height. The SS DFT approach leads to a small (0.1-0.2Å) inwards shift of the peaks, suggesting that some more repulsion would be required in the PBE+LLP functional. On the basis of these data, we consider the SS DFT approach suitable to describe liquid acetonitrile.

The next test was a comparison of the density of states (DOS) of a 2nm-TiO_2 anatase nanoparticle, calculated using SS DFT to the DOS of a full KS DFT calculation (Figure 4.28). The DOS is well reproduced for the occupied states and for the unoccupied states close to the band gap. Therefore, both, the full KS and SS DFT will produce the same band gap. Although the discrepancy in DOS for higher un-
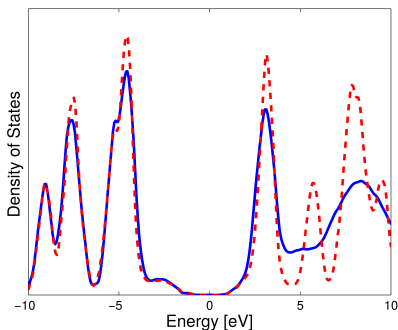
Figure 4.28: The DOS of a 2 nm TiO_2 nanoparticle in explicit acetonitrile solvent. The red dashed line represents the DOS as obtained without partitioning the system, while the blue solid line is the DOS for the SS+LS approach.

occupied states is non-negligible, the similarity of DOS in the region of interest, the band gap region, supports the reliability of this method.

The speedup provided by the SS DFT implementation has been investigated. Despite the relatively small size of the nanoparticles, the number of atoms in the systems is in the order of few thousands due to the solvent. Therefore, the systems greatly benefited from the combination of SS DFT with the linear-scaling framework. The nanoparticles are large enough to be in the linear-scaling regime and SS DFT reduces the cost of the simulation of the acetonitrile solvent. For a small particle (210 atoms particle, 4500 atoms solvent) SS DFT provided a speedup by about a factor of 5.4 and for a larger particle (1233 atoms particle, 3480 atoms solvent) a speedup of 4.4.

Lastly, the band gap was also investigated as a function of the particle size for particles with sizes between 2 and 6 nm (Figure 4.29). At this scale, experimental studies [59–63] have reported that the band gap strongly depends on the size of particles due to the quantum confinement effect which states that the band gap increases with decreasing particle size. However, in our simulations the band gap was found to decrease with decreasing particle size. The decrease in the band gap for small particles could originate from their larger surface to volume
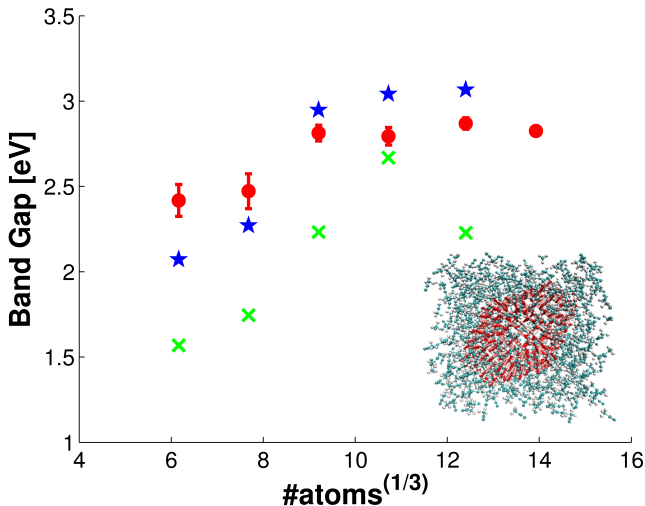
Figure 4.29: The band gap of TiO_2 nanoparticles as a function of the particle size. (red) Solvated particles in acetonitrile. DFT+U (U=6.5eV). The band gap was averaged over an MD simulation. (blue) Particles in vacuum, DFT+U (U=6.5eV), (green) Particles in vacuum, no DFT+U used. Insert shows the solvated system with 1233 atoms in the nanoparticle.

ratio. This surface to volume ratio increases the contribution of the electronic states of the under-coordinated surface atoms, which leads to a smaller band gap. This mechanism has also been suggested based on experiments. [64, 65]

4.8 Satellite Tobacco Mosaic Virus

For the final demonstration of the ground-state combination of the subsystem (SS) and linear scaling (LS) approaches, we compute the electronic structure of the satellite tobacco mosaic virus in solution. This virus was the first entire life form simulated at the atomic level by Freddoline et al in Ref. [67] using empirical force fields. Here, we perform the first ever first principles simulation of such a system. This
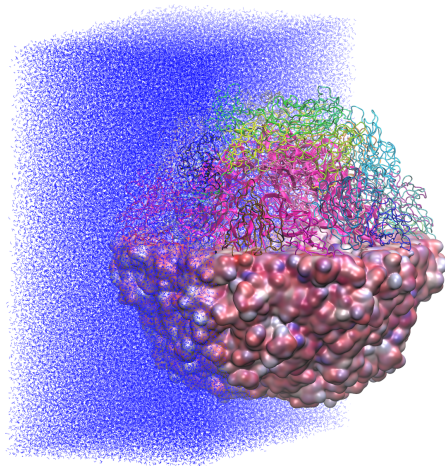
Figure 4.30: Visualisation of the electrostatic properties of the viral capsid as computed with DFT. The solvent (blue sticks), protein and RNA (ribbons) are shown in part, while the surface, which corresponds to the viral capsid, is coloured according to the computed charges of its constituent atoms. The system contains somewhat more than one million atoms. Visualised using VMD. [66]

calculation takes all important interactions (bonding, repulsion, electrostatics and polarisation, van der Waals, etc.) into account based on the electronic structure only. Clearly, our simulations cannot asses the dynamics of the virion at relevant timescales (ns to ms). For these timescales, empirical force fields, running on specialised hardware are more appropriate for the task. [68] With this calculation we illustrate that relaxing such structures and computing their electronic densities is possible and learn how to deal with practical aspects of such calculations. This could be useful to analyse X-ray structure in more detail, [69] which is commonly still based on databases of precomputed densities, [70] even though a subsystem based approach has been applied to proteins already. [71] Furthermore, the computed densities provides a route to obtain electrostatic properties of, and possibly interactions between, large protein assemblies, or the RNA and protein

