Electron Interactions & Quantum Transport

Rostgaard, Carsten; Jacobsen, Karsten Wedel

Publication date:
2010

Document Version
Early version, also known as pre-print

Center for Atomic-scale Materials Design
Department of Physics
Technical University of Denmark

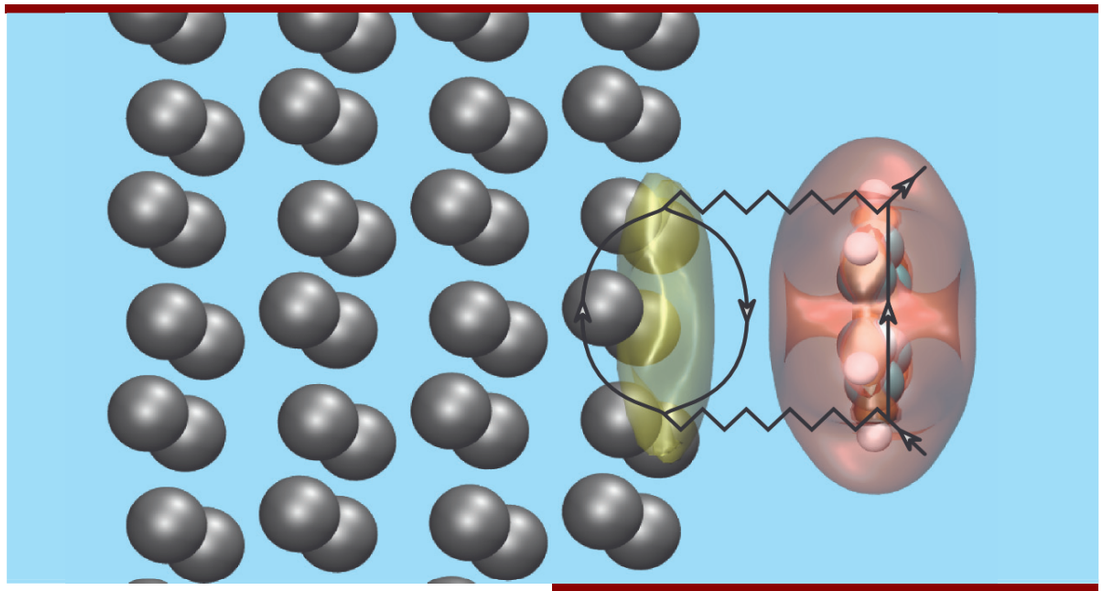
Building 311
DK-2800 Kgs. Lyngby
Denmark

Phone: +45 4525 3224
Fax: +45 4593 2399

www.camd.dtu.dk

Electron Interactions and Quantum Transport

Electron Interactions and Quantum Transport



Carsten Rostgaard
Ph.D. Thesis
November 2009

Carsten Rostgaard

CAMd
Center for Atomic-scale Materials Design

Electron Interactions & Quantum Transport

Ph.D. Dissertation

Carsten Rostgaard
November 2009

Center for Atomic-scale Materials Design
Department of Physics
Technical University of Denmark

Preface

This thesis is submitted in candidacy for the Ph.D. degree from the Technical University of Denmark (DTU). It is based on the work carried out at the Center for Atomic-scale Materials Design (CAMD), Department of Physics at DTU from August 2006 to November 2009 under the supervision of assistent professor Kristian S. Thygesen and professor Karsten W. Jacobsen. Financial support was provided by DTU.

Lyngby, November 16, 2009
Carsten Rostgaard

Abstract

This thesis is concerned with theoretical and numerical modeling of electron transport through molecules coupled to metallic leads. The basic description of the atomic and electronic structure is based on the Kohn-Sham (KS) variant of density functional theory (DFT), which is combined with the non-equilibrium Green function (NEGF) formalism to describe electrical properties of metal-molecule junctions.

The shortcomings of the standard DFT-NEGF approach which originates from the use of ground-state equilibrium DFT and (semi-) local exchange correlation (xc) density functional approximations, are addressed by: i) The use of (hybrid-) exact exchange functional to go beyond the (semi-)local xc functional approximation. ii) The inclusion of many-body non-equilibrium correlation effects in terms of the GW self-energy approximation, and iii) describing the effect of electron-phonon (e-ph) interactions via the first Born e-ph self-energy.

A numerically tractable scheme using an accurate real-space representation of the density and wave-functions has been devised to obtain the all-electron (AE) exact exchange (EXX) energy and the associated non-local EXX or Hartree-Fock (HF) potential operator. The latter allows for self-consistent EXX based calculations within a generalized KS scheme.

The (fractional) inclusion of EXX reduce self-interaction errors of the (semi-) local xc functionals which in turn leads to an enhanced description of the highest occupied molecular orbital (HOMO), and can also be used to improve the calculated atomization energies of molecules. Most DFT/HF codes which support EXX are either based on (incomplete) atomic orbital basis sets, or employ pseudopotential schemes to reduce the numerical effort. In the presented AE scheme it is found that valence-core interactions inaccessible in pseudopotential schemes are important for the absolute position of the HOMO in orbital dependent exchange-correlation functionals, while only having a small impact on atomization energies and the fundamental gap.

The EXX scheme has been generalized to the determination of arbitrary two-electron (four-orbital) matrix elements of the Coulomb operator for both isolated and extended systems. The accurate evaluation of two-electron Coulomb matrix elements, without basis set limitations or pseudopotential approximations, is essential for many-body extensions based on DFT/HF. In this work it is used for linear response time-dependent DFT and GW calculations.

A GW scheme is devised, which contrary to many implementations employs no additional approximations beyond the basis set representation. A highly efficient basis set which consists of generalized Wannier functions augmented by numerical atomic-orbital functions is used. The majority of applications of the GW method in the literature have been for crystalline systems, but with the expanding field of nanoscience, its use on systems of reduced dimensionality has increased, and an

assesment of the validity of the GW approximation for these systems is needed. In this work the applicability of GW for isolated systems is investigated systematically for a large database of small molecules. The predicted position of the highest occupied molecular orbital (HOMO) is found to be consistently improved compared to both DFT and HF. Furthermore, the full GW solution is used to asses the validity of the often invoked linearized quasiparticle equation approximation where off-diagonal elements in the GW self-energy operator are neglected.

A generalization of GW to non-equilibrium conditions is used to calculate the conductance of molecules in metallic junctions. It is demonstrated that GW includes a proper description of dynamical screening effects such as the classical image charge formation. In mean-field descriptions like DFT and HF, such an effect is not included and may be corrected for on an ad-hoc basis. Such a correction is often estimated from classical electrostatic models employing perfect metal interfaces and molecules modeled by point charges. In this work a different approach is demonstrated based on a constrained DFT scheme and total energy differences. This scheme takes into account the full atomistic description of the junction and it is demonstrated how it may be used to efficiently estimate dynamical screening effects observed in GW calculations.

A theoretical characterization of a recently demonstrated break junction experiment in which a Pt wire is stretched in the presence of water vapor [1] has been performed. The most likely atomic configuration is identified, and a lowest order inclusion of e-ph interactions is utilized to provide insight into features of the experimental observed IV characteristics that are believed to be related to inelastic effects.

Resumé

Denne PhD afhandling beskriver teoretisk og numerisk modellering af elektron transport i molekyler koblet til metalliske ledere. Den atomare og elektroniske struktur er beskrevet ved hjælp af tæthedsfunktional teori (DFT), mere præcist er det Kohn-Sham (KS) varianten af denne teori som er anvendt i kombination med ikke-ligevægts Green funktion (NEGF) formalismen til at beskrive de elektriske egenskaber af molekyler i metalliske kontaker.

Begrænsningerne i standard DFT-NEGF metoden, som stammer fra brugen af grundtilstands ligevægt DFT og (semi-) lokal exchange-korrelation (xc) tæthedsfunktional approksimationer, er blevet adresseret ved: i) At bruge (hybrid-) eksakt-exchange funktionaler til at gå ud over den (semi-) lokale xc funktional approksimation, ii) At inkluderere mange-partikel ikke-ligevægt korrelationseffekter ved brug af GW selv-energi approksimation og iii) At beskrive effekten af elektron-fonon (e-ph) vekselvirkninger via Born e-ph selv-energi antagelsen.

En numerisk håndterlig implementering, som bruger en nøjagtig real-rums repræsentation af tæthed og bølgefunktioner er blevet udarbejdet til beregning af alle-elektron (AE) eksakt exchange (EXX) total energier samt den tilhørende ikke-lokale EXX eller Hartree-Fock (HF) potential-operator. Sidstnævnte muliggør selvkonsistente EXX baserede beregninger, indenfor den generaliserede KS-teori.

Den (delvise) inkludering af EXX reducerer selv-vekselvirkningsfejl i de (semi-) lokale xc funktionaler hvilket medfører en bedre beskrivelse af den højeste okkuperede molekylære orbital (HOMO) og kan også anvendes til at forbedre beregning af atomariseringsenergierne i molekyler. De fleste DFT/HF implementeringer som understøtter EXX er enten baseret på (ufuldstændige) atomar-orbital basis sæt, eller anvender pseudopotentialer for at reducere den numeriske belastning. I den heranvendte AE implementering ses det at valens-kerne vekselvirkninger, som uedelades ved brug af pseudopotentialer, er essentielle for den nøjagtige placering af HOMO i orbital-afhængige exchange-korrelations funktionaler, mens de kun har en mindre effekt på atomariseringsenergier og det fundamentale båndgab.

EXX implementeringen er blevet generaliseret til bestemmelsen af arbitrære to-elektron (fire-orbital) matrixelementer af Coulomb operatoren både for isolerede og periodiske systemer. Den nøjagtige evaluering af to-elektron Coulomb matrixelementer, uden basis sæt begrænsninger og pseudopotential approksimationer, er essentiel for mange-partikel udvidelser baseret på DFT/HF. I denne afhandling er dette anvendt til lineær-respons tidsafhængig DFT og GW beregninger.

Et system til GW beregninger er udviklet, som i modsætning til mange implementeringer, ikke involverer ekstra approksimationer udover basissæt repræsentation. Et yderst effektivt basis sæt bestående af generaliserede Wannier funktioner suppleret med numeriske atomar-orbital funktioner er blevet anvendt. Størstedelen af anvendelserne af GW metoden i literaturen er baseret på krystallinske systemer, men

i takt med en øget interesse for nanovidenskab er anvendelsen for lav-dimensionelle systemer tiltaget og der er derfor behov for at evaluere anvendeligheden af GW approksimationen for disse systemer. I denne afhandling er der foretaget en systematisk undersøgelse af anvendelsen af GW for isolerede systemer via en omfattende database for små molekyler. Den beregnede placering af HOMO'en er konsekvent forbedret i forhold til både DFT og HF. Derudover er den komplete GW løsningsmodel brugt til at vurdere gyldigheden af den ofte anvendte lineariserede kvasipartikel lignings approksimation hvor ikke-diagonale elementer i GW selv-energi operatoren er tilsidesat.

En generalisering af GW til ikke-ligevægts betingelser er anvendt til at beregne konduktansen af molekyler i metalliske kontakter. Det vises at GW inkluderer en korrekt beskrivelse af dynamiske screening effekter såsom dannelse af klassisk billedladninger. I middel-felt beskrivelser såsom DFT og HF er en sådan effekt ikke inkluderet, men det er muligt at kompensere for dette ad hoc. Denne kompensation er ofte estimeret ud fra klassike elektrostatiske modeller baseret på perfekte metalliske grænseflader og molekyler modelleret via punkt-ladninger. I dette PhD studium er en alternativ tilgang beskrevet baseret på constrained DFT beregninger og totalenergi forskelle. Dette system tager hensyn til den komplette atomare beskrivelse af kontakten og det vises hvordan systemet kan bruges til effektivt at estimere dynamiske screeningseffekter observeret i GW beregninger.

Der er foretaget en teoretisk karakterisering af det for nylig demonstrerede "break junction" eksperiment hvor en platin tråd bliver strukket under tilstedeværelsen af vanddamp [1]. Den mest sandsynlige atomare konfiguration er blevet identificeret og en inkludering af e-ph vekselvirkninger til laveste orden er anvendt for at give indsigt i egenskaberne af de eksperimentielt observerede IV karakteristika som menes at være relaterede til uelastiske effekter.

List of Included Papers

Paper I
Time-dependent density-functional theory in the projector augmented-wave method
M. Walter, H. Häkkinen, L. Lehtovaara, M. Puska, J. Enkovaara, C. Rostgaard, and J. J. Mortensen
The Journal of Chemical Physics **128**, 244101 (2008).

Paper II
Fully selfconsistent GW calculations for molecules
C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen
Accepted Physical Review B.

Paper III
Polarization-induced renormalization of molecular levels at metallic and semiconducting surfaces
J. M. Garcia-Lastra, C. Rostgaard, A. Rubio, and K. S. Thygesen
Physical Review B **80**, 245427 (2009).

Other Publications

Paper IV
The projector Augmented-wave Method
C. Rostgaard
arXiv:0910.1921 [cond-mat.mtrl-sci] (2009).

Paper V
Density functional theory based screening of ternary alkali-transition metal borohydrides: A computational material design project
J. S. Hummelshøj *et al*
The Journal of Chemical Physics **131**, 014101 (2009).

Paper VI
Electronic structure calculations using GPAW: A real-space implementation of the projector-augmented wave method
J. Enkovaara, C. Rostgaard, J. J. Mortensen, M. Dułak, L. Ferrighi, J. Gavnholt, C. Glinsvad, H. A. Hansen, M. Kuisma, A. Larsen, L. Lehtovaara, M. Ljungberg, O. Lopez, P. G. Moses, J. Ojanen, T. Olsen, N. A. Romero, M. Strange, M. Vanin, M. Walter, H. Häkkinen, R. M. Nieminen, J. Nørskov, M. Puska, T. Rantala, K. Thygesen, K. W. Jacobsen
To be submitted.

Contents

Chapter 1

Introduction

The general theme of the thesis is the development and application of tools for the theoretical modeling of molecular electronics. The concept of "molecular" electronics is generally used to describe all types of nanostructures used as electronic components including actual molecules[2], but also metallic chains of single atoms [3] and larger structures such as nanotubes [4].

The miniaturization of electronic devices is technologically important as it implies faster devices (by reduced response times) and reduced operating currents. The current threshold for operation determines the heating of the device, which is a key limiting factor for many design parameters. In conventional silicon based electronics, the incessant quest for smaller devices has led to an exponential increase in the transistor density on integrated circuits over the last 40 years; a trend commonly referred to as Moore's law [5]. This reflects a corresponding exponential decrease in the feature size of individual components to the point where state-of-the-art commercially available transistors now have features as small as 20 nm (width of the gate pitch in the 32nm-technology field-effect transistors available from Intel in 2011[6]). This corresponds to about 100 atoms and is already approaching device sizes where the quantum mechanical wave nature of the charge carries can no longer be ignored and will have a strong influence on device operation. At the ultimate small device scale, classical continuum models like Ohm's law does no longer apply, and quantum mechanics must be applied to describe the motion of electrons correctly. This makes nano-scaled components conceptually different from larger devices and requires a firm theoretical understanding of the electron transport process. In the down-scaling process of conventional silicon technology such an understanding may be used to try and suppress undesirable effects (from a classical convential viewpoint) introduced by the quantum nature of electrons.

Molecular electronics represents a markedly different approach, namely a bottom up approach, where the quantum nature of charge carriers is used and exploited directly in the design. One way of realizing such an approach is by using organic molecules as the basic building blocks of integrated circuits [7, 8]. Use of molecules that are abundant in nature may also have an economical advantage over expensively synthesized high-purity crystals of silicon based technology. In addition molecules can be self-organizing when contacted by leads, which would be an important step in order to overcome the technological limitations of using lithographic procedures, as used in the conventional technology, at the nanoscale.

Using single organic molecules as transistors was suggested already in 1974 by [9],

but it was not until experimental advances in the mid-1990s [10, 11] made it possible to actually perform electrical characterization of truly nano-scale structuresthat the field of molecular electronics really gained interest. At present, tunneling microscope (STM) imaging technique can be used to scan the IV characteristic of single molecules [12, 13]. STM can also be used to actively create a nanostructure by crashing an STM tip into a surface and then by subsequently slowly retracting the tip it may happen that freely suspended homogeneous atomic chains of up to ten atoms are dragged out from the surface [3, 14, 15]. A related method is the mechanically controlled break-junction (MCBJ) technique [16, 17], in which a notched wire is gently stretched to the breaking point. This can also lead to formation of atomic wires, and further by performing the experiment in a diluted gas, it is possible to capture single molecules in the junction before rapture [18, 19].

Nanostructured systems display a multitude of highly non-classical phenomena such as conductance quantization[20, 21], interference effects[15, 22, 23], strong electron-phonon coupling effects,and also electron-electron correlation effects such as Kondo and Coulomb blockade physics [24–26]. This makes the study of such systems interesting from a fundamental point of view and combines the fields of solid-state physics with the chemistry of molecules, and a direct experimental probe for studying the quantum mechanical wave nature of electrons. A detalied understanding of these quantum effects is also essential for the future technological development of molecular electronics as well as a potential further down scaling of conventional silicon technology.

The non-trivial electric characteristics due to intricate quantum effects, makes nano-scaled devices highly sensitive to the detailed atomic structure. Technologically, this presents a barrier for making reproducible devices. However, if this barrier can be overcome their is great prospect of exploiting the rich features and flexibility available to represent more functionality in a single molecular device than is done with conventional silicon technology. This could be envisioned by operating the nano-device in different bias intervals or gate potentials intervals.

In practice, fabrication of individual molecular devices in a reliable atomic configuration is not presently possible. Experiments performed with the same setup often show considerable variation upon repeated junction formation, and usually a large number of repeated measurements is used to extract a statistical description of the considered junction. In this way the characterization is indirect and the unknown atomic geometry implies that measurements must be complemented by theoretical models and calculations for a full characterization and enhanced understanding.

The theoretical description of electron transport through an open boundary quantum system of mutually interacting electrons and nuclei under non-equilibrium conditions imposed by the applied bias represents a mayor challenge, and always invoke a series of approximations. The validity of such approximations may in turn be tested by comparison to experiment, thus imposing a cycle of mutual comparison between measurements and simulations.

Due to the large difference in mass between electrons and nuclei, it is often a good approximation to solve the separated electronic degrees of freedom subject to the static potential of a fixed nuclei lattice, and in turn regard the nuclei as classical charges moving on the potential energy surface generated by the electron cloud. The isolated electronic structure problem is complicated by the electron-electron (e-e) interactions correlating the electronic degrees of freedom. This may be addressed by replacing the e-e interactions by an effective mean-field potential as is done in Hartree-Fock (HF) and Kohn-Sham (KS) theory, in which case the electronic

structure is determined by solving a set of single particle equations (one for each electron) each subject to the same local KS or non-local HF single particle potential operator.

The independent-particle approximation represents a major simplification. However, the KS-DFT scheme offers a construction from which the exact total energy in principle can be obtained. And due to the existence of efficient and reasonable accurate approximations for energies within KS-DFT a good description of many properties derived from total energies is provided. This includes equilibrium geometry of the junction, mechanical properties, and vibrational frequencies.

Formally the spectral properties of the KS system can not in general be interpreted as physical excitation energies, but it does present a practical simplification and indeed the "standard" approach to first-principles calculations of the conductance of nano-scale junctions is nowadays to combine KS-DFT with non-equilibrium Green function (NEGF) techniques [27–29]. For independent-particle approximations, the general expression for the conductance derived by Meir and Wingreen [30] reduce to the simple Landauer formula [31]where the conductance is conveniently described in terms of a transmission probability of electrons. The low bias and low temperature conductance is then simply the transmission function evaluated at the Fermi level times the unit of quantum conductance $G_0 = 2e^2/h$ while the (low temperature) finite bias current is given by the integral of the transmission probability function over energies corresponding to the bias window.

The electron transfer process of electrons through the junction via an intermediate charged excitation of the molecule depends acutely on the molecular excitation spectrum provided by the underlying theoretical model.

Although not rigorously justified [32–34] the KS-DFT transport scheme have been successfully applied to systems characterized by strong coupling between the molecule and the electrodes, such as homogeneous metal point contacts and mono-atomic wires [23, 35, 36]. This agreement is presumably because the high transmission probability allows a phase-coherent electron transfer consistent with the mean-field potential, and the large hybridization leads to broad resonances at the Fermi level facilitating the current. The bandstructure of such "metallic" systems, or resonances close to the Fermi level are in practice generally well described by KS-DFT. This is in contrast to the rather poor description offered by KS-DFT in the case of quasiparticle excitations of systems with a gap such as semi-conductors or isolated (weakly coupled) molecules.

Indeed, the NEGF-DFT method systematically overestimates the conductance of more weakly coupled system such as organic molecules passivated by anchoring side groups (e.g. thiol or amine) [37–40]. For these systems, KS-DFT yields an incorrect size for the gap which affects strongly affects the transport properties. This class of systems span a potentially technologically important type of junctions. This can be realized by observing that small reactive molecules are prone to large structural and electronic reconfiguration which sometimes involves forming entirely new structures by dynamical molecular decomposition which in turn makes the design process rather complex. Also very weakly coupled molecules which preserves their molecular structure can not be expected to form stable junctions. However, the passivated organic molecules largely preserve both structure and properties of the isolated molecule while at the same time forming reasonable chemical bonds with the metal to allow for stable junctions even under finite bias. The positioning of the passivated molecules in the junction is largely determined by the position of the side groups which also dictates the bonding nature, thus producing more

reliable junctions which may be less sensitive to details in junction geometry[41]. It is therefore unfortunate for the field of molecular electronics in general that these systems are poorly described theoretically by the standard method based on KS-DFT.

The erroneous positioning of molecular energy levels in KS-DFT is related to self-interaction (SI) errors in the approximate exchange-correlation functionals (or rather the lack of correction for the SI in the Hartree term). This can be corrected for by e.g. introducing a degree of exact exchange, but the optimal amount depends on the (local) screening and is therefore system dependent. Introducing SI corrections for the occupied KS-DFT levels and thus the band gap, have been shown to improve agreement with experiments [42]. However, with a parameter controlling the gap and thereby the tunneling current, this is no longer first-principles and compromise the predictive power of such an approach.

Besides the problems related to the position of molecular energy levels, dynamic correlation effects outside the scope of a any (static) single-particle description can be a very important physical effect. An example is the renormalization of molecular levels by the dynamical screening [43, 44], which is a consequence of the correlations induced by e-e interaction and cannot – even in principle, be described in mean-field theory like DFT or HF.

In the work presented here, the limitations of some of the above mentioned approximations in the standard DFT-NEGF approach are addressed. Several techniques for improving the description are implemented and applied to relevant systems. This involves first of all starting from as accurate a DFT description as possible which implies code development such as DFT exchange correlation functionals beyond the (semi-) local approximation. Secondly a GW implementation devoid of the commonly employed technical simplifications and therefore unbiased towards continuum or isolated systems, has been devised. The GW method have been applied to a broad range of molecules as well as junction configurations in both the strongly coupled contact and weakly coupled tunneling regime.

The above considerations are all related to purely electronic structure issues. At a finite bias, an important effect is heating of the junction due to the interaction between lattice vibrations (phonons) and electrons. The effect of inelastic scattering of electrons on the transmission probability is studied in relation to an experimental work on the conductance properties of single water molecules.

Outline of the thesis

Chapter 2 gives an introduction to the theory behind the applied methods, and practical issues related to the numerical implementations. Density-functional theory (DFT) is first introduced followed by a description of the non-equilibrium Green function (NEGF) formalism, and how it can be combined with DFT for the use of ab-initio electron transport calculations

Chapter 3 focus on pure DFT issues. It is demonstrated how quantities can be extracted from a projector augmented-wave (PAW) implementation of DFT, and how to modify the self-consistency scheme. The PAW-DFT development is related to analysis of systems, inclusion of exact-exchange, linear response time-dependent DFT, and evaluation of correlation (image charge) effects on the ionization energies of molecules in proximity to metallic surfaces.

Chapter 4 extends the mean-field description of DFT by addition of the GW correlation self-energy in the NEGF framework. Details of the numerical scheme devised for the GW calculations are first discussed, followed by an application to small molecules. Having confirmed that GW offers a good description in both limits of low electronic screening (molecules) and large screening (metallic systems), the non-equilibrium variant is applied to the hybrid setup of molecules sandwiched between metallic leads, and the conductance is determined.

Chapter 5 presents a study of water in Pt break-junctions. Energetics and vibrational modes as well as the inelastic IV signals are determined and compared to experiment.

Chapter 2

Theory and Methodology

This chapter describes the basic theoretical framework used within the thesis as well as the applied methods.

The text is not intended as a detailed account of all the topics, as most is expected to be well known to the reader. It will however provide a review of the relevant topics, affording reference to more in-depth literature and original contributions when appropriate. The aim is to present the basic equations referred to in the remaining text, and highlight the approximations (both theoretical and technical) that are applied as well as discuss their interpretation and justification.

The chapter is organized as follows: In section 2.1, the general concept of electronic structure calculations is introduced. This is followed by a description of the popular density-functional theory (DFT) description of the electronic structure in sec. 2.2. This description also form the basis of all calculations presented in this work, and some of the more technical aspects of numerical DFT implementations are discussed in sec. 2.3. Section 2.4 describes the Green function approach to the electronic structure problem, and in particular, how this can be used to incorporate non-equilibrium conditions and many-body interactions in the description via so-called self-energy operators. The explicit form of the relevant self-energies are discussed in section 2.5.

Hartree atomic units ($m_e = e^2 = \hbar = 1$) will be used throughout the thesis unless explicitly stated otherwise, e.g. in plots.

2.1 Electronic Structure Calculations

At an atomic scale, all properties of matter are in principle determined by the quantum mechanical wave equation $\widehat{H}\Psi = i\hbar\partial_t\Psi$ known as the Schrödinger equation (in this non-relativistic form). The explicit form of the differential operator $\widehat{H}$, denoted the Hamiltonian is well known, it is linear and of second order. The solution of the equation, many-body wave function Ψ depends in general on both a time variable, a space variable for each nucleus, and a spin and space variable for each electron in the system. The squared modulus of the wave function can be interpreted as a probability distribution for the position of electrons and nuclei, and the expectation value of any physical observable can be determined directly in terms of this quantum state[45].

Based on this, one might claim that with the discovery of the Schrödinger

equation, all problems of chemistry and a large part of physics are formally solved (freely quoted from [46]). The problem is that the complexity of the equations is such that they can not be solved in practise for any realistic system.

In practical simulations of matter, a series of approximations are invoked. The standard approximations are described below (for notational brevity, the spin coordinate for electrons has been suppressed).

Relativistic effects and core electrons For the type of systems considered here, relativistic effects are only expected to be important for the core electrons of heavy elements. These electrons are however also chemically inert and are, for all applications in this thesis, treated in the frozen core approximation. The reference orbitals of the core electrons are obtained from a scalar relativistic description of the isolated atoms.

Time dependence All calculations are based on a determination of the time independent ground state. For the time-dependent applications: the linear response calculations in sec. 3.2.6 and the non-equilibrium electron transport calculations in chapter 4, the stationary ground state act as the initial reference state. For a time independent Hamiltonian, the product solution $\Psi(t) = e^{-i\mathcal{E}_i t}\Psi_i$ solves the Schrödinger equation, where Ψ_i are the solutions of the time *independent* Schrödinger equation

$$\widehat{H}\Psi_i = \mathcal{E}_i\Psi_i \tag{2.1}$$

corresponding to the i'th excited stationary state of total energy $\mathcal{E}_i$.

Ion dynamics Due to the large mass difference between electrons and nuclei ($M_n/m_e \approx 10^3$ – 10^6), it is usually well-justified to decouple the comparably slow dynamics of the nuclei from the faster electron dynamics. This is the basis for the adiabatic approximation of [47], in which the electrons are assumed to react instantaneously to the movements of the nuclei.

Under this assumption, the electron and nuclei dynamics can be decoupled, and the full Hamiltonian can be solved by a product solution of an ionic and an electronic wave function.

The slow nuclei dynamics are governed by the free particle like Hamiltonian $\left(\sum_a \frac{-1}{2M_a}\nabla^2_{\boldsymbol{R}_a} + E(\boldsymbol{R})\right)$, moving on the effective potential energy surface $E(\boldsymbol{R})$ (denoted the Born-Oppenheimer surface, BOS) generated by the electrons, which depends parametrically on the nuclei coordinates themselves.

For molecular dynamics, and for determining the equilibrium geometry, we treat the ions as semi-classical particles moving on the (quantum mechanically determined) BOS, with each atom experiencing the force

$$\boldsymbol{F}_a = -\frac{\partial E(\boldsymbol{R})}{\partial \boldsymbol{R}_a}. \tag{2.2}$$

Although the Born-Oppenheimer approximation is often well justified, lattice-electron interactions can be important. In chapter 5, the effect of electron-phonon interactions on electron transport is examined as a perturbation to the BOS. For a review of lattice dynamics, see [48].

Electronic structure and electron interactions For any given lattice configuration, $\boldsymbol{R}_0$, the system is completely described by the instantaneous electronic structure.

The electronic degrees of freedom are described by the solution of $\widehat{H}^e(\boldsymbol{R}_0)\Psi^e = E_n(\boldsymbol{R}_0)\Psi^e$, where, in the absence of any externally applied fields, the electronic Hamiltonian

$$\widehat{H}^e = \sum_n \frac{-1}{2}\nabla^2_{\boldsymbol{r}_i} + \frac{1}{2}\sum_{i\neq j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|} - \sum_{ia}\frac{Z_a}{|\boldsymbol{r}_i - \boldsymbol{R}_a|} + \frac{1}{2}\sum_{a\neq a'}\frac{Z_a Z_{a'}}{|\boldsymbol{R}_a - \boldsymbol{R}_{a'}|} \quad (2.3)$$

is a sum of the kinetic energy, electron-electron interaction, electron-ion interaction, and the electrostatic lattice energy.

In this expression, all but the second term are single particle operators involving at most one electron coordinate. The two-body electron-electron (e-e) interaction operator $\widehat{V}_{ee} = \frac{1}{2}\sum_{i\neq j}\frac{1}{|\boldsymbol{r}_i - \boldsymbol{r}_j|}$ correlates the different electron coordinates, and presents a major complication for determining the electronic structure.

For all wave function based applications discussed here, the e-e interaction will be accounted for by mapping it onto an *effective* single particle potential $\widehat{V}_{\text{eff}} = \sum_i \hat{v}_i$, which can be achieved via e.g. the Hartree-Fock (HF) or density-functional theory (DFT) schemes. Both of these are discussed in sec. 2.2.

In the Green function based approach, e-e interactions can be included as a perturbation on the non-interacting problem. The theoretical background is discussed in section 2.4, and the effect of e-e inclusion is studied in sections 4.3 and 4.4 for molecular systems and electron transport respectively.

To simplify notation, the superscript e on the electronic Hamiltonian and wave function is dropped in the remainder of this text. As only the electronic degrees of freedom are treated quantum mechanically this should not cause any confusion.

The terms of the electronic Hamiltonian (2.3) are divided into three terms

$$\widehat{H} = \widehat{T} + \widehat{V}_{\text{ext}} + \widehat{V}_{ee}, \quad (2.4)$$

with $\widehat{T}$ denoting the kinetic energy operator, $\widehat{V}_{\text{ext}}$ the single particle part of the potential, i.e. the electron-ion potential, the ion-ion energy, and any possible externally applied static potentials, and $\widehat{V}_{ee}$ describing the (e-e) interactions.

Of the approximations described above, the mean-field description of the e-e interaction is the most crucial, and the one given the most attention in the applications.

The procedure for constructing such a mapping is discussed in the next section.

2.2 Density-functional Theory

As the wave function for an N-electron system is a function of $3N$ independent spatial variables, solving the electronic Hamiltonian is not possible in practice. As acknowledged by Walter Kohn in his nobel lecture [49], even representing the solution numerically on a computer is impossible for any interesting number of electrons. For example, representing a 5-electron wave function in single precision

(4 bytes per number) on a mesh of 10 grid points for each direction would require $4\text{bytes} \times 10^{3\times 5} \approx 4 \times 10^6\text{GB}$ of memory, and that is just storing the solution.

The problematic term is the electron interaction operator $\widehat{V}_{ee}$, which correlates the different electron coordinates. For a non-interacting Hamiltonian (a sum of single particle operators only, e.g. $T + V_{\text{ext}}$) the many-body wave function can be represented as a single Slater determinant[50, 51]. A Slater determinant is an explicit construction of an anti-symmetric N particle wave function from N single particle waves. All quantities can thereby be re-expressed in terms of the single particle waves. In the example used above, the required memory for representing the 5-electron function would thus be reduced to $4 \times 5 \times 10^3 = 19\text{kB}$.

The approximation used for the electronic structure is then to invent a fictitious non-interacting system, where the effective external potential has been modified such that the properties of the model system emulates that of the interacting system.

The most popular way of constructing such an effective mean-field potential is by use of density-functional theory (DFT), which will be reviewed in the next section. DFT offers a scheme for constructing the effective potential, such that the non-interacting system produce the same ground state density as the interacting system. Although the electronic energy levels of this system do not reproduce those of the real system, the theory simultaneously offers a way to determine the real ground state energy from the ground state energy of the model system. This gives an accurate description of the BOS, and hence quantities derived from this, such as equilibrium geometries, reaction barriers etc.

Working with non-interacting systems is a tremendous simplification in all wave function based theories, and several schemes exist for describing the interacting electrons via some mean-field effective potential. The most popular being that of Density-functional Theory (DFT).

Although the construction is only designed to result in non-interacting wave functions that give an accurate description of the true ground state density, it does provide the most efficient and accurate way to construct a sensible mean-field potential, and it is often used, with varying success, to estimate response properties and excitation energies of real systems.

The following will be a compressed review of the theory. For a more detailed description of DFT, see for example the excellent book [52].

2.2.1 Hohenberg-Kohn Theory

The basis of DFT is the two Hohenberg-Kohn (HK) theorems (originally formulated by [53]), here reproduced in the more general formulation by Levy and Lieb [54–56] which solves the ν-representability[57] of the original version.

The first HK theorem is obtained by rewriting the (Rayleigh-Ritz) variational principle for the ground state energy $E_0 = \min_{\Psi\to N}\langle\Psi|\widehat{H}|\Psi\rangle$ [58] into two separate minimizations, one over all densities, and a second over all wave functions producing a given density

$$E_0 = \min_{n(\boldsymbol{r})\to N}\left[\min_{\Psi\to n}\langle\Psi|\widehat{T} + \widehat{V}_{ee}|\Psi\rangle + \int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) v_{\text{ext}}(\boldsymbol{r})\right]. \tag{2.5}$$

Expressing the first term explicitly as a density functional

$$F[n] \equiv \min_{\Psi\to n}\langle\Psi|\widehat{T} + \widehat{V}_{ee}|\Psi\rangle, \tag{2.6}$$

shows that the ground state energy can be considered as the minimum of a functional $E[n] = F[n] + \int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) v_{\mathrm{ext}}(\boldsymbol{r})$ which depends on the density only, and that the minimizing density $n_0(\boldsymbol{r})$ is the ground state density of the system. This is the first HK theorem.

The constrained minimization in equation (2.5) is achieved by solving the associated Euler-Lagrange equation [59]

$$\left.\frac{\delta F[n]}{\delta n(\boldsymbol{r})}\right|_{n=n_0} + v_{\mathrm{ext}}(\boldsymbol{r}) - \mu = 0 \tag{2.7}$$

where the conservation of electrons is enforced by the introduced Lagrange multiplier μ, which is adjusted such that $\int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) = N$, and is recognized as the chemical potential of the system.

Since the HK functional $F[n]$ is universal (i.e. system independent), equation (2.7) provides a one-to-one correspondence between the external potential and the ground state density. This constitutes the second HK theorem. Since all observables are in principle functionals of the external potential (via the Schrödinger equation), they can also be considered as functionals of the ground state density.

2.2.2 Generalized Kohn-Sham Scheme

In principle, the term density-functional theory refers to all theories based on approximating the HK density functional $F[n]$.

Explicitly density-dependent approximations and direct determination of the ground state density via (2.7) are denoted Thomas-Fermi type models as this leads to expressions similar to those already examined independently by [60, 61] and later extended by [62] some 30 years before the HK theorem. These kinds of models have however never led to a description accurate enough to have predictive power[63, 64]. The subject is however still an active research area due to the extreme simplicity of working only with a single function in $\mathbb{R}^3$.

A different route was taken by [65]. Recognizing that the energy functional only requires an accurate density, and is independent of how this density is constructed, they proposed to utilize an auxiliary model system governed by a simpler Hamiltonian, but with an effective external potential designed such that the system would produce the correct ground state density. The presentation below is based on the generalization of the original procedure by [66, 67].

Equation (2.7) provides a direct recipe for designing an effective potential of a model system such that this will produce the correct ground state density.

For a model system with an effective potential $v_{\mathrm{eff}}(\boldsymbol{r})$ to produce the same density as the real system, the Euler-Lagrange equations (2.7) of the two systems must be identical. This implies that

$$v_{\mathrm{eff}} = v_{\mathrm{ext}} + \partial \Delta F[n] / \partial n(\boldsymbol{r}). \tag{2.8}$$

where $\Delta F[n] = F[n] - F_s[n]$ is the difference between the HK functional of the real system F and that of the model system F_s. This functional must be approximated. For a given approximated functional, the ground state density can be determined by minimizing the energy of the model system $E_s[n] = F_s[n] + \int \mathrm{d}\boldsymbol{r} v_{\mathrm{eff}}(\boldsymbol{r}) n(\boldsymbol{r})$. Which

can then be used to determine the real ground state energy via

$$\begin{aligned} E_0 &= F_s[n_0] + \Delta F[n_0] + \int \mathrm{d}\boldsymbol{r} n_0(\boldsymbol{r}) v_{\mathrm{ext}}(\boldsymbol{r}) \\ &= E_s[n_0] + \Delta F[n_0] - \int \mathrm{d}\boldsymbol{r} n_0(\boldsymbol{r}) \left.\frac{\partial \Delta F[n]}{\partial n(\boldsymbol{r})}\right|_{n=n_0}. \end{aligned} \tag{2.9}$$

The advantage of this indirect approach to the HK functional, is that the main parts of $F[n]$ are determined by the auxiliary system (thereby implicitly making F orbital dependent and thus much more flexible), while an explicit density functional approximation (which is easier to evaluate but much harder to make good approximations for) is only needed for the remaining part $\Delta F[n]$ which usually only accounts for a small part of the energy.