106

capsid.[72]

The starting point for our calculations is the fully solvated STMV structure that is provided as a benchmark for the NAMD program, and available online.[73] Each of the 60 proteins (2200 atoms) and the RNA fragment (30300 atoms) are considered a molecule in the SS scheme. The latter is highly charged (949 e-, corresponding to the number of nucleotides), which is correctly captured in the SS+LS calculation. The system is neutralised by magnesium and chloride ions, solvated in water. Somewhat more than 50 geometry relaxation steps have been performed to relax the structure. This relaxation is intended to reduce the gradients and total energy, but ideally the system would be equilibrated at a finite temperature, which is still out of reach. The resulting electronic structure has been analysed in terms of the Mulliken charges, as shown in Figure 4.30. During optimisation, total energy is lowered by approx. 2000 Hartree, only 2 mHt/atom, and supports the quality of the initial structure. The root mean square gradient is reduced by roughly one order of magnitude to 7×10^{-4} a.u. . However, the relaxation of the structure is rather challenging in this context. Large gradients are present in the structure, and a trust radius approach, limiting the atomic displacements to 0.05Å turned out to be essential to obtain a stable optimisation in combination with a standard LBFGS[74, 75] optimiser. Without trust radius, large local distortions that result from attempted optimisation steps, would lead to an instable SCF procedure. Linear scaling preconditioning of the geometry optimisation problem might be a complementary approach to further improve the geometry optimisation algorithm, but is not yet available in CP2K. An interesting quantity of the system is the HOMO-LUMO gap. This gap increases from 0.7eV in the original structure to 1.2eV after relaxation. Even though GGA DFT underestimates band gaps,[76, 77] a finite gap is thus obtained for this large structure. This is important as it shows that a properly solvated protein/DNA structure can indeed be described with GGA DFT, even in the presence of various charged residues and nucleotides. Computing the corresponding orbitals is not yet possible, and we therefore do not know if the two orbitals that contribute to this small HOMO-LUMO gap are spatially close, or rather far apart, but this seems an interesting question for future research. Indeed, fluctuations in these large, disordered, systems should modulate the

electrostatic potential in such a way that the orbital energies can be locally pushed up or down significantly, so that the energies of HOMO orbitals of molecular fragments in one region come closer to the energies of LUMO orbitals of molecular fragments in faraway regions. In extreme cases where these levels would cross, Born-Oppenheimer simulations would become problematic. Finally, we believe that the STMV structure is an excellent benchmark system for linear scaling electronic structure calculations. We have established that it is at the same time challenging, but also feasible to deal with this system. It complements the usual benchmarks in that it is fully disordered, three dimensional, and chemically sufficiently rich. The computation of the HOMO-LUMO gap of the original structure, and the relaxation of that structure seem two reasonable targets of different complexity for such a benchmark.

With the SS + LS approach, the computational cost of this calculation is manageable on today's supercomputing resources. Using 2450 compute nodes (each 8 cores, no GPU employed) of a Cray XC30 system at CSCS, one geometry optimisation step requires roughly 1h, while the memory peaks at 13GB per node. The speedup over a purely LS approach is significant, as measured for the first SCF step. SS+LS required 233s on 2450 nodes, while LS required 1385s on 4626 nodes, an 11-fold speedup. Memory usage per node is more similar, as both KS+LS and LS only required 9GB per node for the first SCF step, despite the fact that the occupation of the matrices (8.4M $\times$ 8.4M in size) is 0.019% in the SS+LS and 0.267% in the LS case. The similarity might be due to the fact that the implementation of the SS approach stores several density grids, which are large (2304 $\times$ 2304 $\times$ 2304 points). The usage of GPUs (one NVIDIA K20X with 6GB memory, per node) results in a 30% speedup in the SS+LS case, which mostly results from the accelerated sparse matrix matrix multiplications, as performed by the GPU accelerated version of the DBCSR library.[78] A GPU accelerated run was not possible for the LS only approach, as insufficient GPU memory was available. Interestingly, in the SS+LS approach, the computer time is approximately evenly divided between the linear algebra implied by the LS approach, and the evaluation of the SS embedding terms for the Hamiltonian. The latter is dominated by the needed Fast Fourier Transforms (FFT) to compute the exchange and correlation functionals.

4.9 N3 Dye: Electronic Dynamics in the Linear Response Regime

As a test of the density matrix based EMD and EMD + SS methods, electron dynamics in the linear response regime is performed. The spectrum is then compared to linear response TDDFT (LR-TDDFT), which is an alternative perturbative approach to calculate the absorption spectrum.
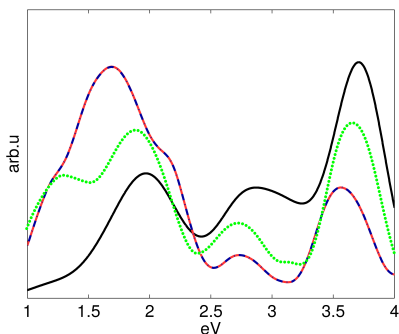

Figure 4.31: The pseudo absorption spectra of an N3 dye in acetonitrile solution (box of 2x2x2 nm^3 with 407 atoms). The spectra were obtained from a 250 fs trajectory, with a time step of 5 as. For the Carbon, Hydrogen, Sulphur, and Oxygen in the dye, TZV2P-MOLOPT-GTH basis sets[3] were used. The spectra were smoothed using the LOESS[79, 80] method, with a span of 1.1 eV. (red, solid) Molecular orbital based EMD, (blue, dash-dot) Density Matrix based EMD, (green, dots) SS-DFT-method with EMD, (black, solid) the TDDFT spectrum of N3 dye in implicit ethanol, calculated by Fantacci et al.[81] The spectrum of the molecular orbital based EMD is exactly on top of a density matrix based EMD simulation.

Using as a test system the N3 dye molecule in acetonitrile, which is a typical sensitiser and solvent in dye sensitiser solar cells (DSSCs), EMD dynamics starting from a ground-state electronic configuration, but at finite ionic temperature, is performed. From this dynamics a pseudo UV-VIS spectrum can be obtained through a Fourier transfor-

mation of the time-dependent electronic and nuclear dipole moment. This spectrum is employed to compare wave function and density matrix EMD, and to study the influence of subsystem DFT model. Furthermore, this spectrum can be compared to a traditional LR-TDDFT calculation, as the peak positions (at fixed nuclei) coincide with that of LR-TDDFT. A correct reproduction of the intensities would require an initial electronic excitation, i.e. as obtained by applying a delta pulse.[82] The performed GGA simulations are expected to underestimate the excitation energies as compared to experiment, which is due to the fact that GGA functionals systematically underestimate the band gap.[76, 77]

The results of these simulations are shown in Figure 4.31. As expected, the density matrix based implementation of EMD was found to produce indistinguishable results from the traditional wave function based EMD implementation in CP2K. Comparing EMD+SS and EMD simulations, fair agreement is found, with peak positions at similar locations, and with similar intensities. Clearly, as the ionic dynamics between the two models differs, some differences are to be expected. We note that the sampling of ionic configurations during the EMD or EMD+SS runs is a useful aspect that is not present in typical, static, RTP or LR-TDDFT calculations, and that this will naturally broaden the peaks in the spectrum. However, the dynamics is roughly 250fs (up to 1ps with SS DFT), so that only the fastest vibrations are averaged over. The fact that simulations of up to 1ps can be performed demonstrates the long-term stability of the method. Finally, the results compare favourably with LR-TDDFT calculations by Fantacci et al.[81] These calculations employ an implicit ethanol solvent model, and employ a different xc functional (BPW91). Despite these differences in computational setup, a similar spectrum is obtained.

Bibliography

[1] Jürg Hutter, Marcella Iannuzzi, Florian Schiffmann, and Joost VandeVondele. CP2K: atomistic simulations of condensed matter systems. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 4(1):15–25, 2014.

[2] Joost VandeVondele, Matthias Krack, Fawzi Mohamed, Michele Parrinello, Thomas Chassaing, and Jürg Hutter. Quickstep: Fast and accurate density functional calculations using a mixed gaussian and plane waves approach. *Comput. Phys. Commun.*, 167(2):103 – 128, 2005.

[3] Joost VandeVondele and Jürg Hutter. Gaussian basis sets for accurate calculations on molecular systems in gas and condensed phases. *J. Chem. Phys*, 127(11), 2007.

[4] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[5] Stefan Goedecker, M. Teter, and J. Hutter. Separable dual-space Gaussian pseudopotentials. *Phys. Rev. B*, 54:1703–1710, Jul 1996.

[6] Alberto Castro, Miguel A. L. Marques, and Angel Rubio. Propagators for the time-dependent Kohn–Sham equations. *J. Chem. Phys*, 121(8):3425–3433, 2004.

[7] Samuel Andermatt, Jinwoong Cha, Florian Schiffmann, and Joost VandeVondele. Combining linear-scaling DFT with subsystem DFT in Born-Oppenheimer and Ehrenfest molecular dynamics

simulations: From molecules to a virus in solution. *J. Chem. Theory Comput.*, 12:3214–3227, 2016.

[8] Samuel Andermatt, Mohammad Hossein Bani-Hashemian, Sascha Brück, Joost VandeVondele, and Mathieu Luisier. Transport simulations with density-matrix-based real-time time-dependant density functional theory. In *International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pages 177–180. IEEE, 2017.

[9] Samuel Andermatt, Mohammad Hossein Bani-Hashemian, Fabian Ducry, Sascha Brück, Sergiu Clima, Geoffrey Pourtois, Joost VandeVondele, and Mathieu Luisier. Microcanonical rt-tddft simulations of realistically extended devices. *J. Chem. Phys*, 149(12):124701, 2018.

[10] Walter R. Duncan and Oleg V. Prezhdo. Theoretical studies of photoinduced electron transfer in dye-sensitized TiO2. *Annu. Rev. Phys. Chem.*, 58(1):143–184, 2007.

[11] Alexey V. Akimov, Amanda J. Neukirch, and Oleg V. Prezhdo. Theoretical insights into photoinduced charge transfer and catalysis at oxide interfaces. *Chem. Rev.*, 113(6):4496–4565, 2013.

[12] Oleg V. Prezhdo, Walter R. Duncan, and Victor V. Prezhdo. Photoinduced electron dynamics at the chromophore-semiconductor interface: A time-domain ab initio perspective. *Prog. Surf. Sci.*, 84(12):30 – 68, 2009.

[13] William M. Stier Walter R. Duncan and Oleg V. Prezhdo. Ab initio nonadiabatic molecular dynamics of the ultrafast electron injection across the alizarinTiO2 interface. *J. Am. Chem. Soc.*, 127(21):7941–7951, 2005.

[14] Colleen F. Craig Walter R. Duncan and Oleg V. Prezhdo. Time-domain ab initio study of charge relaxation and recombination in dye-sensitized TiO2. *J. Am. Chem. Soc.*, 129(27):8528–8543, 2007.

[15] William Stier and Oleg V. Prezhdo. Non-adiabatic molecular dynamics simulation of ultrafast solar cell electron transfer. *J. Mol. Struct.: Theochem*, 630(13):33 – 43, 2003.

[16] Christian F. A. Negre, Valeria C. Fuertes, M. Belén Oviedo, Fabiana Y. Oliva, and Cristián G. Sánchez. Quantum dynamics of light-induced charge injection in a model dye-nanoparticle complex. *J. Phys. Chem. C*, 116(28):14748–14753, 2012.

[17] Roco Snchez de Armas, Jaime O. Lpez, Miguel A. San-Miguel, Javier F. Sanz, Pablo Ordejn, and Miguel Pruneda. Real-time TD-DFT simulations in dye sensitized solar cells: The electronic absorption spectrum of alizarin supported on TiO2 nanoclusters. *J. Chem. Theory Comput.*, 6(9):2856–2865, 2010.

[18] Jin Sun, Jian Song, Yi Zhao, and Wan-Zhen Liang. Real-time propagation of the reduced one-electron density matrix in atom-centered Gaussian orbitals: Application to absorption spectra of silicon clusters. *J. Chem. Phys.*, 127(23), 2007.

[19] M. Belén Oviedo and Bryan M. Wong. Real-time quantum dynamics reveals complex, many-body interactions in solvated nanodroplets. *J. Chem. Theory Comput.*, ASAP(0):null, 2016.

[20] J. T. Frey and D. J. Doren. Tubegen 3.4 web-interface. http://turin.nss.udel.edu/research/tubegenonline.html, 2011.