One such model system is to choose the full Hamiltonian, but require that the model system can be described by a Slater type wave function, Ψ_s. The HK functional for such a system can be determined explicitly by

$$F_s[n] = \min_{\Psi_s \to n} \langle \Psi_s | \widehat{T} + \widehat{V}_{ee} | \Psi_s \rangle = T_s + U_H + E_x. \tag{2.10}$$

where each term is expressible by the set of the single particle waves $\{\psi\}$ constituting Ψ_s

$$T_s = \sum_n \langle \psi_i | \tfrac{1}{2} | \psi_i \rangle, \tag{2.11}$$

$$U_H = \frac{1}{2} \iint \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{r}' \frac{n(\boldsymbol{r}) n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}, \tag{2.12}$$

$$E_x = -\frac{1}{2} \sum_{nn'} \iint \mathrm{d}\boldsymbol{r} \mathrm{d}\boldsymbol{r}' \frac{\phi_n^*(\boldsymbol{r}) \phi_{n'}(\boldsymbol{r}) \phi_{n'}^*(\boldsymbol{r}') \phi_n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{2.13}$$

U_H is the classical electrostatic energy of the charge distribution $n(\boldsymbol{r})$, which includes a spurious interaction of each electron with itself. The exchange energy E_x is directly related to the anti symmetric form of the Slater determinant, and hence the Pauli exclusion[68]. E_x also removes the orbital self-interaction of the Hartree energy.

The part of the HK functional not captured by this model is denoted the correlation energy

$$E_c[n] \equiv F[n] - F_s[n] = \min_{\Psi_0 \to n} \langle \Psi | \widehat{T} + \widehat{V}_{ee} | \Psi \rangle - \min_{\Psi_s \to n} \langle \Psi_s | \widehat{T} + \widehat{V}_{ee} | \Psi_s \rangle \tag{2.14}$$

and its functional derivative, the correlation potential $v_c(\boldsymbol{r}) = \partial E_c / \partial n(\boldsymbol{r})$.

To determine the ground state density, the energy of the model system $E_s[n] = \langle \Psi_s | \widehat{T} + \widehat{V}_{ee} + \widehat{V}_{\mathrm{eff}} | \Psi_s \rangle$ should be minimized with respect to the N single particle orbitals $\{\psi_n\}$ subject to the constraint $\langle \psi_n | \psi_m \rangle = \delta_{nm}$, which leads to the single particle equations

$$\left[-\tfrac{1}{2} \nabla^2 + v_H(\boldsymbol{r}) + \hat{v}_x(\boldsymbol{r}) + v_c(\boldsymbol{r}) + v_{\mathrm{ext}}(\boldsymbol{r}) \right] \psi_n(\boldsymbol{r}) = \epsilon_n \psi_n(\boldsymbol{r}), \tag{2.15}$$

where the orbitals have been rotated to diagonalize the matrix of Lagrange multipliers ϵ_{nm}. The Hartree potential $v_H(\boldsymbol{r}) = \int \mathrm{d}\boldsymbol{r}' n(\boldsymbol{r}) / |\boldsymbol{r} - \boldsymbol{r}'|$, is the density derivative of the Hartree energy U_H, and the Fock operator $\hat{v}_x$ comes from the orbital derivative of the exchange energy, and is given by

$$\hat{v}_x \psi_n(\boldsymbol{r}) = \frac{\partial E_x}{\partial \psi_n^*(\boldsymbol{r})} = -\sum_m \int \mathrm{d}\boldsymbol{r}' \frac{\phi_m^*(\boldsymbol{r}') \phi_n(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|} \phi_m(\boldsymbol{r}). \tag{2.16}$$

Note that (2.15) has been presented in the specific choice of wave functions that diagonalize the matrix of Lagrange multiplier. This set is called the canonical orbitals and are convenient, as each orbital is then associated with a single eigen energy ϵ_n and the total energy is just the sum of these. One could also choose a different linear combination, like e.g. those that minimize the spread functional (maximally localized Wannier functions) which are convenient for some purposes. Notice that the Hamiltonian in (2.15) is invariant under rotations of the (occupied) wave functions. Some model systems do not satisfy this, e.g. Perdew-Zunger type self-interaction corrected functionals [69], in which case the appropriate wave functions are those that minimize the expectation value of the Hamiltonian.

The equations (2.15) can be considered as a generalization of the Hartree-Fock scheme which is recovered by the approximation $E_c = 0$. Note that the equations must be solved self-consistently, as the Hartree, exchange and correlation potentials are all functionals of the resulting wave functions via the density.

In principle, if the approximation of the correlation functional is correct, the self-consistent solution of the single particle equations for the model system will produce the exact ground state density. One should however remember that this is the only real property the model is designed to reproduce. Even the total energy of the model system is wrong, although the true ground state energy can be recovered from (2.9), which for the generalized HF system, reduces to

$$E_0 = \sum_n \epsilon_n + E_c - \int \mathrm{d}\boldsymbol{r} n_0(\boldsymbol{r}) v_c(\boldsymbol{r}). \tag{2.17}$$

In the case of pure HF, in which the correlation part is completely neglected, the total energy reduce to the energy of the Fock model system, i.e. the total energy is simply the sum of single particle eigenvalues.

The determination of the exact exchange energy and Fock operator is computationally very demanding. In the original KS scheme, the model system was therefore chosen as completely non-interacting electrons, such that $F_s = \min_{\Psi_s \to n} \langle \Psi_s | \widehat{T} | \Psi_s \rangle$, leaving $\Delta F[n] = E_{Hxc}$ to be approximated by some density functional. The reconstructed ground state energy, and the single particle equations for this model system are

$$E_0 = \sum_n \epsilon_n + E_{Hxc} - \int \mathrm{d}\boldsymbol{r} n_0(\boldsymbol{r}) v_{Hxc}(\boldsymbol{r}), \tag{2.18}$$

$$\Big[-\tfrac{1}{2}\nabla^2 + \hat{v}_{Hxc}(\boldsymbol{r}) + v_{\mathrm{ext}}(\boldsymbol{r}) \Big] \psi_n(\boldsymbol{r}) = \epsilon_n \psi_n(\boldsymbol{r}). \tag{2.19}$$

Note that as the Hartree energy is already an explicit density functional, the only change compared to the previous model is that the exchange is now treated approximately. This has the implication that the exchange potential in now a multiplicative (approximate) potential $v_x(\boldsymbol{r}) = \partial E_x[n]/\partial n(\boldsymbol{r})$, which makes the problem much easier.

2.2.3 Approximating Exchange and Correlation

The exchange-correlation functional $E_{xc}[n(\boldsymbol{r})]$, is a complicated quantity, which has to account for all e-e interaction effects for the KS-DFT scheme to be exact. Consequentially, a lot of effort has been, and still is, put into the construction of accurate functionals, which has led to a wide range of different xc-functional of

varying accuracy, complexity, and type of systems for which they are well-suited. Here only a brief description will be given for the most common types.

The simplest type of functional, suggested already by Kohn and Sham in their original work [65] is the local density approximation (LDA), in which the energy density is simply related to the local density in each point of space

$$E_{xc}^{LDA}[n] = \int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) \epsilon_{xc}^{\mathrm{hom}}(n(\boldsymbol{r})). \tag{2.20}$$

Here $\epsilon_{xc}^{\mathrm{hom}}$ is the xc energy density of a homogeneous electron gas (HEG) with the constant density $n^{\mathrm{hom}} = n(\boldsymbol{r})$. The exact expression for the exchange part is well-known from Thomas-Fermi-Dirac theory[62], while the correlattion part is fitted to different quantum Monte-Carlo simulations of the spin-paired HEG [69, 70] based on various parametrizations[70–72], to arbitrary densities and spin states. Resulting in a multitude of LDA type functionals.

These kinds of functionals work surprisingly well, also for systems with large density gradients. This can be attributed to the LDA form satisfying many of the known constraints of the exact xc functional [73, 74] which should be conserved when constructing approximate functionals.

A natural improvement on the LDA is obtained by adding a dependence on the local density-gradient

$$E_{xc}^{\mathrm{GGA}}[n] = \int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) f(n(\boldsymbol{r}), \nabla n(\boldsymbol{r})), \tag{2.21}$$

which makes the xc-functional a semi local functional of the density. The minimal set of constraints this type of functional should satisfy to give sensible results, is to reduce to LDA for a HEG, and satisfy the sum rule. Functionals satisfying this are termed generalized gradient approximations (GGA). A large variety of these exist, but the PBE[75] functional is generally considered reliable and is consistently used throughout the work presented here as the only (semi-) local xc approximation.

In general, purely density dependent functionals are convenient, as they are easily evaluated, and the corresponding xc potential can be expressed in a closed analytic form. More generally, a large number of extra ingredients, besides the local density and gradient, can be used, and the added flexibility used to satisfy additional constraints, leading to a ladder of increasingly complex functionals [76] which depend on e.g. the kinetic energy density, KS eigenvalues, or KS orbitals.

One example of an orbital dependent functional is the exact exchange (2.13). Although the exact expression is known, the orbital dependent expression leads to computationally harder equations to evaluate / solve than an explicit density functional. In addition, a combined local approximation for both exchange and correlation is often better than the exact exchange potential combined with a local correlation approximation, due to a large degree of cancellation between the two terms. Especially for extended systems, the self-interaction problem cured by exchange is less significant, while screening of the long range electrostatics of the exchange (which should be contained in the exact correlation) *is* important.

Exact exchange does improve some properties of the non-interacting system (compared to the real system), and knowledge of the exact expression is valuable when constructing approximate density functionals. For improved comparison between the multiplicative potential induced by a density approximation of the exchange energy, and the potential induced by the exact orbital dependent expression, there exists

several ways of converting the Fock operator into a local form[77, 78], most notable of which is the optimized effective potential (OEP) method, which can be used to construct multiplicative xc-potentials from general orbital dependent functionals [79]. A particulaly simple iterative construction has recently been proposed by [80, 81].

The only other xc functional than PBE used here, is the hybrid variant PBE0 [82], which includes 25 % exact exchange in the PBE functional (reducing the local exchange part by a corresponding amount), which can be argued to be a good aproximation for certain systems, and has on an empirical basis been shown to provide an accurate description of both energies and bandstructures for a broad range of systems.

2.3 DFT Implementations

The fundamental approximation in DFT is the choice of exchange-correlation functional, but for all practical implementations, a number of additional numerical considerations are of importance. In this section, some of the more important approximations will be discussed, such as the treatment of infinite systems, the tratment of core electrons, and the representation of wave functions. For a more detailed description of these aspects as well as many other numerical details, the papers [83–85] can be recommended.

All code development and implementations presented in this thesis, have been done in the real-space DFT code GPAW[86, 87] and the simulations interface ASE[88, 89]. Most results are also based on these, although some have been performed using the plane-wave code DACAPOand the atomic-orbital based code SIESTA[90].

2.3.1 Periodic Systems

For isolated systems, the domain in which the KS equations should be solved can be restricted to a finite region determined by the decay of the effective potential (infinite in principle, but in practice truncted to a few Å of vacuum), and the appropriate boundary conditions are simple Dirichlet boundary conditions.

For infinite systems the domain is in principle unrestricted, but if the system can be described by a periodic lattice, Bloch's theorem[91] states that the wave functions can be described by a product of a function with the same periodicity as the lattice $u_{n\boldsymbol{k}}(\boldsymbol{r})$, and a phase factor

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}), \tag{2.22}$$

where $\mathbf{k}$ is a wave vector belonging to the first Brillouin zone (BZ) of the lattice.

The advantage of this representation is that the Hamiltonian is diagonal with respect to the reciprocal lattice vector such that the eigenstates can be determined separately for each $\mathbf{k}$-point, and secondly that although the wave function itself is not periodic, the boundary conditions imposed by (2.22) imply that the wave function needs only be solved within the supercell defining the lattice.

In practice expectation values related to the super cell are determined by averaging over a discrete sampling of the first BZ. Efficient schemes for chosing the set of $\boldsymbol{k}$-points have been devised [92, 93], and accurate expectation values can be determined using relatively coarse $\boldsymbol{k}$-points grids. Note however that the required

density of $\boldsymbol{k}$-points for a given accuracy is highly system dependent and can severely affect the quality of the averaging if not chosen appropriately.

2.3.2 Treating Core Electrons

The chemically inert core electrons are usually unaffected by the assembly of molecules and solids, and can be frozen in the corresponding states of the isolated atoms. The KS equations are thus only solved explicitly for the valence states, but as these must still be orthogonal to the core states, and thereby feature rapid variations in the core region, making them difficult to represent in an accurate way numerically.

Several methods exist for handling this issue. GPAW is based on the projector augmented-wave (PAW) method[94, 95], in which an auxiliary atomic basis set is employed to project out the core structure. This basis is non-zero only in the core region and well-suited to describe the nodal structure of the valence state there, while a different basis can be used to represent the long range structure of the valence states.

The effect is that the smooth part of the valence states, denoted the pseudo waves, experience a non-local potential operator of the form

$$\hat{v} = \tilde{v}(\boldsymbol{r}) + \sum_a \sum_{ij} |\tilde{p}_i^a\rangle v_{ij}^a \langle \tilde{p}_j^a|, \tag{2.23}$$

where $\tilde{v}(\boldsymbol{r})$ is a local smooth potential, $|\tilde{p}_i^a\rangle$ projector functions localized in the core region of atom a, and v_{ij}^a are coefficients determined by the expansion of the pseudo waves in the auxhilliary atomic basis. This representation is in principle exact, provided that the set of projector functions span a complete representation of the core region, and likewise that the pseudo waves are represented in an accurate way.

The PAW method is discussed in some detail in section 3.1. The DACAPO and SIESTA codes are based on the Vanderbilt ultra-soft pseudopotentials [96] and Norm-conserving pseudopotentials [97, 98] respectively. These both represent approximate ways of evaluating the v_{ij}^a coefficients, and can be derived directly by suitable approximations to the PAW scheme.

2.3.3 Representing the Wavefunctions

The three codes used, GPAW, DACAPO, and SIESTA, each use different methods to represent the pseudo wave functions. DACAPO is based on plane waves, GPAW and SIESTA can both use numerical atomic orbitals, while GPAW also has the possibility to use real-space grids. Each representation has its own advantages and disadvantages that I will try to summarize here.

Plane Waves

Due to the periodicity of the Bloch function $u_{n,\boldsymbol{k}}(\boldsymbol{r})$, the KS wave functions can be expanded in a discrete set of plane waves

$$\psi_{n,\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} c_{n,\boldsymbol{k}+\boldsymbol{G}} \cdot e^{i(\boldsymbol{k}+\boldsymbol{G})\cdot\boldsymbol{r}}, \tag{2.24}$$

where $\boldsymbol{G}$ are reciprocal lattice vectors (i.e. $\boldsymbol{G}\cdot\boldsymbol{R} = 2\pi p$, $p \in \mathbb{Z}$, where $\boldsymbol{R}$ is the lattice vector). For large reciprocal lattice vectors, the kinetic energy contribution

to the Hamiltonian is dominant, and the truncation of the plane wave expansion is done by choosing a cutoff energy $E_c = \frac{1}{2}|\boldsymbol{k} + \boldsymbol{G}_c|^2$ which determines the maximal plane wave coefficient, $\boldsymbol{G}_c$, used in the expansion.

The advantage of a plane waves basis is that it can be systematically converged by increasing the cutoff energy and is not biased towards a specific form of the wave function.

The drawbacks are that a large number of plane waves are generally needed to aschive convergence, only periodic boundary conditions are supported implying that large supercells must by used for finite systems to avoid spurious interactions between repeated images, and the plane wave representation relies heavily on the fast Fourier transform (FFT) [99], which is a non-local operation and difficult to parallelize efficiently.

While the kinetic energy operator is diagonal in reciprocal space, implying that both the kinetic energy operator and solving the Poisson equation for the Hartree potential is trivial in this space, an FFT to real space is needed to evaluate the xc-potential, and to avoid the convolution otherwise required to apply the multiplicative (in real-space) effective KS potential.

Real Space Grids

One can also choose to represent the wave functions directly in real space, using a uniform mesh of grid points. This allows for flexible boundary conditions, a systematic convergence by reducing the grid spacing, and the code can be very efficiently parallelized over many processors using domain decomposition.

In real-space the kinetic energy operator is represented by a semi-local finite-difference stencil, and also the Poisson equation can be solved using purely grid based algortihms [100, 101], thus requiring very little communication amongst procesors.

The disadvantages are that the representation requires even more degrees of freedom than the plain waves for the same accuracy, and the discrete grid gives rise to the so-called egg-box effect which is an artificial variation in observables by shifting the grid.

Numerical atomic orbitals

A different possibility is to work with linear combinations of atomic orbitals (LCAO), in which the KS wave function is written as

$$\psi_{\mathbf{k}n} = \sum_{\mathbf{R}} \sum_{\nu} C_{\mathbf{k}n\nu} \exp(i\mathbf{k} \cdot \mathbf{R}) \Phi_\nu(\mathbf{r} - \mathbf{R}), \tag{2.25}$$

where the first sum extends over all lattice vectors for which the the atomic orbitals Φ_ν have overlap with the super cell in which $\psi_{n\boldsymbol{k}}$ is represented, and $\nu = a, n, l, m$ is a combined atomic orbital index for orbitals centered on atom a

$$\Phi^a_{nlm}(\mathbf{r}) = \zeta^a_{nl}(|\mathbf{r} - \mathbf{R}^a|) Y_{lm}(\mathbf{r} - \mathbf{R}^a), \tag{2.26}$$

where ζ^a_{nl} is a numerical radial function which vanishes beyond a certain cut-off radius, and Y_{lm} is a spherical Harmonic.

A minimal basis set consists of one atomic orbital-like function for each valence state of the atom. Extra radial functions can be added to improve the span of the basis; basis sets are denoted single-zeta (sz), double-zeta (dz) and so on, depending

on the number of such radial functions per valence state. It is normally desirable to add a basis function corresponding to the lowest unoccupied angular momentum quantum number. This is called a polarization function.

The advantage is that very few atomic orbitals are needed for a sensible description of the wave functions, and the much reduced basis size implies that the Hamiltonian can be directly diagonalized in this basis, while iterative minimization algorithms are needed for plane waves and grid representations to extract only the lowest eigenstates.

The disadvantage is that there is, in general, no systematic way of converging the basis, and a truely accurate representation is hard to obtain.

2.3.4 Nuclei Dynamics

Irrespective of the code used for solving the electronic structure problem, the interface program ASE has been used to simulate nuclei dynamics.

Nuclei dynamics can be simulated be treating the nuclei as (semi-) classical particles moving on the (quantum mechanically determined) Born-Oppenheimer surface (BOS). The force on each nucleus can be determined via the Hellmann-Feynman theorem[102]

$$\boldsymbol{F}_a = -\frac{\partial E(\boldsymbol{R})}{\partial \boldsymbol{R}_a} = -\langle\Psi|\frac{\partial \widehat{H}(\boldsymbol{R})}{\partial \boldsymbol{R}_a}|\Psi\rangle. \tag{2.27}$$

In equation (2.3) only the electron-ion potential and ion-ion energy depends on R, but in pseudopotential schemes other terms will also contribute to the force. See section 3.1.

Of special interest is the equilibrium geometry $\boldsymbol{R} = \boldsymbol{R}^0$. determined by the condition of a vanishing force

$$\boldsymbol{F}_a = -\left.\frac{\partial E(\boldsymbol{R})}{\partial \boldsymbol{R}_a}\right|_{\boldsymbol{R}=\boldsymbol{R}^0} = 0, \text{ for all } a. \tag{2.28}$$

The vibrational modes and frequncies in the equilibrium configuration, are determined from a harmonic approximation $\boldsymbol{R}_a(t) = \boldsymbol{R}_a^0 + \tilde{v}_a e^{i\Omega t}$. Inserting this in Newton's second law of motion $M_a\partial_t^2\boldsymbol{R}_a = F(\boldsymbol{R}_a)$ leads to the eigenvalue problem

$$\boldsymbol{C}\boldsymbol{v} = \Omega^2\boldsymbol{v} \tag{2.29}$$

for determination of the vibrations. Here Ω are the eigen frequncies, and the normalized eigenvectors are related to the physical eigenmodes via $\boldsymbol{v}_a = \sqrt{M_a}\tilde{\boldsymbol{v}}_a$. The mass scaled Hessian, or dynamical matrix

$$C_{a\nu,a'\nu'} = \frac{1}{\sqrt{M_a M_{a'}}}\frac{\partial^2 E}{\partial R_{a\nu}\partial R_{a'\nu'}}, \tag{2.30}$$

can be estimated by a finite difference approximation of the forces when the atoms are slightly displaced from their equilibrium sites

$$C_{a\nu,a'\nu'} \approx \frac{-1}{\sqrt{M_a M_{a'}}}\frac{F_{a\nu}(\Delta R_{a'\nu'}) - F_{a\nu}(-\Delta R_{a'\nu'})}{2\Delta R_{a'\nu'}}. \tag{2.31}$$

A displacement distance of $\Delta R_{a'\nu'} = 0.02$ is usually appropriate [103], and is used in all calculations presented here.

For an excellent review of simulations of lattice dynamics, and the effects of phonons, see [48].

2.4 Non-equilibrium Green Function Theory

In this section, the basics of non-equilibrium Green function theory (NEGF) are presented. After a short introduction of the difference in concept compared to wave function based theories, the non-interacting equilibrium solution is presented in section 2.4.1, followed by a general presentation of Green function based perturbation theory in sec. 2.4.2. This is then used to allow for a current flow through the device, and also for a perturbation based description of electron-electron (e-e) or electron-phonon (e-ph) interactions in sec. 2.4.3. The section concludes by a summary of the relation to observables in sec. 2.4.4.

For the treatment here, observables are, instead of real space coordinates, represented in a basis of single-particle orbitals $\{\phi_i\}$. These are not to be confused with the single particle eigen states $\{\psi_i\}$ of the wave function approach, here the orbitals are simply any complete set of orthonormal basis functions.

In second quantization, the system Hamiltonian is described by the operator

$$\hat{H} = \sum_{ij} h_{ij} \hat{c}_i^\dagger \hat{c}_j + \hat{V},$$

where $\hat{c}_i^\dagger$ ($\hat{c}_i$) creates (annihilates) a state in orbital ϕ_i, the single particle part of the Hamiltonian is described by the matrix $\boldsymbol{h}$ with elements $h_{ij} = \langle \phi_i | \hat{h} | \phi_j \rangle$, and $\hat{V}$ describes the interacting part of the Hamiltonian. In this chapter, matrices denoted by lower case letters will consistently refer to single particle (non-interacting) operators, while matrices related to interacting operators use capital letters. For the treatment here, $\hat{V}$ can be any generic type of interaction, but for the applications, only the two-body eletron-electron interaction operator $\hat{V}_{ee}$ and the electron-phonon operator $\hat{V}_{e\text{-ph}}$ are used.

In section 2.2, the aim was to make an accurate description of the many-body wave function (in terms of optimized single particle waves). In Green function theory, focus is not on the wave function itself, but rather on the value of the related expectation values.

Expectation values are determined by the trace $\langle \hat{O}_H(t) \rangle = \mathrm{Tr}[\hat{\rho} \hat{O}_H(t)]$, where $\hat{\rho}$ is the state operator at time t_0, and $\hat{O}_H(t) = \hat{U}_H(t_0, t) \hat{O}_H \hat{U}_H(t, t_0)$, with the time evolution operator given by (for forward propagation)

$$\hat{U}_H(t, t') = \hat{\mathcal{T}}_t e^{-i \int_{t'}^{t} \mathrm{d}t_1 \hat{H}(t_1)}, \quad t > t' \tag{2.32}$$

where $\hat{\mathcal{T}}_t$ is the chronological time-ordering operator ordering late times left. Backwards propagation is given by the unitarity relation $\hat{U}_H(t, t') = \hat{U}_H^\dagger(t', t)$.

For the pure states discussed in section 2.2, the state operator would be $\hat{\rho} = |\Psi\rangle\langle\Psi|$, but the states considered here will not have a fixed particle number but rather be in thermal equilibrium at temperature T with an electron reservoir at chemical potential μ. Such a state is described by the grand canonical state operator

$$\hat{\rho} = e^{-\beta(\hat{H} - \mu \hat{N})} / Z, \tag{2.33}$$

with $\hat{N} = \sum_{ij} c_i^\dagger c_j$ being the number operator, and $Z = \mathrm{Tr}\left[e^{-\beta(\hat{H} - \mu \hat{N})} \right]$ the grand canonical partition function.

In general, the expectation value of any generic single particle operator can be expressed in terms of the functions

$$G_{ij}^<(t, t') = i \langle \hat{c}_{H,i}^\dagger(t) \hat{c}_{H,j}(t') \rangle, \qquad G_{ij}^>(t, t') = -i \langle \hat{c}_{H,i}(t) \hat{c}_{H,j}^\dagger(t') \rangle, \tag{2.34}$$

called the lesser and greater Green function respectively. From the definition, these are seen to correspond to electron and hole propagation respectively, with the lesser (greater) removing an electron (hole) from state (j, t') and reinserting it in state (i, t). The total single particle excitation spectrum is then determined by the spectral function

$$\boldsymbol{A}(t,t') = i[\boldsymbol{G}^{>}(t,t') - \boldsymbol{G}^{<}(t,t')]. \tag{2.35}$$

For the purpose of calculating the physical response of a system, it is sometimes convenient to work with the causal Green functions

$$\boldsymbol{G}^{r}(t,t') = \theta(t-t')\left[\boldsymbol{G}^{>}(t,t') - \boldsymbol{G}^{<}(t,t')\right] \tag{2.36a}$$

$$\boldsymbol{G}^{a}(t,t') = \theta(t'-t)\left[\boldsymbol{G}^{<}(t,t') - \boldsymbol{G}^{>}(t,t')\right] \tag{2.36b}$$

known as the retarded and advanced Green function. From the definition, it follows directly that the retarded and advanced functions can be related to the lesser and greater via the spectral function

$$\boldsymbol{A} = i\left[\boldsymbol{G}^{>} - \boldsymbol{G}^{<}\right] = i\left[\boldsymbol{G}^{r} - \boldsymbol{G}^{a}\right]. \tag{2.37}$$

For the treatment here, only systems that are stationary in time, and therefore translationally invariant on the time axis will be considered.[1] In this case, the Green functions can only depend on the time difference $t - t'$, which will for notational brevity simply be denoted t. The translational invariance implies the symmetry relations

$$\left[\boldsymbol{G}^{\lessgtr}(t)\right]^{\dagger} = -\boldsymbol{G}^{\lessgtr}(-t) \qquad \left[\boldsymbol{G}^{\lessgtr}(\omega)\right]^{\dagger} = -\boldsymbol{G}^{\lessgtr}(\omega) \tag{2.38a}$$

$$[\boldsymbol{G}^{r}(t)]^{\dagger} = \boldsymbol{G}^{a}(-t) \qquad [\boldsymbol{G}^{r}(\omega)]^{\dagger} = \boldsymbol{G}^{a}(\omega) \tag{2.38b}$$

where the frequency argument ω is used to indicate the Fourier transform of $\boldsymbol{G}(t)$.

From the above symmetry relations, it follows that for stationary states, it is sufficient to work explicitly with the retarded and lesser Green functions only. The advanced then follows from the hermitian conjugate of the retarded, and the greater component from $\boldsymbol{G}^{>} = \boldsymbol{G}^{<} + \boldsymbol{G}^{r} - \boldsymbol{G}^{a}$, thus completely defining the system. Note that other choices are possible, but it turns out that the governing equation for $\boldsymbol{G}^{r}$ and $\boldsymbol{G}^{<}$ are particularly simple, so these will be used in general throughtout the thesis.

2.4.1 Equilibirum

For a system in equilibrium with a single particle reservoir, i.e. characterized by a chemical potential μ, it is possible to employ the Lehmann representation (i.e. using an expansion in the, unknown, many-body eigenstates of $\hat{H}$). In this representation, the spectral function can be expressed as

$$A_i(\omega) = 2\pi \sum_n \Big[|\langle \Psi_n^{N+1} | \hat{c}_i^{\dagger} | \Psi_0^N \rangle|^2 \delta(\omega - E_n^{N+1} + E_0^N) + |\langle \Psi_n^{N-1} | \hat{c}_i | \Psi_0^N \rangle|^2 \delta(\omega - E_0^N + E_n^{N-1}) \Big], \tag{2.39}$$

[1] In principle, oscillating stationary states could exist, thereby compromising translational invariance, but assuming some sort of finite "friction" such solution will decay. The assumed time symmetry is a major simplification and will be adopted here as a plausible assumption.

where $|\Psi_n^N\rangle$ denotes the n'th excited state of the N electron system, and E_n^N the corresponding energy. This is seen to have peaks at the possible excitation energies, provided that there is finite overlap between the ground state with an electron added in / removed from state i, and the possible states of the charged system.

This leads to the definition of a generalized density of states (DOS) $\rho(\omega)$ and projected DOS $\rho_i(\omega)$

$$\rho(\omega) = \mathrm{Tr}[\boldsymbol{A}(\omega)]/2\pi, \tag{2.40a}$$

$$\rho_i(\omega) = A_{ii}(\omega)/2\pi, \tag{2.40b}$$

which are used to discern the excitation spectrum of the system.

An important consequence of the Lehmann representation is that the lesser Green function can be related to the Fermi-Dirac distribution function $f(\omega) = 1/(e^{\beta\omega} + 1)$ via the fluctuation-dissipation theorem

$$\boldsymbol{G}^<(\omega) = if(\omega - \mu)\boldsymbol{A}(\omega). \tag{2.41}$$

As the spectral function is uniquely determined by the retarded Green function from the relation $\boldsymbol{A} = i(\boldsymbol{G}^r - \boldsymbol{G}^{r\dagger})$, it is clear that $\boldsymbol{G}^r$ is sufficient to completely characterize the equilibrium system.

For a system described by a non-interacting time-independent Hamiltonian $\boldsymbol{h}_0$, the retarded Green function is easily determined by (see e.g. [104])

$$\boldsymbol{g}_0^r(\omega) = \left[(\omega + i\eta)\boldsymbol{I} - \boldsymbol{h}_0\right]^{-1}, \tag{2.42}$$

where the positive infinitesimal $\eta = 0^+$ has been introduced to make the Fourier transform converge. This shifts the poles at the eigenvalues of h_0 slightly below the real frequency axis, such the inverted matrix is non-singular, and makes g_0^r analytic in the upper complex half plane.

For the non-interacting Hamiltonian, the generalized DOS reduce to $\mathrm{Tr}(\boldsymbol{a}_0)/2\pi = \sum_i \delta(\omega - \varepsilon_i)$, where $\{\varepsilon\}$ are the eigen values of $\boldsymbol{h}_0$, and $\boldsymbol{a}_0$ the spectral function. This is consistent with the usual definition of the DOS.

From the non-interacting equilibrium system described by $\boldsymbol{g}_0$, time-dependence (needed for electron transport) and/or interactions can then be included by a perturbation expansion, which is the topic of the following section.

2.4.2 Perturbation Theory

Consider a Hamiltonian $H(t) = h(t) + V(t)$, which consist of a single particle (but possibly time dependent) part h for which the Green functions can be determined exactly, and a perturbation V. In the definition of the, say, greater Green function $G_{ij}^>(t,t') = -i\,\mathrm{Tr}\{\rho c_{H,i}(t)c_{H,j}^\dagger\}$, the reference to the full Hamiltonian H in both the state operator and the time evolution can be changed to h by switching to the interaction picture. This implies substituting $Z\rho = Z_0\rho_0 U_{V_h}(t_0 - i\beta, t_0)$ and e.g. $c_{H,i}(t) = U_{V_h}(t_0, t)c_{h,i}(t)U_{V_h}(t, t_0)$, where Z_0 and ρ_0 are the partition function and state operator governed by h, and U_{V_h} is the time evolution operator with respect to the perturbation $V_h(t) = U_h(t_0, t)V(t)U_h(t, t_0)$.

All of the evolution operators can be collected into a single exponential by introducing the contour $\mathcal{C}_I$ depicted in figure 2.1a, a contour ordering operator $\hat{\mathcal{T}}_\mathcal{C}$ that orders operators according to this contour, and defining the generalized contour-ordered Green function

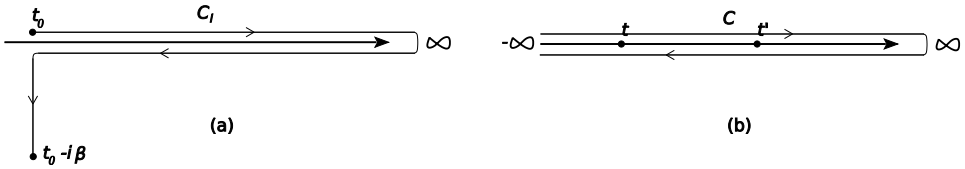


Figure 2.1: The Keldysh contour.

$$G_{ij}(\tau,\tau') = -i\frac{\mathrm{Tr}\left\{\rho_0\hat{\mathcal{T}}_{\mathcal{C}_I}\left[e^{-i\int_{\mathcal{C}}\mathrm{d}\bar{\tau}\hat{V}_h(\bar{\tau})}\hat{c}_{h,i}(\tau)\hat{c}^\dagger_{h,j}(\tau')\right]\right\}}{\mathrm{Tr}\left\{\rho_0\hat{\mathcal{T}}_{\mathcal{C}_I}\hat{c}_{h,i}(\tau)\hat{c}^\dagger_{h,j}(\tau')\right\}}. \tag{2.43}$$

The contour ordered Green function (GF) is obviously a more complicated object than its real-time counterparts, but because all reference to the perturbation is contained in the single exponential, insertion of the taylor expansion of this leads directly to the desired perturbation expansion. In addition, all of the real-time Green functions are contained within the single contour GF, and can be recovered by a suitable combination of the contour time on the real axis. For example choosing τ on the forward (backward) horizontal branch, and τ' on the backward (forward) branch leads to the the lesser (greater) GF respectively. The contour GF also contains multiple other variants, that are needed to perform the perturbation expansion.

The vertical part of the contour in figure 2.1a is related to the effect of the perturbation on the initial state at $t = t_0$. Here, the assumption will be made that the perturbation is switched on at t_0, such that $\rho = \rho_0$, and the vertical branch can be neglected. By setting the reference time $t_0 \to -\infty$ and studying only the steady-state system reached after a long time, it seems plausible that this choice will not affect the steady-state result. The contour used in practice is thus the contour shown in figure 2.1b with no vertical branch, and no reference time. Notice that setting $t_0 \to -\infty$ restores translational invariance on the time axis, such that the real-time components of the GF can be represented by a single time difference or by a frequency argument. On the contour, two time coordinates are still needed to distinguish times on different branches.

Inserting the Taylor expansion of the exponential in the contour GF, each term in the expansion will contain an increasing number of interaction and annihilation / creation operators. Since both time-evolution and statistical averaging are related to a non-interacting Hamiltonian ($\hat{h}$), Wicks theorem[105] can be applied to contract all multiple operator expectation values to products of single pairings. Each term will then be a series of products of a single particle operator and the non-interacting Green function $\boldsymbol{g}$.

All terms in the expansion which involves the perturbation operator can formally be collected into a single two-index function called the self-energy Σ_{ij}. Representing this as the matrix $\boldsymbol{\Sigma}$, the expansion can be collapsed in the Dyson equation

$$\boldsymbol{G}(\tau,\tau') = \boldsymbol{g}(\tau,\tau') + \iint \mathrm{d}\tau_1\tau_2\boldsymbol{g}(\tau,\tau_1)\boldsymbol{\Sigma}(\tau_1,\tau_2)\boldsymbol{G}(\tau_2,\tau'), \tag{2.44}$$

with matrix multiplication implied. The self-energy can be represented by sums of Feynman diagrams representing all possible combinations of free propagation by g with multiple scattering events on the perturbation V. The included diagrams should be "connected" and "irreducible". All disconnected diagrams are cancelled

by the expansion of the denominator in (2.43)), and the reducible diagrams (i.e. diagrams that can be divided into two parts only connected by a free propagation line) are automatically included by iteration of the Dyson equation (2.44).

The explicit form of the self-energy depends on the type of perturbation. The perturbations relevant for this text are discussed in the next section.

2.4.3 Non-equilibrium and Interactions

Two types of perturbations will be considered in this section. The first is a time-dependent but quadratic perturbation, which is needed to describe transport of (non-interacting) electrons in non-equilibrium systems. The second perturbation is a many-body interaction operator (in this case e-e or e-ph). As the non-equilibrium perturbation is quadratic, the many-body interaction can be added on top of this, or it can act as the sole perturbation.

The non-equilibrium transport setup is an open boundary non-periodic problem, with two semi-infinite leads (L and R) coupled to a finite central region. By employing a localized basis set, the (non-interacting) Hamiltonian can be decomposed as

$$\begin{bmatrix} \boldsymbol{h}_L & \boldsymbol{h}_{LC} & \mathbf{0} \\ \boldsymbol{h}_{LC}^\dagger & \boldsymbol{h}_c & \boldsymbol{h}_{RC}^\dagger \\ \mathbf{0} & \boldsymbol{h}_{RC} & \boldsymbol{h}_R \end{bmatrix}, \tag{2.45}$$

The vanishing coupling between L and R can always be achieved by defining a large enough central region, i.e. including more of the leads in the central region.

Initially (at $t = t_0$) the three regions are in contact, but electron transfer is prohibited (i.e. the elements $\boldsymbol{h}_{\alpha C}$, $\alpha \in L, R$ are set to zero). Due to the partitioning in a local basis set, the initial state operator is simply the product $\hat{\rho} = \hat{\rho}_L \hat{\rho}_C \hat{\rho}_R$ of the state operators of the individual subdomains, which are all in thermal equilibrium at the same temperature, but possibly different chemical potentials. The perturbation is then at $t = t_0$, to switch on the coupling between leads and central region, which will induce a current as the lead of highest chemical potential discharges into the one with the lowest. The leads are kept at their respective equilibrium potentials μ_α and eventually a steady-state current will develop through the central region. As the initial Hamiltonian is simply the diagonal of (2.45), the central region is initially completely defined in terms of the retarded equilibrium GF $\boldsymbol{g}_0^r = [(\omega + i\eta) - \boldsymbol{h}_C]^{-1}$. The perturbation expansion leads to the contour ordered non-equilibrium GF

$$\boldsymbol{g}(\tau, \tau') = \boldsymbol{g}_0(\tau, \tau') + \iint \mathrm{d}\tau_1 \tau_2 \boldsymbol{g}_0(\tau, \tau_1) \boldsymbol{\Sigma}_{\text{leads}}(\tau_1, \tau_2) \boldsymbol{g}(\tau_2, \tau'), \tag{2.46}$$

where $\boldsymbol{\Sigma}_{\text{leads}} = \boldsymbol{\Sigma}_L + \boldsymbol{\Sigma}_R$ is the perturbation to the central region by the coupling to the left and right leads. This perturbation is particularly simple, as it is both quadratic, and time-independent on the interval defined by the contour ($t > t_0$). The perturbation expansion can therefore be summed to infinite order, yielding the exact expression[106]

$$\boldsymbol{\Sigma}_\alpha(\tau_1, \tau_2) = \boldsymbol{h}_{\alpha C}^\dagger \boldsymbol{g}_{0,\alpha}(\tau, \tau') \boldsymbol{h}_{\alpha C} \quad \alpha \in L, R. \tag{2.47}$$

Here, $\boldsymbol{g}_{0,\alpha}$ is the equilibrium GF of lead α, which is completely defined by the retarded component $\boldsymbol{g}_{0,\alpha}^r = [(\omega + i\eta) - \boldsymbol{h}_\alpha]^{-1}$. As the leads, and therefore the dimension of $\boldsymbol{h}_\alpha$, are infinite, direct inversion is impossible, but in a localized

basis set, the couplings matrices $\boldsymbol{h}_{\alpha C}$ only couple the central region to the *surface* elements of the leads, and hence only the surface elements of the lead Green functions are needed. For a semi-infinite periodic lead Hamiltonian, a rapidly converging iterative scheme can be devised to determine the surface Green function without ever inverting matrices larger than the size corresponding to the periodicity of the lead Hamiltonian.[107, 108]

Although the non-equilibrium Hamiltonian is time-dependent (it contains a step function at $t = t_0$ switching on the lead coupling), it is non-interacting, and in the general formulation by [105], Wicks theorem still applies. The non-interacting non-equilibrium system can thus be used as the unperturbed system forming the basis of a further perturbation expansion.