[21] Vasili Perebeinos, J. Tersoff, and Phaedon Avouris. Electron-phonon interaction and transport in semiconducting carbon nanotubes. *Phys. Rev. Lett.*, 94:086802, Mar 2005.

[22] G. Pennington and N. Goldsman. Semiclassical transport and phonon scattering of electrons in semiconducting carbon nanotubes. *Phys. Rev. B*, 68:045426, Jul 2003.

[23] Yung-Fu Chen and Michael S. Fuhrer. Electric-field-dependent charge-carrier velocity in semiconducting carbon nanotubes. *Phys. Rev. Lett.*, 95:236803, Nov 2005.

[24] Mohammad Hossein Bani-Hashemian, Sascha Brück, Mathieu Luisier, and Joost VandeVondele. A generalized Poisson solver for first-principles device simulations. *J. Chem. Phys.*, 144(4), 2016.

[25] Massimiliano Di Ventra and Tchavdar N. Todorov. Transport in nanoscale systems: the microcanonical versus grand-canonical picture. *J. Phys-Condens. Mat.*, 16(45):8025, 2004.

[26] Neil Bushong, Na Sai, and Massimiliano Di Ventra. Approach to steady-state transport in nanoscale conductors. *Nano Lett.*, 5(12):2569–2572, 2005.

[27] Mathieu Luisier, Andreas Schenk, Wolfgang Fichtner, and Gerhard Klimeck. Atomistic simulation of nanowires in the sp^{3}d^{5}s* tight-binding formalism: From boundary conditions to strain calculations. *Phys. Rev. B*, 74:205323, 2006.

[28] Sascha Brück, Mauro Calderara, Mohammad Hossein Bani-Hashemian, Joost VandeVondele, and Mathieu Luisier. Towards ab-initio simulations of nanowire field-effect transistors. In *Computational Electronics (IWCE), 2014 International Workshop on*, pages 1–3. IEEE, 2014.

[29] Mohammad Hossein Bani-Hashemian. *Large-Scale Nanoelectronic Device Simulation from First Principles.* PhD thesis, Department of Materials, ETH Zürich, 2016.

[30] Anthony J. Stone and David J. Wales. Theoretical studies of icosahedral C60 and some related species. *Chem. Phys. Lett.*, 128(5):501–503, 1986.

[31] Chiao-Lun Cheng, Jeremy S. Evans, and Troy Van Voorhis. Simulating molecular conductance using real-time density functional theory. *Phys. Rev. B*, 74(15):155112, 2006.

[32] Jeremy S. Evans, Oleg A. Vydrov, and Troy Van Voorhis. Exchange and correlation in molecular wire conductance: Nonlocality is the key. *J. Chem. Phys.*, 131(3):034106, 2009.

[33] Na Sai, Neil Bushong, Ryan Hatcher, and Massimiliano Di Ventra. Microscopic current dynamics in nanoscale junctions. *Phys. Rev. B*, 75(11):115410, 2007.

[34] Jeremy S. Evans, Chiao-Lun Cheng, and Troy Van Voorhis. Spin-charge separation in molecular wire conductance simulations. *Phys. Rev. B*, 78(16):165108, 2008.

[35] Xiaofeng Qian, Ju Li, Xi Lin, and Sidney Yip. Time-dependent density functional theory with ultrasoft pseudopotentials: Real-time electron propagation across a molecular junction. *Phys. Rev. B*, 73(3):035408, 2006.

[36] Roi Baer, Tamar Seideman, Shahal Ilani, and Daniel Neuhauser. Ab initio study of the alternating current impedance of a molecular junction. *J. Chem. Phys.*, 120(7):3387–3396, 2004.

[37] Y Wang, C-Y Yam, Thomas Frauenheim, G. H. Chen, and Thomas A. Niehaus. An efficient method for quantum transport simulations in the time domain. *Chem. Phys.*, 391(1):69–77, 2011.

[38] ChiYung Yam, Xiao Zheng, GuanHua Chen, Yong Wang, Thomas Frauenheim, and Thomas A Niehaus. Time-dependent versus static quantum transport simulations beyond linear response. *Phys. Rev. B*, 83(24):245448, 2011.

[39] Kálmán Varga. Time-dependent density functional study of transport in molecular junctions. *Phys. Rev. B*, 83(19):195130, 2011.

[40] Philipp Schaffhauser and Stephan Kümmel. Using time-dependent density functional theory in real time for calculating electronic transport. *Phys. Rev. B*, 93(3):035115, 2016.

[41] Max M. Shulaker, Gage Hills, Nishant Patil, Hai Wei, Hong-Yu Chen, H-S Philip Wong, and Subhasish Mitra. Carbon nanotube computer. *Nature*, 501(7468):526–530, 2013.

[42] Sune R. Bahn and Karsten W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4(3):56–66, 2002.

[43] Aaron D. Franklin, Mathieu Luisier, Shu-Jen Han, George Tulevski, Chris M. Breslin, Lynne Gignac, Mark S. Lundstrom, and Wilfried Haensch. Sub-10 nm carbon nanotube transistor. *Nano lett.*, 12(2):758–762, 2012.

[44] M. Büttiker, Y. Imry, Rolf Landauer, and S. Pinhas. Generalized many-channel conductance formula with application to small rings. *Phys. Rev. B*, 31(10):6207, 1985.

[45] B. Govoreanu, G. L. Donadio, K. Opsomer, W. Devulder, V. V. Afanas'ev, T. Witters, Sergiu Clima, N. S. Avasarala, A. Redolfi, S. Kundu, et al. Thermally stable integrated Se-based OTS selectors with¿ 20 MA/cm 2 current drive,¿ 3.10^3 half-bias nonlinearity, tunable threshold voltage and excellent endurance. In *VLSI Technology, 2017 Symposium on*, pages T92–T93. IEEE, 2017.

[46] Rainer Waser, Regina Dittmann, Georgi Staikov, and Kristof Szot. Redox-based resistive switching memories–nanoionic mechanisms, prospects, and challenges. *Adv. mater.*, 21(25-26):2632–2663, 2009.

[47] Ilia Valov, Rainer Waser, John R Jameson, and Michael N. Kozicki. Electrochemical metallization memoriesfundamentals, applications, prospects. *Nanotechnology*, 22(25):254003, 2011.