If many-body interactions are only included in the central region, the perturbative inclusion of these leads to the Dyson equation

$$\boldsymbol{G}(\tau,\tau') = \boldsymbol{g}(\tau,\tau') + \iint \mathrm{d}\tau_1\tau_2 \boldsymbol{g}(\tau,\tau_1)\boldsymbol{\Sigma}_{\mathrm{int}}(\tau_1,\tau_2)\boldsymbol{G}(\tau_2,\tau'), \tag{2.48}$$

which when combined with (2.47) can be related to the equilibrium GF of the central region by

$$\boldsymbol{G}(\tau,\tau') = \boldsymbol{g}_0(\tau,\tau') + \iint \mathrm{d}\tau_1\tau_2 \boldsymbol{g}_0(\tau,\tau_1)\boldsymbol{\Sigma}_{\mathrm{tot}}(\tau_1,\tau_2)\boldsymbol{G}(\tau_2,\tau'), \tag{2.49}$$

where $\boldsymbol{\Sigma}_{\mathrm{tot}} = \boldsymbol{\Sigma}_{\mathrm{int}} + \boldsymbol{\Sigma}_{\mathrm{leads}}$.

The explicit form of the interaction self-energy is discussed for e-e and e-ph interactions in section 2.5.

Note that inclusion of interactions for an isolated central region without coupling to leads follows directly by omission of Σ_{leads} in (2.49).

2.4.4 Real-time Functions and Relations to Observables

Analytic Continuation

The Dyson equation, and the self-energy expansion are both derived on the contour, as it is for the contour ordered GF that a perturbation expansion can be derived. When converting these equations back to real-time expressions, the contour integrations and multiplications which mix different time arguments will couple the expressions for retarded, advanced, lesser, and greater components. A set of rules for conversion of the most commonly encountered products and convolutions can be derived, see e.g. [106], and are referred to as the Langreth rules[109].

For convenience, some of the most used rules are listed listed in table 2.1. In this table, the time arguments of the real time components are reduced to the single time-difference argument needed in steady-state, and a Fourier transform has been performed when convenient to avoid convolutions.

An important application of the Langreth rules, is to convert the Dyson equation (2.49) to a set of coupled equations for the real-time components (here subsequently transformed to frequency space)

$$\boldsymbol{G}^r(\omega) = \boldsymbol{g}_0^r(\omega) + \boldsymbol{g}_0^r(\omega)\boldsymbol{\Sigma}_{\mathrm{tot}}^r(\omega)\boldsymbol{G}^r(\omega), \tag{2.50a}$$

$$\boldsymbol{G}^<(\omega) = \boldsymbol{G}^r(\omega)\boldsymbol{\Sigma}_{\mathrm{tot}}^<(\omega)\boldsymbol{G}^{r\dagger}(\omega) + \boldsymbol{\Delta}^<(\omega), \tag{2.50b}$$

	$A(\tau_1,\tau_2)B(\tau_1,\tau_2)$	$A(\tau_1,\tau_2)B(\tau_2,\tau_1)$
$X^r(t)$	$A^rB^r + A^rB^< + A^<B^r$	$A^rB^{<\dagger} + A^<B^{r\dagger}$
$X^<(t)$	$A^<B^<$	$A^<B^{>\dagger}$

	$\int_\mathcal{C} \mathrm{d}\tau A(\tau_1,\tau)B(\tau,\tau_2)$	$\iint_\mathcal{C} \mathrm{d}\tau\mathrm{d}\tau' A(\tau_1,\tau)B(\tau,\tau')C(\tau',\tau_2)$
$X^r(\omega)$	A^rB^r	$A^rB^rC^r$
$X^<(\omega)$	$A^rB^< + A^<B^{r\dagger}$	$A^rB^rC^< + A^rB^<C^r + A^<B^rC^r$

Table 2.1: The Langreth conversion rules, here reduced to the single time-difference argument required in steady-state, and Fourier transformed when convenient.

with

$$\boldsymbol{\Delta}^<(\omega) = [\boldsymbol{I} + \boldsymbol{G}^r(\omega))\boldsymbol{\Sigma}^r_{\mathrm{tot}}(\omega)]\, \boldsymbol{g}_0^<(\omega)\, [\boldsymbol{I} + \boldsymbol{G}^r(\omega))\boldsymbol{\Sigma}^r_{\mathrm{tot}}(\omega)]^\dagger . \tag{2.51}$$

The above term can be shown to vanish everywhere in the case of an interacting system, and in the case of non-interacting systems it is only non-zero at frequencies corresponding to bound states[107]. It is thus irrelevant for the work presented here, and will be ignored in the following.

Notice that the equation for the retarded component, (2.50a) only couples to the other components through a possible dependence in the retarded self-energy. When expressed in frequency space, it does not involve any convolutions, and can be solved directly, yielding

$$\boldsymbol{G}^r(\omega) = \left[(\boldsymbol{g}_0^r(\omega))^{-1} - \boldsymbol{\Sigma}^r_{\mathrm{tot}}(\omega)\right]^{-1} . \tag{2.52}$$

This is the prime reason for choosing to work with the retarded instead of one of the other possible real-time Green functions.

Relation To Observables

In the previous, it has been shown how to determine the Green function, but a link to the desired observables has yet to be established.

From the definition of the lesser Green function, eq. (2.34), it is seen that the electron density matrix is given by

$$\boldsymbol{D} = -i\boldsymbol{G}^<(t=0) = -i\int \frac{\mathrm{d}\omega}{2\pi}\boldsymbol{G}^<(\omega). \tag{2.53}$$

The spectral properties of the system are given by the spectral function $\boldsymbol{A}(\omega) = i\left[\boldsymbol{G}^r - \boldsymbol{G}^{r\dagger}\right]$,which have peaks at the quasi-particle excitation energies. The diagonal elements of this defines a projected density of states (DOS) $\rho_i(\omega)$, and the trace gives the total DOS $\rho(\omega)$

$$\rho_i(\omega) = A_{ii}(\omega)/2\pi, \tag{2.54}$$
$$\rho(\omega) = \mathrm{Tr}\,\boldsymbol{A}(\omega)/2\pi. \tag{2.55}$$

The electron flow from lead α can be obtained from the time-derivative of the number operator in the lead $\partial_t\hat{N}_\alpha = -i\langle[\hat{H},\hat{N}_\alpha]\rangle_0$ using that the only part of the

Hamiltonian which does not commute with the number operator is the tunneling Hamiltonian. The result is the Meir-Wingreen current formula [30]

$$I_\alpha = \int \frac{\mathrm{d}\omega}{2\pi} \mathrm{Tr}\left[\boldsymbol{\Sigma}_\alpha^<(\omega)\boldsymbol{G}^>(\omega) - \boldsymbol{\Sigma}_\alpha^>(\omega)\boldsymbol{G}^<(\omega)\right], \tag{2.56}$$

where all matrices, and hence the trace, relate to basis functions in the central region only. The lesser and greater GF components reflect the density of electron and hole states at energy ω in the central region respectively and the lead self-energies are simply products the corresponding lead GF and the coupling between lead and the central region. Thus the above form of the current formula has the appealing physical interpretation of a balence between the electron and hole transfer between lead α and the central region.

For non-interacting systems, the fluctuation-dissipation theorem for the lead self-energies can be used to express the lesser and greater GF, via the Keldysh equation, in terms of the Fermi distribution relevant for the two leads respectively. After some rearrangements, this leads to the particularly simple expression for phase-coherent transport (i.e. with no quasiparticle scattering)

$$I = \frac{2e^2}{h}\int \mathrm{d}\omega \left[f_L(\omega) - f_R(\omega)\right] T(\omega), \tag{2.57}$$

where units have been re-introduced to highlight the natural apperance of the quantum unit of conductance $G_0 = 2e^2/h$, and the transmission function

$$T(\omega) = \mathrm{Tr}\left\{\boldsymbol{G}^r(\omega)[\boldsymbol{\Sigma}_L^r - \boldsymbol{\Sigma}_L^{r\dagger}(\omega)]\boldsymbol{G}^{r\dagger}(\omega)[\boldsymbol{\Sigma}_R^{r\dagger}(\omega) - \boldsymbol{\Sigma}_R^r]\right\}. \tag{2.58}$$

In the form of (2.57), the low temperature current is simply given by an integral over the energy window defined by the difference in chemical potential of left and right lead, i.e. the bias. In the limit of low bias, the conductance $G = I/V$ thus becomes

$$G = G_0 T(\epsilon_F), \tag{2.59}$$

which is identical to the Landauer-Büttiker result $G = G_0 \sum_t |t_n|^2$ with the squared transmission probabilities t_n given by the eigenvalues of the matrix in (2.58).

Note that the formulation of the current in terms of a transmission function and an integral limited to energies in the bias window is only possible for non-interacting systems. If interactions are present, quasiparticle scattering leads to contributions from outside of the bias window, and equation (2.56) should be utilized.

2.4.5 NEGF Calculations in Practice

The preceding sections where purely based on the theoretical aspect of NEGF theory, with no reference, as to how the involved quantities are obtained. Two types of input is needed for the NEGF construction: i) A representation of a mean-field Hamiltonian in a basis with finite support in the transport direction such that a left, central, and right part can be distinguished as well as the coupling between these, and ii) a representation of the interaction operator, if any, in the central region. Based on these, and a specification for the chemical potentials of the leads, the excitation spectrum (DOS), non-equilibrium density matrix, and the current can be determined.

The source of the mean-field Hamiltonian H_0 and the interaction matrix V, is in principle arbitrary, but for an ab-initio description of an atomistic junction, the best available source of a mean-field Hamiltonian is by use of KS-DFT. Although KS-DFT provides good total energies, there is no good reason why it should lead to accurate values of the current. Putting this issue aside and accepting that it in practice often produce good results, and it is anyway the best source of a mean-field description of real systems, this section will describe how the different parts of the principally infinite dimensional, open boundary, Hamiltonian can be extracted from DFT in practice.

The representation of the many-body interactions is delayed to the chapters in which they are used. Note that when KS-DFT is used to describe the mean-field Hamiltonian, e-e interaction self-energies should only be used to represent the *change* in Hartree, exchange, and correlation relative to the approximations of these inherent in KS-DFT.

Non-orthognoal Basis sets

Up to now, all expressions for Green functions and related quantities have been derived for an orthogonal basis. In practical implementations, the employed basis is however often non-orthogonal., The generalization of NEGF theory to non-orthogonal basis sets has been formulated by [110], and the practical implications are summarized here for convenience.

The primary implication of the non trivial overlap matrix $\boldsymbol{S}$ is that one works with the dual Green function

$$\boldsymbol{G}^r(\omega) = \left[(\omega + i\eta)\boldsymbol{S} - \boldsymbol{H}_0 - \boldsymbol{\Sigma}\right]^{-1}, \tag{2.60}$$

which is related to the physical Green function via $\boldsymbol{S}\boldsymbol{G}^r\boldsymbol{S}$. This also entails a modified expression for the retarded lead self-energies

$$\boldsymbol{\Sigma}^r_\alpha(\omega) = \left[(\omega + i\eta)\boldsymbol{S}_{\alpha C} - \boldsymbol{H}_{\alpha C}\right]^\dagger g^r_\alpha(\omega) \left[(\omega + i\eta)\boldsymbol{S}_{\alpha C} - \boldsymbol{H}_{\alpha C}\right]. \tag{2.61}$$

Using the dual Green function defined by the above in all expressions, most equations remain structurally unchanged, namely: the Keldysh equation, the fluctuation-dissipation equations (both for the lesser GF and the lesser lead self-energies), the current formula, and the expansion series of the e-e and e-ph interactions (and thereby all self-energies related to these). The only other modification, besides the expressions for the retarded GF and lead self-energies, relates to the projected density of states, which for orbital ϕ_i becomes

$$D_i(\varepsilon) = [\boldsymbol{S}\boldsymbol{A}(\varepsilon)\boldsymbol{S}]_{ii} \;/\; 2\pi S_{ii}. \tag{2.62}$$

Correspondingly, the total density of states, or quasiparticle spectrum, is given by the sum of these, i.e.

$$D(\varepsilon) = \mathrm{Tr}(\boldsymbol{A}(\varepsilon)\boldsymbol{S})/2\pi. \tag{2.63}$$

Boundary Conditions

A generic, but typical, example of the type of junction considered in the applications presented in this thesis is shown in figure 2.2.

By assuming the system is periodic in the plane transverse to the transport direction, the transverse Bloch vector $\boldsymbol{k}_\perp$ becomes a good quantum number. This

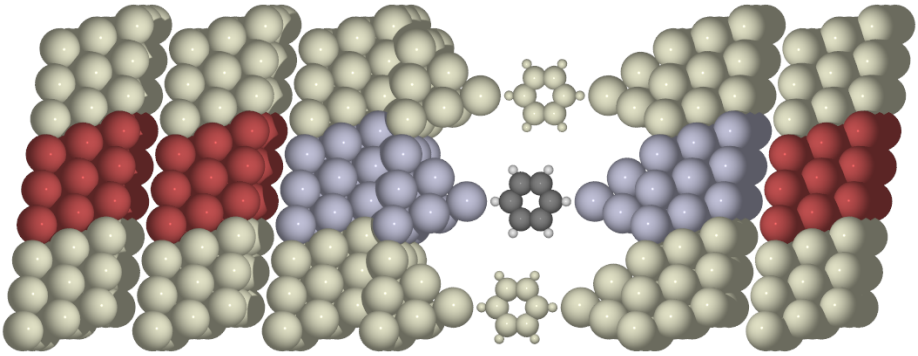


Figure 2.2: The setup used in practice to extract parameters for non-equilibrium transport. The colors and gap between principal layers are intended as a visual guide only. Red atoms indicate the principal layers of the lead, blue the central region, and yellowish the periodic images induced in the transverse plane by the supercell approach.

implies that all matrices are diagonal in $\boldsymbol{k}_\perp$, and can be referenced by a single such index and basis functions belonging to the confined transverse region determined by the periodicity, which presents a major simplification. Expectation values, such as the current, density-matrix, transmission function etc. should then be averaged over $\boldsymbol{k}_\perp$-vectors in the first BZ of the transverse supercell. In practice, the mean-field Hamiltonian matrices extracted from DFT are simply based on a discrete sampling as described in sec. 2.3.1, and these are then treated independently in the NEGF scheme and the final output is averaged. From a practical point of view, the inclusion of k-point sampling is thereby trivial. For notational simplicity, the $\boldsymbol{k}_\perp$ index on quantities will be suppressed in the following, but it is noted that a sufficient sampling of $\boldsymbol{k}_\perp$ is crucial for obtaining reliable results [111]. The assumed periodicity can cause interferrence between the repeated junctions, but the effect is in practice negligible for molecules seperated by about 5 Å or more from their periodic images. Note that the Bloch sums of localized orbitals in the transverse direction do *not* have finite support, and nor is this required of the formalism. Locality is only required in the transport direction, to facilitate the left, central, right partitioning in (2.45):

$$\boldsymbol{h}_0 = \begin{bmatrix} \boldsymbol{h}_L & \boldsymbol{h}_{LC} & \boldsymbol{0} \\ \boldsymbol{h}_{LC}^\dagger & \boldsymbol{h}_c & \boldsymbol{h}_{RC}^\dagger \\ \boldsymbol{0} & \boldsymbol{h}_{RC} & \boldsymbol{h}_R \end{bmatrix}.$$

The finite support of the basis in the transport direction also implies that the lead Hamiltonian can be divided into principal layers which, if chosen large enough, only have nearest neighbor coupling. In combination with the assumed periodicity of the lead indicated in fig. 2.2 this has the consequence, that the lead Hamiltonian

may be written in a block-tridiagonal form

$$\boldsymbol{h}_L = \begin{bmatrix} \ddots & \vdots & \vdots & \vdots \\ \cdots & \boldsymbol{\epsilon}_L & \boldsymbol{t}_l & 0 \\ \cdots & \boldsymbol{t}_L^\dagger & \boldsymbol{\epsilon}_L & \boldsymbol{t}_L \\ \cdots & 0 & \boldsymbol{t}_L^\dagger & \boldsymbol{\epsilon}_L \end{bmatrix} \tag{2.64}$$

where $\boldsymbol{\epsilon}_L$ is the (onsite) Hamiltonian matrix of a single principal layer in the left lead, and $\boldsymbol{t}_L$ is the coupling between neighboring layers. A similar relation holds for the right lead, and the overlap matrices of both.

Note that the partitioning in (2.64) assumes that the defined central region is so large that the potential of the surface principal layer has converged to its bulk value. If also the coupling between has converged at the interface, the (non-zero part of) the coupling matrices $\boldsymbol{H}_{\alpha C}$ is also identical to $\boldsymbol{t}_\alpha$. Thus the system is completely defined in terms of the matrices $\boldsymbol{\epsilon}_\alpha$, $\boldsymbol{t}_\alpha$, $\boldsymbol{h}_C$, and the corresponding overlap matrices.

A description of the leads can be extracted from DFT by performing a calculation for a single principle layer using a sufficiently dense $\boldsymbol{k}$-point sampling in the transport direction, and then construct the onsite and hopping matrices by

$$\boldsymbol{\epsilon}_\alpha = \frac{1}{N_k}\sum_{k_\parallel} e^{ik_\parallel \cdot 0}\boldsymbol{h}_\alpha(k_\parallel), \qquad \boldsymbol{t}_\alpha = \frac{1}{N_k}\sum_{k_\parallel} e^{ik_\parallel \cdot 1}\boldsymbol{h}_\alpha(k_\parallel), \tag{2.65}$$

where the $\boldsymbol{k}$-vector is given in scaled coordinates, and $\boldsymbol{h}_\alpha(k_\parallel)$ is the k dependent output of the DFT calculation.

The Hamiltonian of the central region indicated in fig. 2.2 should in principle be determined subject to open boundary conditions. As the potential is assumed to have converged to the bulk lead values at the end planes of the region, the Hamiltonian is instead determined assuming periodic boundary conditions (and a Γ-point approximation in the transport direction). The effects of the periodic boundary conditions are subsequently removed by manually setting all matrix elements in the Hamiltonian and overlap matrices to zero if the corresponding basis functions are separated by more than half the length of the super-cell. The effect of the semi-infinite leads are included in the Green function via the lead self-energies (2.61). From the short range of the coupling between central and lead regions, only the part of the lead GF coresponding to the surface principal layer needs be determined. From the particularly simple form of (2.64) this surface GF can be calculated using the efficient iterative decimation technique [108].

In practice, due to the finite size of the central region, the Fermi level of the bulk lead and central region may differ, and the Hamiltonian of the central region should be shifted to align these to avoid spurious electron scattering at the interface between the regions.

2.5 Self-energy Approximations

The self-energy can be deduced by application of Wicks theorem on the expansion of the contour Green function (2.43). The final expressions are however much simpler to derive by drawing diagrams.

In this section, the most common self-energy approximations to the e-e interaction, which are also used extensively throughout this thesis, are presented, and the

physical content discussed. A short description of the e-ph self-energy diagrams is also presented, as these will be used in the final chapter of the thesis in simulations of inelastic transport experiments.

2.5.1 Electron-electron Interactions

This section is dedicated to the electron-electron pair interaction operator

$$\hat{V} = \sum_{ijkl} V_{ij,kl} \hat{c}_j^\dagger \hat{c}_k^\dagger \hat{c}_l \hat{c}_i, \quad V_{ij,kl} = \iint \mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r}' \phi_i(\boldsymbol{r}) \phi_j^*(\boldsymbol{r}) \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \phi_k^*(\boldsymbol{r}') \phi_l(\boldsymbol{r}'). \quad (2.66)$$

The rules for drawing the Feynman diagrams are simple: Green functions are represented by lines, and interactions by zig-zag lines, as shown in figure 2.3.

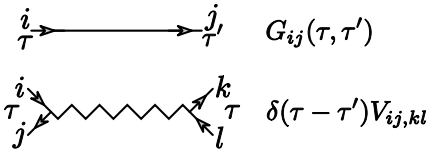


Figure 2.3: The base elements of the diagrams: The propagation line and the interaction line.

A diagram representing an n'th order expansion term should contain n interaction lines. An interaction line has four terminals, all of these except two (the external terminals) should be connected by free propagation lines.

To obtain all possible terms at a given order, one should make all possible topologically different irreducible drawings. Irreducible means that it should not be possible to cut the drawing into two disconnected part by removing a single propagation line. All such terms are automatically included by iteration of the Dyson equation.

To convert the drawings into equations, one multiplies all interaction and Green function matrices and sum over all connected terminals indices. In addition, the term should be multiplied by i^n and an additional factor of -1 for every Fermion loop[106].

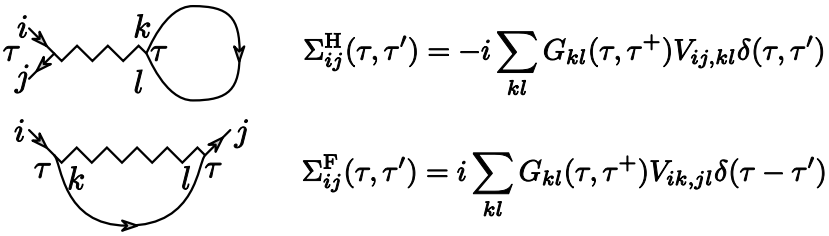


Figure 2.4: The two possible first order diagrams: The Hartree bubble, and the Fock oyster.

The system is best demonstrated by examples. Figure 2.4 shows the two possible first order diagrams; the Hartree bubble, and the Fock oyster, with their respective conversions into equations. The interpretation of the Hartree diagram is that an

electron interacts instantaneously with the density of all electrons (including itself). This is a classical electrostatic interaction. In the Fock diagram, the electron, taking all possible routes from i to j, will interact with its own final destination. This diagram will reinforce the Pauli exclusion principle missing in the Hartree diagram, which implies removing the spurious self-interaction inherent in the Hartree diagram.

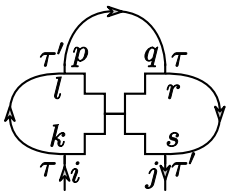
Both first order diagrams are instantaneous in time, implying that when converted to real-time, their lesser component will be zero. The retarded components are[2]

$$\Sigma^r_H(t)_{ij} = \delta(t) \sum_{kl} D_{kl} V_{ij,kl}, \tag{2.67a}$$

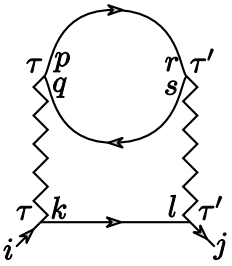
$$\Sigma^r_F(t)_{ij} = -\delta(t) \sum_{kl} D_{kl} V_{ik,jl}, \tag{2.67b}$$

where $\boldsymbol{D} = -i\boldsymbol{G}^<(t=0)$ is the density matrix. Inclusion of of both first order diagrams is denoted the Hartree-Fock (HF) approximation.

The two possible second order diagrams are shown in figure 2.5. These are the cross diagram, and the polarization bubble diagrams. Inclusion of both first order and both second order diagrams is denoted the second Born approximation (for e-e interactions).



$$\Sigma^{\mathrm{C}}_{ij}(\tau,\tau') = -\sum_{klpqrs} G_{kl}(\tau,\tau')G_{pq}(\tau',\tau)G_{rs}(\tau,\tau')V_{ik,rq}V_{lp,js}$$



$$\Sigma^{\mathrm{P}}_{ij}(\tau,\tau') = +\sum_{klpqrs} G_{kl}(\tau,\tau')G_{pr}(\tau,\tau')G_{sq}(\tau',\tau)V_{ik,pq}V_{rs,jl}$$

Figure 2.5: The two possible second order diagrams: The cross diagram, and the polarization bubble.

Especially the polarization bubble self-energy Σ^P has interest. It resembles the Fock diagram, but while the virtually excited electron propagates and later returns to the Fermi sea, its self-interaction is screened by the intermediate generation of an electron-hole pair elsewhere in the system. The loop of two connected Green functions represents an electron-hole pair, because the two Green functions propagate in different directions in time, which can be viewed as propagating an electron and a hole in the same direction. Alternatively, application of the Langreth rules, table

[2]Note that if including a spin component in the oribital indices, the Coulomb matrix $V_{ij,kl}$ is non-zero only if i,j are same-spin, as well as the k,l pair. This implies that in the Fock diagram, all spin projections must equal that of the external vertices, while the Hartree diagram includes an internal spin summation.

2.1, shows that conversion of a pair of Green functions anti-parallel in time will involve pairings of lesser and greater Green functions, which represent electron and hole propagators.

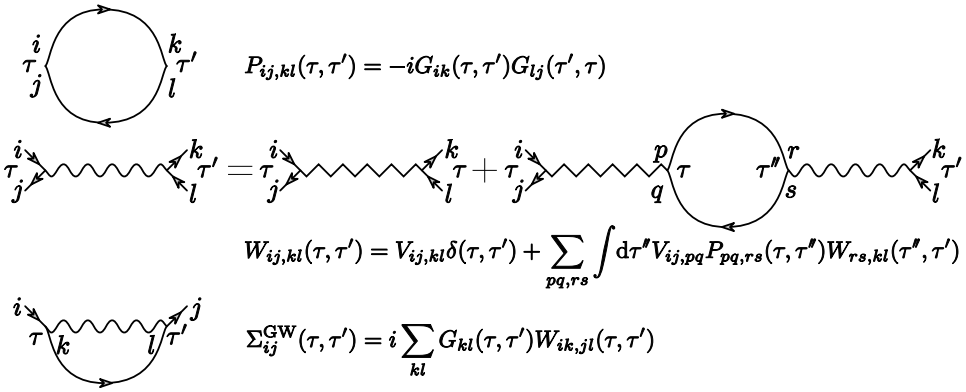


Figure 2.6: The GW self-energy diagram.

The screening of the Fock interaction by a sequence of electron-hole pair generations can be summed to infinite order by introducing the screened interaction W represented by the wobbly line in figure 2.6, and the associated GW self-energy also shown in the figure. The screened interaction satisfies a Dyson-like equation that will produce all orders of polarization when iterated.

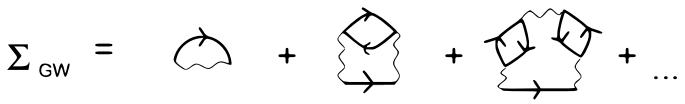


Figure 2.7: Expansion of the terms included in the GW self-energy

The diagrams included in the GW self-energy can be argued to be the most important of each order for a Fermi liquid using a divergence number argument[104]. On the other hand, the different diagrams of a given order represent different pairings in the Wicks contraction of an expansion term, and neglecting some while including other will introduce a form of self-interaction. I.e. matrix elements which should be zero due to cancellation of different pairings in the contraction, will not vanish. On the other hand the GW construction sums an important set of diagrams to infinite order and is known to perform well for especially extended systems of metallic or semiconducting character.

The GW approximation can also be derived from the formally exact set of self-consistent equations by [112] sketched in figure 2.8a. The GW approximation fig. 2.8b follows by ignoring the vertex corrections Γ.

2.5.2 Electron-phonon Interactions

In a harmonic approximation of the BO surface, the isolated coupling between electrons and phonons is in second quantization, described by the Hamiltonian

$$\hat{H}_{e\text{-ph}} = \sum_{\lambda}\sum_{ij} M^{\lambda}_{ij}\hat{c}^{\dagger}_{i}\hat{c}_{j}(\hat{b}^{\dagger}_{\lambda} + \hat{b}_{\lambda}), \tag{2.68}$$

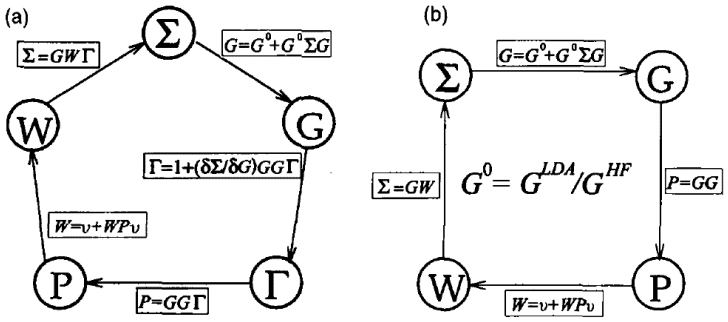


Figure 2.8: The formaly exact coupled equations for the electron-electron interaction self-energy knowns as Hedins wheel. Reproduced from [113].

where $\hat{b}^\dagger_\lambda/\hat{b}_\lambda$ creates / annihilates a phonon in eigen mode λ. The coupling matrix M^λ_{ij} is related to the directional derivative of the Hamiltonian along the eigen mode $\tilde{\boldsymbol{v}}_a$ (see section 2.3.4), or in terms of the normalized mass scaled modes $\boldsymbol{v}_a$:

$$M^\lambda_{ij} = \sum_{a\nu} \sqrt{\frac{\hbar}{2M_a\Omega^\lambda}} v^\lambda_{a\nu} W_{ij,a\nu}, \tag{2.69}$$

where Ω^λ is the phonon energy and

$$W_{ij,a\nu} = \langle\phi_i|\frac{\partial v_{\text{eff}}(\boldsymbol{r})}{\partial R_{a\nu}}|\phi_j\rangle. \tag{2.70}$$

The diagramatic expansion of this interaction leads to exactly the same diagrams as for the e-e-interaction, except in this case, the interaction (zig-zag) lines do not represent $V_{ij,kl}$, but rather $M^\lambda_{ij}D_{0,\lambda}(\tau,\tau')M^\lambda_{kl}$, where $D_{0,\lambda}$ is the free phonon propagator.

In the applications studied here, phonons are assumed to dissipate immediately into the leads when excited (the strongly damped, or zero life-time limit), and the mean occupation statistics are also assumed to be zero. In this limit, the real-time free phonon GFs are simply

$$D^r_{0,\lambda}(\omega) = \frac{1}{\omega - \Omega^\lambda + i\eta} - \frac{1}{\omega + \Omega^\lambda + i\eta}, \tag{2.71a}$$

$$D^<_{0,\lambda}(\omega) = -2\pi i\delta(\omega \pm \Omega^\lambda). \tag{2.71b}$$

The first Born approximation for the e-ph interaction is the sum of first order terms, i.e. the equivalent of the Hartree-Fock approximation for e-e interactions. The Hartree term however merely induce a static phonon-induced shift of the electronic mean-field potential. It can be argued that the dynamics are better described by omitting this term[103, 114], as it would have been at least partially screened had it the electronic Hartree term been allowed to respond to this (when including e-ph effects I do not also include e-e effect). In any case it does not lead to any signal in the I/V characteristic and is not considered here. This leaves the Fock term, which is given by

$$\boldsymbol{\Sigma}^{\lessgtr}_\lambda(\omega) = \boldsymbol{M}^\lambda \boldsymbol{G}^{\lessgtr}(\omega \pm \Omega^\lambda)\boldsymbol{M}^\lambda, \tag{2.72a}$$

$$\boldsymbol{\Sigma}^r_\lambda(\omega) = \tfrac{1}{2}[\boldsymbol{\Sigma}^>_\lambda(\omega) - \boldsymbol{\Sigma}^<_\lambda(\omega)] - i\int \frac{\mathrm{d}\omega'}{2\pi}\frac{\boldsymbol{\Sigma}^>_\lambda(\omega') - \boldsymbol{\Sigma}^<_\lambda(\omega')}{\omega' - \omega}. \tag{2.72b}$$

Chapter 3

Developments in gpaw

Most of the work presented in this thesis has been done based on results from the PAW-DFT code GPAW [86, 87]. Even the Green function (GF) based calculations are based on a mean-field Hamiltonian, and Coulomb matrix extracted from GPAW. The code is based on a frozen core approximation, thus treating only the valence states explicitly. This also implies that the GF based methods will represent the valence part of the Hilbert space only. It is however important to acknowledge that the code, due to the use of PAW, is an all-electron (AE) method, such that extracted matrix elements and expectation values correspond to AE states subject to the full AE potential. For the self-energies used in the GF method it is important that they are based on AE quantities, and that although operating in the valence space only, the explicit orbital valence-core interaction can also be extracted from PAW.

The chapter starts with a short introduction to the PAW methodology in sec. 3.1, followed by a more detailed account in sec.3.2 of how to extract various AE quantities from the PAW code in practise, including how to implement exact-exchange based functionals in the PAW framework. The remaining two sections are devoted to applications. Section 3.3 studies the application of exact-exchange and hybrid functionals to obtain improved atomization and ionization energies over standard local xc approximations. Section 3.4 demonstrates the use of a constrained DFT approach to study the image charge effect on ionization of a molecule at a metal surface.

Both of the applications chosen to illustrate this section are related to describing the charged excitation of systems in DFT. This is motivated by the general topic of the thesis on molecular electronics. To describe the electron transport through a molecular device, an accurate description of the quasiparticle excitation spectrum is essential, as driving a current involves (for weakly bound molecules) sequential tunneling of electrons from one lead onto the molecule and from the molecule to the other lead, both processes having charged intermediate states. In the next chapter, 4, the GF based perturbation expansion of the e-e interactions are applied to improve the spectrum, but before venturing into this, it is interesting to study first how far one can get in terms of accuracy of the predicted excitation spectrum using purely DFT based methods. This is the purpose of sections 3.3 and 3.4.

The chapter is based on the papers Paper IV on extensions to PAW, Paper I on time-dependent DFT in PAW, and Paper VI on general GPAW development. Most of the applications in sec. 3.3 and sec. 3.4 are however unpublished.

3.1 The Projector Augmented-wave Method

The central idea in the projector augmented-wave (PAW) method is to express the all-electron valence states ψ_n in terms of a smooth pseudo wave $\tilde{\psi}_n$ augmented by a local basis set expansion restricted to a small region Ω^a, called the augmentation sphere, around each atom a. This is achieved through the definition of a linear transformation operator $\hat{\tau}$

$$\psi_n = \hat{\tau}\tilde{\psi}_n, \quad \hat{\tau} = 1 + \sum_a \sum_i |\Delta\phi_i^a\rangle\langle\tilde{p}_i^a|, \tag{3.1}$$

where $\Delta\phi_i^a$ and $\tilde{p}_i^a$ are atom specific but system independent functions which are only non-zero inside the augmentation sphere.

$\Delta\phi_i^a = \phi_i^a - \tilde{\phi}_i^a$ is formed as the difference between a basis set well suited for expansion of ψ and $\tilde{\psi}$ respectively. To ensure a vanishing $\Delta\phi_i^a$ outside the sphere, these partial waves must be identical beyond Ω^a, but individually, they do not need finite support. Thus ϕ_i^a is usually chosen as solutions (possibly unbound) of the atomic KS equation. The choice of $\tilde{\phi}_i^a$ dictates the behavior of the solution $\tilde{\psi}$ inside of Ω^a, and is therefore chosen as a smooth continuation of ϕ_i^a in this region. $\tilde{p}_i^a$ must be dual to $\tilde{\phi}_i^a$, but can within this constraint be optimized for rapid convergence of the expansion. In practice, these projector functions are chosen such that the PAW transformation is exact for the isolated atom despite the truncation of the expansion in (3.1).

Note that the partial waves $\Delta\phi_i^a$ are always only represented inside the augmentation sphere. They are therefore not required to be smooth, as they can be efficiently represented using spherical harmonics and fine radial grids. The projector functions however must be smooth, as they are used to form the inner product with $\tilde{\psi}_n$ and must therefore be representable on the uniform real-space mesh.

Although not native to PAW, the frozen-core approximation is usually applied, in which the core electrons are fixed to the frozen unpolarizable orbitals of the free atom. The KS equations are then solved for the valence states only, and consequently the PAW transformation is only applied to these.

From (3.1) it can be seen that in PAW the KS equations are re-expressed and solved in terms of $\tilde{\psi}_n$, and the expansion coefficients of these in the atomic basis set

$$P_{ni}^a = \langle\tilde{p}_i^a|\tilde{\psi}_n\rangle. \tag{3.2}$$

I will here sketch the implications of the PAW transformation on the ground state problem, as this is instructive for the developments in the following sections. For an in-depth description of PAW, see [94, 95, 115] or Paper IV.

In terms of the all-electron wave functions ψ_n and density $n(\boldsymbol{r}) = \sum_n f_n|\psi_n(\boldsymbol{r})|^2$, the KS energy functional reads

$$E[n] = T_s[\{\psi_n\}] + U'_H[\rho] + E_{xc}[n]$$

where $\rho(\boldsymbol{r}) = n(\boldsymbol{r}) + \sum_a Z^a(\boldsymbol{r})$ is the total charge density of electrons plus nuclear charges $Z^a(\boldsymbol{r}) = -\mathcal{Z}^a\delta(\boldsymbol{r} - \boldsymbol{R}^a)$, with $\mathcal{Z}^a$ being the atomic number of the nucleus. The prime on the Hartree functional indicates that one should remember to remove the self-interaction error of the nuclear charges introduced by expressing the total energy in this way. The KS equations then follow by minimizing E with respect to $\{\psi_n\}$ under the constraint $\langle\psi_n|\psi_m\rangle = \delta_{nm}$ which yields a set of equations of the

form $\widehat{H}\psi_n(\boldsymbol{r}) = \epsilon_n\psi_n(\boldsymbol{r})$, where ϵ_n are the Lagrange multipliers used to keep the orbitals normalized, and $\widehat{H} = \partial E/f_n\psi_n\partial\psi_n^*$ is the Hamiltonian

$$\widehat{H} = -\tfrac{1}{2}\nabla^2 + u_H[\rho] + v_{xc}[n]. \tag{3.3}$$

The PAW expressions follow along the same lines but with the replacement $\psi_n = \hat{\tau}\tilde{\psi}$. There are however a few technicalities that I would like to highlight.

Re-expressing the KS energy functional results in

$$E = T_s[\{\tilde{\psi}_n\}] + U_H[\tilde{\rho}] + E_{xc}[\tilde{n}] + \sum_a \Delta E^a[\{D^a_{ij}\}]. \tag{3.4}$$

The arguments of the kinetic and xc functionals are simply replaced with the pseudo waves and a pseudo electron density $\tilde{n} = \sum_n f_n|\tilde{\psi}_n(\boldsymbol{r})|^2 + \tilde{n}_c(\boldsymbol{r})$, where the last term is just a smoothened version of the core electron density. The Hartree functional contains a pseudized total density $\tilde{\rho}(\boldsymbol{r}) = \tilde{n}(\boldsymbol{r}) + \sum_a \tilde{Z}^a(\boldsymbol{r})$, but due to the non-locality of the Hartree functional, direct insertion of the PAW transformation leads to cross terms between different augmentation spheres. To avoid these, the term $\tilde{Z}^a(\boldsymbol{r})$ contains not only a smoothened version of the nuclear charge but also a set of state dependent compensation charges designed such that the mentioned cross terms disappear. The last term of eq. (3.4), $\sum_a \Delta E^a$, is a sum of atomic corrections for the use of pseudo quantities in the first terms, as well as a correction due to the introduced compensation charges and a correction for the nuclear self-interaction. These atomic corrections as well as the compensation charges depend on the instantaneous state of the system via the atomic density matrix

$$D^a_{ij} = \sum_n P^{a*}_{ni} f_n P^a_{nj}. \tag{3.5}$$

In the PAW transformation, pseudo waves are only orthogonal with respect to the metric operator $\hat{S} = \hat{\tau}^\dagger\hat{\tau}$, which implies that the PAW KS equations become a generalized eigenvalue problem $\widehat{\widetilde{H}}\tilde{\psi}_n(\boldsymbol{r}) = \epsilon_n\hat{S}\tilde{\psi}_n(\boldsymbol{r})$, where the PAW Hamiltonian follows from the derivative $\widehat{\widetilde{H}} = \partial E/f_n\tilde{\psi}_n\partial\tilde{\psi}_n^*$

$$\widehat{\widetilde{H}} = -\tfrac{1}{2}\nabla^2 + u_H[\tilde{\rho}] + v_{xc}[\tilde{n}] + \sum_a\sum_{ij}|\tilde{p}^a_i\rangle\Delta H^a_{ij}\langle\tilde{p}^a_j|. \tag{3.6}$$

The non-local atomic part of the Hamiltonian

$$\Delta H^a_{ij} = \int \mathrm{d}\boldsymbol{r}\, u_H[\tilde{\rho}](\boldsymbol{r})\frac{\delta\tilde{Z}^a(\boldsymbol{r})}{\delta D^a_{ij}} + \frac{\delta\Delta E^a}{\delta D^a_{ij}} \tag{3.7}$$

includes both a term due to the state dependence of the compensation charges and of the atomic energy corrections.

The explicit expression for ΔE^a, ΔH^a_{ij}, and $\tilde{Z}^a[\{D^a_{ij}\}](\boldsymbol{r})$ can be found in Paper IV (which corrects the erroneous corresponding terms in the appendix of [87]).

In passing I note that due to the dependence of the augmentation on the ion coordinates, the force expression in PAW

$$\boldsymbol{F}^a = -\frac{dE}{d\boldsymbol{R}^a} = -\frac{\partial E}{\partial\boldsymbol{R}^a} + \sum_n f_n\epsilon_n\langle\tilde{\psi}_n|\frac{d\hat{S}}{d\boldsymbol{R}^a}|\tilde{\psi}_n\rangle \tag{3.8}$$

has contributions both from the shape functions defining the compensation charges (in the pseudo Hartree energy), the pseudo core densities (in both pseudo Hartree and xc energies), and the projector functions (in both atomic corrections to the energy and the metric operator $\hat{S}$). The term related to the compensation charges also relays the force from the lattice potential.

The Vanderbilt type of ultra-soft pseudopotentials[96] (used in my DACAPO calculations) can be obtained by a linearization of the dependence of the atomic energy corrections on D^a_{ij} around the density matrix of the reference atom[115] and is therefore no longer an exact transformation. Norm-conserving Troullier-Martins [97, 116] type pseudopotentials (used in my SIESTA calculations) follow from the further restriction that the pseudo partial waves should have the same norm as the AE partial waves. Both ultra-soft and norm-conserving pseudo potentials are more difficult to construct than PAW setups, and both suffer from larger transferability problems. The norm-conserving restriction generally leads to very hard pseudopotentials for elements with strongly localized orbitals like the late transition metals, although the Troullier-Martins flavor are usually somewhat softer than the originally suggested scheme by [98].

3.2 Extracting Quantities From PAW

Obviously it is a convenient property of PAW that one is able to make an accurate representation of the all-electron valence states without the need of introducing a pseudopotential to describe an effective core potential. The transformation operator does however complicate the bookkeeping when one needs to extend the scheme or extract quantities for post processing purposes.

This section provides a few examples of accommodating expressions to the PAW formalism, that I have found useful, and implemented in GPAW. Focus is restricted to quantities used in the various applications presented here and in other chapters. These will then serve to exemplify the procedure of conforming to the PAW framework.

Applications generally fall into two categories: Modifications to the standard KS SCF cycle and post processing on the converged KS states. The former includes all self-consistent treatment of e.g. external potentials, non-local exact exchange, constrained DFT, LDA + U calculations, etc. Post processing includes extraction of all-electron matrix elements like the Hamiltonian and overlap matrices (for e.g. transport calculations), the xc-potential and Coulomb operator (for GW and lrTDDFT), the spread operator (for maximally localized Wannier functions), projections of eigenstates onto a basis set (for projected Wannier functions) and the dipole vector (valence-valence for lrTDDFT and valence-core for XAS calculations). For other post processing purposes it is not always enough simply to extract matrix elements, but one might need to reconstruct all-electron quantities directly on a real-space grid. This could for example be the all-electron density.

In connection with analysis, an additional feature of the PAW formalism is the inherent partitioning of space into atomic regions, which supplies a direct way to probe the local atomic structure via quantities already calculated and used in the scheme. Of special interest are the projector overlaps P^a_{ni}, eq. (3.2), which are by construction simultaneous expansion coefficients of both the pseudo and the all-electron wave functions inside the augmentation spheres in the set of partial waves. They can thereby be used to probe the local orbital character of the full

wave functions. For example I define an orbital projected density of states by

$$n_i(\varepsilon) = \sum_n \delta(\varepsilon - \epsilon_n) \left| P_{ni}^a \right|^2 , \tag{3.9}$$

which does not require any extra projections, as P_{ni}^a are directly available. It can be shown that $P_{ni}^a = \langle \tilde{p}_i^a | \tilde{\psi}_n \rangle \approx \langle \phi_i^a | \psi_n \rangle$ (see Paper IV). The projector overlaps also define the atomic density matrix D_{ij}^a, eq. (3.5), which represent the full AE density within the augmentation sphere and can be used for e.g. Mulliken charge analysis, to define local magnetic moments, or to make LDA + U [117, 118] corrections to the SCF cycle.

All examples given here are implemented in GPAW and described in Paper VI or Paper IV. The following is a more detailed account of a selection of these examples, used directly in applications presented here or in later chapters.

3.2.1 External Potentials

As the simplest possible example of modifying the SCF, consider the application of an external potential in DFT.

Without the PAW transformation, this addition is trivial, as the desired potential, $v_{\text{ext}}(\boldsymbol{r})$, should simply be added to the effective KS potential, and the total energy adjusted by the energy associated with the external potential $E_{\text{ext}} = \int \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) v_{\text{ext}}(\boldsymbol{r})$.

To conform to the PAW expressions, the total energy is rewritten in terms of pseudo quantities, according to

$$E_{\text{ext}} = \int \mathrm{d}\boldsymbol{r} \tilde{n}(\boldsymbol{r}) v_{\text{ext}}(\boldsymbol{r}) + \sum_a \left[\int \mathrm{d}\boldsymbol{r} v_{\text{ext}}(\boldsymbol{r}) \left[n_c^a(\boldsymbol{r}) - \tilde{n}_c^a(\boldsymbol{r}) \right] + \sum_{ij} D_{ij}^a \Delta V_{ij}^{a,\text{ext}} \right] ,$$

where

$$\Delta V_{ij}^{a,\text{ext}} = \int \mathrm{d}\boldsymbol{r} v_{\text{ext}}(\boldsymbol{r}) \left\{ \phi_i^a(\boldsymbol{r}) \phi_j^a(\boldsymbol{r}) - \tilde{\phi}_i^a(\boldsymbol{r}) \tilde{\phi}_j^a(\boldsymbol{r}) \right\} . \tag{3.10}$$

The contribution to the PAW Hamiltonian from v_{ext} then follows from the derivative with respect to $\tilde{\psi}_n^*$ which yields

$$\widetilde{H}_{\text{ext}}(\boldsymbol{r}) = v_{\text{ext}}(\boldsymbol{r}) + \sum_a \sum_{ij} |\tilde{p}_i^a\rangle \Delta V_{ij}^{a,\text{ext}} \langle \tilde{p}_j^a| .$$

To evaluate the atomic corrections to both Hamiltonian and energy, the inner product with all of the partial waves must be determined. This computational inconvenience can be avoided by assuming slow variation over the extent of the augmentation sphere, and setting $v_{\text{ext}}(\boldsymbol{r}) \approx v_{\text{ext}}(\boldsymbol{R}^a)$. This simplifies the atomic correction to $\Delta V_{ij}^{a,\text{ext}} \approx v_{\text{ext}}(\boldsymbol{R}^a) \Delta S_{ij}^a$ and also the term related to the core densities reduce to already calculated quantities.

The approximation is exact for piecewise constant potentials, which is what I have used the implementation for. An example of the application of such potentials to study image charge formation is given in section 3.4.

The implementation has later been generalized to first order Taylor expansions within each augmentation sphere, such that linear external potentials (corresponding to a homogeneous applied electric field) can be handled exactly. This has for example been used in GPAW by [119] to study the field effects in STM simulations. Linear fields can also be used in TDDFT calculations to drive an electron current.

3.2.2 All-electron Density

In PAW, the all-electron density is expressed by

$$n(\boldsymbol{r}) = \tilde{n}(\boldsymbol{r}) + \sum_a \left[n_c^a(\boldsymbol{r}) + \sum_{ij} D_{ij}^a \left(\phi_i^a(\boldsymbol{r})\phi_j^a(\boldsymbol{r}) - \tilde{\phi}_i^a(\boldsymbol{r})\tilde{\phi}_j^a(\boldsymbol{r}) \right) \right].$$

In actual PAW calculations all-electron quantities are at all times available in principle, but are in practice never handled directly. Instead they are expressed in the composite basis representation of a global pseudo description augmented by local atomic basis functions, as this offers the most accurate description of the core region.

For post processing purposes I find that it is sometimes useful to reconstruct the all-electron quantities on a single regular grid, despite the inaccurate description in the core region.

To minimize the effect of an insufficient grid description the density is interpolated to an extra fine grid, and the summed partial waves are then mapped onto this grid. Usually an 2^3 times finer grid than the regular wave function grid can easily be afforded, as the array is only used for post-processing and only for a single array (as opposed to the large number of wave functions used in every iteration).

My main use of the reconstructed all-electron density is for the application of Bader analysis [120]. Application of this partitioning scheme to the pseudo density is problematic. As the pseudo density can feature local maxima not associated with an atom (this is not possible for the physical electron density), the positioning of the dividing surfaces will be wrong if these intersect the augmentation sphere, and the integrated charge within each partition does not reflect the number of electrons. These artifacts are mostly cured by using the reconstructed all-electron density. In practice the insufficient grid in the core region can cause inaccuracies in the integrated value within the augmentation sphere, especially in the presence of heavy chemical elements. From the accurate atomic orbital description of the density corrections, the error induced by putting them on a regular grid is however known, and a practical fix is to simply adjust the value of the density at the grid point closest to each nucleus such that the corrections integrate correctly. The procedure is shown on the left of figure 3.1. This fix will never influence the determination of the dividing surfaces, and in practice it will also never cause a spurious local minimum, as the adjustment is minute compared to the exponential nature of the core density.

The right panel of figure 3.1 shows the dividing surfaces of a water molecule found using the Bader program of [121]. In this case, more than half an electron (0.54) is transferred from each hydrogen to oxygen, which causes the saddle node plane of the partitioning to move very close to, and well within the augmentation sphere of the hydrogen atoms.

I apply this scheme of reconstructing the AE density plus Bader analysis to study image charge formation in section 3.4.

3.2.3 Calculating Coulomb Matrix Elements

When trying to describe electron interactions beyond the level of standard (semi-) local density approximations, one will often need Coulomb matrix elements of the

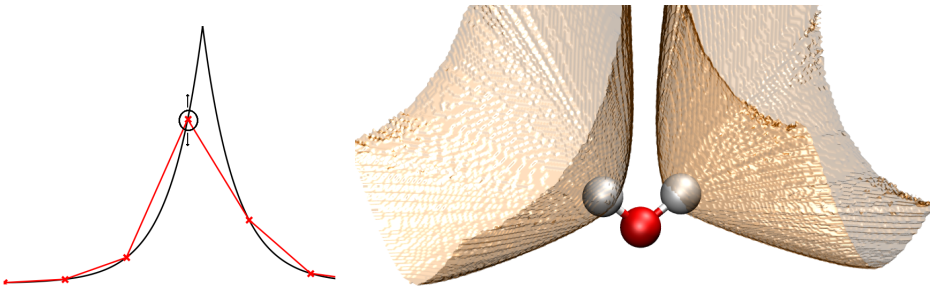


Figure 3.1: *Left*: Adjusting the central grid point. *Right*: The dividing surfaces of the all-electron density of a water molecule, as determined by a Bader partitioning scheme.

type

$$V_{nn',mm'} = (n_{nn'}|n_{mm'}) := \iint \frac{d\boldsymbol{r}d\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|} n^*_{nn'}(\boldsymbol{r}) n_{mm'}(\boldsymbol{r}'), \tag{3.11}$$

where the orbital pair density $n_{nn'}(\boldsymbol{r}) = \psi^*_n(\boldsymbol{r})\psi_{n'}(\boldsymbol{r})$.

Such elements are needed in some formulations of vdW functionals (although not the one implemented in GPAW), in linear-response TDDFT (see sec. 3.2.6, or Paper I) where only pair densities corresponding to electron-hole pairs are needed, in exact exchange or hybrid functionals (see next section) where only elements of the form $V_{nn',nn'}$ with both indices correspond to occupied states, are needed, and for GW calculations (see chap. 4, or Paper II), where all elements are needed.

Introducing the PAW transformation in (3.11), the pair densities partition according to

$$n_{nn'}(\boldsymbol{r}) = \tilde{n}_{nn'}(\boldsymbol{r}) + \sum_a \left(n^a_{nn'}(\boldsymbol{r}) - \tilde{n}^a_{nn'}(\boldsymbol{r})\right) \tag{3.12}$$

with the pseudo and one-center pair densities defined by

$$\tilde{n}_{nn'} = \tilde{\psi}^*_n \tilde{\psi}_{n'}, \quad n^a_{nn'} = \sum_{ij} P^{a*}_{ni} P^a_{n'j} \phi^a_i \phi^{a*}_j, \quad \tilde{n}^a_{nn'} = \sum_{ij} P^{a*}_{ni} P^a_{n'j} \tilde{\phi}^a_i \tilde{\phi}^{a*}_j. \tag{3.13}$$

Similar to the problem with the Hartree potential in the standard PAW scheme, direct insertion of the partitioned pairdensity (3.12) in the integral (3.11) would, due to the non-local nature of the Coulomb kernel, lead to undesired cross terms between different augmentation spheres and between quantities represented on incompatible grids, i.e. the pseudo pair density on the regular grid, and the one-center expansions on the radial support grid. Such terms can be avoided by introducing some compensation charges, $\tilde{Z}^a_{nn'}$, chosen such that the potential of $n^a_{nn'} - \tilde{n}^a_{nn'} - \tilde{Z}^a_{nn'}$ is zero outside the respective augmentation sphere of atom a. This is achieved by doing a multipole expansion of the summed terms in a set of suitable functions $\tilde{g}^a_L(\boldsymbol{r})$ and requiring all expansion coefficients (multipole moments) to be zero, and entails a compensation of the form

$$\tilde{Z}^a_{nn'}(\boldsymbol{r}) = \sum_L \tilde{g}^a_L(\boldsymbol{r}) \sum_{ij} \Delta^a_{L,ij} P^{a*}_{ni} P^a_{n'j}, \tag{3.14}$$

which is very similar to the regular compensation charges in PAW

$$\tilde{Z}^a(\boldsymbol{r}) = \sum_L \tilde{g}_L^a(\boldsymbol{r}) \left(\Delta^a \delta_{l,0} + \sum_{ij} \Delta_{Lij}^a D_{ij}^a \right). \tag{3.15}$$

The main differenc is that the standard compensation charges have an additional monopole term which corrects for the core densities. The remaining constants $\Delta_{L,ij}^a$ are identical but the state dependent coefficients $P_{ni}^{a*} P_{n'j}^a$ for the pair-density compensation charges differ from the coefficients D_{ij}^a (the atomic density matrix) of the total density compensation charges.

Introduction of such compensation charges makes it possible to obtain the clean partitioning

$$V_{nn',mm'} = (\tilde{\rho}_{nn'}|\tilde{\rho}_{mm'}) + 2 \sum_a \sum_{i_1 i_2 i_3 i_4} P_{mi_1}^a P_{ni_2}^{a*} \Delta C_{i_1 i_2 i_3 i_4}^a P_{n'i_3}^{a*} P_{m'i_4}^a. \tag{3.16}$$

Here the last term is a trivial functional of the expansion coefficients P_{ni}^a involving only the constants $\Delta C_{i_1 i_2 i_3 i_4}^a$ which are again identical to corrections used in the expression for the Hartree potential and thus already precalculated (see Paper IV for a derivation and the definition of C^a). The only computationally demanding term relates to the Coulomb matrix element of the smooth compensated pair densities $\tilde{\rho}_{ij} = \tilde{n}_{ij} + \sum_a \tilde{Z}_{ij}^a$, which are expressible on coarse grids.

The formally exact partitioning (3.16) makes it possible, at moderate computational effort, to obtain Coulomb matrix elements in a representation approaching the complete basis set limit. Such elements are easily evaluated for atomic orbital based implementations, where the analytic form of the basis functions can be exploited, but such representations are not easily converged to accurate numbers. Compared to standard pseudopotential schemes, it is important to note that in the PAW decomposition, all information on the nodal structure of the all-electron wave functions in the core region is retained, which is important due the non-local probing of the Coulomb operator.

Integration over the the Coulomb kernel $1/|\boldsymbol{r}-\boldsymbol{r}'|$ is done by solving the associated Poisson equation, i.e.

$$(\tilde{\rho}_{nn'}|\tilde{\rho}_{mm'}) = \int \mathrm{d}\boldsymbol{r} \tilde{\rho}_{nn'}(\boldsymbol{r}) \tilde{v}_{mm'}(\boldsymbol{r}), \quad \nabla^2 \tilde{v}_{mm'}(\boldsymbol{r}) = -4\pi \tilde{\rho}_{mm'}(\boldsymbol{r}). \tag{3.17}$$

The Poisson equation can be solved in reciprocal space, using FFTs to transform between real- and reciprocal space, but I usually stick to solving the equation directly in real-space using the efficient multi-grid solver implemented in GPAW (see e.g. [87]). This allows for an efficient parallelization scheme using domain decomposition, which requires little communication across processors due to the semi-local structure of the Laplacian.

Note that the pair-density is not a density, and can be complex valued in general. For real-valued wave functions, it can still have negative values. Another important difference compared to the evaluation of the Hartree energy is, that while $\tilde{\rho}(\boldsymbol{r})$ integrates to zero for neutral systems, the integral $\int \mathrm{d}\boldsymbol{r} \tilde{\rho}_{nn'}(\boldsymbol{r}) = \delta_{nn'}$ shows that the compensated pair densities $\tilde{\rho}_{nn}$ have a non-zero total charge.

A total charge is problematic for the determination of the associated potential, and is handled differently for different boundary conditions. For periodic systems, charge neutrality is enforced by subtracting a homogeneous background charge, and

the error so introduced is removed to leading order ($V^{-1/3}$ where V is the volume of the simulation box) by adding the potential of a missing probe charge in an otherwise periodically repeated array of probe charges embedded in a compensating homogeneous background charge[1]. This can be determined using the standard Ewald technique, and corresponds to a rigid shift of the potential. The procedure is identical to that suggested by [124].

For non-periodic, isolated, systems a different strategy is used. As the charge is known to have a well-defined center in this case, and that this center will not be close to the boundaries of the simulation box, a multipole expansion of the pair density with respect to the center of this box is performed. The Poisson equation can then be solved subject to the boundary values dictated by the multipole expansion. In practice a monopole correction is used. This is correct to the same order as the correction for periodic cells, but the prefactor on the error is much smaller, and leads to converged potentials even for small cells.

This implementation of calculating Coulomb matrix elements is used extensively in the work presented here, both in GW calculations (chapter 4), lrTDDFT (sec. 3.2.6), and for exact exchange and hybrid functionals (sec. 3.3).

An added benefit of the implemented scheme to determine potentials of charged densities presented above, is that it opens for the possibility of doing calculations on charged systems. The only additional modification is that the occupation statistics should be changed to satisfy $\sum f_n = N - C$ where C is the charge (remember that the chemical potential μ is a Lagrange multiplier used to enforce that the density integrates to the total electron number). The energetics of charged molecules are studied in section 3.3 to estimate ionization potentials from total energy differences.

A detail on the use of the above method to study charged periodic systems is that the correcting potential shift in this case should be scaled down by the (static macroscopic) dielectric function of the crystal. This should not be done when using the method to calculate matrix elements of the bare Coulomb repulsion.

3.2.4 Exact Exchange Energy

Expressed in the KS eigenstates (characterized by diagonalizing the density matrix), the exact exchange energy functional involves only two-index Coulomb elements

$$E_{\mathrm{xx}} = -\frac{1}{2}\sum_{nm} f_n f_m \delta_{\sigma_n,\sigma_m} V_{nm,nm}. \tag{3.18}$$

Rewriting this in terms of pseudo waves and atomic corrections follows directly from eq. (3.16), for the terms in the sum where n and m both refer to valence states. Terms where either index refers to a core orbital can be reduced to trivial functionals of the projector overlaps P^a_{ni}, as also shown in [124]. The final expression reads

$$E_{\mathrm{xx}} = -\frac{1}{2}\sum_{nm}^{\mathrm{val}} f_n f_m \delta_{\sigma_n,\sigma_m} (\tilde{\rho}_{nm}|\tilde{\rho}_{nm}) - \sum_a \Delta E^a_{\mathrm{xx}}[\{D^a_{ij}\}]. \tag{3.19}$$

The only computationally demanding term is the first, which is the exchange energy of the compensated pair-orbitals. The atomic corrections

$$\Delta E^a_{\mathrm{xx}} = \sum_\sigma \sum_{i_1 i_2 i_3 i_4} D^{a*}_{i_1 i_3}(\sigma) \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_2 i_4}(\sigma) + \sum_{i_1 i_2} D^a_{i_1 i_2} X^a_{i_1 i_2} + E^{a,\text{c-c}}_{\mathrm{xx}} \tag{3.20}$$

[1] The rigid potential shift is given by $E = \sum_{\boldsymbol{R}\neq(0,0,0)} \frac{1}{|\boldsymbol{R}|} - \int_V \mathrm{d}\boldsymbol{r} \frac{1}{|\boldsymbol{r}|}$.

divide into a correction for the pseudization and compensation charges in the valence-valence interaction (proportional to ΔC^a), a term describing the valence-core exchange interaction energy (proportional to X^a), and lastly the (constant) exchange energy of the core electrons $E_{\mathrm{xx}}^{a,\text{c-c}}$.

The orbital interaction of valence and core thus comes out directly from the PAW transformation, and can be accounted for by the simple product of the atomic density matrix with the system independent tensor

$$X^a_{i_1 i_2} = \frac{1}{2}\sum_{\alpha}^{\text{core}} \iint d\boldsymbol{r} d\boldsymbol{r}' \frac{\phi^a_{i_1}(\boldsymbol{r})\phi^{a,\text{core}}_{\alpha}(\boldsymbol{r})\phi^a_{i_2}(\boldsymbol{r}')\phi^{a,\text{core}}_{\alpha}(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{3.21}$$

Although the valence-core interaction is computationally trivial to include, it is not unimportant as will be shown in section 3.3. The inclusion stems from PAW being an all-electron method, and this kind of valence-core interaction would be difficult to estimate in pseudopotential schemes.

3.2.5 The Fock Operator

To include exact exchange in the self-consistent KS scheme, the form of the non-local Fock operator must first be determined. The Fock operator can then be used directly as described in sec. 2.2.2 within the generalized KS scheme as is the usual convention in Hybrid functionals, or it can subsequently be converted into a multiplicative potential via the OEP[79, 80] equation (or one of the approximations to this, LHF[77] and KLI[125]). Here, only the non-local Fock operator is used. For a generalization of the OEP equation to PAW, see Paper IV. The use of the non-local operator corresponds, if correlation is completely neglected, to the HF scheme and is equivalent to the GF based perturbation expansion of e-e interactions, if only the first order self-energy diagrams are included.

For the iterative minimization schemes used in real-space (and plane wave) codes, the explicit form of the non-local Fock operator $v^{\mathrm{NL}}(\boldsymbol{r},\boldsymbol{r}')$ is never needed, and would indeed be impossible to represent on any realistic grid. Instead only the action of the operator on a state is needed. As with the Hamiltonian operator, the action on the pseudo waves is derived via the relation $f_n\hat{v}^{\mathrm{NL}}|\tilde{\psi}_n\rangle = \partial E_{\mathrm{xx}}/\partial\langle\tilde{\psi}_n|$, which yields

$$\hat{v}^{\mathrm{NL}}|\tilde{\psi}_n\rangle = \sum_m \delta_{\sigma_n,\sigma_m} f_m \tilde{v}_{nm}(\boldsymbol{r})|\tilde{\psi}_m\rangle + \sum_a\sum_i |\tilde{p}^a_i\rangle \Delta V^a_{ni}, \tag{3.22}$$

where $\tilde{v}_{nm}$ is the solution of $\nabla^2\tilde{v}_{nm}(\boldsymbol{r}) = -4\pi\tilde{\rho}_{nm}(\boldsymbol{r})$, and the Fock atomic corrections

$$\Delta V^a_{ni} = \sum_m \delta_{\sigma_n,\sigma_m} f_m \int \mathrm{d}\boldsymbol{r}'\tilde{v}_{nm}(\boldsymbol{r}')\frac{\partial\tilde{Z}^a_{nm}(\boldsymbol{r}')}{\partial P^{a*}_{ni}} + \sum_j \frac{\partial\Delta E^a_{\mathrm{xx}}}{\partial D^a_{ij}(\sigma_n)}P^a_{nj}, \tag{3.23}$$

contain both a term related to the state dependence of the introduced compensation charges and of the total energy corrections.

The expressions (3.22) and (3.23) are equivalent to those derived by [124].

The computational bottleneck is to determine the potential of all pair-orbitals $\tilde{v}_{nm}(\boldsymbol{r})$, which makes the exact-exchange potential relatively costly compared to standard local xc approximations. It is however important to note that the only computationally demanding step is related to pseudized quantities, which can be

accurately represented on coarse grids. This makes it possible to do basis set converged self-consistent EXX calculations at a relatively modest cost. Without the PAW transformation a real-space based HF implementation would be impossible.

The need to evaluate potentials for each pair of orbitals makes the computational cost scale quadratically with the number of bands in the calculation, whereas KS-DFT calculations based on local xc approximations scale considerably better for small systems. For very large systems, even standard KS scale as N^3 due to the subdiagonalization of the Hamiltonian within the basis of trial wave functions, but this term only dominates for systems with more than a few hundred atoms. In practice I find EXX based calculations to be a about 10-60 times slower than GGA based calculations for molecules in the range 1-30 atoms. Exactly like for local xc based KS-DFT calculations, exploitation of orbital locality in real-space e.g. by transforming to Wannier functions can be used to make linear-scaling algorithms for exact-exchange in large extended system [126, 127]. In atomic orbital basis sets, construction of Wannier functions is not needed, and HF has been applied to systems containing hundreds of atoms [128].

As a technical consideration, it should be mentioned that the augmentation sphere dependence in both the compensation charges of the pair-densities and in the atomic corrections leads to contributions to the Hellmann-Feynman forces. See Paper IV for an explicit expression, and a derivation, of these.

3.2.6 Linear Response Time-dependent DFT

The spectrum of HF calculations reflects the spectrum for charged excitations in the approximation of no orbital relaxation, where charged excitations mean electron addition and removal. These are also the type of excitations described by GW calculations and those relevant for electron transport, where the current flow induces charge fluctuations on the molecule during electron hopping events.

A different kind of excitation is the neutral excitations usually probed by optical spectroscopy, where electrons are excited from the ground state to an excited state. In the Green function framework such excitations are described by the Bethe-Salpeter equation. This kind of excitation can however also be treated within the time-dependent DFT [129] framework.

One approach is to study the linear response of the KS system to a time-dependent perturbation of the potential, and use this to estimate the spectrum of possible optical excitations.

In the linear response regime, and using the adiabatic approximation for the xc-potential of TDDFT, [130, 131, 132] showed that the poles of the density response function, and hence the spectrum of allowed excitations, is equivalent to solving the algebraic eigenvalue problem

$$\mathbf{\Omega} F_I = \omega_I^2 F_I. \tag{3.24}$$

The eigenvalues of $\mathbf{\Omega}$ determine the transition energies ω_I from the ground state to the excited state I, and the eigenvectors F_I can be related to the corresponding dipole oscillator strengths.

The basis of the matrix and eigenvectors in (3.24) is the set of same-spin electron-hole pair orbitals $n_{ij\sigma}$, i.e. i and j are band indices referring to an occupied and an unoccupied state respectively, and σ denotes the spin projection index. The elements of $\mathbf{\Omega}$ are then

$$\Omega_{ij\sigma,kq\tau} = \delta_{ij\sigma,kq\tau}\varepsilon_{ij\sigma}^2 + 2\sqrt{\varepsilon_{ij\sigma}\varepsilon_{kq\tau}}K_{ij\sigma,kq\tau}, \tag{3.25}$$

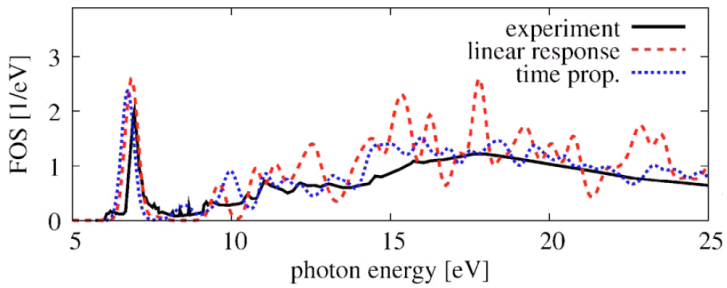


Figure 3.2: Calculated optical absorbtion spectrum of benzene using linear response and time-propagation TDDFT compared to experiment. From [122].

where $\varepsilon_{ij\sigma} = \varepsilon_{j\sigma} - \varepsilon_{i\sigma}$ are the KS eigenenergy differences, and the coupling matrix

$$K_{ij\sigma,kq\tau} = \langle n_{ij\sigma}|1/|\boldsymbol{r} - \boldsymbol{r}'| + \delta v_{xc}(\boldsymbol{r})/\delta n(\boldsymbol{r}')|n_{kq\tau}\rangle \tag{3.26}$$

describes the linear response of the Hartree and xc potentials to the perturbation. The first term is the Coulomb matrix element $V_{ij\sigma,kq\tau}$, which is obtained in PAW via eq. (3.11). The second term can for local xc functionals be determined by a finite difference scheme and partitions due to the locality into pseudo and atomic correction terms in the usual way, see Paper I for details.

The time limiting step in the Casida equation (3.24) is the determination of the Coulomb matrix elements. Only pair orbitals corresponding to electron/hole combinations are needed, and efficient algorithms exist for extracting just the lowest transitions [133]. Still, the procedure generally requires a large number of unoccupied states to converge the (lowest) excitation energies.

Both the linear response and time-propagation TDDFT schemes have been implemented in GPAW. The implementation and several applications are described in Paper I. In this, also a time-propagation TDDFT implementation is demonstrated. The time-propagation scheme can in general also be used to study non-linear phenomena, and the linear response can be extracted by applying a weak perturbation in the form of a delta pulse in the dipole field and study the following free propagation. A comparison of the optical absorption spectrum predicted for the standard benchmark molecule benzene is shown on figure 3.2, which also shows the experimental spectrum. The excitation energies found by diagonalization of the Ω matrix in the Casida equation have here been folded with Gaussians to give a continuous spectrum. The weight of each excitation is determined via the eigenvectors, and a common smearing width of 0.2 eV for the Gaussians has been used.

The agreement between the two very different implementations is striking. Also the agreement with experiment is good, both implementations reproducing the spectral peak at 6.9 and the overall shape of the broad feature from 10 eV and up. Both positioning and relative strength of the peaks are reproduced.

3.3 Application of Exact Exchange

Although the explicit expression for the exact exchange energy in DFT is well known, it is rarely used. The reason is in part that the expression is only an

implicit functional of the density via an explicit orbital dependence which makes it much more computationally demanding than conventional density-functional approximations, and in part because the full inclusion of exact exchange actually degrades the quality of the estimated total energy. The second issue is due to the lack of accurate correlation energy functionals that are compatible with exact exchange. The *fractional* inclusion of EXX in local xc approximations can however be used to improve atomization energies. Energetics are however already well-described by local xc approximations.

This section will focus on the application of exact exchange to molecules. In part to benchmark the implementation, and in part to investigate the effects of an exact treatment of exchange on energetics and spectral properties of the KS system.

3.3.1 Benchmarking and Atomization Energies

The implementation described here is probably the only full-potential real-space implementation of exact-exchange, but there exists a fully equivalent PAW implementation in the VASP code [134] based on a plane wave basis set. A plane wave basis can be systematically improved towards a complete representation of space in an equivalent fashion to reducing the grid spacing in a real-space code, and as both are PAW implementations of DFT, they should be directly comparable.

In [124], the VASP EXX implementation is benchmarked by calculating the heat of formation of a large test-set of small molecules (the G2-1 test set [135]) and comparing to accurate numbers determined using the GAUSSIAN03 code. While GPAW and VASP provide an accurate representation of the valence wave functions, but assume frozen core orbitals and employs a truncated PAW expansion to describe the valence states, GAUSSIAN03 is an all-electron code using no pseudization schemes, but on the other hand employs a local atomic orbital basis set which can not be systematically converged.

In figure 3.3 atomization energies obtained using the present implementation in GPAW (blue), are compared with those reported in [124] for VASP (green) and GAUSSIAN03 (red). The energies are calculated using the GGA type PBE[75] xc-functional (dotted with squares) and the hybrid PBE0[82] functional (solid with triangles) which includes a 25% fraction of exact exchange. The atomization energy is the negative dissociation energy, i.e. the energy required to split the molecule, implying that it is positive for stable molecules. Atomization energies are reported in kcal/mol (=43meV).

In the top panel of figure 3.3, the calculated energies are plotted relative to experimental values derived from measured heats of formation extrapolated to zero temperature and corrected for zero-point energies using calculated vibration frequencies following [135, 136]. Note that both measurement, extrapolation, and estimation of zero-point energies induce some variation in reported experimental values. Comparing the experimental values reported in [124] to numbers from [137] I find a mean absolute variation of 1.3 kcal = 55meV and a max deviation of 0.17eV. From the plot, it is seen that the inclusion of exact exchange in the PBE functional markedly improves the predicted atomization energies from a mean absolute error relative to experiment of 8.6kcal/mol to 3.7 when going from PBE to PBE0, which brings the predictions very close to the limit of experimental accuracy.

On the scale of the top panel in figure 3.3, all three codes gives numbers almost on top of each other. Focusing on the differences between the codes, the lower panel shows the calculated energies from GPAW and GAUSSIAN03 relative to the numbers

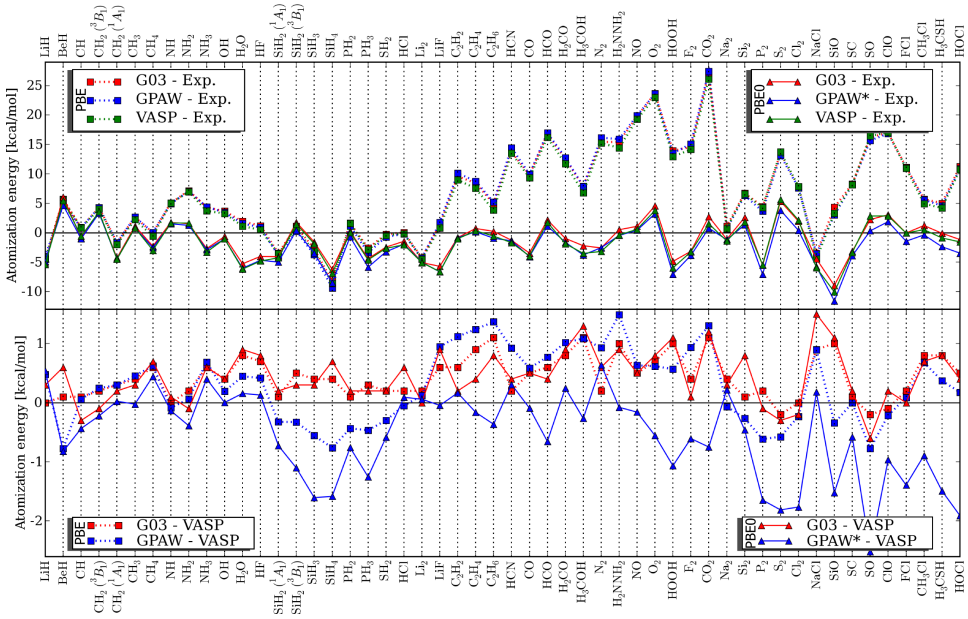


Figure 3.3: Calculated atomization energies. Top: calculated energies relative to experiment, bottom: calculated energies relative to each other. The asterisk, *, indicates PBE0 values evaluated at PBE geometries. All other numbers have been fully structurally relaxed.

from VASP. For comparison of the codes, it is relevant to first consider the accuracy of the calculations.

[124] report that they have converged the size of the basis sets in both GAUSSIAN03 and VASP to within at most 10meV by using an aug-cc-pV5Z basis set and an energy cutoff of 1000eV in the two codes respectively. In VASP care was also taken to increase the accuracy of the PAW data set which implies choosing small augmentation spheres and extended partial wave expansions. The geometries used in GAUSSIAN03 and VASP have been optimized for both PBE and PBE0 calculations. In my GPAW calculations I use a well converged grid spacing of 0.16Å, but a standard PAW data set and I have not done any geometry optimization for the PBE0 calculations, but simply used the PBE geometries.

The mean absolute deviation between the results from the three codes shown in the lower panel of fig.3.3 is less than 0.5 kcal/mol for PBE, which is actually quite impressive given the very different implementations. The PBE0 values from GPAW are generally too small. This is attributed to the lacking geometry optimization which would lower the energy of the molecules, and hence increase the atomization energy. There is actually a nice correspondence between those systems reported in [124] to have the largest restructuring between PBE and PBE0 to also be those with the worst PBE0 energies calculated by GPAW. Even with these small deviations, the variations between the three codes are much smaller than the level needed for so-called “chemical accuracy” of 1 kcal/mol.