[48] A. Emboras, B. Cheng, P. Ma, Y. Salamin, C. Haffner, M. Luisier, C. Hafner, and J. Leuthold. Atomic photodetection. In *2016 Conference on Lasers and Electro-Optics (CLEO)*, pages 1–2, June 2016.

[49] Alexandros Emboras, Alessandro Alabastri, Fabian Ducry, Bojun Cheng, Yannick Salamin, Ping Ma, Samuel Andermatt, Benedikt Baeuerle, Arne Josten, Christian Hafner, et al. Atomic scale photodetection enabled by a memristive junction. *ACS nano*, 12(7):6706–6713, 2018.

[50] Arthur Ashkin, James M. Dziedzic, J. E. Bjorkholm, and Steven Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Opt.lett.*, 11(5):288–290, 1986.

[51] Eeuwe S. Zijlstra, Nils Huntemann, Alan Kalitsov, Martin E. Garcia, and Ulf Von Barth. Optimized gaussian basis sets for Goedecker–Teter–Hutter pseudopotentials. *Motel. Simul. Mater. Sc.*, 17(1):015009, 2008.

[52] Yasuhiro Harada and Toshimitsu Asakura. Radiation forces on a dielectric sphere in the rayleigh scattering regime. *Opt. Commun.*, 124(5):529 – 541, 1996.

[53] GE Norman and VV Stegailov. Stochastic theory of the classical molecular dynamics method. *Mathematical Models and Computer Simulations*, 5(4):305–333, 2013.

[54] Fabian Ducry, Alexandros Emboras, Samuel Andermatt, Mohammad Hossein Bani-Hashemian, Bojun Cheng, Juerg Leuthold, and Mathieu Luisier. Ab-initio modeling of CBRAM cells: from ballistic transport properties to electro-thermal effects. *arXiv preprint arXiv:1712.06681*, 2017.

[55] Evgueni A. Chagarov and Andrew C. Kummel. Ab initio molecular dynamics simulations of properties of a-Al$_2$ o$_3$/vacuum and a-Zr o$_2$/vacuum vs a-Al$_2$ o$_3$/ Ge (100)(2× 1) and a-Zr o$_2$/ Ge (100)(2× 1) interfaces. *J. Chem. Phys.*, 130(12):124717, 2009.

[56] M. Bass, P. A. Franken, J. F. Ward, and G. Weinreich. Optical rectification. *Phys. Rev. Lett.*, 9(11):446, 1962.

[57] Daniel R Ward, Falco Hüser, Fabian Pauly, Juan Carlos Cuevas, and Douglas Natelson. Optical rectification and field enhancement in a plasmonic nanogap. *Nat. Nanotechnol*, 5(10):732, 2010.

[58] Brian Oregan and Michael Grätzel. A low-cost, high-efficiency solar cell based on dye-sensitized. *Nature*, 353(6346):737–740, 1991.

[59] K. Madhusudan Reddy, Sunkara V. Manorama, and A. Ramachandra Reddy. Bandgap studies on anatase titanium dioxide nanoparticles. *Mater. Chem. Phys.*, 78(1):239 – 245, 2003.

[60] S. Monticone, R. Tufeu, A. V. Kanaev, E. Scolan, and C. Sanchez. Quantum size effect in TiO2 nanoparticles: does it exist? *Appl. Surf. Sci.*, 162163:565 – 570, 2000.

[61] Norifusa Satoh, Toshio Nakashima, Kenta Kamikura, and Kimihisa Yamamoto. Quantum size effect in TiO2 nanoparticles prepared by finely controlled metal assembly on dendrimer templates. *Nat. Nanotechnol.*, 3(2):106–111, 2008.

[62] Claudius Kormann, Detlef W. Bahnemann, and Michael R. Hoffmann. Preparation and characterization of quantum-size titanium dioxide. *J. Phys. Chem.*, 92(18):5196–5201, 1988.

[63] Masakazu. Anpo, Takahito. Shima, Sukeya. Kodama, and Yutaka. Kubokawa. Photocatalytic hydrogenation of propyne with water on small-particle titania: size quantization effects and reaction intermediates. *J. Phys. Chem.*, 91(16):4305–4310, 1987.

[64] Junguang Tao, Tim Luttrell, and Matthias Batzill. A two-dimensional phase of TiO2 with a reduced bandgap. *Nat. Chem.*, 3(4):296–300, 2011.

[65] Hiroko Ariga, Toshiaki Taniike, Harumo Morikawa, Mizuki Tada, Byoung Koun Min, Kazuya Watanabe, Yoshiyasu Matsumoto, Susumu Ikeda, Koichiro Saiki, and Yasuhiro Iwasawa. Surface-mediated visible-light photo-oxidation on pure TiO2(001). *J. Am. Chem. Soc.*, 131(41):14670–14672, 2009.

[66] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graphics*, 14:33–38, 1996.

[67] Peter L. Freddolino, Anton S. Arkhipov, Steven B. Larson, Alexander McPherson, and Klaus Schulten. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure*, 14(3):437 – 449, 2006.

[68] David E. Shaw, Paul Maragakis, Kresten Lindorff-Larsen, Stefano Piana, Ron O. Dror, Michael P. Eastwood, Joseph A. Bank, John M. Jumper, John K. Salmon, Yibing Shan, and Willy Wriggers. Atomic-level characterization of the structural dynamics of proteins. *Science*, 330(6002):341–346, OCT 15 2010.

[69] Tibor S. Koritsanszky and Philip Coppens. Chemical applications of X-ray charge-density analysis. *Chem. Rev.*, 101(6):1583–1628, 2001.

[70] Slawomir Domagala, Bertrand Fournier, Dorothee Liebschner, Benoit Guillot, and Christian Jelsch. An improved experimental databank of transferable multipolar atom models-ELMAM2. construction details and applications. *Acta Crystallogr. Sect. A*, 68(3):337–351, MAY 2012.

[71] Karin Kiewisch, Christoph R. Jacob, and Lucas Visscher. Quantum-chemical electron densities of proteins and of selected

protein sites from subsystem density functional theory. *J. Chem. Theo. Comput.*, 9(5):2425–2440, MAY 2013.