From the excellent agreement with GAUSSIAN03 I conclude that the frozen core approximation used in the two PAW implementation does not have any significant impact on the energetics. The frozen core approximation is not a prerequisite in PAW

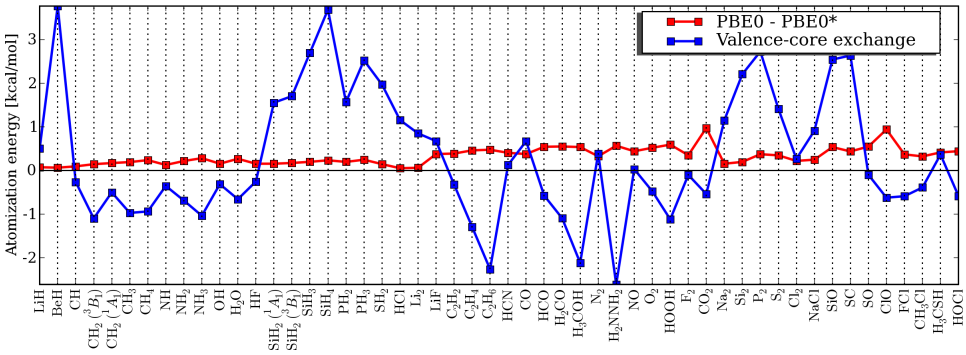


Figure 3.4: Importance of orbital self-consistency (red) and valence-core inclusion (blue) in calculated PBE0 atomization energies. The asterisk, *, indicates PBE0 energies evaluated non-selfconsistently using PBE orbitals.

and has been relaxed in the work done by [138] who came to the same conclusion; that this approximation is well-justified. The level of agreement with an all-electron code is only possible because PAW is a full potential code. It must be expected that orbital-dependent functionals, to a higher degree than density functionals, probe the qualitative nodal structure of the individual orbitals in the core region. Especially so for the Coulomb kernel which forms the basis of exact-exchange. When applying the PAW transformation to the EXX energy and potential, the resulting PAW expressions had the structure of the usual expressions applied to the (compensated) pseudo waves and in addition, a set of atomic corrections for the pseudization. These corrections can be split in three terms: 1) a term related to valence orbitals correcting for the difference between pseudo and AE waves in the augmentation spheres and the influence of the compensation charges. 2) a static energy shift related to the exchange energy of the core orbitals, and 3) a term describing the orbital interaction between valence and core states. The importance of this last term in the atomization energy is shown (in blue) in figure 3.4 for the entire G2-1 molecular database.

The effect of the valence-core exchange interaction for this test suite is seen to be several kcal/mol for the atomization energy. If one looks closer at the eigenvalue spectrum of the individual calculations (not shown here), it reveals that the valence-core interaction can induce a shift of several eV in the eigenvalues of the frontier orbitals. This shows that for the exchange process, it is not enough to consider in isolation the valence electrons, which interact with the core orbitals only via the effective potential generated by these, but that explicit orbital interactions between valence and core are important. Both for accurate atomization energies but especially for spectral properties. The effect is largest for the absolute positioning of the HOMO, while the gap is affected to a lesser degree, about 0.1 eV. Note that it has previously been concluded that the neglect of valence-core interaction in pseudopotential codes is a minor effect [139, 140] based on calculations of band gaps (they report an effect of 0.1 eV).

An interesting topic is the importance of self-consistency in the KS orbitals. The difference in atomization energy between PBE0 evaluated using PBE orbitals and self-consistent PBE0 orbitals is shown in red on figure 3.4. The effect of self-consistency is less than 0.3 kcal/mol in average suggesting that PBE and PBE0

orbitals are very similar. This is interesting as it allows accurate energetics to be determined by a simple energy evaluation using PBE0 xc following a self-consistent PBE calculation. This can save a lot of time. For the small molecules studied here, EXX based calculations are about 5 times slower per SCF than pure GGA type, which gets worse with bigger systems due to the inherent quadratic scaling of EXX with the number of bands. From the figure, the difference induced by orbital self-consistency is seen to be largest (1 kcal/mol) for CO_2 which, not surprisingly, also shows the largest difference in self-consistent PBE and PBE0 energies in figure 3.3. This implies that the predicted atomization energy for this molecule can be improved relative to experiment from an error of 27 kcal/mol in PBE (see fig. 3.3) to within 1 kcal/mol using a non-selfconsistent PBE0 calculation. Note that in this specific case, the prediction is actually worsened by self-consistency. The huge improvement for CO_2 is largely fortuitous, but the trend is general across the database.

3.3.2 Ionization Potentials

While hybrid functionals can clearly improve the accuracy of DFT predicted energetics, these are already well described by local approximations. A more promising aspect of hybrid functionals is the improvement in the spectral properties of the KS system relative to real excitations.

The DFT formalism only provides an explicit formula for converting the *energy* of the KS system to an energy of the real system. No such formulas exist for interpreting the KS eigenvalue spectrum. Despite the lacking theoretical foundation, the raw KS eigenvalues are often used, and frequently very successfully, to estimate excitation levels of the real system.

In Hartree-Fock theory, Koopmans' theorem [141, 142] states that the eigenvalues can be interpreted as electron addition and removal energies in the approximation of no following orbital relaxation. Specifically, the HOMO is an approximation to the (negative) ionization potential (IP). In DFT, Janack's theorem implies that the DFT-HOMO (but only the HOMO) has a physical interpretation, namely as the exact (for the exact xc functional) ionization potential, including orbital relaxation [143]. In principle, the HOMO of DFT and HF should therefore be directly comparable, both giving an estimate of the ionization potential (IP).

Note that in all calculations done here, the exact exchange operator is included in the form of the non-local Fock operator consistent with the convention in hybrid functional calculations. In this form, pure EXX is therefore equivalent to standard HF.

Non-local operators are formally not within the framework of standard KS-DFT, although they are allowed in this form in the generalized KS framework (see section 2.2.2). For some purposes, e.g. for comparing density-functional approximations to the exact expression, it is advantageous to use a local form of the exact exchange operator (i.e. solve the OEP equations [78]), but that is not done here.

Figure 3.5 shows the calculated ionization potential of all atoms and molecules in the G2-1 database plus a few larger molecules (benzene, biphenyl, and naphthalene), determined by the position of the HOMO using the three functionals PBE, PBE0, and EXX (HF) compared to experimental (vertical) ionization potentials from [137].

In general 100% inclusion of exact exchange (HF) performs markedly better for positioning of the HOMO level than 25% in PBE0 which is optimal for the energetics. PBE performs the worst, with a systematic underestimation of the HOMO by a

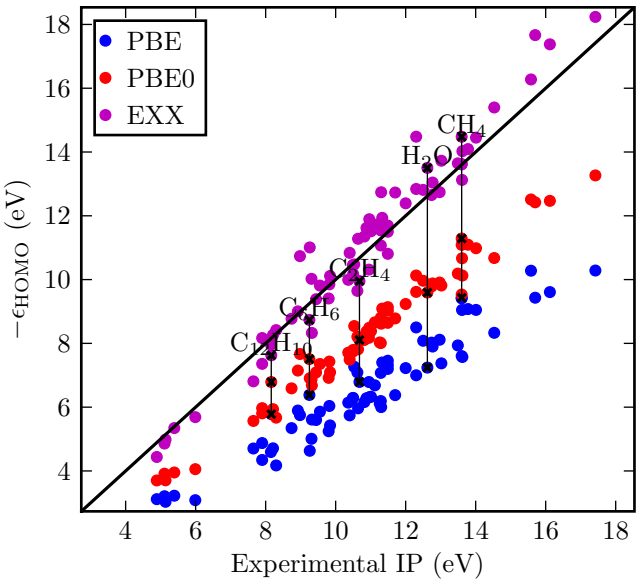


Figure 3.5: Ionization potentials estimated by the negative HOMO eigenvalue of DFT using different xc approximations.

factor of two on average. The optimal fraction of exact exchange is related to electronic screening of the system, and is much reduced in metallic systems. I will return to this aspect in chapter 4 where the use of GW allows for a system- and even spatially dependent dynamical screening, and leave the subject for now.

Motivated by the small impact of orbital self-consistency in fig. 3.4, I have calculated the eigenvalues of the EXX (or PBE0) Hamiltonian using the PBE orbitals (i.e. doing a single SCF iteration). This is again done for CO_2, as this molecule shows the largest difference between PBE and PBE0 orbitals. The result of this "one-shot" estimation of the eigenvalues is shown in figure 3.6 compared to the fully self-consistent spectrum at each level of xc approximation.

The negative of the experimental IP is marked by the dashed line on figure 3.6. Again it is seen that for CO_2 self-consistency has almost no effect on the eigenvalue spectrum; diagonalizing the EXX Hamiltonian using PBE orbitals is just as good.

As the total energies of DFT are generally very good, it is of course possible to calculate the IP by total energy differences between the molecule in its neutral state, and with one electron removed. The result of this approach is shown in figure 3.7 for a subset of the molecules (those that have a spin-paired ground state), and is seen to be in excellent agreement with experimental numbers.

In contrast, total energy differences from HF theory severely underestimates the ionization energies, and are actually better estimated from the HOMO energy than the total energy difference. As an example, consider the experimental IP of 12.6 and 10.7 eV for water and ethylene respectively. These come out nicely from DFT energy differences at 12.8 and 10.7eV, whereas HF yields 10.5 and 8.8 (in comparison the negative HOMO is 13.5 and 10.0).

Although PBE DFT can be used to determine IPs via total energy differences,

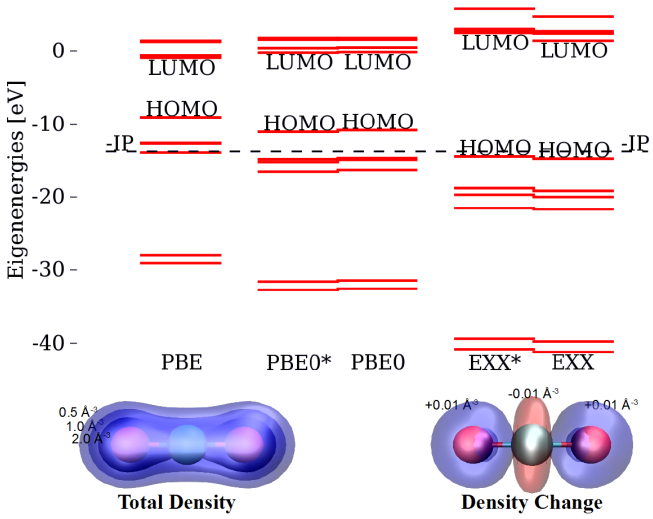


Figure 3.6: Importance of self-consistency on the eigenvalues of CO_2. The asterisk, *, indicates values evaluated non-selfconsistently on PBE orbitals. The dashed line indicates the experimental IP. The two plots in the bottom show isosurfaces of the PBE density and the density change induced by self-consistency in EXX respectively. Self-consistency induces a density change by less than 2%, with electrons moving from carbon towards oxygen in EXX

some applications rely directly on spectral properties, and for these, the incorrect positioning of the HOMO by a factor of two for molecules can be critical.

3.4 Image Charge Formation

From the previous application, it seems that an accurate description of the excitation spectrum of (small) isolated molecules can be obtained from DFT by the use of exact exchange.

In many applications of interest, molecules do however not appear in isolation. For example in catalysis and molecular electronics, the molecule will always be in close proximity to a crystal surface. This can have a large influence on the molecular excitation spectrum. For chemisorbed molecules, the renormalization of the molecular states will be governed by hybridization with the crystal surface states, but even for very weakly bound molecules, there can be proximity effects via image charge formation in the crystal as a response to the electric field of the charged molecule. This effect is dictated by the long range electrostatics and persists beyond the reach of orbital overlap. The effect is to reduce the energy cost of charging the molecule, as the electric field is (partially) screened at the crystal surface. The magnitude of the reduction depends on the ability of the crystal surface to polarize, i.e. the mobility of charge carriers at the Fermi level. Experimentally, this effect has been observed by [44] who argues that image charge formation is the main cause of the reduction in charging energy by an order of magnitude of an OPV5 molecule in a single-electron transistor (SET) setup compared to the free molecule. The effect has also been studied on the basis of model calculations by [144] within the GW

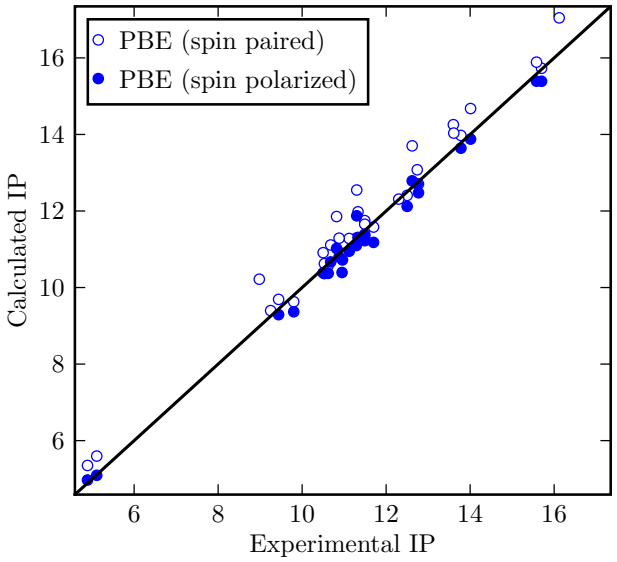


Figure 3.7: Ionization potentials estimated from total energy differences using the PBE xc-functional. The molecules considered her all have a spin paired ground state, and the empty (filled) circles indicates that the charged state has been treated as spin paired (polarized).

approach, and by [43] using *ab-initio* GW calculations of benzene on graphite.

Image charge formation is a dynamic correlation effect, related to the (time-dependent) response of the metal electron cloud during the virtual excitation of the molecule. This kind of effect can not be captured, even in principle, by a mean-field theory like DFT or HF.

The effect is included in the GW description, which is the primary advantage of GW over HF. It is an important effect for electron transport through molecules, and will be scrutinized in large detail on the basis of GW calculations in the next chapter. Here however, it is illustrated how to estimate the effect of image charge formation on the molecular ionization potential using a DFT total energy approach with a standard local (PBE) xc functional. This provides a very illustrative and physically intuitive description of the phenomenon.

3.4.1 Constrained DFT

From classical electrostatic arguments it is expected that the infinite metallic surface will react to a charging of the molecule by forming an image charge of same magnitude but opposite sign. The charged state can therefore by modeled by a charge neutral total system, in which the charge on the molecule is constrained to be $+1e$. The removed electron is allowed to settle in the most energetically favorable position which is not on the molecule.

Mathematically, the constraint can be formulated by the requirement

$$\int_{\Omega} \mathrm{d}\boldsymbol{r} n(\boldsymbol{r}) = M, \tag{3.27}$$

where M is the total number of electrons which should be contained within the domain Ω chosen to encompass the molecule. Formally, minimizing the energy functional under this constraint is equivalent to applying the external potential

$$v_{\text{ext}}(\boldsymbol{r}) = \begin{cases} \lambda, & \text{for } \boldsymbol{r} \in \Omega \\ 0 & \text{otherwise} \end{cases} \tag{3.28}$$

to the Hamiltonian, where λ is an unknown Lagrange multiplier chosen to satisfy eq. (3.27). I.e. a piecewise constant gate potential which is zero outside some domain Ω. In practice, the strength of the gate potential (bias) is simply increased until a total charge of $+1e$ is observed in the molecule.

For the method to be applicable, the separation of molecule and metal must be large enough that a domain Ω can be positioned such that no states overlap (significantly) with the domain boundary.

3.4.2 Results

The case study is based on a simple molecule with a large HOMO-LUMO gap (benzene) adjacent to a simple metallic crystal surface with high mobility charge carriers (sodium).

The setup can be seen from figure 3.8d. The plane of the benzene atoms is parallel to the metal surface, at a separation of 5Å. As bonding of benzene to metal surfaces is typically dominated by weak van der Waals forces which can not be accounted for by simple GGA xc functionals, the distance is simply chosen (read: no force relaxation) at a realistic size large enough to avoid hybridization and close enough to allow electrostatic interaction, such that the image charge effect can be studied in isolation. The yellow box indicates the domain used for the gate potential. One side of the box is midway between metal surface plane and benzene plane, i.e. 2.5Å from both. The other sides of the box gate are all at least 3Å from the closest benzene atom, and their position have no influence on the calculation.

As a reference, the ionization potential of gas phase benzene molecule (also indicated on fig. 3.5) is 9.25 eV both experimentally and when estimated by DFT total energy differences.

The gate potential is applied to the system and the bias gradually increased until one electron has been pushed off the molecule. During this process, the system response is monitored by following the three quantities plotted in figure 3.8: a) The change in total energy of the system excluding the energy of the gate potential itself, b) the charge on the molecule, and c) the position of the molecular HOMO.

Application of the gate potential in PAW follows be the description in sec. 3.2.2. The total charge associated with benzene is determined by Bader analysis on the all-electron density as described in sec. 3.2.2. The molecular HOMO is determined by the local density of states formed by projecting the orbitals onto the domain Ω.

At zero bias, the LUMO level is not far above the Fermi level, and due to the finite width it is slightly occupied, and the net charge on the molecule estimated by Bader analysis is -0.04e. From the orbital resolved DOS as defined in (3.9), the HOMO is seen to have p_z symmetry as expected from the gas phase, the peaks have finite width primarily due to the finite electronic temperature used in the calculation and to a lesser degree by very weak hybridization with the metal.

The first 3 eV applied bias does not lead to any significant charge transfer, the system energy is virtually unchanged, and the HOMO level is shifted upwards by

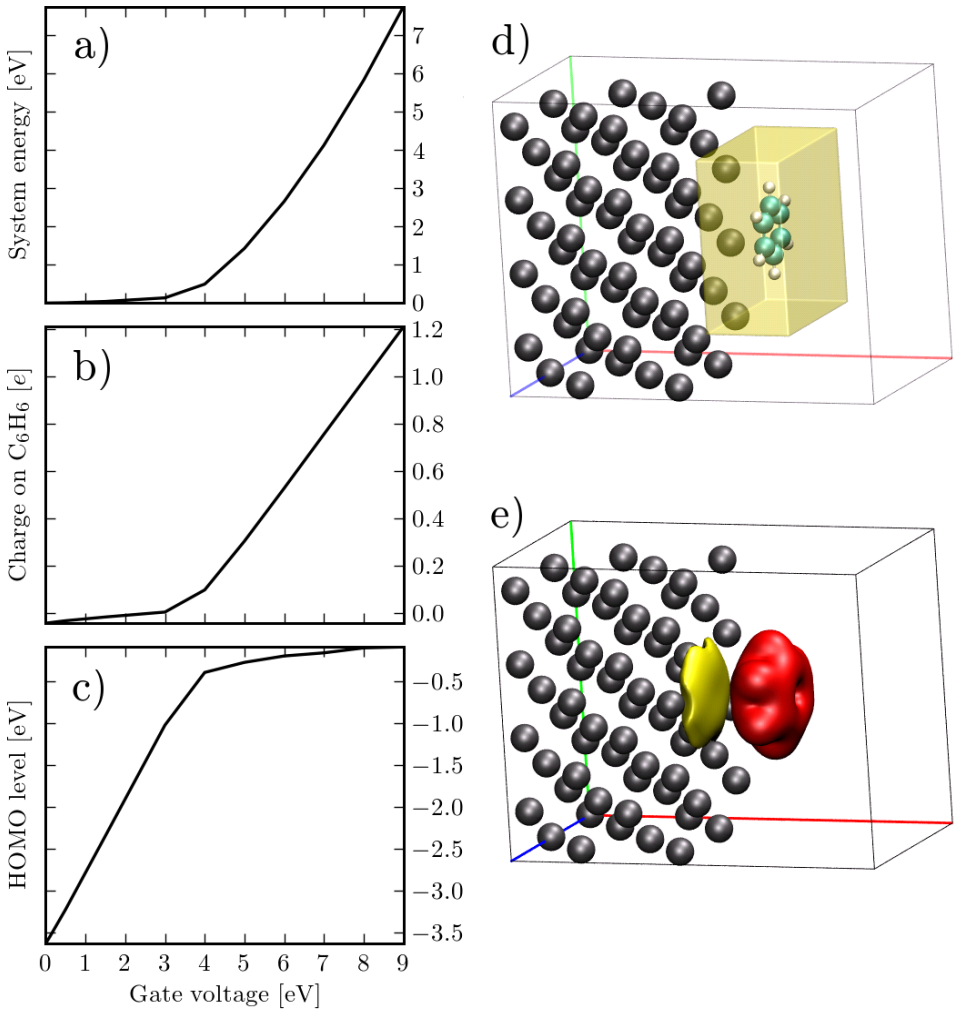


Figure 3.8: Application of a box gate potential to benzene 5 Å from a sodium surface. The figure shows the change in the system energy (a), total charge of the molecule (b) and position of the molecular HOMO relative to the Fermi level (c) upon varying the strength of the box gate potential shown in (d). Figure (e) shows two isosurfaces of the density change induced by applying a gate potential of 8 eV corresponding to removing one electron from the molecule. The red (yellow) isosurface corresponds to electron deficiency (surplus) compared to the neutral ground state. Isosurfaces are at $\pm 0.005 e/^3$.

almost the same amount (2.6 eV). As the HOMO level gets within 0.7 eV of the Fermi level, electrons start to move from the molecule to the sodium surface. During this process the HOMO is more or less pinned to the Fermi level, as the reduction in Hartree potential compensates the increasing gate potential.

At a bias of 8 eV, approximately one electron has been removed from the benzene molecule. The energy change in the system is at this point only 5.8 eV, which is almost a 40% reduction compared to the gas phase IP of 9.25 eV. The density change at this bias relative to the neutral ground states is shown in fig. 3.8e. The electron is seen to be removed from the p_z dominated HOMO state of the benzene and forms a screening cloud immediately in front of the sodium surface. At smaller isosurface values, one can also see a similar charge accumulation at the backside of the metal slab. Charge transfer proceeds after a gate bias of 8 eV, as the HOMO is doubly occupied.

3.4.3 Summary

This small example illustrates the mechanism of one particular dynamical screening effect; the image charge formation. The nature of the effect as well as the actual size can be estimated by the proposed constrained approach for weakly coupled molecules. Clearly this is not an effect accounted for by the spectrum of KS eigenvalues.

The effect is however automatically included for all levels of electron addition and removal in the spectral function derived from GW based Green function calculations. The GW self-energy includes polarization effects in the spectrum, without necessitating a calculation for each charged state, via the dynamical dependence of the screened interaction on all orders of pair-bubble (particle hole formation) diagrams, during the virtual excitations. The study of GW based excitation spectra is the subject of the following chapter.

I note that the method used in this section can not be applied to estimate the image charge effect on the electron affinity, as a negative gate potential leads to unphysical states defined by the gate domain boundary. This is not a problem at positive bias.

Chapter 4

The GW Approximation for *ab-initio* Calculations

The standard approach to first-principles conductance calculations of molecular junctions is today based on the combination of DFT and NEGF. While properties derived from total energies (or rather total energy differences) are accurately predicted by DFT, it is well known that DFT suffers from a band gap problem implying that the single-particle KS eigenvalues cannot in general be interpreted as real quasiparticle (QP) excitation energies. In particular, semilocal exchange-correlation functionals severely underestimates the fundamental gaps of both insulators, semi-conductors, and molecules.[69, 145–147]

Transport through metallic systems, and strongly bound molecules with high transmission probabilities are generally well-described at the mean-field DFT level. For the technologically important class of weakly bound molecules however, the poor representation of the QP excitations offered by DFT leads to very low agreement with experiment, as the tunneling mediated transport between leads via the resonant states of the molecule depends exponentially on the absolute position of the frontier orbitals relative to the Fermi level.

As shown in the last chapter (sec. 3.3), HF yields much improved values for the HOMO eigenvalue compared to DFT for isolated molecules. For bulk semi-conductors and insulators, the hybrid[82, 148] and screened hybrid[149, 150] functionals, which admix around 25% of the (screened) Fock exchange with the local DFT exchange are generally found to offer the best agreement with experimental band gaps[145, 146], while pure HF dramatically overestimates the gap[151, 152]. For metals HF breaks down completely.

The good agreement for molecules is due to the perfect cancellation of the Hartree self-interaction by the corresponding terms in the Fock (exact) exchange, while screening of the exchange interaction is a relatively weak effect in molecular systems[69, 153]. On the other hand, in extended systems the effect of self-interaction is less important and the long range Coulomb interaction becomes short ranged due to dynamical screening. This also indicates that the optimal fraction and screening length in (screened) hybrid functionals is system dependent, and for the composite junction setup with metallic leads and a weakly coupled molecule, these parameters would be ill defined. Also, due to the static nature of the HF self-energy, it can not describe effects like the image charge formation studied in section 3.4.

The many-body GW approximation of Hedin[112] was originally derived to correct for the deficient description of metals by HF theory, and the structure of the approximation is identical to HF save the replacement of the bare Coulomb operator $\hat{V} = 1/|\boldsymbol{r} - \boldsymbol{r}'|$ by the dynamically screened interaction $\hat{W} = \hat{\epsilon}^{-1}\hat{V}$, where $\epsilon(\boldsymbol{r}, \boldsymbol{r}', \omega)$ is the (frequency dependent) microscopic dielectric function. The approximation has been widely and successfully used to calculate QP band structures in metals, semiconductors, and insulators [113, 147, 154]. The fact that the screening is determined by the system itself instead of being fixed a priori as in the screened hybrid schemes, suggests that the GW method should be applicable to a broad class of systems ranging from metals with strong screening to molecules with weak screening. From the diagrammatic expansion of the screened interaction shown in sec. 2.5, it is expected to give an accurate description of the long range electrostatic image charge formation via all orders of electron-hole pair generation in the metal as a response to the molecular excitation. It is therefore ideally suited for the description of molecular electronics, and the present chapter presents an implementation and several applications of the GW approximation.

The chapter proceeds by first describing in detail the numerical implementation and the computational procedure in general in sec. 4.1. The GW approximation has already been extensively scrutinized in the context of extended systems, but no systematic study of the performance for molecules seems to be available. For this reason, the scheme is first applied to a database of molecules in sec. 4.3, which also serves to benchmark the implementation and the simple systems allow for an easier assessment of the physical content of the approximation. In the final section, 4.4, the scheme is applied to the two molecular junctions in the opposite limits of contact and tunneling regime.

To highlight the effect of *e*-*e* interactions, all calculations are performed at both the DFT, HF, and GW level and compared.

4.1 Computational Procedure

This section provides a detailed account of the implemented GW scheme.

The GW perturbation expansion for the inclusion of correlation effects in the Green function leads to a set of numerically very involved equations. This has the consequence that most applications of GW makes a series of assumptions and/or approximations on the form of the involved quantities. The most commonly used approximations will be discussed in section 4.1.5.

For the GW calculations done here however, such simplifications are avoided, and the equations solved directly in their most basic from. Avoiding assumptions based on the expected properties of specific physical systems, leads to an unbiased description, which is essential for the predictive power of the method when applied to new types of systems. As a consequence of the direct description, only very limited basis sets can be afforded numerically. Since this is the main limitation of the GW scheme presented here, some effort will be spend on the issue of constructing an effective, but compact, basis. This is done in section 4.1.1.

Section 4.1.2 describes an efficient technique for determining and storing the large tensors used to represent the interaction, screened interaction and polarization operators. This is followed by some details related to the underlying DFT code and the treatment of core states before summarizing in sec. 4.1.4 which presents the broad overview of the procedure.

4.1.1 Choice of Basis Set

The chosen basis should be minimal, while providing an accurate representation of the occupied subspace of H_0 (and thereby also g_0) plus the first few relevant excited states. Additional degrees of freedom are needed to represent the perturbations due the leads and e-e interactions. The basis should also provide a sufficient description of the bare (screened) interaction V (W) as well as the polarizability P. For the use in electron transport simulations, the basis must in addition be localized to facilitate the system partitioning.

To meet these requirements, the basis set used here is designed to: i) mimic a numerical atomic orbital (NOA) basis $\{\Phi\}$ (these are described in sec. 2.3.3) and ii) to contain an exact span of the lowest KS eigen states $\{\psi_n\}$, which are in turn determined from a high accuracy grid based DFT calculation in GPAW. The resulting basis functions, $\{\phi_i\}$, are denoted projected Wannier functions (PWF). The construction is based on the work by [155] but is similar to the quasi-atomic orbital (QO) method of [156]. In practice, the PWF come out as the linear combination

$$\phi_i(\boldsymbol{r}) = \sum_n \psi_n(\boldsymbol{r}) A_{ni} + \sum_\mu \Phi_\mu(\boldsymbol{r}) B_{\mu i}, \tag{4.1}$$

with the rotation matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ chosen to satisfy the mentioned constraints (span the lowest ψ, and mimic Φ). For completeness the exact procedure for their construction is given at the end of this section. The rotation matrices are uniquely determined from the projection and overlap matrices $V_{ni} = \langle\psi_n|\Phi_i\rangle$ and $S_{ij} = \langle\Phi_i|\Phi_j\rangle$ via a simple algebraic expression.

It is found to improve the quality of the basis to span not only the occupied subspace, but also states a few eV higher in energy. This energy threshold should not be increased too much, as that will only include additional irrelevant KS states in the span, while reducing the resemblence to the target NAO. By resembling the NAO, a more localized basis is achieved, and a better representation of the deviation between g_0 and G within the relevant Hilbert space of ground-state plus low energy excitations.

Compared to standard maximally localized Wannier functions (MLWF), the PWF are much easier to construct, and can be systematically augmented by additional radial (zeta-) functions and angular channels (polarization functions). It is also an advantage that the center and form of the PWF are known *a-priori* in constrast to MLWF. This implies that the coupling between lead and central region can be taken from the lead principal layer coupling, as the basis is expected to be similar in the two regions. For the purpose of analysis of bond character, the minimal, and optimally localized MLWF basis is superior, but as a generic basis set, PWF is the better choice.

Figure 4.1 shows the difference between a set of partly occupied MLWF [157, 158] for benzene and a SZ-PWF basis. In this figure, the PWF Hamiltonian has been subdiagonalized in the space of orbitals on each atom, to better display the symmetry. The figure displays the pseudo part of the functions only. The implementation of the partly occupied MLWF follows the generalization to PAW by [159]. Note that unlike the MLWF, the PWF basis is non-othogonal.

Compared to a pure NAO basis, the PWF basis is more effective because it contains an exact span of the occupied KS subspace, but the price paid is that they only *resemble* NAOs, thus the orbitals do not have well-defined angular momentum, and the radial part is not exactly zero beyond any cutoff radius. Thus one cannot

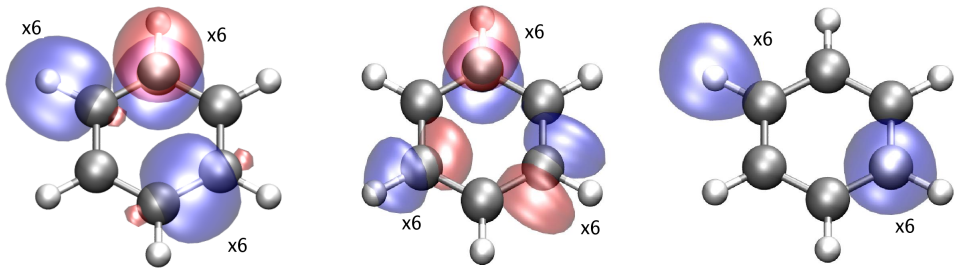


Figure 4.1: *Left*: Partly occupied WF; 6 C-H σ bonds, 6 C-C σ bonds, and 6 p_z bonds on C. *Right*: Projected WF; 12 s orbitals on C and H respectively, and 3 p-type orbitals on each C, one p_z, one along the C-H bond, and one perpendicular to the C-H bond.

utilize two- and four-center integral formulas to evaluate matrix elements of the overlap-, kinetic- and Coulomb operators. Instead the orbitals are constructed on the real-space grid, and subsequently the operators, likewise represented on the grid, are projected onto the PWF basis.

In practice everything is represented within the PAW framework to allow for an accurate description on a realistic grid. This implies that all wave functions are defined by a pseudo part and a set of atomic expansion coefficients. This affects calculations of projections and expectation values in a trivial fashoin, and when contructing the PWF basis, the rotation matrices should simply be applied to both the pseudo part and the expansion coefficients. Note that due to the use of PAW, the operators obtained from GPAW are full-potential quantities, and the wave functions from which the PWF basis is constructed correspond to the all-electron valence states. The interaction of the valence states with the core is discussed in sec. 4.1.3.

For technical reasons I find it most convenient to construct the PWF based on a single-zeta NAO, and then subsequently augment the basis by unmodified additional NAO zeta- and polarization functions. This removes problems with linear dependence in the basis generation, and additionally, when increasing the basis size in this way, the original basis is left unchanged, whereas a full PWF construction for the enlarged target NAO would produce slightly different basis functions for the entire set.

Construction of Projected Wannier Functions

Constructing N_{AO} PWFs, which span the $N_{\mathrm{occ}} < N_{AO}$ lowest KS eigen states is done by the linear combination

$$\phi_i(\boldsymbol{r}) = \sum_{n=1}^{N_{\mathrm{occ}}} \psi_n(\boldsymbol{r}) U^o_{ni} + \sum_{l=1}^{N_l} f_l(\boldsymbol{r}) U^u_{li}, \tag{4.2}$$

where the $N_l = N_{AO} - N_{\mathrm{occ}}$ extra degrees of freedom $f_l(\boldsymbol{r})$ needed to supplement the KS states, are extracted from the orthogonal complement to the occupied KS

space

$$f_l(\boldsymbol{r}) = \sum_{\mu=1}^{N_{AO}} \left(\Phi_\mu(\boldsymbol{r}) - \sum_n \psi_n(\boldsymbol{r}) \langle \psi_n | \Phi_\mu \rangle \right) U^l_{\mu l}. \tag{4.3}$$

For the PWFs ϕ to mimic the target NAOs Φ, the expansion coefficients are then chosen to minimize the summed squared norms of the residuals $r_i = \phi_i - \Phi_i$ under the constraint that the set f_l are orthonormal, i.e. from the unconstrained minimization of

$$\Omega = \sum_{i_1}^{N_{AO}} \langle r_i | r_i \rangle - \sum_{ll'=1}^{N_l} \Lambda_{ll'} (\langle f_l | f_{l'} \rangle - \delta_{ll'}). \tag{4.4}$$

Writing out the expressions shows that they depend only on the projection $V_{ni} = \langle \psi_n | \Phi_i \rangle$ and the overlap matrix $S_{ij} = \langle \Phi_i | \Phi_j \rangle$. After some algebra, and collecting the expressions, the two matrices $\boldsymbol{A} = \boldsymbol{U}^o(\boldsymbol{I} - \boldsymbol{U}^l \boldsymbol{U}^u)$ and $\boldsymbol{B} = \boldsymbol{U}^l \boldsymbol{U}^u$ in equation (4.1) should be chosen according to

$$\boldsymbol{A} = \boldsymbol{V}(\boldsymbol{I} - \boldsymbol{U}^l \boldsymbol{U}^{l\dagger} \boldsymbol{F}), \tag{4.5a}$$

$$\boldsymbol{B} = \boldsymbol{U}^l \boldsymbol{U}^{l\dagger} \boldsymbol{F}, \tag{4.5b}$$

with $\boldsymbol{U}^l$ being the matrix that diagonalize $\boldsymbol{F} = \boldsymbol{S} - \boldsymbol{V}^\dagger \boldsymbol{V}$, truncated to the N_l columns corresponding to the eigenvectors of highest eigenvalue, and normalized such that $\boldsymbol{U}^{l\dagger} \boldsymbol{F} \boldsymbol{U}^l$ is the $N_l \times N_l$ identity matrix, i.e. $U^l_{\mu l} = v_{\mu l}/\sqrt{\sigma_l}$, where $\boldsymbol{F}\boldsymbol{v}_l = \sigma_l \boldsymbol{v}_l$.

4.1.2 Calculating the Coulomb Matrix

For a basis set based on analytic atomic orbitals, matrix elements of the Coulomb operator can be efficiently calculated by explicit analytical expressions[160]. For NAO, analytic expressions exist for the angular part (as this is simply products of spherical harmonics) [161] and a numerical Fourier-Bessel transform technique can be used for the 1D radial problem. For the Wannier function based scheme employed here, this is not possible, and the calculation is done by the grid-based PAW technique described in section 3.2.3.

Direct calculation on the pseudo grid of all required elements was possible for HF and lrTDDFT calculations in the KS eigen basis, as these required only 2-index elements for HF and pair orbitals restricted to electron-hole pairs (which decayed rapidly in importance as the energy difference between these increased) for the lrTDDFT calculations.

For the Green function based schemes, all of the Coulomb matrix elements, $V_{ij,kl}$, are needed, which becomes prohibitively costly to compute for larger basis sets. In addition, once calculated, storing the matrix is by itself prohibitive. For 60 orbitals, storing the full $V_{ij,kl}$ matrix requires 200 MB, which implies that for a typical energy grid with 10^4 grid points (e.g. from -100 to 100 eV with a spacing of 0.02eV), each of the matrices $\boldsymbol{P}^{r/<\!/>}(\omega)$ and $\boldsymbol{W}^{r/<\!/>}(\omega)$ would require 2 TB to store, and determining $\boldsymbol{W}^r$ would require inverting a $60^2 \times 60^2$ matrix at each energy point.

A scheme for a compact representation of V describing only distinct and non-negligible elements is therefore needed for any practical calculations. To this end, the product basis technique of [162, 163] is implemented. In this approach, the similarity between the elements of the non-local Coulomb operator

$$V_{ij,kl} = \langle n_{ij} | \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} | n_{kl} \rangle, \tag{4.6}$$

where $n_{ij}(\mathbf{r}) = \phi_i^*(\mathbf{r})\phi_j(\mathbf{r})$, and the much simpler local pair orbital overlap

$$S_{ij,kl} = \langle n_{ij}|n_{kl}\rangle, \tag{4.7}$$

is used to screen for the significant elements of V. Notice that pair orbitals made from orbitals with no overlap will also not contribute to V, and secondly that two pair orbitals with small overlap in S are expected to give a small contribution in V. In principle, two pair orbitals with zero direct overlap in S can have an overlap via the long range coulomb kernel, but this contribution is presumably much smaller than for pairs which also have direct overlap.

Let $\boldsymbol{U}$ denote the rotation matrix that diagonalize $\boldsymbol{S}$, and $\boldsymbol{\sigma}$ the corresponding eigenvalues, i.e.

$$\boldsymbol{U}^\dagger \boldsymbol{S}\boldsymbol{U} = \boldsymbol{\sigma}\boldsymbol{I}. \tag{4.8}$$

This defines an optimized set of pair orbitals

$$n_q(\boldsymbol{r}) = \sum_{ij} n_{ij}(\boldsymbol{r})U_{ij,q}/\sqrt{\sigma_q}, \tag{4.9}$$

which are mutually orthonormal, i.e. $\langle n_q|n_{q'}\rangle = \delta_{qq'}$. In basis of these, the Coulomb matrix elements are

$$V_{qq'} = \langle n_q|\frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|}|n_{q'}\rangle, \tag{4.10}$$

which are related to the original matrix elements via

$$V_{ij,kl} = \sum_{qq'} U_{ij,q}\sqrt{\sigma_q}V_{qq'}\sqrt{\sigma_{q'}}U^*_{kl,q'}. \tag{4.11}$$

From this equality, only elements of $V_{qq'}$ corresponding to elements for which both σ_q and $\sigma_{q'}$ are non-negligible are expected to be of physical significance.