[72] Pengyu Ren, Jaehun Chun, Dennis G. Thomas, Michael J. Schnieders, Marcelo Marucho, Jiajing Zhang, and Nathan A. Baker. Biomolecular electrostatics and solvation: a computational perspective. *Quart. Rev. Biophys.*, 45(4):427–491, NOV 2012.

[73] STMV. NAMD benchmark input, January 2016.

[74] Ciyou Zhu, Richard H. Byrd, Peihuang Lu, and Jorge Nocedal. L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Trans. Math. Softw.*, 23(4):550–560, December 1997.

[75] José Luis Morales and Jorge Nocedal. Remark on: L-BFGS-B: Fortran subroutines for large-scale bound constrained optimization;. *ACM Trans. Math. Softw.*, 38(1):7:1–7:4, December 2011.

[76] John P. Perdew. Density functional theory and the band gap problem. *Int. J. Quant. Chem.*, 28(S19):497–523, 1985.

[77] Paula Mori-Sánchez, Aron J. Cohen, and Weitao Yang. Localization and delocalization errors in density functional theory and implications for band-gap prediction. *Phys. Rev. Lett.*, 100:146401, Apr 2008.

[78] Ole Schütt, Peter Messmer, Jürg Hutter, and Joost VandeVondele. *GPU-Accelerated Sparse Matrix-Matrix Multiplication for Linear Scaling Density Functional Theory*, pages 173–190. John Wiley & Sons, Ltd, 2016.

[79] William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, 74(368):829–836, 1979.

[80] William S. Cleveland and Susan J. Devlin. Locally weighted regression: An approach to regression analysis by local fitting. *J. Amer. Statist. Assoc.*, 83(403):596–610, 1988.

[81] Filippo De Angelis Simona Fantacci and Annabella Selloni. Absorption spectrum and solvatochromism of the [Ru(4,4-COOH-2,2-bpy)2(NCS)2] molecular dye by time dependent density functional theory. *J. Am. Chem. Soc.*, 125(14):4381–4387, 2003.

[82] Alisa Krishtal, Davide Ceresoli, and Michele Pavanello. Subsystem real-time time dependent density functional theory. *J. Chem. Phys.*, 142(15), 2015.

Chapter 5

Conclusions and Outlook

To enable large scale atomistic simulations beyond the ground-state approximation, an efficient density matrix based version of real-time propagation (RTP) and Ehrenfest molecular dynamics (EMD) was implemented.[1] Sparse linear algebra was employed as it offers the prospect to achieve linear scaling computational cost. The numerical stability and the accuracy of the calculations were further enhanced by removing numerical noise via a two level filtering procedure and a McWeeny purification scheme of the density matrix.

Linear scaling was only partially achieved. Kohn's near-sightedness principle was only proven to hold for the electronic ground-state. It is not generally valid for non-ground-state electronic configurations where a low-magnitude, long-ranged tail of the density matrix was occasionally found. As a result, the density matrix becomes non-sparse at low filtering thresholds, which prevents linear scaling in high accuracy simulations. This limitation is physical in nature and not related to the algorithms or the implementation. Nonetheless, even for a non-sparse density matrix, sparsity occurs during the calculation of higher order terms in the propagation, which can be exploited for performance. Furthermore, RTP and EMD were also combined with subsystem DFT which intrinsically enforces the

near-sightedness of the electrons and therefore circumvents the afore-mentioned issue.

The stability, performance, and correctness of the implementation were demonstrated on a number of test systems.[1] Additionally, RTP and EMD simulations were used to support experimentalists in the investigation of conductive bridging random access memories (CBRAM).[2, 3] Lastly, RTP simulations where used to demonstrate the feasibility of large scale transport simulations of nanoelectronic devices.[4]

With RTP becoming more popular in the field of transport simulations, it would be a natural extension of this work to implement open boundary conditions (OBCs) as an alternative to the microcanonical boundary conditions employed herein. Open boundary conditions can be implemented through complex absorbing potentials (CAPs)[5] at the system boundaries that mimic the interaction with infinite electron reservoirs from which electrons can either enter or leave. By adding OBCs, it would become possible to simulate non-zero currents in the long time limit and to obtain a true steady-state, which would, however, come at a significant increase in computational cost.

This increase could be counteracted by further improving the performance of the current implementation of RTP and EMD. One possibility to accomplish this goal would be to implement the leap-frog propagation algorithm as an alternative to the exponential-midpoint (EM) and enforced time-reversal symmetry (ETRS) algorithms. The leap-frog algorithm is a time-reversible explicit propagation scheme that avoids the inner loop of the implicit schemes currently used. Another approach to enhance the computational performance would be the implementation of explicit complex linear algebra routines in the DBCSR library used to handle the matrix multiplications. This would lead to a reduction of the required bandwidth at the cost of more floating point operations (FLOPs), see section 3.3. As our applications are typically limited by the bandwidth, this could enable a non-negligible speed up.

To summarise, the benefits of this work are threefold. First, the open source nature of the CP2K code and its availability to all of the scientific community make the methods implemented in this work applicable, for instance, to photo absorption or electron transport calculations in atomic scale systems. Second, the insights gained in

this thesis on the limit of achieving a linear scaling regime within RTP/EMD schemes together with the improvements of their stability and performance will serve as a reference for future implementations of such ab-initio techniques. Lastly, as demonstrated with several practical examples, the simulations conducted using the methods developed in this work can accompany experimental efforts on novel nano-device technologies and pave the way for better efficient component generations.

Bibliography

[1] Samuel Andermatt, Jinwoong Cha, Florian Schiffmann, and Joost VandeVondele. Combining linear-scaling DFT with subsystem DFT in Born–Oppenheimer and Ehrenfest molecular dynamics simulations: From molecules to a virus in solution. *J. Chem. Theory Comput.*, 12:3214–3227, 2016.

[2] Bojun Cheng, Alexandros Emboras, Yannick Salamin, Fabian Ducry, Ping Ma, Yuri Fedoryshyn, Samuel Andermatt, Mathieu Luisier, and Juerg Leuthold. Ultra-compact electrochemical metalization cells offering reproducible atomic scale memristive switching. *Adv Electron Mater*, 2018. submitted.