The idea then, is to calculate $\boldsymbol{V}$ in the basis of n_q according to (4.10), but only for pair orbitals with an eigenvalue larger than some threshold. This will reduce the computational time to determine $\boldsymbol{V}$, but more importantly, the reduced basis representation V_{qq}, will also imply that the much larger dynamic quantities, $\boldsymbol{P}$ and $\boldsymbol{W}$ can be kept in this reduced representation.

The truncation scheme is analyzed in figure 4.2 for a SZ-PWF basis for Benzene. A SZ basis for Benzene consists of 30 orbitals, resulting in 900 pair orbitals. The left plot shows the sorted eigenvalues σ_q and the right plot the error induced in the calculated HF HOMO eigen value by truncation of $V_{qq'}$ relative to using the full matrix. It is seen that an error less than 0.05eV is reached at 150 included pair orbitals, which corresponds to 17% of the total number, and an eigen value of $\sigma = 10^{-4}a_0^{-3}$

A truncation of pair orbitals with an eigenvalue less than $\sigma = 10^{-5}a_0^{-3}$ is used in all calculations presented in this thesis, and is quite generally found to cause an error that is an order of magnitude smaller than that caused by the finite basis set (0.1eV).

The modest reduction of pair orbitals for Benzene is increased when going to larger systems, as shown in figure 4.3 for a SZP basis set on BDA. Here the eigenvalue threshold $\sigma = 10^{-5}a_0^{-3}$ reduces the number of pair orbitals to only 7%, which implies a reduction of V by more than two orders of magnitude.

Note that in an AO basis, the number of pair orbitals with a norm (eigenvalue of S) larger than a given threshold is related to the number nearest neighbor atoms

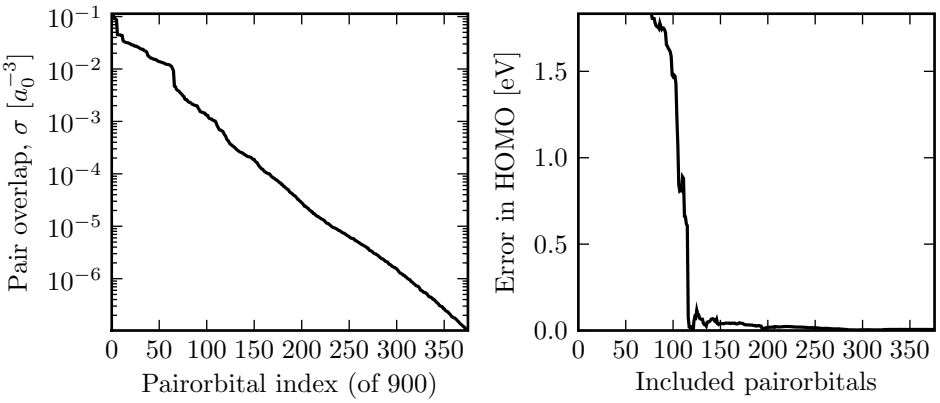


Figure 4.2: Convergence study of the pair-orbital truncation scheme for Benzene (SZ). Left: eigenvalues of the pair-orbital overlap matrix. Right: the induced (signed) error in the HOMO eigen value by truncation of the Coulomb matrix.

within a distance determined by the norm threshold. From this argument it follows that the number of significant n_q will scale linearly with system size and $V_{qq'}$ therefore quadratically as opposed to the N^4 scaling of the full Coulomb matrix.

4.1.3 PAW Quantities and Core Interactions

All inputs to the NEGF code, are extracted from the real-space PAW code GPAW[86] as described in section 3.2. The mean-field Hamiltonian H_0 used in the NEGF code is thus the self-consistent Kohn-Sham Hamiltonian. Since this includes a Hartree contribution corresponding to the KS-DFT density distribution, and a static, local approximation to the exchange and correlation potential, these should be subtracted from the interaction self-energy to avoid double counting

$$\mathbf{\Sigma}^r_{\text{tot}} = \mathbf{\Sigma}^r_{\text{leads}} + \mathbf{\Sigma}^r_{ee} - V_{Hxc}, \tag{4.12}$$

where $V_{Hxc,ij} = \langle \Phi_i | v_H(\boldsymbol{r}) + v_{xc}(\boldsymbol{r}) | \Phi_j \rangle$ is a static self-energy operator, and thus has no lesser or greater components.

The non-trivial determination of the AE two-particle operator $V_{ij,kl}$ from a PAW framework was described in section 3.2.3. The use of PAW imply that all operators are full-potential quantities. The constructed PWF basis onto which these are projected, span the all-electron KS valence states, and consists of a smooth pseudo NAO-like part augmented by local atomic partial waves.

The core-electrons are assumed frozen in their respective reference states, and are not represented in the basis. As demonstrated in section 3.2.5 for the Fock operator, interactions between valence and core states should be included in the self-energy operators for an accurate description of these.

The Hartree self-energy only interacts with the core electrons via the total density of these. This effect is included in the mean-field Hamiltonian, as only the valence part is removed in (4.12), and the core electron density remains frozen. For the

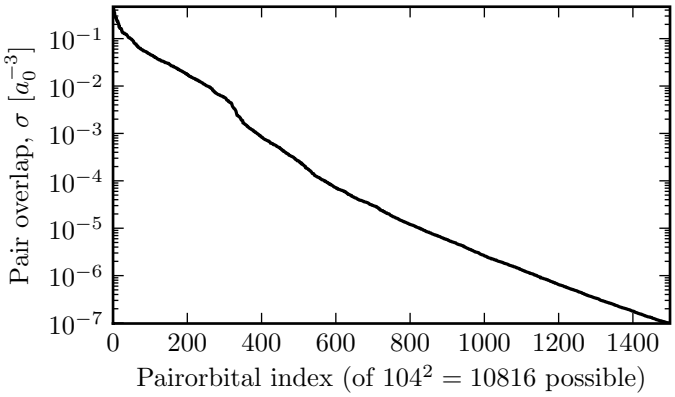


Figure 4.3: Optimized pair orbital norm a SZP-PWF basis on a BDA molecule. For larger systems, the truncation to $\sigma = 10^{-5}a_0^{-3}$ leads to a larger percentage wise reduction of pair orbitals. Here reduced to only 7% of the total number of pairs. In general the number of significant optimized pair orbitals scales linearly with system size, whereas the total number scales quadratically.

Fock self-energy, the valence-core interaction is

$$\Sigma^{\text{core}}_{x,ij} = -\sum_{k}^{\text{core}} V_{ik,jk}, \tag{4.13}$$

where it has been used that the density matrix is simply the identity matrix in the subspace of atomic core states (these are doubly occupied, but as the operator is diagonal in spin, no factor of two appears). This is related to the valence-core exchange in section 3.22, and can in PAW be extracted via

$$\Sigma^{\text{core}}_{x,nm} = -\sum_{a}\sum_{ij}\langle\tilde{\Phi}_n|\tilde{p}^a_i\rangle X^a_{ij}\langle\tilde{p}^a_j|\tilde{\Phi}_m\rangle. \tag{4.14}$$

with the tensor X^a_{ij} given by eq. (3.21).

This core contribution to the valence Fock self-energy typically has elements on the order of 1eV, and can therefore not be ignored. Note that these interactions are not accessible in pseudopotential approximations, where the valence wave functions are not physically meaningful in the core region.

For the correlation part of the GW self-energy, valence-core interactions are neglected. This is reasonable, because the polarization bubble, $\boldsymbol{P}$, involving core and valence states will be small due to the large energy difference and small spatial overlap of the valence and core states. The same procedure is used by [164, 165], who report that including more states in the valence has no significant effect, if valence-core interactions are treated at the HF level. If these are not included, it is necessary include more core states in the valence description.

4.1.4 Summary and Program Flow

In this section the procedure for performing a non-equilibrium GW calculation is detailed.

To characterize the non-interacting system, the Hamiltonian and overlap matrices of the central and lead (if any) regions must be specified. The interaction are determined by the specification of the coulomb matrix V. For model calculations, this can typically be represented directly in the four index form $V_{ij,kl}$. For *ab-initio* calculations, the Coulomb matrix V is determined in the reduced basis of optimized pair-orbitals defined by the (truncated) matrix of eigenvectors U and eigenvalues σ of the pair-orbital overlap matrix, as described in section 4.1.2. When using DFT to describe the mean-field Hamiltonian, the DFT xc matrix for the central region should also be determined (used to avoid double counting), as well as the core exchange self-energy $\Sigma_{F,\text{core}}$.

The program flow is skecthed in figure 4.4. The components of the Green function are determined in the red box, with the retarded component given by the Dyson equation, and the lesser by either the Keldysh or the fluctuation-dissipation equation depending on whether the system is out of equilibrium or not.

For the initial estimate of the Green function, the total self-energy is for isolated systems assumed to be zero, and for extended systems given by the lead self-energies only (upper left corner of the figure). The initial Green function if based on a DFT Hamiltonian, includes electron interactions at the mean-field Hxc level. Once this is constructed, the interaction self-energies can be calculated. The static Hartree and Fock components are shown on the left, and the dynamic correlations on the right. Once determined they can by used to form a new estimate for the total self-energy (bottom of the figure) which is then used as input to a new evaluation of the Green function (red box). This loop is iterated until self-consistency in the Green function using a Pulay[167] mixing scheme to update the retarded GF[107]. When converged, the density matrix, projected- and total density of states, current, and non-interacting transmission function can be evaluated and analysed.

The code assumes a spin-paired ground state, which gives rise to the factor of 2 on the Hartree self-energy and polarization matrix in the flowchart (the Fock self-energy is diagonal in spin).

In the following some of the more technical details are listed. These are not essential and can be skipped by the impatient reader.

Technical Details

The Langreth rules of tab. 2.1 have been used to convert the interaction self-energies defined in the contour in sec. 2.5 to real time components. Note however that the GW self-energy is first separated into a static exchange part and the dynamical correlations $\Sigma_{GW}(\tau,\tau') = \Sigma_F\delta_{\mathcal{C}}(\tau,\tau') + \Sigma_{\text{corr}}(\tau,\tau')$, which is not only convenient for analysis, but also for numerical reasons as will be explained shortly. Since the Fock diagram is static, the lesser / greater components of GW relate purely to correlations, while the retarded part decompose as $\Sigma^r_{GW}(t) = \Sigma_F\delta(t) + \Sigma^r_{\text{corr}}(t)$. Having removed the singular δ function for the correlation self-energy, this can be determined from the relation

$$\Sigma^r_{\text{corr}}(t) = \theta(t)\left[\Sigma^>_{\text{corr}}(t) - \Sigma^<_{\text{corr}}(t)\right]. \tag{4.15}$$

This indirect route is numerically preferable to the direct application of the Langreth rules to get Σ^r_{GW}, because use of G^r and W^r in the time domain is avoided. These functions are constructed in frequency space, where they decay like $1/\omega$ (due to the inherent step function in time in retarded components), and are therefore not easily Fourier transformed to the time domain. The inverse problem arises when

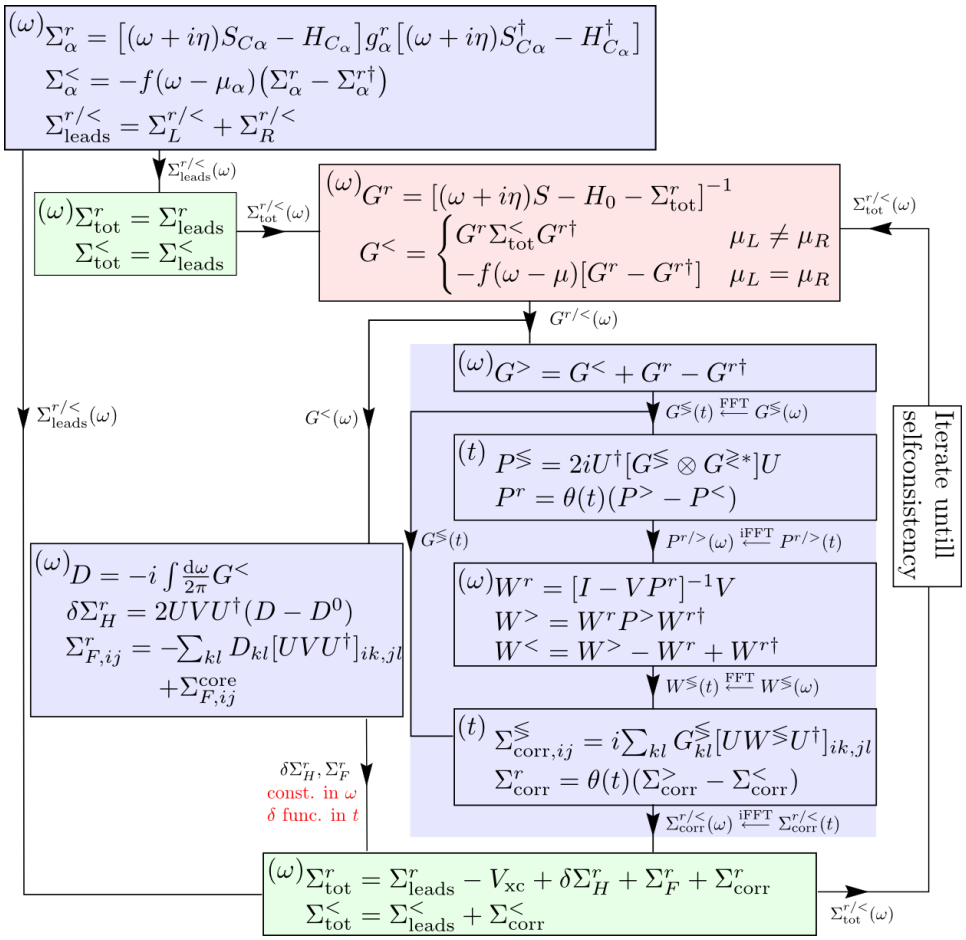


Figure 4.4: Flowchart for the NEGF based GW code KeldyshGF[166] starting from the upper left corner. Blue boxes indicates self-energy approximations, green boxes are the total self-energy. The components of the Green function are determined in the red box, with the lesser GF determined from *either* the Keldysh *or* the fluctuation-dissipation equation for non-equlibibrium and equilibrium systems respectively. The operation $U^\dagger XU$ downfolds the 4-index polarization to the optimized pairorbital representation, while $UXU^\dagger$ reconstructs the bare and screened interactions. The Kronecker product $[A \otimes B]_{ij,kl} = A_{ik}B_{jl}$.

transforming $\Sigma^r_{\text{corr}}(t)$ back to frequency domain, but in this case, the function can simply by padded by zeros before doing the inverse transform, to obtain a reliable transform. In practice, the array is padded to ten times the original size before being transformed. In addition, the inverse FFT from time to energy is done using endpoint corrections as described in [168, Chap. 13.9] to correct for the use of FFT on a non-periodic function. The reasoning and procedure is the same for the calculation of the retarded polarization function. The extended grid and end-point corrections are only used in conjunction with the FFTs of step functions.

The dynamical quantities G, W, P, Σ_{leads}, and Σ_{corr} are expressed directly on the real frequency/time axis using a dense uniform grid, thus avoiding analytic continuations from the imaginary axis. In practice, the energy grid should extend about 60 eV beyond the lowest and highest spectral peaks for HF, while GW calculations, for an accurate representation of the polarization, requires a grid that extend (about 80 eV) beyond the largest possible energy difference. For the calculations done here, a typical energy grid range from -150 to 150 eV. A grid spacing of 0.02 eV seems to give converged values.

The lead coupling and correlation gives natural life-time reduction of excitation, while the parameter η, provides an artificial broadening of the discrete levels, and is be reduced until the results have converged. The energy grid spacing determines the lower limit of η, which is fixed to $\eta = 2\Delta E$ for all calculations presented here.

In equilibrium, the chemical potential, μ, used in the fluctuation-dissipation theorem is well-defined for extended systems only. For isolated systems, μ is chosen to lie in the HOMO-LUMO gap. In the limit of small η the result of the GW calculation becomes independent of the precise position of μ inside the gap.

Due to the finite width of the spectral peaks, the truncation of the spectrum at the lower limit of the energy grid and the exact positioning of the Fermi level will influence the numerical value of the total electron number. The Hartree potential is particularly sensitive to changes in the total electron number, and it is therefore important that the change in Hartree potential is *not* calculated by the difference to an initial Hartree potential evaluated by projecting the static KS Hartree Potential onto the PWF basis but rather evaluated using the *change* in density matrix only, as indicated in figure 4.4. The initial density matrix is, like the self-consisten density matrix, based on numerical integration of the lesser Green function. In this way, the change in Hartree potential is formed from matrices made from compatible numerical approximations, i.e. basis set, Coulomb matrix truncation, and energy grid. Specifically, the change in Hartree potential is exactly zero if the Green function does not change.

By switching between energy and time domains using the FFT algorithm[99, 169], expensive convolutions are avoided whereby all operations in the chart are local operations on the energy / time grid.

Parallelization is done by distributing the energy / time grid of all dynamic matrices G, P, W, and Σ_{corr} over the available processors. Due to the locality of all matrix operations, no communication is required for these operations. Only when changing between energy and time domain is communication needed, as the FFT algorithm is highly nonlocal. To switch between energy and time domain, the matrices are first redistributed by changing to a matrix index division, the FFT is then performed for each matrix index independently, and lastly, the matrices are re-collected while distributing over the energy /time axis again. In practice this scheme works very well, showing a close to linear scaling with the number of processors, and has been used in calculations with more than 450 CPU's.

4.1.5 Common Approximations to GW

In conventional GW calculations, the full Green function is not calculated, but one focus instead only on the poles of the retarded Green function $\widehat{G}^r = [\omega - \widehat{H}_0 - \widehat{\Sigma}(\omega)]^{-1}$, giving the quasiparticle energies

$$\varepsilon_n^{\mathrm{QP}} = \langle \psi_n^{\mathrm{QP}} | \widehat{H}_0 + \widehat{\Sigma}(\varepsilon_n^{\mathrm{QP}}) | \psi_n^{\mathrm{QP}} \rangle,$$

where the quasiparticle states $|\psi_n^{\mathrm{QP}}\rangle$ are the eigenstates of $\widehat{H}_0 + \widehat{\Sigma}$ evaluated at the pole energies.

The self-energy $\widehat{\Sigma}$ is non-local, complex, energy-dependent and in general non-Hermitian, which implies that the quasiparticle energies are generally complex reflecting that the quasiparticles are only approximate eigenstates of the many-body wave function. The imaginary part of $\varepsilon_n^{\mathrm{QP}}$ is inversely proportional to the quasiparticle lifetime, and the real-part reflects the actual excitation energy.

The standard approach is to assume that the quasiparticle states, are well-described by the eigenstates ψ_n^0 of $\widehat{H}_0$, which implies that the excitation energies can be determined via

$$\varepsilon_n^{\mathrm{QP}} \approx \varepsilon_n^0 + \Re \langle \psi_n^0 | \widehat{\Sigma}_{\mathrm{GW}}(\varepsilon_n^{\mathrm{QP}}) | \psi_n^0 \rangle, \tag{4.16}$$

where ε_n^0 are the eigenvalues of $\widehat{H}_0$. Since the mean-field and quasiparticle states are assumed to be identical, changes in the Hartree potential and the off diagonal elements of the self-energy are ignored. Further assuming that the frequency dependence of $\widehat{\Sigma}_{\mathrm{GW}}$ can be approximated by its first order Taylor expansion in a sufficiently large neighborhood of ε_n^0 yields the linearized quasiparticle equation

$$\varepsilon_n^{\mathrm{QP}} \approx \varepsilon_n^0 + Z_n \Re \langle \psi_n^0 | \widehat{\Sigma}_{\mathrm{GW}}(\varepsilon_n^0) | \psi_n^0 \rangle, \tag{4.17}$$

where the so-called quasiparticle weight

$$Z_n = \left[1 - \frac{\partial \langle \psi_n^0 | \widehat{\Sigma}_{\mathrm{GW}}(\varepsilon) | \psi_n^0 \rangle}{\partial \varepsilon} \Big|_{\varepsilon_n^0} \right]^{-1}. \tag{4.18}$$

Typically, calculations reported as selfconsistent[164, 165] iterate (4.17) and (4.18) using the estimated QP energy to re-evaluate the GW self-energy. This makes the operators self-consistent in the energy (i.e. equivalent to (4.16)), but any orbital renormalization are still neglected, i.e. off-diagonal elements of the self-energy and Hartree contributions. In addition, the new energy is often only used to re-evaluate the screened interaction and not the single-particle Green-function, i.e. $\Sigma_{\mathrm{GW}} = iG_0W[G_0]$, with $G_0(z) = (z - H_{\mathrm{KS}})^{-1}$. The computational expensive RPA screening used in W is often avoided using model dielectric funtions, or by the plasmon-pole approximation[113]. Pseudopotential approaches often approximate valence-core interactions or neglect them altogether (here, these are included exactly at the HF level, as motivated in [164]. The convolution of the Green function and screened interaction on the energy axis requires a very fine grid spacing due to the poles of the Green function on (close to) the real axis. This is typically avoided by evaluating the quantities on the imaginary frequency axis, on which these are more well-behaved, and then relying on analytic continuation to transform back to the real axis.

These approximations have mainly been developed and tested for solids, and less is known about their applicability in other systems. The implementation of the GW method presented here avoids all of these technical approximations allowing for a direct and unbiased assessment of the GW approximation itself.

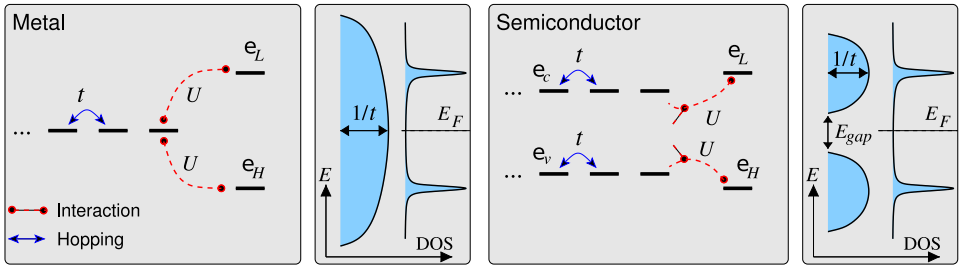


Figure 4.5: The lattice models representing a metal-molecule and semiconductor-molecule interface, respectively. Interaction are represented by the non-local Coulomb interaction U between substrate and molecule.

4.2 Model Study of the Image Charge Effect in GW

This section describes a model study of the GW based description of image charge formation.

The psysical system modeled is a molecule on a crystalline substrate. The molecule is assumed to be weakly coupled such that hybridization effect can be ignored. All interactions internally on the molecule and substrate are neglected, such that the interactions between substrate surface and molecule, and hence the image charge effect, can be studied in isolation.

The molecule is represented by its HOMO and LUMO levels, i.e.

$$\hat{H}_{\text{mol}} = \xi_H \hat{n}_H + \xi_L \hat{n}_L \tag{4.19}$$

where e.g. $\hat{n}_H = c^\dagger_{H\uparrow} c_{H\uparrow} + c^\dagger_{H\downarrow} c_{H\downarrow}$, is the number operator of the HOMO level. A metallic substrate was studied by [144] using a semi-infinite tight-binding chain as model Hamiltonian

$$\hat{H}_{\text{met}} = \sum_{i=-\infty}^{0} t(c^\dagger_i c_{i-1} + c^\dagger_{i-1} c_i). \tag{4.20}$$

In the case of a metallic substrate, the metal-surface interaction was modeled by the interaction operator

$$\hat{U}_{\text{met}} = U \hat{n}_0 \hat{N}_{\text{mol}}$$

where $\hat{N}_{\text{mol}} = \hat{n}_H + \hat{n}_L$ is the number operator of the molecule, and n_0 that of the metal surface site. In this case, the interaction operator could be modeled by an effective Coulomb interaction, with only two external vertices [144].

In this work, the model is extended to a semiconducting substrate modeled by

$$\hat{H}_{\text{sc}} = \sum_{\alpha=c,v} \sum_{i=-\infty}^{0} \varepsilon_\alpha \hat{n}_{\alpha i} + t(c^\dagger_{\alpha i} c_{\alpha i-1} + c^\dagger_{\alpha i-1} c_{\alpha i}), \tag{4.21}$$

where $\alpha = c, v$ refers to conduction and valence bands, respectively. The polarization effect of the metal in response to charging of the molecule appear via (to first order) the pair bubble diagram describing electron-hole pair generation. For a

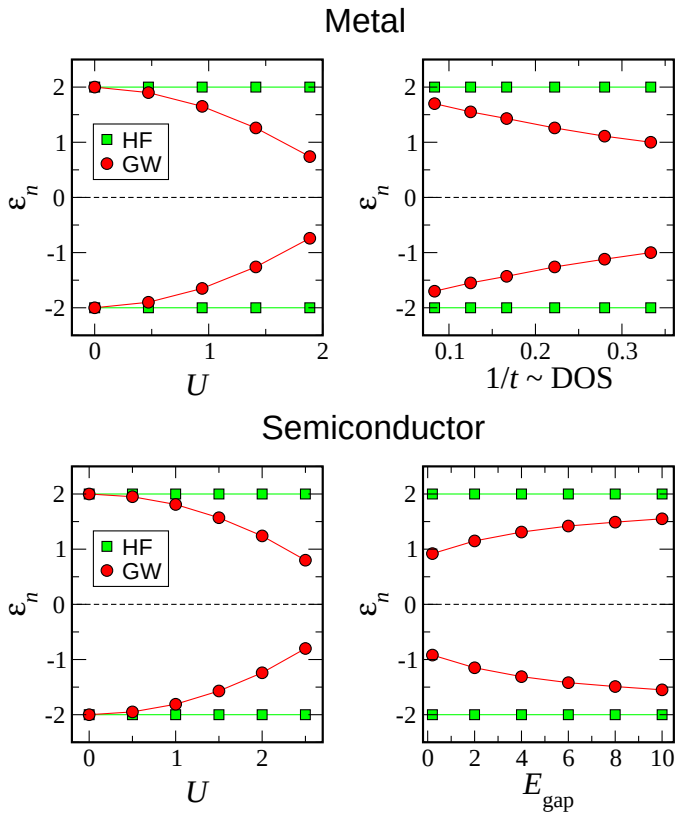


Figure 4.6: Peak position of the molecular resonances for the TB models of metallic (upper panel) and semiconducting (lower panel) substrate as a function of interaction strength (left) and polarizability of the substrate (right). The polarizability is controled by varying the band width for the metal, and the band gap for the semiconductor.

semiconductor, screening is thereby mediated by transitions between valence and conduction bands, and the interaction operator takes the form

$$\hat{U}_{\text{sc}} = U \sum_{\sigma} (c^{\dagger}_{c0,\sigma} c_{v0,\sigma} + c^{\dagger}_{v0,\sigma} c_{c0,\sigma}) \hat{N}_{\text{mol}}, \tag{4.22}$$

which must be described by a full four-vertex interaction operator.

The models describing the metallic and semiconducting systems are depicted in figure 4.5. The chosen Fermi level is indicated in the schematic DOS; mid-gap of the semi-conductor and in the center of the metallic DOS. The HOMO and LUMO level are centered around the Fermi level.

The peak position of the molecular resonances are shown in figure 4.6 at both the HF and GW level. It is seen the the image charge induced gap reduction is quadratic in the interaction strength. The polarization response of the substrate can be adjusted by varying the band width for the metal (thus adjusting the DOS at the Fermi level and therby the mobility of charge carriers) and by changing the band gap of the semiconductor. The HF spectrum does not include polarization effects (it

corresponds to *unrelaxed* addition / removal energies, following Koopmans theorem) and hence do not shown any effect when varying the mentioned parameters.

The effect is understood by studying the first order term of GW rsulting in polarization effects. That is the 2. order polarization bubble diagram of fig. 2.5 (the first order diagram is the static Fock operator). Neglecting hybridization effect, the system can be described by the eigenstates of the isolated molecule $\{\psi_\nu\}$ and suurface $\{\psi_k\}$ respectively. In terms of these, the time-ordered polarization self energy is

$$\Sigma^{(2)}_{\nu\nu} = \sum_{k}^{\text{occ}} \sum_{k'}^{\text{empty}} \int i G_{0,\nu\nu}(\omega') V_{kk'} \; P_{kk'}(\omega' - \omega) V_{k'k} \mathrm{d}\omega', \tag{4.23}$$

with the polarization and Coulomb matrices are defined by

$$P_{kk'}(\omega) = \frac{1}{\omega - \omega_{kk'} + i\eta} - \frac{1}{\omega + \omega_{kk'} - i\eta} \tag{4.24a}$$

$$V_{kk',\nu\nu} = \iint \frac{\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r'}}{|\boldsymbol{r} - \boldsymbol{r'}|} \psi^*_k(\boldsymbol{r}) |\psi_\nu(\boldsymbol{r'})|^2 \psi_{k'}(\boldsymbol{r}) \tag{4.24b}$$

where $\omega_{kk'} = \varepsilon_{k'} - \varepsilon_k \geq 0$.

Inserting the time-ordered GF of the isolated molecule $G_{0,\nu\nu}(\omega) = 1/(\omega - \varepsilon_\nu + \mathrm{sgn}(\varepsilon_\nu - \mu) i\eta)$ in the self-energy (4.23) results in

$$\Sigma^{(2)}_{\nu\nu}(\omega) = \frac{1}{\pi} \int \frac{\Delta(\omega')}{\omega - \omega' + \mathrm{sgn}(\mu - \varepsilon_\nu) i\eta} d\omega' \tag{4.25}$$

where Δ defines a dynamic interaction strength

$$\Delta = \pi \sum_{k}^{\text{occ}} \sum_{k'}^{\text{empty}} |V_{kk',\nu\nu}|^2 \delta(\omega_{kk'} - \mathrm{sgn}(\varepsilon_\nu - \mu)(\omega - \varepsilon_\nu)). \tag{4.26}$$

The physically relevant retarded self-energy follows from the time-ordered component via the Kramer-Kronig realtion [170]

$$\Sigma^{(2),r}_{\nu\nu} = \mathcal{H}[\Delta(\omega')](\omega) - i\Delta(\omega), \tag{4.27}$$

where $\mathcal{H}$ denotes the Hilbert transform.

In the limit where $V_{kk',\nu\nu}$ is independent of k and k' as in the model interaction, the interaction strength Δ the shifted joint density of states (JDOS) of the substrate, shifted by the unperturbed molecular resonance ε_ν (and scaled by U). In this case the quasiparticle peak in the GW approximation can be determined by the graphical construct shown in figure 4.7b+c for metals and semiconductors respectively.

The solution relies on the linearized quasiparticle equation to obtain the peaks of the GW resonances

$$\varepsilon^{\mathrm{QP}}_\nu = \varepsilon_\nu + \Re\langle\psi_\nu|\Sigma^{(2),r}(\varepsilon^{\mathrm{QP}}_\nu)|\psi_\nu\rangle, \tag{4.28}$$

i.e. the QP is determined by the intersection of the line $y = \omega - \varepsilon_\nu$ and the Hilbert transform of the JDOS (the real part of $\Sigma^{(2),r}_{\nu\nu}$).

This trends observed on figure 4.6 can be explanation in terms of the graphical solution on fig. 4.7. For a metal, the JDOS raises linearly at $\omega = 0$ with a slope given by the metal's DOS at E_F. This suggests that the level shift should increase with the substrate DOS at the Fermi level. For a semiconductor, the JDOS raises smoothly at $\omega = E_{\mathrm{gap}}$, suggesting that the level shift should decrease with E_{gap}.

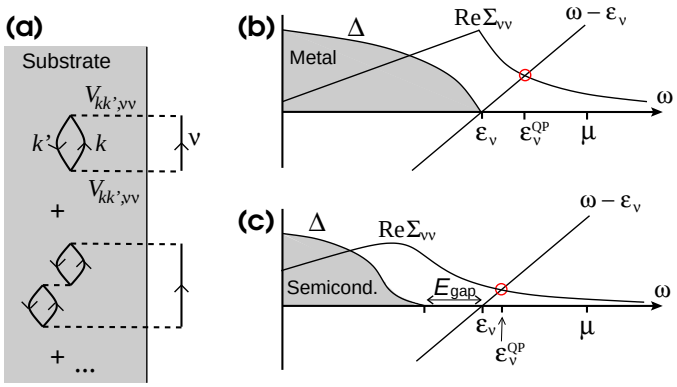


Figure 4.7: (a) Feynman diagrams representing dynamic polarization of the substrate induced by an electron propagating in the molecule. (b) and (c): Generic shapes of the imaginary and real parts of the self-energy of Eq. (4.27) for a metallic and semi-conducting substrate assuming $V_{kk',\nu\nu}$ to be energy independent. From Paper III.

4.3 Application to Molecules

In the field of computational surface science, there is a growing tendency to apply the GW approximation to low-dimensional systems and nanostructures[43, 144, 171–175] and the method seems especially promising for the use in quantum transport[176, 177]. In view of this trend it is important to establish the performance of the GW approximation for other systems than the crystalline solids. In this section first-principles benchmark GW calculations are presented for a series of small molecules. In the literature not many systematic tests of GW appear for molecules, and for those that exist, a series of additional approximations are typically invoked. In this work only the limited basis affects the results, and this approximation is in principle simple to check for convergence, whereas the influence of e.g. the 2e-h approximation or neglect of off-diagonal elements in the self-energy are not easily converged. The relatively large number of atoms does provide some overlap with existing calculations, which are compared to, thereby presenting a testing ground for the implementation itself, and an opportunity to test convergence behavior in basis, energy grid, pairorbital truncation etc.

4.3.1 Basis Set Convergence

As the operators are resolved in a minimal basis set, the accuracy of this should first be validated.

The convergence on adding additional radial and ζ functions to the SZ projected Wannier basis is shown in figure 4.8. Interestingly the addition of polarization functions to the PWF basis does not seem to influence the results, whereas added polarization functions has a significant impact. This is in contrast to common experience for the use of pure NAO basis sets in standard DFT calculations, where polarization functions are usually more important than ζ functions.

In general it is found that the augmented projected Wannier basis is significantly

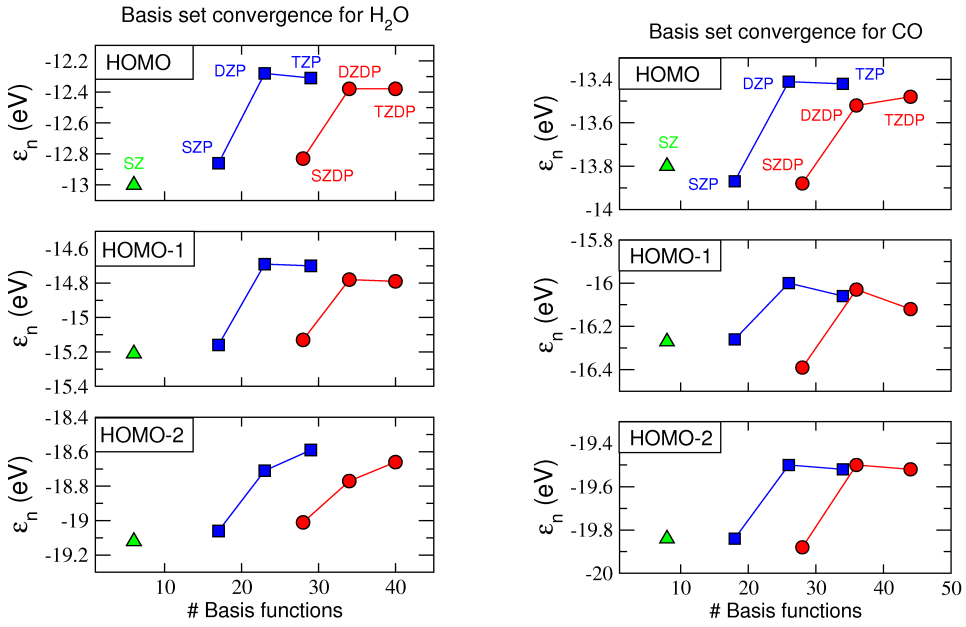


Figure 4.8: Convergence of the three highest occupied levels of H_2O (left) and CO (right) obtained from the spectral peaks of PWF-GW calculations with a variable number of NAO augmentation functions added to the SZ Wannier function basis.

better for HF and GW calculations than the corresponding pure NAO basis. For the peaks of the spectral functions, a 0.1eV convergence is reached at a DZP PWF basis for the GW calculations. Comparing the peaks of basis function resolved Green function HF calculations to the basis set free grid-implementation of HF discussed in section 3.2.5 yields a mean absolute error of 0.09eV for a DZP-PWF basis.

An accuracy of 0.1eV is more than sufficient for the studies done here, which aims at correcting DFT spectra that are 4 eV off and HF spectra that are about 1 eV off. It is also the same accuracy as other published full GW calculations with to the results are compared.

4.3.2 Results

The curent GW scheme assumes a spin-paired ground state, and the molecules considered here therefore counts the 33 molecules of the G2-1 test set [135] (studied in sec. 3.3) satisfying this constraint. Figure 4.9 (left) shows the calculated HOMO energies compared experimental ionization potentials for these molecules.

The geometries are fixed at the values reported in [135]. For the correlated GW self-energy, single eigenvalues are not defined, and the indicated values are obtained from the peaks in the corresponding density of states. It is found that the finite η used, giving the artificial broadening of the spectral peaks, induce a slight dependence in the peak positions on this value. The dependence is linear and the peak position easily extrapolated to $\eta = 0$.

As was observed in section 3.3, the dominant error in the spectrum of the local PBE xc potential stems from the insufficient cancellation of the self-interaction in

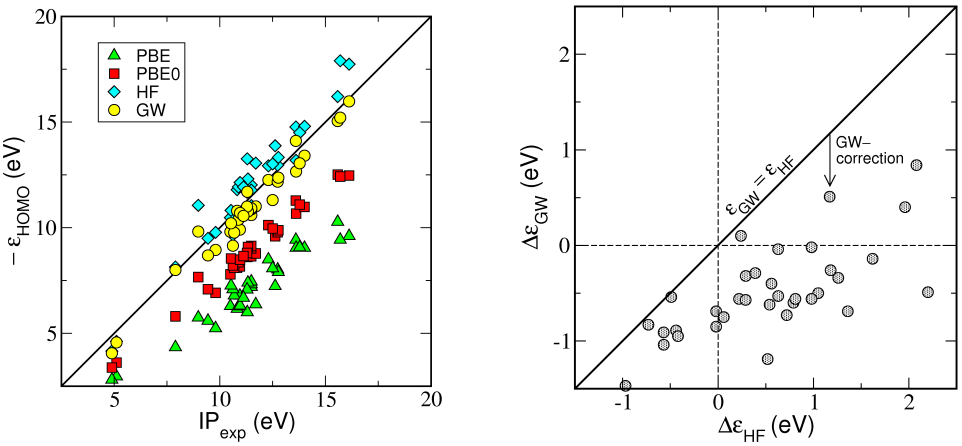


Figure 4.9: Left: calculated ionization potentials compared to experiment. Right: The deviation of the calculated GW (ordinate) and HF (abscissa) ionization potentials relative to experiment. Negative values imply an underestimated IP, i.e. too high HOMO energy.

the Hartree potential, and is gradually improved on going to the hybrid PBE0 and to HF. In this case, the mean absolute error (MAE) relative to experiment is 4.4, 2.6, and 0.8 eV in PBE, PBE0, and HF respectively. In this application, also GW values are included, which gives a further improvement on HF to a MAE of 0.5 eV.

The difference between HF and GW is highlighted in fig. 4.9 (right). The first order term in the GW expansion is identical to HF, and the effect of the dynamical correlation is represented by the vertical displacement of points from the diagonal line in the figure. As is evident from the figure, the effects of adding screening in GW to the bare interaction of HF, has only a minor effect on an absolute scale, but does on a relative scale result in an almost 40% reduced MAE compared to HF.

From Koopmans' theorem [141, 142] the HF eigenvalues represent electron addition / removal energies when orbital relaxations in the final state are neglected. The effect of GW is then in part to describe the response of the electrons to the added / removed electron, i.e. include orbital relaxations in the final state, and in part to introduce correlations into the many-body eigenstates. For isolated molecules the dominant effect is the orbital relaxations, which explains the systematic reduction of the IP in GW compared to HF, as orbital relaxations imply a reduced energy cost of removing an electron. Note that the reduction in the HF IP is not just a rigid shift, but is larger for molecules where HF presents the largest overestimation of the ionization potential. GW generally overcorrects the HF values slightly.

In Paper II, the GW values are compared to results in the literature, and are sound to compare quite well the above results given the differences in the implementation of the GW approximation.

For the molecules studied here, the approximations of the linearized quasiparticle equation (4.17), i.e. that off-diagonal matrix elements of the self-energy can be neglected, and that the frequency dependence can be approximated by its first order Taylor expansion in a sufficiently large neighborhood of $\varepsilon_n^{\mathrm{DFT}}$, are found to be well justified. An example is presented in figure 4.10 which shows the full HF and GW density of states for NH_3 together with the real part of $\langle\psi_{\mathrm{HOMO}}^{\mathrm{DFT}}|\Sigma_{\mathrm{corr}}(\varepsilon)|\psi_{\mathrm{HOMO}}^{\mathrm{DFT}}\rangle$,

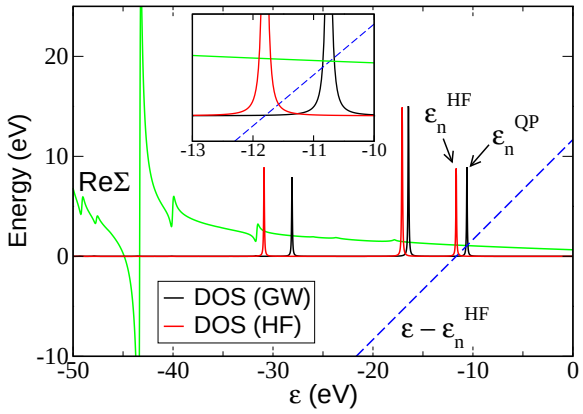


Figure 4.10: Density of states for the NH_3 molecule calculated in HF and GW, respectively. Arrows mark the level corresponding to the HOMO in the two calculations. In the linearized QP equation, the intersection between the line $y = \varepsilon - \varepsilon_n^{\mathrm{HF}}$ and the real part of $\langle \psi_{\mathrm{HOMO}}^{\mathrm{DFT}} | \Sigma_{\mathrm{corr}}(\varepsilon) | \psi_{\mathrm{HOMO}}^{\mathrm{DFT}} \rangle$ (green curve) determines the position of the GW level.

which (since the Fock part is static and thus does not contribute to the first order expansion) gives the QP (GW) energy by the intersection with the line $y = \varepsilon - \varepsilon_n^{\mathrm{HF}}$. Clearly the intersection is right at the GW peak, thus justifying the approximation.

4.4 Application to Electron Transport

In the preceding section it was demonstrated that the GW approximation, originally derived for metallic systems to fix the deficiency of Hartree-Fock theory, could also be applied to isolated systems due to the inherent adaptive screening. For bulk extended systems, GW has over the last many years been extensively tested and is known to give accurate band structures, greatly improving the dramatically underestimated band gaps of DFT, and closing the overestimated gaps of HF. From the application to molecules, GW is seen to be very close to HF, which provides an excellent description of molecular spectra, and even manages to improve slightly on these. The GW scheme is therefore expected to perform well also for the hybrid transport setup, where an accurate description of the molecular resonances as well as the electrode response to the charging of the molecule, is crucial to describe the process of electron transfer between the Fermi sea of the leads, and the molecular resonances. From the diagrammatic expansion of the GW screening operator, it is expected to be able to account for the long range electrostatic image charge formation via the infinite order electron-hole pair generation in the metal as a response to the molecular excitation.

For the strongly coupled molecules, where hybridization between metal and molecular states leads to broad resonance supporting direct charge transfer and high transmission probabilities, DFT has already been successfully applied to a wide range of systems. It is for the weakly coupled systems, where the severe underestimation of the fundamental gap in DFT leads to a corresponding overestimation of the transmission probability of electrons, sometimes by several orders of magnitude,

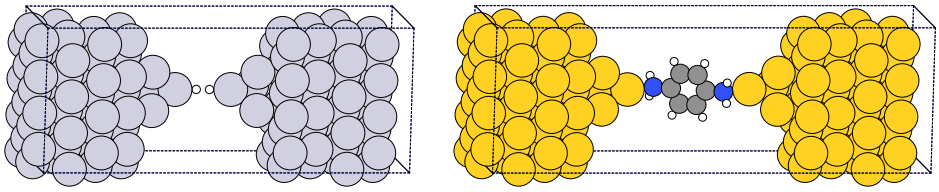


Figure 4.11: Setup used to model the Pt/H_2 (left) and Au/BDA junction (right). Structures are optimized using the PAW-DFT code GPAW with the PBE xc potential. Bond lengths are: Pt–H = 1.7 Å, H–H is 0.9 Å, and Au–N = 2.54 Å. The (orthorhombic) cell dimensions are $28.5 \times 11.3 \times 9.8$ Å for the Pt system and $35.6 \times 11.8 \times 10.2$ Å for the Au system.

that the real promise of GW lies. Despite of this, to test the range of applicability, the GW scheme is in the following applied both to transport through a strongly coupled, highly conducting molecule and a weakly coupled molecule where transport is governed by single electron tunneling and long excitation lifetimes. As a prototype of the former hydrogen in a platinum junction (Pt/H_2) is used, as it has been intensively studied both experimentally and theoretically, and as a typical example of an experimentally well characterized system for which DFT is known to fail, benzenediamine[1] coupled to gold electrodes (Au/BDA) is considered.

The simulation setups used for the two junctions are shown in figure 4.11. As DFT gives an accurate description of the (Born-Oppenheimer) potential energy surface, and therefore also the forces and equilibrium structures, both have been relaxed using DFT (with PBE xc) in GPAW. For the much studied Pt/H_2 system, it is generally accepted that the shown geometry offers a good description of the actual experimental configuration. This junction has an experimental conductance of about $1G_0$ which is also the prediction of DFT. Less is known about the precise geometry of the Au/BDA junction, but it has been shown by [39] based on both measurements and DFT calculations that conductance variation with different geometries is small, due to the nature of the amine-gold bond. Typical DFT conductance calculations are in the range 0.02-0.06 G_0 which is almost an order of magnitude too large compared the experimental value of 0.0064 [178], and for this reason only the single geometry shown in 4.11 is considered, without further variation of configuration space.

Structure optimization was done in GPAW using the PBE xc functional, and a converged grid-spacing and k-point sampling. Due to the computational effort of the GW calculations, all transport calculations are done in the Γ point approximation (also DFT and HF calculations for consistency) which is somewhat crude for the 10 Å cell size. The leads are described at the DFT-PBE level using principal layers three atomic layers deep.

Before showing the GW based transport calculations of the two bulk junctions in sections 4.4.4 and 4.4.5 respectively, some initial calculations are presented in sections 4.4.1 on the isolated H_2 and BDA molecules, and sec. 4.4.2 on some technical convergence properties of GW. These initial calculations are relevant for the interpretation of the bulk GW transport calculations.

[1] Specifically the considered molecule is 1,4-benzenediaminie also known as p-phenylenediamine (PPD) or p-aminoaniline

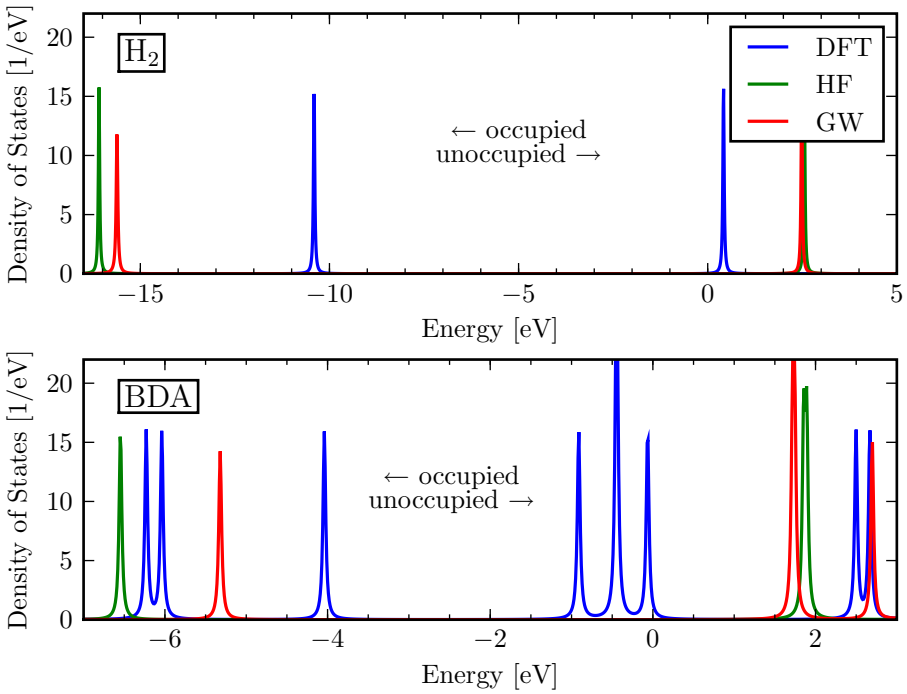


Figure 4.12: Density of states for the isolated H_2 (top) and BDA (bottom) molecules evaluated at the DFT, GF, and GW level. The energy zero is the vacuum level. Basis is a dzp PWF (spanning DFT states below 3 eV) for H_2 and dz PWF (spanning DFT states below 0 eV) for BDA. $E = [-150; 150]$ eV, $\eta = 0.02$ eV. The H_2 HOMO is -10.4, -16.1, and -15.6 eV in DFT, HF, and GW respectively corresponding values for BDA are -4.0, -6.7, and -5.3 eV. The experimental IP of the two molecules is 15.4 and 6.9 eV [137, 179].

4.4.1 Gas Phase Molecules

Before applying the HF and GW approximations to the contacted molecules, it is relevant to consider them in their gas phase. Figure 4.12 shows the DFT, HF, and GW DOS for the isolated H_2 and BDA molecules. The molecules follow the same trend as those in the database of the molecular study (hydrogen and BDA are not part of that database), with HF opening the gap compared to DFT, and inclusion of correlations in GW then closes it a bit again. The effect of screening is less pronounced for H_2 than BDA, due to the reduced electron number, which implies a weaker density response to excitations and a lesser degree of orbital relaxation.

Focusing first on the H_2 spectrum, this behaves as expected. Compared to the experimental IP of 15.4 eV, the DFT HOMO is much too large. Removal of self-interaction errors in HF brings the HOMO close to experiment, but overcorrects the DFT value. The small reduction in ionization cost of 0.5 eV by allowing dynamical orbital relaxation (self-screening) of the HF system, brings the GW value almost on par with experiment, with a deviation of only 0.3eV. Note that for both molecules, the LUMO states in HF and GW are very close. Apparently the closed shell states

of the ground state does not allow much redistribution in response to an added electron.

The situation is different for BDA. The experimental data for the IP of BDA in ref. [179], varies between 6.9 and 7.6 eV. Estimating it by a PBE-DFT total energy difference in GPAW between neutral and (spin-polarized) charged state yields 6.64 eV, which is presumably close to the real vertical ionization energy for this geometry. The HF calculation for BDA shows a peak at exactly this value, and the considerable shift in GW (by 1.3 eV) thus severely degrades the agreement with the expected ionization energy.

The calculation is well converged in energy grid and η, the only concern is the limited dz-PWF basis for BDA. Polarization functions are usually included, but a dzp basis for BDA contains 144 orbitals which is numerically intractable. From the basis test in section 4.3.1 polarization functions are however not expected to be important for a PWF description of molecules. Comparing the peaks corresponding to occupied states with an accurate grid based HF calculation in GPAW shows agreement within 0.15 eV (unoccupied HF eigen energies are overestimated in the dz basis, as expected when describing delocalized states in an atomic basis). This is maybe not surprising as the Wannier basis is an exact span of the (occupied) DFT states, which is a good span of HF/GW MOs. This does not prove that the dz-PWF basis provides an accurate span of the polarization or screened interaction. A comparison of sz-PWF and dz-PWF GW results shows only a minor effect (0.1 eV) on the occupied states, and from the experience of section 4.3, it seems plausible that this is *not* a numerical problem, but rather that GW does not give a good description of BDA for some reason. This would then be due to the neglected diagrams in the approximation, most notably, the second order cross diagram of figure 2.5. Note that G_0W_0(LDA) calculations on the very similar benzene molecule by J.-M. Garcia-Lastra [Paper III] performed in a plane wave code using the plasmon pole approximation, gives a similar underestimation. In comparison, the GW HOMO of benzene is found to be -7.9 eV (see fig. 7 in Paper III) and thus underestimates the experimental IP of benzene of 9.2 eV by exactly the same amount (1.3 eV) as observed here for the BDA molecule.

4.4.2 Hydrogen in 1D Platinum Wires

When applying the GW self-energy to the transport setup, there are a few additional technical details besides those for the isolated system. This section is focused on the study of these, using a more simple wire system as a test base for the bulk junction of interest. The reduced system consists of monoatomic platinum wires connected to a hydrogen molecule, as shown in figure 4.13.

When using interaction self-energies in the transport setup, a limited region in which to described *e*-*e* interactions must be defined. First of all because the system partitioning and hence the Meir-Wingreen current expression requires a finite interaction region, and secondly the computational cost of the GW scheme is too high to apply to a large region. Limiting the region for which the interaction self-energy should be applied implies a restriction of the orbital indices i, j in $\Sigma_{ij}(\tau, \tau') = \sum_{kl} W_{ik,jl}(\tau, \tau') G_{kl}(\tau, \tau')$ to this region. For the self-energy to be well-defined the k, l indices must be included in a larger region. In general large enough for the pairobitals defined by ik and jl to no longer have overlap with the screened interaction $\epsilon^{-1}V$ beyond this extended region. For HF the enlarged region must be large enough to include all overlaps with the bare Coulomb operator.

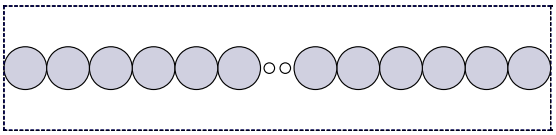


Figure 4.13: Scattering region of the wire Pt/H_2 system. Black box indicates the simulation cell. Transverse directions are treated non-periodic. Transport direction is extended by infinite leads described by a 5 Pt atom principal layer. Bond lenths are Pt–Pt = 2.4 Å, Pt–H = 1.7 Å, and H–H = 0.9 Å.

As the metal is well-described by DFT, the region of the GW self-energy will in all cases considered here be restricted to the molecule only. To include more basis functions in the sum as described above in practice; the self-energy is first determined in the extended region which includes several metal atoms, and is consecutively truncated to the molecule by replacing the values of the self-energy at the metal atoms by the DFT xc self-energy. Also the coupling between the active GW region (the molecule) and the inactive GW region (the additional metal atoms) will be replaced by the mean-field xc approximation, such that

$$\mathbf{\Sigma} = \begin{bmatrix} \Sigma_{\mathrm{xc}} & \Sigma_{\mathrm{xc}} & \Sigma_{\mathrm{xc}} \\ \Sigma_{\mathrm{xc}} & \Sigma_{GW} & \Sigma_{\mathrm{xc}} \\ \Sigma_{\mathrm{xc}} & \Sigma_{\mathrm{xc}} & \Sigma_{\mathrm{xc}} \end{bmatrix}, \tag{4.29}$$

where the subdivision is in the left inactive GW region, the active GW region (the molecule), and the right inactive GW region respectively.

It is important to stress, that even-though the GW self-energy is only used on the molecule, it is defined in terms of a larger region, which in a diagrammatic phrasing means that the external vertices are restricted to the molecule while the polarization bubbles can exist in the metal, and screen excitations in the molecule by electrostatic interaction. Thus the image charge effect is fully accounted for by this approximation.

To test the range of the (screened) Coulomb operator in the Fock (GW) self-energy, the calculated DOS and transmission function are determined using a variable number of additional Pt atoms on each side of the molecule in the construction of the interaction self-energy (which only applies to the molecule itself), as shown in fig. 4.14. Clearly the first Pt atom is of importance, especially for the bare Coulomb operator in HF, while the GW self-energy, which includes HF as the first order term, is somewhat less influenced by this inclusion. When one or more Pt atoms are included in the interaction self-energy, the molecular states hybridize more strongly with the metal *d*-band, resulting in an increased spectral weight in the range -6.5 to 0 eV and by extension the transmission function becomes more sensitive to the van-Hove singularity at the upper *d*-band edge. This issue will be revisited based on more detailed calculation in fig. 4.16. From the calculations shown in fig. 4.14 it is concluded that 4 Pt atoms are sufficient for an accurate description of the self-energy, and this number will be used throughout the remaining calculations.

The transmission function shown in figure 4.14 is in principle only relevant for non-interacting systems (and also in the case of static interactions like HF), in which case the Meir-Wingreen expression for the current reduce to an integral over the transmission function within the bias window, and the zero-bias linear conductance

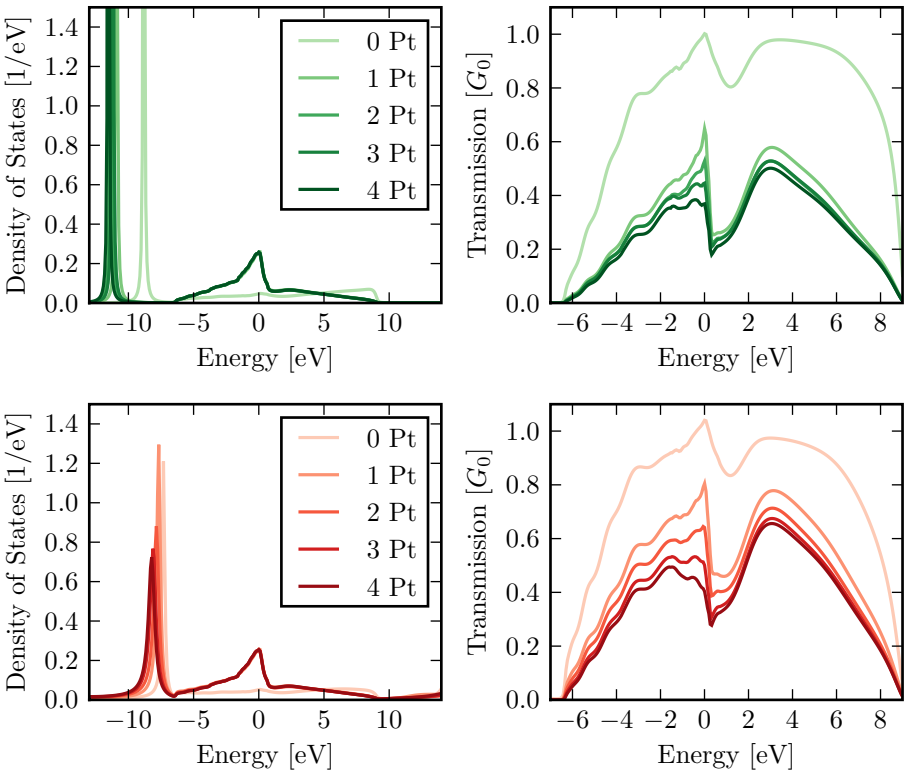


Figure 4.14: Convergence of the molecular DOS (left) and transmission function (right) for the HF (top, green) and GW (bottom, red) approximations on including additional Pt atoms on each side of the molecule in the construction of the self-energy.

is given by the value of the transmission function at the chemical potential. For dynamic interaction self-energies (like GW), the integrand of the Meir-Wingreen formula will in general have weight outside of the bias window. To test if at least the linear conductance can still be related to the transmission function evaluated with the GW self-energies and Green functions a comparison of a finite bias and transmission function approach is given in fig. 4.15.

The smallest possible bias that can be applied is given by the grid spacing of the energy grid, and thereby related to the size of η giving the artificial broadening of spectral peaks. The transmission function also depends indirectly on this parameter via the Green function. As is evident from figure 4.15, both transmission function and finite bias conductance are linear in η, and converge towards the same limiting value. Based on this, the transmission function will be depicted and interpreted for all three xc approximations DFT, HF, and GW, although not formally justified for the latter. As a comment on the η dependence, I find based on extensive testing not shown here, that for sufficiently small eta (in practice around 0.02eV for all systems considered here) the transmission function is point-wise (in the energy grid) linear in η and the entire curve can therefore be extrapolated to $\eta = 0$ from two calculations at different η. As was also observed in the study of the molecular database, this

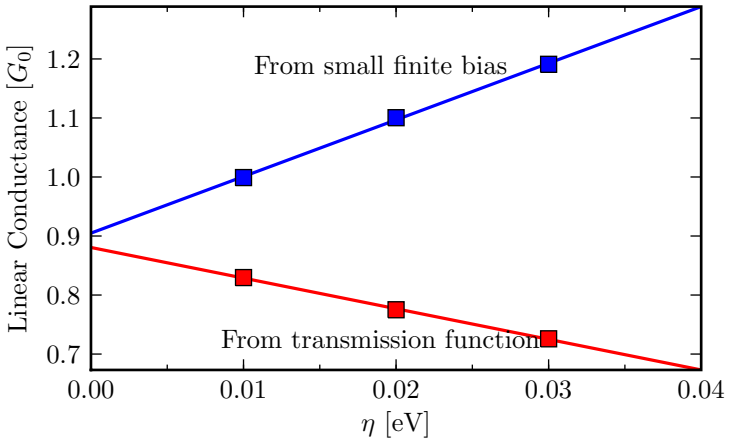


Figure 4.15: The zero-bias conductance using the GW approximation, as estimated by evaluation the transmission function at the Fermi energy and by calculating I/V for a small but finite bias ($V = \Delta E = 0.5\eta$).

linear relation also holds for the peak positions of spectral function. For coupled systems this is found to remain valid for the peaks of sharp resonances, but the general functional behavior of the DOS can *not* be extrapolated in this fashion.

To conclude the study of the Pt wire / H_2 system, figure 4.16 gives a comparison of the DFT, HF, and GW density of states and transmission function. Looking first at the spectral peaks associated with the molecule (full lines in top plot), the usual trend is seen repeated; HF opens the gap compared to DFT, and GW then close it a bit again. This is then directly reflected in the transmission function, which is largest in DFT, significantly reduced in HF, and in between for GW.

The metal DOS (included in the total DOS, i.e. dotted lines) has a broad low s-band which extends up to 9 eV, and a higher more narrow d-band ranging from -6.5 to 1. As the system is one dimensional, van-Hove singularities appear, and the z axis is not symmetry related to the xy directions thus splitting the d band. The metal band restricts the range in which the transmission function can be finite, as this is related to elastic coherent transmission, and metal states must therefore be available at a given energy to give any transmission.

A useful tool is to look at the diagonal matrix elements of the static (i.e. non-correlation related) Hamiltonian spanned by the renormalized molecular eigenstates of DFT (determined by subdiagonalization of the DFT Hamiltonian in the space of orbitals on the molecule). These will in the following be referred to as RMO eigenvalues, and are indicated with circles in fig. 4.16. The DFT HOMO-LUMO states at -7 and 4eV are separated considerably to the HF/GW RMOs at -11 and 9 eV. Since the static part of GW, for which the RMOs are constructed is the HF Hamiltonian, these are almost on top of each other. The slight difference is due to redistribution in the density matrix during self-consistency in the GW Green function.

Comparing the RMO eigenvalues to the corresponding spectral peaks, it is seen that hybridization with the metal states, causes the LUMOs, which lie within the metal s band, to split into broad bonding resonances at the Fermi level, and much

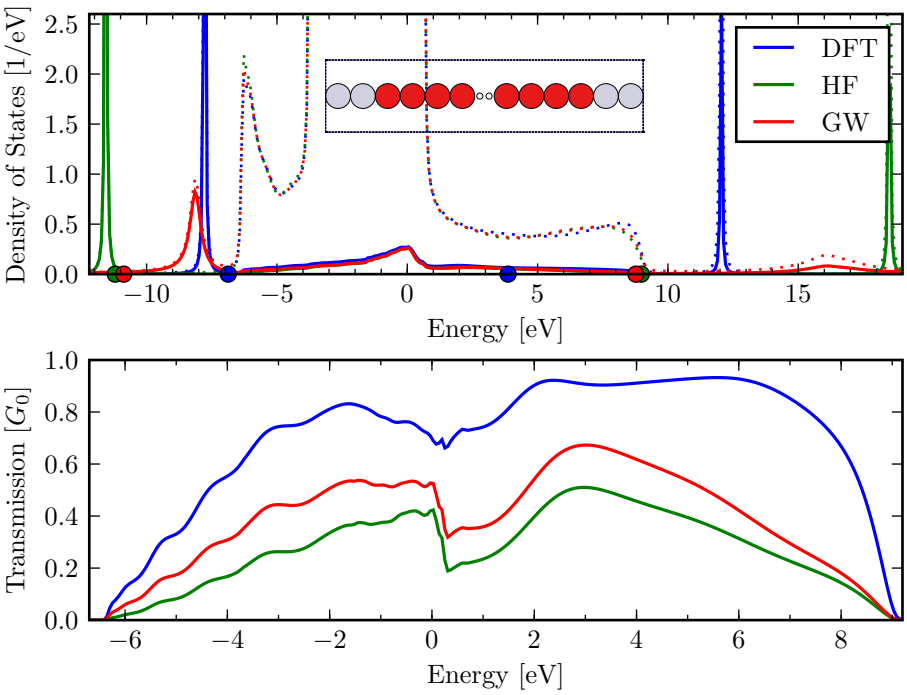


Figure 4.16: Insert shows the Pt wire / H_2 system, red Pt atoms indicate atoms included in the interaction self-energies in the molecule. Top plot shows the total DOS (dotted) of the system and the projected DOS on the molecule (full lines) in the DFT, HF, and GW xc approximations. Circles indicate the matrix elements of the static Hamiltonian (i.e. DFT or HF part) in the DFT RMO basis. Bottom plot shows the transmission function in the three xc approximations (extrapolated to $\eta = 0$). Basis is a PWF szp, $E = [-160; 160]$ eV, $\eta = 0.03$ eV.

higher anti-bonding resonances to the other side of the band edge. The HOMOs are all below the metal band edge, and the DFT and HF (bonding) spectral peaks are therefore almost on top of the eigenvalues with a slight downshift as expected from a Newns-Anderson [180] construction (an anti-bonding hybridization may also form for the HOMO contributing to the plateau around ϵ_F). The GW resonance not only shifts the wrong way, but also by a significant amount. This is due to the dynamic correlation not included in the RMO, which reduce the removal energy of HOMO electron. The upshift of the GW HOMO resonance can thus be taken as a measure of the strength of the image charge effect (and to a smaller extent the molecular self-screening) which is clearly very strong as expected from the short metal molecule separation (1.7 Å) and large metal DOS at ϵ_F. The strong correlation can also be recognized from the much reduced lifetime (increased spectral width) of the GW peaks compared to the DFT and HF peaks which for resonances outside of the metal band is only limited by the finite η. From the shift of the GW spectral peak compared to the RMO HOMO, the image charge effect can be estimated to

about 3.5 eV, bringing GW almost up to DFT, assuming a downshift due to metal hybridization of about 1 eV.

Integration of the DFT molecular DOS yields a molecule occupation of 3.24 electrons. The electron pileup is caused by the broad resonance at the Fermi level due to hybridization of the molecular LUMO. Hybridization of the HOMO causes this to be slightly depleted.

Opening the gap in GW and even more in the HF spectrum, *de*populates the LUMO, while increasing the occupation of the HOMO. For this system the dominant effect is the former, which leads to a reduced Hartree repulsion for HF compared to GW, and hence the lower RMO HOMO eigenvalue. For the LUMO, an increased Fock contribution cancels this effect and moves HF LUMO element to the other side of GW.

For this system, the HOMO is mostly occupied in all three approximations. For DFT, the RMO LUMO is occupied by 1.24 electrons. Going from DFT to GW to HF, the LUMO is gradually pushed up in energy, depopulating the LUMO by 0.04 at each step. The higher molecular occupation in GW compared to HF leads to a 0.25 eV larger Hartree repulsion, which reduces the electron removal cost, and the binding energy of an added electron, and hence move the GW RMO HOMO and LUMO closer to the Fermi level by this amount compared to HF.

4.4.3 Basis Set Issues

Moving from the wire to the more realistic systems of figure 4.11 with bulk electrodes, it turns out to be difficult to avoid linear dependencies in the PWF basis set. Also the lack of bounded support of the PWF basis can cause problems. For this reason, the bulk electrodes will be treated in a plain LCAO basis. This section will address to which degree the reduced quality basis affects the presented results.

Figure 4.17 shows the transmission function of the two systems in consideration in three different basis set sizes calculated at the DFT level, for which large basis sets do not present a computational burden. Here the large dzp basis is expected to give an accurate description of the system, while the smallest basis set ("szp/sz") is the only manageable basis size for GW applications. In this basis, all atoms which are included in the description of the GW self-energy are described at a sz level, while the remaining atoms described at the mean-field level can be any size; here restricted to szp.

For the Pt/H_2 system, the low quality basis seems to give a quite good resemblance to dzp, while it is a bit more dubious for the Au/BDA system, giving a reduced peak at the LUMO resonance and a slightly too large transmission at the Fermi level. Despite this minor discrepancy, the level of agreement is actually surprising, as a sz LCAO basis is a very coarse representation of the wave functions. At least for molecules it is known to give qualitative wrong results. Figure 4.18 (left) shows the eigenvalues of the isolated BDA molecule in different basis sets compared to an accurate grid calculation. To make the eigenvalues comparable, the values from the LCAO calculations have been shifted to align the value of the HOMO to the grid calculation. For the sz calculation this entails a shift of 1.6 eV.

In contrast, when subdiagonalizing the molecular Hamiltonian of the contacted BDA, as shown in figure 4.18 (right), the agreement is much better, and does not require a shift. Note that in the LCAO basis, the Hamiltonian spanned by orbitals on the molecule can be diagonalized to obtain the RMO values, which is not possible in the grid representation, hence the missing grid values. In the contacted system,

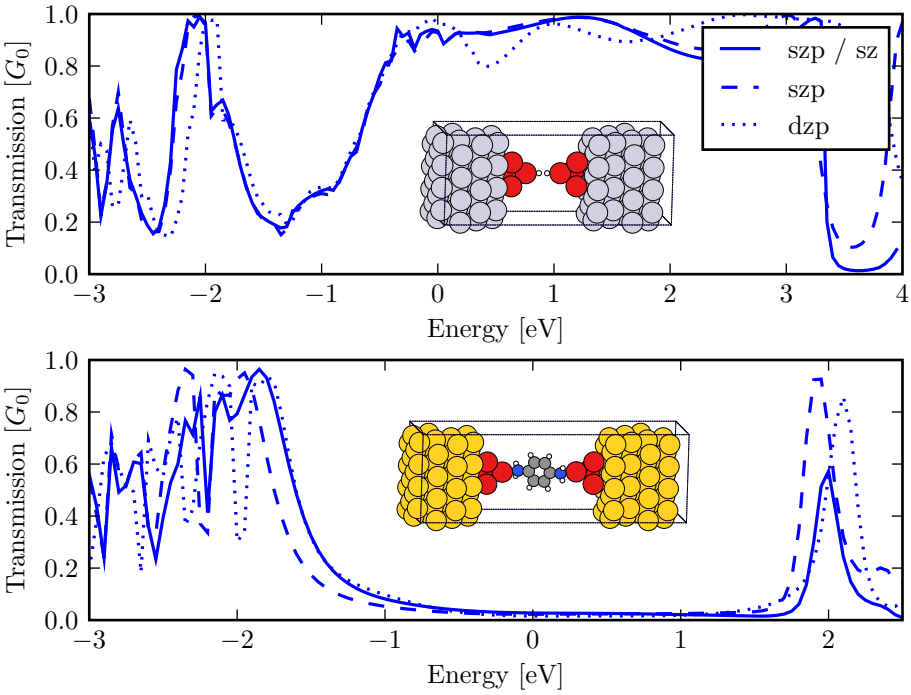


Figure 4.17: Basis size convergence for the DFT transmission of Pt/H_2 (top) and Au/BDA (bottom). The different curves indicate different LCAO basis sets, where "szp/sz" is a combined basis for which szp is used for the electrodes, and sz for the metal pyramids (marked in red) and the molecule.

hybridization with metal states causes a finite spectral width of the RMO HOMO and LUMO which consequently have finite overlap with the Fermi level. This results in a pinning of the absolute position of the HOMO. Interestingly the energetically lower molecular states are also in close agreement between sz and dzp, which may be due to overlap of the nearby metal sz basis with the molecular region resulting in an improved description via basis set superposition. The cutoff radius of gold s and d orbitals are 4.5 and 3.1 Å respectively, which gives substantial overlap with the BDA molecule (the Pt tip – amine separation is 2.5 Å).

4.4.4 Hydrogen in a Platinum Junction

The experiments performed by [181] on the conductance of a single hydrogen molecule, have been the subject of intense theoretical scrutiny [181–186]. The interest stems amongst other things from the simplicity of the system which constitutes the smallest imaginable molecular contact. An interesting point is that the measured conduction is close to 1 G_0. The perfect transmission is surprising for the closed shell, normally inert, molecule.

The structure shown in the insert of fig. 4.19 is found to reproduce many of the experimentally observed properties of the Pt/H_2 junction such as linear conductance [184], stretching dependence of vibrational modes [185] and to some degree the

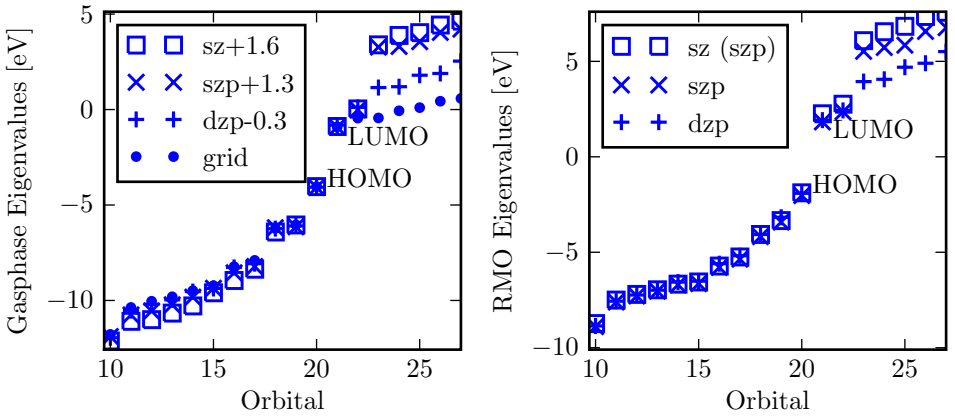


Figure 4.18: Left: Eigenvalues af gas phase BDA relative to the vacuum level. Right: RMO eigenvalues of the contacted BDA molecule relative to the Fermi level. In the gas phase, inadequate basis sets cause a more or less rigid shift of the eigen values. For the contacted molecule, peak broadening due to hybridization causes a pinning of the RMO-HOMO resonance. The parenthesis (szp) indicates that the electrode is described by a szp basis, but the molecule (and pyramids) are in a pure sz description.

inelastic transmission [186].

The DFT transmission shown in the fig. 4.19 (blue curve, bottom plot) has a wide plateau around the Fermi level, which is attributed to a wide resonance due to hybridization of the anti-bonding H_2 orbital with the metal states leading to a fully open transmission channel [184]. The transmission predicted from plain DFT is in good agreement with the experimental conductance, which is generally true of high transmission systems.

For construction of the interaction self-energies, orbitals on the pyramid atoms (marked red in the insert) have been included in the sum, but as before only the description of the molecule itself is modified, with the platinum and the coupling between platinum and molecule taken from the local xc DFT description. The basis set is sz LCAO on the pyramids, and szp elsewhere. A dzp description of hydrogen or full szp description everywhere does not affect the results significantly.

The conclusions are much the same as for the wire system. The HF and GW RMO HOMOs are very close (indicating small density redistribution between the two) at -10 eV, which is 3 eV below the DFT eigenvalue. Hybridization shifts the spectral peak of the HOMO down by 1 eV compared to the eigenvalue, while polarization effects for GW move the spectral peak up by 3 eV, in close agreement with DFT, although with a reduced lifetime. The hybridization of the molecular orbitals with the metal gives a wide resonance around the Fermi level. An anti-bonding hybrid of the LUMO is formed beyond the shown energy window.

The HF transmission is reduced by a factor of two at the Fermi level compared to DFT due the shifted LUMO. GW recovers somewhat, but unfortunately does not bring the transmission back to 1 G_0 which is the observed conductance in both experiment and DFT based simulations.