[3] Alexandros Emboras, Alessandro Alabastri, Fabian Ducry, Bojun Cheng, Yannick Salamin, Ping Ma, Samuel Andermatt, Benedikt Baeuerle, Arne Josten, Christian Hafner, et al. Atomic scale photodetection enabled by a memristive junction. *ACS nano*, 12(7):6706–6713, 2018.

[4] Samuel Andermatt, Mohammad Hossein Bani-Hashemian, Fabian Ducry, Sascha Brück, Sergiu Clima, Geoffrey Pourtois, Joost VandeVondele, and Mathieu Luisier. Microcanonical rt-tddft simulations of realistically extended devices. *J. Chem. Phys*, 149(12):124701, 2018.

[5] J. G. Muga, J. P. Palao, B. Navarro, and I. L. Egusquiza. Complex absorbing potentials. *Phys. Rep.*, 395(6):357–426, 2004.

Appendix

Proof of the Series Expansion of Density Matrix based EMD

In the following Appendix, we will show how the propagation of the density matrix can be expanded into a series. It has been previously published in Ref. [Andermatt et al., J. Chem. Theory Comput., 12:3214–3227, 2016].

$$e^X P(e^X)^* = (\sum_i \frac{1}{i!} X^i) P (\sum_i \frac{1}{i!} X^i)^* = \sum_n Z_n \qquad (5.1)$$

Where Z_n are the contribution of all terms depending on the nth order of the exponential of X:

$$Z_n = \sum_{i=0}^{n} \frac{1}{(n-i)!i!} X^{n-i} P(X^i)^*. \qquad (5.2)$$

We will prove the following solution, in which each term Z_n can be calculated using a single matrix multiplication:

$$Z_n = \frac{1}{n}(XZ_{n-1} + (XZ_{n-1})^*), Z_0 = P \qquad (5.3)$$

For Hermitian Z_n this equation is equivalent to:

$$Z_n = \frac{1}{n}(XZ_{n-1} + (XZ_{n-1}^*)^*), Z_0 = P \qquad (5.4)$$

Inserting Z_{n-1} into the formula yields.

$$
\begin{aligned}
Z_n = &\frac{1}{n}\bigg(X \sum_{i=0}^{n-1} \frac{1}{(n-i-1)!i!} X^{n-i-1} P(X^i)^* \\
&+ \bigg(X \bigg(\sum_{i=0}^{n-1} \frac{1}{(n-i-1)!i!} X^{n-i-1} P(X^i)^* \bigg)^* \bigg)^* \bigg) \\
= &\frac{1}{n}\bigg(\sum_{i=0}^{n-1} \frac{n-i}{(n-i)!i!} X^{n-i} P(X^i)^* \\
&+ \sum_{i=0}^{n-1} \frac{1}{(n-i-1)!i!} X^{n-1-i} P(X^{i+1})^* \bigg).
\end{aligned}
\tag{5.5}
$$

Shifting the indices of the second term:

$$
Z_n = \frac{1}{n} \sum_{i=0}^{n-1} \frac{n-i}{(n-i)!i!} X^{n-i} P(X^i)^* + \frac{1}{n} \sum_{i=1}^{n} \frac{i}{(n-i)!i!} X^{n-i} P(X^i)^*).
\tag{5.6}
$$

The common terms of both series can be added

$$
\begin{aligned}
Z_n = &\frac{1}{n!} X^n P + \frac{1}{n!} P(X^n)^* + \frac{1}{n} \sum_{i=1}^{n-1} \frac{n-i+i}{(n-i)!i!} X^{n-i} P(X^i)^* \\
= &\sum_{i=0}^{n} \frac{1}{(n-i)!i!} X^{n-i} P(X^i)^*
\end{aligned}
\tag{5.7}
$$

Which is equal to Eq. (5.2) and therefore proves Eq. (5.4). To prove the equality of the Eqs. (5.3) and (5.4), all Z_n need to be Hermitian, which can be proven by induction. The first term $Z_0 = P$ is Hermitian. Inserting Z_0 into Eq. (5.4) we obtain:

$$
Z_1 = \frac{1}{n}(XP + (XP^*)^*) = \frac{1}{n}(XP + (XP)^*)
\tag{5.8}
$$

As the sum of a matrix and its adjoint is always Hermitian, Z_1 is Hermitian. It follows by induction that all Z_n are Hermitian.
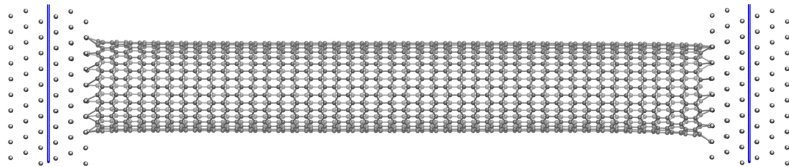
Figure 5.1: Extraction of the voltage. The change in potential as compared to the initial state is calculated in central planes in the contacts. The voltage is then defined as the average potential difference between the two planes

Extracting the Current from RT-TDDFT Simulations

The following section is based on a previous publication [Andermatt et al., J. Chem. Phys, 149(12):124701, 2018] and describes how we obtained the current from an RT-TDDFT simulation at the example of the semiconducting CNT.

To calculate the resistance of a device, the voltage and the current need to be calculated. Neither of them is trivially defined in an RT-TDDFT simulation. The electrostatic potential and the electronic density are three dimensional objects and need to be broken down to single numbers. To calculate the voltage we chose to average the potential in a plane through the centre of each contact (see Fig. 5.1).

For the extraction of the current, four different definitions were investigated. First, the current was interpreted as the time derivative of the sum of the Mulliken charges of the contacts. Second, the contact charges are calculated through an integration of the electronic real-space density over the region of the contacts and the beginning/end of the nanotube and the current is interpreted as the time-derivative of it. Third, the centre of mass of all electrons was calculated and the current interpreted as its movement. Lastly, the current density was directly calculated from the imaginary part of the density matrix. The current was then obtained by integrating the current density through a plane in the middle of the CNT. For the semiconducting nanotube the results of all four approaches are illustrated in figure 5.2.
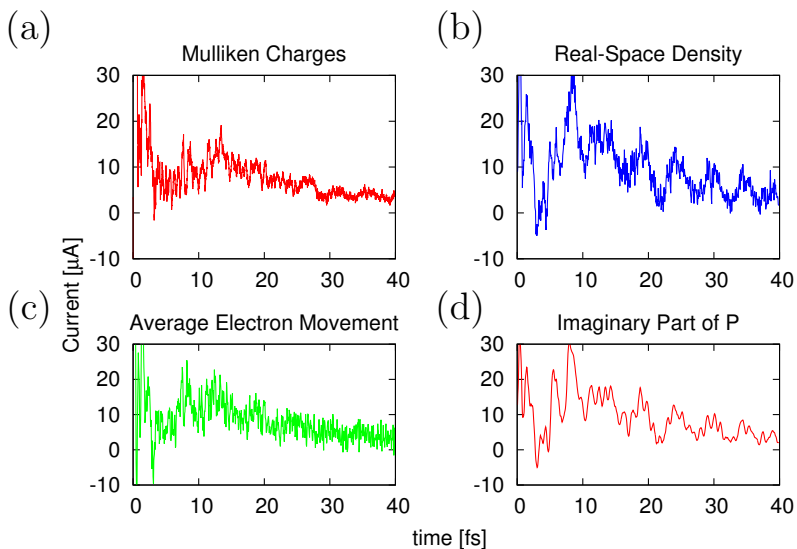
Figure 5.2: Current through the semiconducting nanotube as extracted from four different methods. (a) Change in Mulliken charges from the slab atoms. (b) Calculated from the change in the total real-space electronic density of the slabs and the beginning/end of the nanotube. (c) Calculated based on the movement of the centre of mass of all electrons. (d) Current obtained by calculating the current density from the imaginary part of the density matrix P and subsequent integration of the current density in a plane in the centre of the CNT.
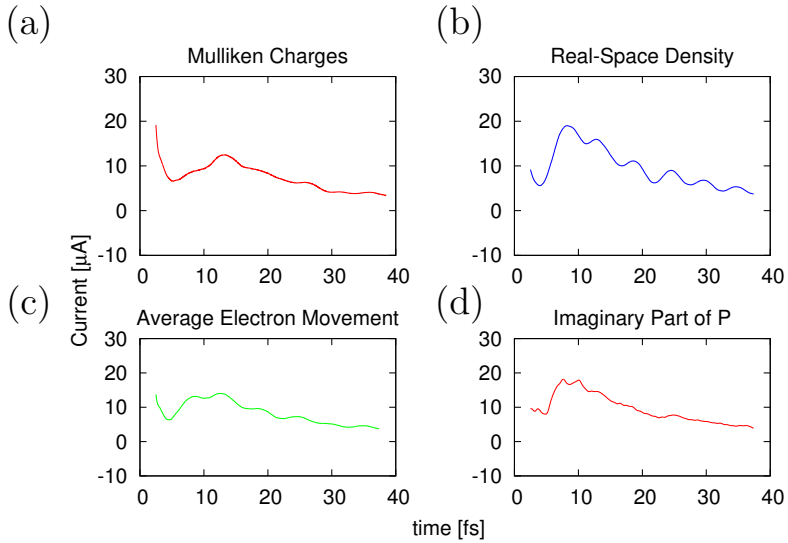
Figure 5.3: As in Fig. 5.2, but taking the derivative of the charge over a 5 fs wide window for (a),(b) and (c) and for taking a moving average with a width of 5 fs for (d).

For all approaches the current exhibits oscillations over time. To smooth these oscillations, the current can be extracted through a linear interpolation of the transmitted charge over time, around each point in time. By widening the range of interpolation a more stable results is obtained (see Fig 5.3).

After the comparison of all four approaches we decided to use the Mulliken charges of the slabs to calculate the current and consequently the resistance.

Curriculum vitae

Personal Data

Name	Andermatt Samuel
Address	Ruchackerstrasse 3, 8046 Zürich
Telephone	+41 79 101 03 17
E-Mail	samuel.t.andermatt@gmail.com
Nationality	Swiss
Civil Status	Single
Date of Birth	21.08.1987
Native Tongue	German

Education

Sept. 2013 - today	PhD, ETH Zürich, planned completion: May 2018
Mar. 2017	Certificate of advanced studies in computer science, ETH Zürich (specialization: Information Systems) (average grade 5.43)
Sept. 2010 - Mar. 2013	Master's studies materials science, ETH Zürich (average grade 5.62)
Sept. 2007 - Feb. 2011	Bachelor studies materials science, ETH Zürich (average grade 5.25)
Aug. 2002 - Sept. 2006	Mathematisch Naturwissenschaftliches Gymnasium, Kantonsschule Im Lee, Winterthur

Work Experience

Sept. 2013 – today	Research Assistant, ETH Zürich
	Activities: Development and implementation of algorithms in the quantum chemistry code CP2K, enabling simulations one hundred times larger than the state-of-the-art. Conducting research in the field of nanoelectronics in collaboration with experimental groups.
Aug. 2010 - Jan. 2012	R&D Assistant at Vistaprint (online printing company), parallel to studies 10%
	Activities: Creation of an image recognition program, analysis of a systematic machine error and creation of a model for correction, and executing measurements for quality control.
Jun. 2010 - Aug. 2010	Industrial Internship at Vistaprint Winterthur
	Activities: Feasibility study on Direct Dye Sublimation.
Feb. 2007 - Apr. 2007	Letter sorter for the Post (50%)

Computer Skills

Programming:	Python, Fortran, bash, C++, MATLAB, C#, Java
Applications:	CP2K, svn, git, VMD, Latex, Mathematica

134

Projects

PhD Dissertation	Combining Ehrenfest Molecular Dynamics with Linear Scaling and Subsystem DFT: Implementation and Application in Device Simulation (Supervisors: Prof. Mathieu Luisier; Dr. Joost VandeVondele)
Master Thesis	Complex Dynamics of Composite Materials with Multilayered Inclusions (Supervisor: Prof Dr. Andrei Gusev, Mark 6)
Bachelor Thesis	Molecular Dynamics Simulation of the Elastic Lennard-Jones Model (Supervisor: Prof. Dr. Martin Kröger, Mark: 6)
IYPT	International Young Physicists' Tournament 2005

Languages

German	Native Language
English	Fluent
French	School Knowledge Gymnasium

Hobbies

Martial Arts: Wing Chun (6 years), Sanda (1 year), Muay Thai (5 years), Brazilian Jiu-Jitsu (2 years)

Jogging

References

Available on Request