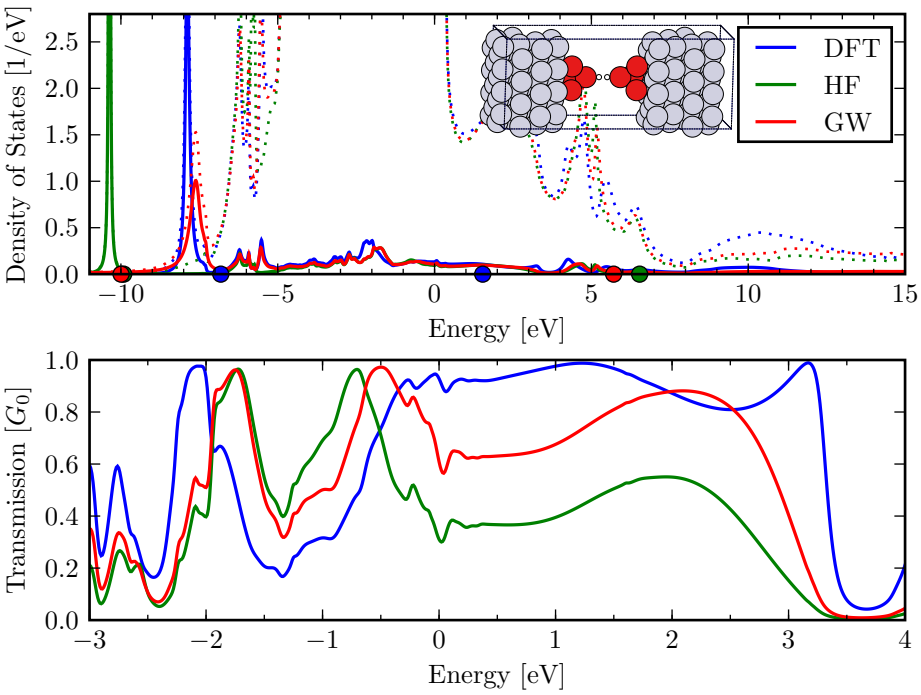


Figure 4.19: Insert shows the Pt/ H_2 system, red Pt atoms indicate atoms included in the interaction self-energies in the molecule. Top plot shows the total DOS (dotted) of the system and the projected DOS on the molecule (full lines) in the DFT, HF, and GW xc approximations. Circles indicate the matrix elements of the static Hamiltonian (i.e. DFT or HF part) in the DFT RMO basis. Bottom plot shows the transmission function in the three xc approximations (extrapolated to $\eta = 0$). Basis is an LCAO szp on electrodes + H_2 and sz on the pyramids (red atoms), $E = [-150; 150]$ eV, $\eta = 0.02$ eV.

4.4.5 Benzenediamine in a Gold Junction

The second system considered here is the Au/BDA junction shown in figure 4.11 right. Experimentally this junction is found to have a linear conductance of 0.0064 G_0 [178], putting it in the opposite limit of the Pt/H_2 system, with a low tunneling transmission probability. In this case DFT based transport calculations typically overestimate the conductance by an order of magnitude. In [39] they argue from the low spread of the measured conductance on repeated break junction experiments, and a similar small variation in values predicted by DFT when considering a wide range of different stable junction geometries, that the disagreement between calculated and measured values is not due to geometrical effects, but rather electron correlations. The transport is found to be determined by the slowly decaying tail of the molecular HOMO resonance, and hence the positioning of this relative to the Fermi level is crucial for an accurate prediction of the conductance. In both the case of BDA [39] and the larger class of BDA and BDT (benzenedithiol) based molecules decorated by various functional groups, as studied by [187], a good agreement with experiment

can be obtained from the DFT calculation, by an ad-hoc adjustment of the DFT gap to correct for self-interaction errors (which leads to an overestimation of the gas phase HOMO), and lack of dynamical screening in the DFT description. The gap is corrected by manually shifting all occupied orbitals down by a fixed energy, and all unoccupied orbitals up by the same amount (denoted a scissor operator). The size of the energy shift is determined by two things: 1) the energy shift required to shift the HOMO to be in accordance with the ionization energy (determined by DFT total energy differences) and 2) a reduction of this shift by the energy of the induced image charge in the metal due to charging of the molecule (determined by modeling the metal surfaces as perfect conductors and the molecule by a series of point charges). In both [39] and [187] the model predicts a shift due to image charge formation of about 1 eV.

Since GW both removes self-interaction in the Hartree term, giving good molecular spectra, and includes a description of dynamical screening, the hope is that this can produce accurate predictions of the conductance by direct application without the need for an ad-hoc postprocessing correction scheme as presented in [39, 187].

The density of states and transmission function determined for the Au/BDA by application of the here presented GW scheme are shown in fig. 4.20 along with DFT and HF results for comparison. The calculational details are similar to those for the Pt/H_2 system, except in this case a sz basis is used for both pyramids and molecule, due to the computational effort associated with the increased number of basis functions for BDA compared to H_2. The electrodes are still treated in a szp basis.

In the case of gold, the *d*-band is fully occupied and is not particularly reactive. Also hybridization with the *s*-band seems to be weak for all but the HOMO state. Thus the non-HOMO resonances are all sharp with long peaks (long life-time excitations), and for the non-correlated self-energy approximations DFT and HF, the resonances are exactly at the position of the RMO eigenvalues. The GW resonances are all shifted by about 2 eV compared to the (static) RMO values due to electron correlation. These correlations are related to a reduced ionization energy on removing an electron from occupied states or adding one for the unoccupied states, and is therefore largely independent on which state is ionized, and changes sign for occupied / unoccupied states. Although the LUMO state in DFT at 2.2 eV does seem to hybridize weakly giving a smaller resonance at 1.9 eV also visible in the transmission function, transport is largely dominated by the HOMO, and for GW and HF the LUMO is shifted far from the metal band to 6.6 and 9 eV respectively. The HOMO resonance is quite broad in all three cases, and lies about -2.2/-2.4 eV for DFT/GW and at -3.5 for HF. These resonances are directly recognizable in the transmission function. Notice that in this case GW does not only move the HF peak towards, but actually beyond the DFT peak. This is probably related to the exaggerated polarization effect also seen for gas phase BDA. Subtracting the superfluous polarization shift seen for the molecule (1 eV), from the total correlation shift of the GW peaks of 2 eV observed here, gives an isolated image charge effect of 1 eV in close agreement with the model calculations of [39] (1 eV) and [187] (0.9 eV).

In HF, the gap is large and the molecular resonances sharp. Integrating the molecular DOS thus gives a total occupation of exactly 42 electrons (neutral molecule). In DFT and GW, the proximity and resonance width of the molecular HOMO to the Fermi level leads to a slight depopulation of the molecule by 0.1 (DFT) and 0.4 (GW) electrons. Consequently the Hartree contribution is decreased, which is the

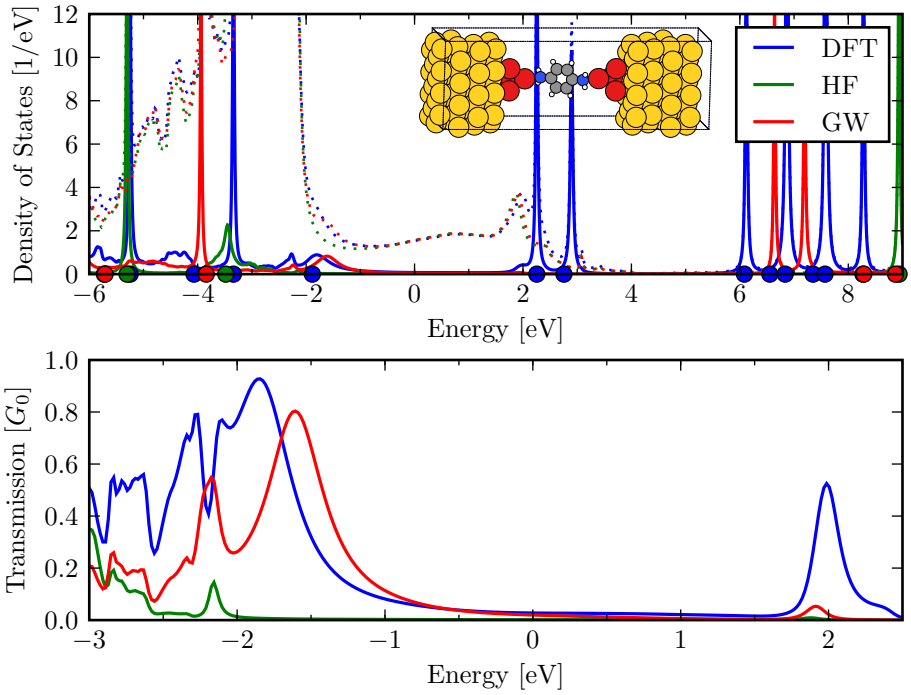


Figure 4.20: Insert shows the Au/BDA system, red Au atoms indicate atoms included in the interaction self-energies in the molecule. Top plot shows the total DOS (dotted) of the system and the projected DOS on the molecule (full lines) in the DFT, HF, and GW xc approximations. Circles indicate the matrix elements of the static Hamiltonian (i.e. DFT or HF part) in the DFT RMO basis. Bottom plot shows the transmission function in the three xc approximations (extrapolated to $\eta = 0$). Basis is an LCAO szp on electrodes and sz on BDA + pyramids (red atoms), $E = [-150; 150]$ eV, $\eta = 0.02$ eV.

primary cause of the largely static down shift of GW RMO eigenvalues compared to HF.

Despite the HOMO resonance moving closer to the Fermi level in GW compared to DFT, the zero-bias conductance is a bit smaller due to the complete lack of a nearby LUMO. The value of the transmission function at the Fermi level is better distinguished on a logarithmic scale as in figure 4.21. This shows a zero-bias transmission of 0.028, 0.019 and 0.0016 G_0 for DFT, GW, and HF respectively. The HF result is 4 times too small compared to experiment, and DFT 4 times too large. GW is an improvement, but is still to close to the DFT value, which is presumably due to the faulty description of the BDA molecule.

It is likely that instead of summing polarization bubbles to all orders as done in GW, it may be better to use all second order diagrams (a 2. Born approximation[188]). This includes the first order polarization term and the cross diagram as discussed in section 2.5. As it includes all (both) terms of second order,

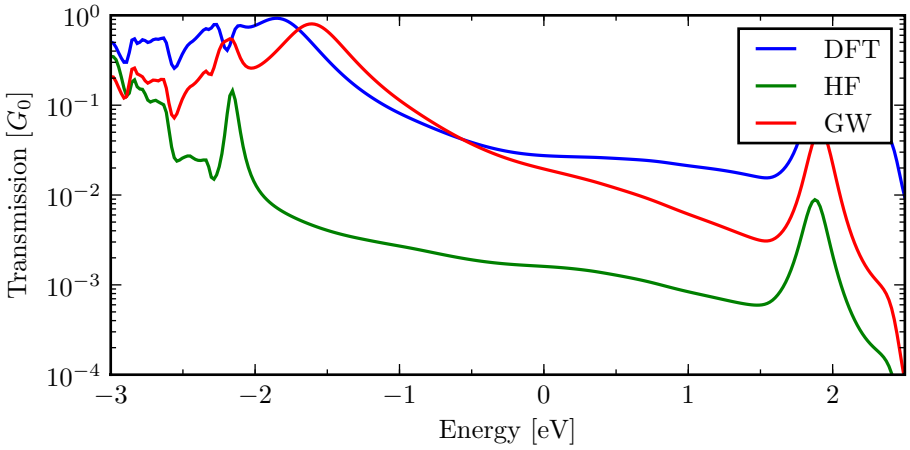


Figure 4.21: Transmission functions of fig. 4.20 on a logarithmic scale.

it is completely self-interaction free also in the polarization. On the other hand 2. Born only treats polarization to lowest order. For bulk metals, the infinite order polarizations in the GW approximation is vital, but as it is only used to describe the metal polarization effect on the molecule in this application, the lowest order inclusion may be sufficient. Note that although the second order self-energy diagrams look similar in form to the second order Møllet-Plesset (MP2) energy, which is known to diverge for metals, the Green function variant is different, and the inclusion in the Dyson equation implicitly sum all reducible diagrams to infinite order [188]. In practice summing diagrams to second order in NEGF schemes does not lead to divergent terms for metals like MP2.

Accepting that GW does not describe gas phase BDA correctly, this can be corrected for by applying a scissor-operator (SO), which corrects for the 1.3 eV overestimation of the molecular HOMO. In practice this can be done by adding the static retarded SO self-energy operator

$$\widehat{\Sigma}_{\mathrm{SO}} = -1.3\mathrm{eV}\sum_{n}^{\mathrm{occ}} |\psi_n\rangle\langle\psi_n|, \tag{4.30}$$

where ψ_n are the renormalized molecular eigenstates of the junction, and the sum is restricted to the occupied molecular states.

Resolved in the non-orthogonal basis LCAO basis, this operator is equivalent to the matrix which in the subspace of orbitals on the molecule is given by

$$\boldsymbol{\Sigma}_{\mathrm{SO}} = -1.3\mathrm{eV}\boldsymbol{S}\boldsymbol{U}\boldsymbol{U}^{\dagger}\boldsymbol{S}. \tag{4.31}$$

Here $\boldsymbol{U}$ is the rotation matrix satisfying $\boldsymbol{U}^{\dagger}\boldsymbol{H}_0^{\mathrm{mol}}\boldsymbol{U} = \boldsymbol{\epsilon}_{\mathrm{RMO}}\boldsymbol{S}$, (where all quantities refer to orbitals on the molecule) but truncated to the columns corresponding to occupied states only. With this additional self-energy, the DOS and transmission changes as shown on fig. 4.22

After applying the scissor operator (SO), there are exactly 42 electrons on BDA (neutral), while without it the tail of the HOMO resonance led to a $0.04e$ depopulation. This gives a rigid increase of all MOs by 0.5eV due to an increased Hartree

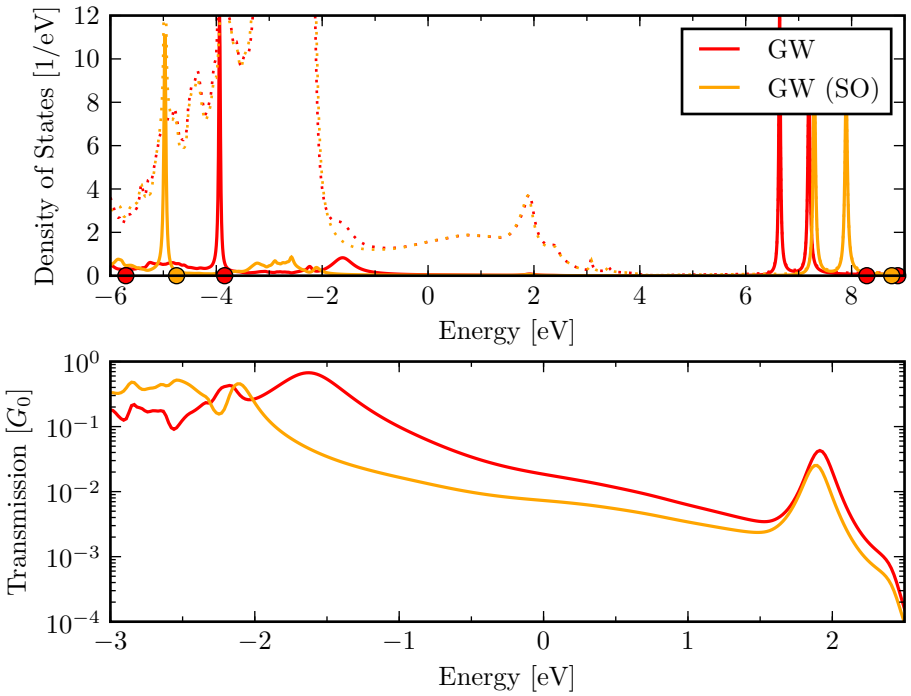


Figure 4.22: Red curves as in fig. 4.20. Orange curves utilize the additional scissor-operator (SO) self-energy.

energy. The Fock self-energy is unchanged indicating negligable orbital relaxation. The increased molecular gap leads to a reduced correlation shift (presumably a reduced internal polarization of the molecule as the metal coupling is unchanged) by 0.3 eV, recognized by the reduced shift between RMO eigenvalues and spectral peaks. Thus all occupied molecular resonances are shifted by -1.3 (SO) + 0.5 (Hartree) - 0.3 (screening) = -0.9 eV, while all unoccpied states shift by +0.5 (Hartree) + 0.3 (screening) = 0.8 eV. Note that the reduced screening has the opposite effect on occupied and unoccupied states, thus opening the gap additionally.[2]

With the applied SO, the transmission function at the Fermi level is reduced to $0.0073G_0$ in close agreement with the experimental value of $0.0064G_0$. It is important to note that although an SO has been applied for correcting the gas-phase spectrum, no correction is made for the polarization effects. These are included naturally in the GW approximation. This is a major improvement over the approach used in [39, 187] as the molecular shift is easily estimated from atomistic calculations on the isolate molecule, while the estimated shift due to image charge formation is based on simplified models and requires a well-defined distinction between metal and molecule (no hybridization).

[2] Applying the suggested SO to the gas phase molecule, only the occupied states are affected, and the shift of these corresponds exactly to the strength of the SO.

4.5 Summary

An GW scheme has been implemented which support non-equilibrium calculations. The implementation makes none of the common approximations, and the accuracy of the calculation is only limited (besides the choice of self-energy operator) by the quality of the input matrices representing the Hamiltonian and interaction operators in some basis, and the undelying energy grid used to represent dynamical quantities. The latter can be systematically converged by tuning a single parameter: the grid spacing of the energy grid.

The issue of basis set representation has addressed by construction of highly accurate projected Wannier functions augmented by numerical atomic orbitals. An scheme has been devised to represent the interactions efficiently using an optimized pairorbital basis.

The capabilities of the GW approximation to represent screening effects have been tested by model studies of semi-conducting surfaces and *ab-initio* calculations on both isolated molecules and molecular junctions.

For the molecules, GW was found to systematically improve the predicted position of the highest occupied molecular orbital compared to both DFT and HF. In the application to bulk junctions, GW was found to give a good description of screening effects. For the weakly coupled Au/BDA system, close agreement with experiment was found without using any ad hoc correction for polarization effects. Such an manual correction based on simplified models is the only way previous calculations have been made to agree with experiment. In GW the relevant screening effect are automatically included in the dynamical self-energy operator.

Chapter 5

Inelastic Transport in a Pt / H_2O Junction

With on one hand the theoretical tools for describing the electronic transport properties relying on a series of approximations, and being too computationally demanding to perform an exhaustive sampling of configuration space, and on the other hand the experimental measurements being done with an unknown atomic configuration, a comparison of experiment and calculations is vital for the understanding of molecular devices.

The work presented in this chapter aims at describing, from a theoretical point of view, the experimental work of [1] on the conductance properties of water. The first section, 5.1 presents a review of the experimental results, and the following section 5.2 describe computational simulations of the experiments.

5.1 Experimental Measurements on Water

The experimental setup used by [1] for probing the electrical properties of individual water molecules, is the mechanically controlled break-junction (MCBJ) technique [16, 17].

The procedure is sketched in figure 5.1. A single piece of metallic wire (in

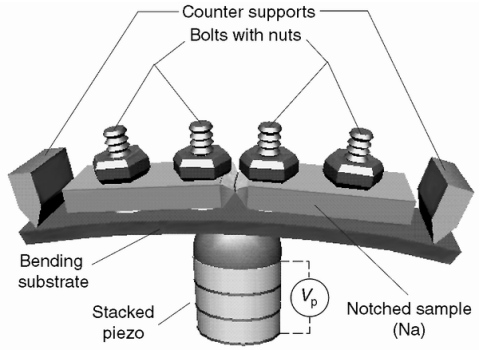


Figure 5.1: Setup for mechanically controlled break-junction experiments.

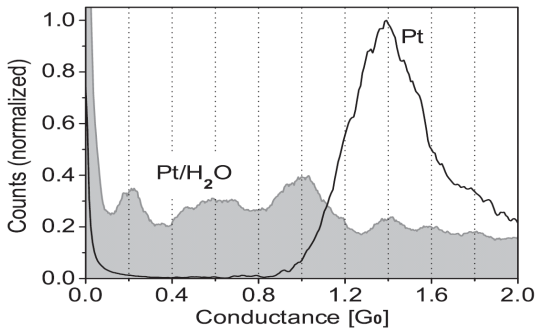


Figure 5.2: Conductance histogram for a Pt junction before (black curve) and after (solid curve) inclusion of H_2O. The histograms are made from 1000 conductance traces recorded at a bias of 0.2V. Reproduced from [1].

this case Platinum) is notched in the middle, and glued onto an elastic substrate. The piezo element can be expanded by adjusting the piezo bias voltage V_p thus stretching the wire. Using this setup it is possible to stretch the wire with subatomic precision. As it is stretched, the wire will grow thinner and eventually break at the position of the notch. With careful arrangements, in high vacuum and at cryogenic temperatures, it is possible to pull out an atomic wire of several nm for certain metals, before the wire finally breaks.

Due to the high reactivity of the severely under-coordinated tips formed upon stretching, it is likely that by admitting a diluted gas of molecules (in this case water vapor) into the chamber when the wire is stretched, one or more of these may bridge the two sides of the metals before rapture.

Electrical characterization of the junction is done by recording the I/V characteristics during stretching. This approach is obviously highly stochastic, and to obtain a statistical description of the system, the break-junction is repeatedly broken and reformed, by cycling the piezo bias. From the ensemble of individual measurements, the features of the most likely (i.e. chemically stable) configurations can be extracted.

5.1.1 Conductance Measurements

One type of electrical characterization of the junction is to simply record the conduction at a fixed (low) bias during stretching. From each of these conduction traces, a conductance histogram can be made showing the most likely conductance values. Such a histogram is shown in figure 5.2 for the Pt junction, before (black curve) and after (solid curve) admission of a diluted gas of de-ionized water vapor in the chamber. Only conductance values in the range 0-2 G_0 are shown, which is the range associated with a nano-scaled junction.

The black curve is characteristic of a clean Platinum contact, and clearly the inclusion of water influence the recorded conductance histogram implying that the molecules must somehow participate in the contact. The histogram does not show the distinct features usually seen in such experiments[2], but displays a rather featureless histogram. Weak peaks around 0.2, 0.6, and 1.0 G_0 are observed. The flat continuum of likely conductance values, suggesting a variety of stable junction configurations, is an important feature that can be exploited, as will become apparent

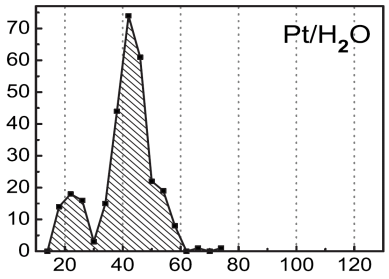


Figure 5.3: Measured phonon distribution for the Pt/H_2O junction. The abscissa indicates possible phonon energies in meV, the ordinate shows the (normalized) number of observations of a given phonon frequency. From [2].

later. One should note the relatively high experimental bias voltage of 200 mV, as the computational predictions will be based on the linear conductance in the zero-bias limit.

5.1.2 Inelastic Spectroscopy

A complementary method of characterizing the junction is by use of inelastic spectroscopy methods, in this case point contact spectroscopy (PCS). The principle is to record the current (or rather its derivative) as a function of the bias voltage across the lead, for a given configuration and stretching distance. When the bias is increased beyond the energy required to excite a phonon $\hbar\Omega$, one would expect a sudden change in the slope of the I/V curve (a step in the differential conductance), as the current-carrying electrons then experience an extra transmission pathway via inelastic scattering on the phonon. Experimentally one can discern if a sudden change in conductance (a step in the differential conductance) is related to electron-phonon interaction by isotope substitution. The mass difference between isotopes should induce a corresponding shift in the step position. The hereby isolated molecular vibration energies supplies a second fingerprint (along with the conductance histogram) of the experimental geometry that can be compared to when doing simulations. The phonon distribution of the Pt/H_2O system as determined by [2] (same group and setup, as in ref. [1]) are shown in figure 5.3.

Interestingly, the sign of the electron-phonon related conductance-step changes from a conductance enhancement in the tunneling regime (low transmission) to a reduction in the contact regime (high transmission).

In a simple picture of the fully open channel, the reduction can be explained by backscattering of electrons. Since at perfect transmission, all forward momentum states are fully occupied at the reduced energy, the electron must by Fermi statistics take on the opposite momentum. In the tunneling regime, the current is mediated by single electron hopping, and the enhancement can be explained by the opening of an additional tunneling channel.

In a combined view of the two limits, based on the lowest order expansion of the e-ph interaction, [189] derived a set of propensity rules describing the sign of the phonon induced conduction step, for a single level model coupled to wide band leads. In the case of a symmetric coupling to the two electrodes, the step should change from enhancement to reduction at a transmission of 0.5.

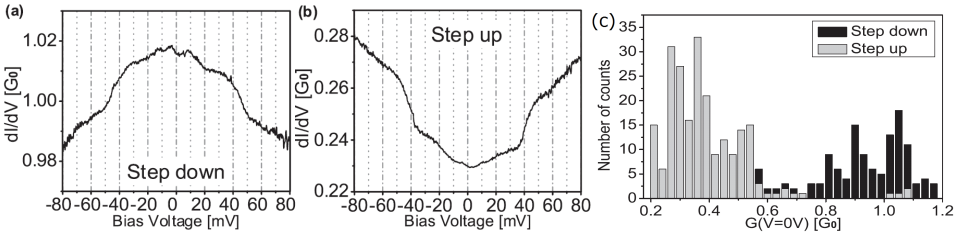


Figure 5.4: Measurements of inelastic features. a and b shows individual $\mathrm{d}I/\mathrm{d}V$ measurements on configurations with a high (a) and low (b) zeros-bias conductance respectively. (c) histogram of step-up / step-down occurrences as a function of zero-bias conductance. From [1].

A set of distinct properties of the Pt/H_2O system makes it possible to study this prediction on an experimental basis. Firstly, the featureless conductance histogram makes is possible to form stable junctions with a zero-bias conductance in a wide range around $0.5G_0$, and secondly the system exhibits a vibrational mode at about 42meV which persists in all configurations.These conditions makes it possible to probe the e-ph signal for the same system, experimental setup, and vibrational mode in the entire range from tunneling to contact regime, and provides a unique possibility to investigate the theoretical predictions.

Figure 5.4a+b shows the differential conductance recorded for two different stable configurations. One at a high zero-bias conductance, fig. 5.4a, showing a conduction reduction at about 46 mV, and one at a low zero-bias conductance, fig. 5.4b, showing a conduction enhancement at about 42 mV.

By measuring the zero-bias conductance, and the sign of the $\mathrm{d}I/\mathrm{d}V$ step from a large number of repeated experiments, the group reports the distribution of step-up and step-down occurrences reproduced in figure 5.4c. The figure shows a cross-over from an enhancement to a reduction at a zero-bias conductance of about $0.65G_0$.

5.1.3 Transmission Channel Analysis

In the Landauer conductance formula [31, 190] $G = G_0 \sum_i \tau_i$, the linear conductance is given by the sum of transmission probabilities for the individual transmission channels, τ_i, times the fundamental conductance unit $G_0 = e^2/2h$.

Due to the discrete nature of electrons, each scattering event can either transmit *or* reflect a charge-quantum of $1e$. This gives rise to a distinct fundamental fluctuation statistics for the measured conductance of any non-perfect conductor, related to the Fano factor $\sum_i \tau_i(1-\tau_i)/\sum_i \tau_i$. Using shot-noise experiments [191] to measure the quantum fluctuations, it is thereby possible determine the number of open transmission channels and their individual transmission probability, by fitting to the predictions of Landauer theory [192]. The group finds that the current is dominated by a single channel, although a second channel accounts for up to 20% of the transmission for high transmission configurations. From the transmission channel analysis they find that the crossover between phonon induced reduction and enhance of $G = 0.65G_0$ corresponds exactly to the value at which the transmission probability of the dominant channel is $\tau_1 = 0.5$.

They conclude by suggesting that the predicted crossover at $\tau = 0.5$ for a single level model can be generalized to apply for the dominant channel for a multichannel

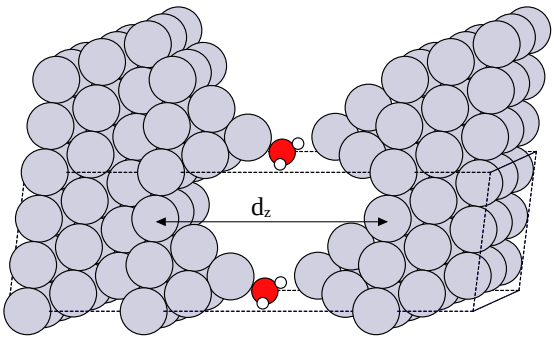


Figure 5.5: Model structure used for simulations of the experiment. The box indicates the unit cell of the scattering region. d_z indicates the electrode surface separation.

system (at least when a single channel is dominant).

5.2 Computational Simulations

In this section the experiment is analysed based on computational modeling. The first step is to determine a realistic atomic configuration, based on the stability of different equilibrium geometries. This is done in section 5.2.1. The following sections are devoted to zero-bias conduction calculations, sec. 5.2.2, analysis of the transmission channels, sec. 5.2.3, and calculation of the inelastic transmission signal, section 5.2.4.

5.2.1 Stability

To make any computational simulations of the experiment, one must first determine a plausible atomic configuration.

The experiment is modelled using the supercell shown in fig. 5.5, i.e. a 3x3 (111) Pt surface with two aligned pyramidal tips and a single H_2O molecule. The close-packed (111) plane is the most stable surface plane, and therefore expected in the experiment. The pyramids are used to emulate the constriction at the notch formed upon stretching of the break-junction. I investigate configuration space for the molecule at various electrode separations d_z, allowing the Pt pyramids to relax in each case, but fixing the Pt surface planes in their bulk equilibrium structure.

The calculations are done with DACAPO using the unitcell marked in the figure with periodic boundary conditions, and a 4×4 kpoint mesh in the transverse plane.

To determine the most likely geometry of the molecule, the system is relaxed starting from several symmetry-distinct initial configurations of the water molecule, and the binding energy relative to the isolated molecule and clean junction are determined. The resulting binding energies are shown in fig. 5.6. To give a broad sampling of configuration space, also high temperature (80 K) Langevin molecular dynamics (MD) have been performed, using low quality numerical parameters. The MD calculations where first allowed 1000 time steps of 2 fs at $d_z = 13.2$ Å to equilibrate, before gently pushing the two electrodes together at a speed of $2 \cdot 10^{-4}$Å/fs.

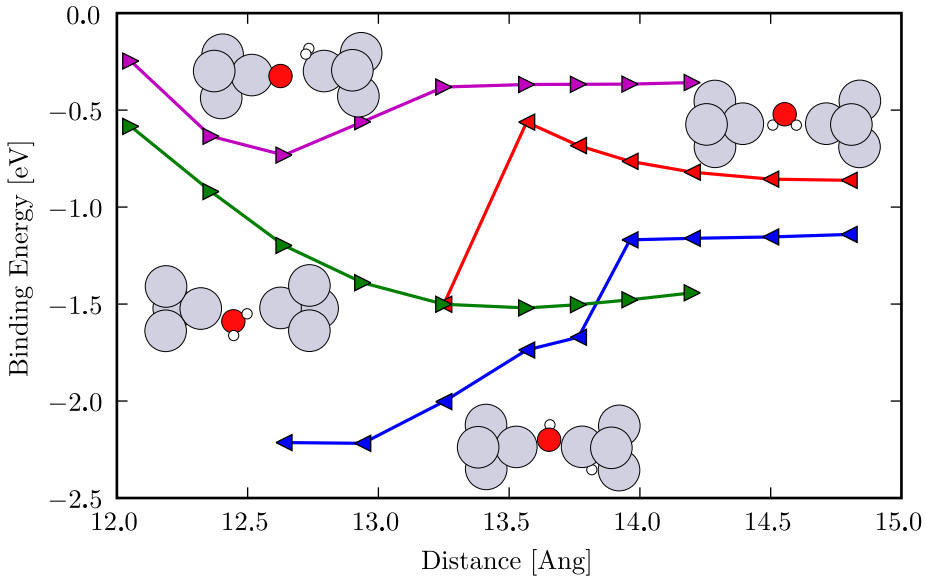


Figure 5.6: Calculated binding energy for different molecule configurations.

From both relaxations and MD simulation, the conclusion is that the most likely configuration of the water in this setup geometry (with two opposing pyramids and a single water molecule) is the asymmetric situation shown in figure 5.5 (green curve in fig. 5.6), where the molecule binds with the Oxygen to one tip, and one OH bond pointing towards the other tip. The second OH bond can more or less freely rotate in the plane transverse to the junction.

Based on the energetics in fig. 5.6, the symmetric positioning of the molecule (red curve) and the dissociated oxygen / hydrogen configuration (blue curve) are excluded based on the low binding energy. The symmetric setup is also unstable to small perturbations, collapsing to the asymmetric case for small displacements of the equilibrium coordinates.

Dissociation of a single Hydrogen atom, leaving behind an OH compound in the middle of the contact (blue curve), with the OH bond perpendicular to the contact axis, is energetically favorable, but in a full analysis of OH in the junction several reasons to discard this solution are found: i) The dissociation has a large energy barrier, and as the inlet gas in the experiment is deionized there should be no OH in the initial water vapor. ii) The calculated conductance trace shows a wide plateau at 1.3 G_0 which lasts until the breaking point of the contact. This is incompatible with the experimental conductance histogram. iii) The calculated phonon spectrum of OH in the junction is strongly stretching dependent, and so do not have stable modes at 20 and 40 meV as seen experimentally.

This analysis leaves only the asymmetric configuration as a possible candidate. This single configuration is studied in more detail in the following sections.

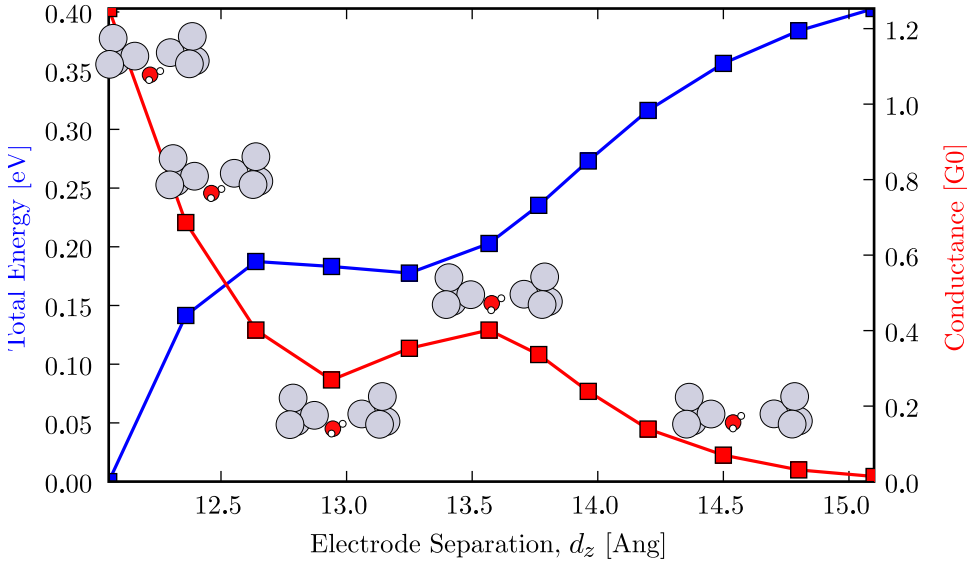


Figure 5.7: Calculated total energy (left axis, blue curve) and linear conductance (right axis, red curve) as a function of electrode separation. Inserts show the atomic configuration at each separation.

5.2.2 Linear Conductance

To calculate the transport properties of the system, a matrix description of the Hamiltonian in a localized basis is required. To this end, SIESTA calculation are performed for each configuration using a DZP basis, and based on the relaxed geometries from the more accurate DACAPO code. The periodic boundary conditions in the transport direction are extended by infinite lead potential as described in section 2.4.5. The leads are described by principal layers 3 atoms long in the transport direction. The calculated linear conductance is shown in fig. 5.7 for a range of electrode displacements, along with the total energies of each junction (from the DACAPO calculations). The insets show the atomic configuration for each distance.

The Pt-O bond is at all separations about 2.2 Å. The H_2O bond lengths and angle are at, or close, to their gas phase values. For $d_z = 12.6$ Å and below, the water molecule sits below the contact. From 12.6 to 13.3 Å it moves into the contact, leading to a more or less stable energy and conductance in this region. Beyond 13.3, the water molecule remains static, with oxygen binding to the left Pt apex atom on the OH bond pointing towards the right tip apex. On stretching beyond this point, the conductance drops exponentially and the total energy has a negative curvature, indicating a broken contact.

The conductance curve in fig. 5.7 indicates that there are several electrode distances, which all lead to a conductance of around 0.4, and one might therefore expect a peak in the experimental histogram at this value. In the experimental setup, adjusting the piezo bias does however *not* translate directly to the distance indicated in the calculation, which is the distance between the surface planes of the two bulk electrodes. In the experiment, the varied distance can be thought

of as the distance between two points in the wires far from the contact region, and deviates from the electrode surface separation due to the finite elasticity of the system [193, 194]. The conversion from surface distance to the experimentally relevant wire stretching depends on the stiffness used to model the leads. Regardless of this parameter though, the experimental setup will never probe the region in figure 5.7 between 12.6 and 13.4 Å, because the slope of the energy is negative in this. Disregarding the mentioned region, leaves a featureless conduction trace consistent with the experimental results.

5.2.3 Transmission Channel Analysis

In the mean-field DFT description of the electronic structure, the Meir-Wingreen current formula [30] reduce to the Landauer expression, with transmission probabilities of the individual channels given by the eigenvalues of the transmission function matrix in the NEGF formalism (see sec. 2.4.5).

The top plot of figure 5.8 shows the calculated transmission function in an interval around the Fermi level along with the two dominant transmission eigenchannels determined by diagonalization of the transmission matrix. The left and right plot shows the transmission for the compressed junction with H_2O pushed below the junction, and a stretched junction where H_2O has moved into the junction, respectively. As a general trend, the dominant channel in the simulations are found to carry 75, 85, and $> 95\%$ of the current at junction separations of 12.0, 12.6, and ≥ 12.9 Årespectively. This in in general agreement with the experimentally determined transmission probability, deduced by fitting shot-noise measurements to predictions of Landauer theory. In [1, table I] they report that transport that the dominant transmission channel, accounts for 80-97 % of the total transmission probability, which is slightly larger than for the simulations at relevant electrode separations.

A useful tool for analysis of transport through molecules, is the concept of renormalized molecular orbital (RMO), determined by diagonalizing the system Hamiltonian in the subspace spanned by orbitals on the molecule. This provides a tool for describing hybridization effects by comparing to the gas phase molecular orbitals. The RMO eigenvalues of the two junctions of fig. 5.8 are compared to the gas-phase spectrum in table 5.1. The occupied canonical MO are only slightly perturbed, indicating weak hybridization of these states despite the relatively large binding energy. The LUMO RMO is pushed up by almost 2 eV. The HOMO has p_z character located primarily on the oxygen, in both gas phase and in the junction, while the LUMO, which in the gas phase is a symmetric anti-bonding σ^* bond between O and H_2, is shifted to have most weight on the hydrogen bridging the two leads. The lower graphs on figure 5.8 show the total and projected DOS for the pyramid tip atoms, and the RMO HOMO and LUMO.

The structure of the transmission function is for the compressed junction dominated by the local DOS of the Pt tip atoms, which due to the narrow d band of the Pt atoms gives rise to a sharp peak in the transmission at the Fermi level. This indicates that the conductance is very sensitive to the applied bias, as the finite bias conductance is (neglecting density redistribution) given by the average of the transmission function over energies in the bias window. The calculation is, due to the same effect, sensitive to the exact position of the Fermi level. At larger electrode separations, the transmission is to a higher degree influenced by the water LUMO, which regains a peaked structure in the tunneling regime, but is very broad for the

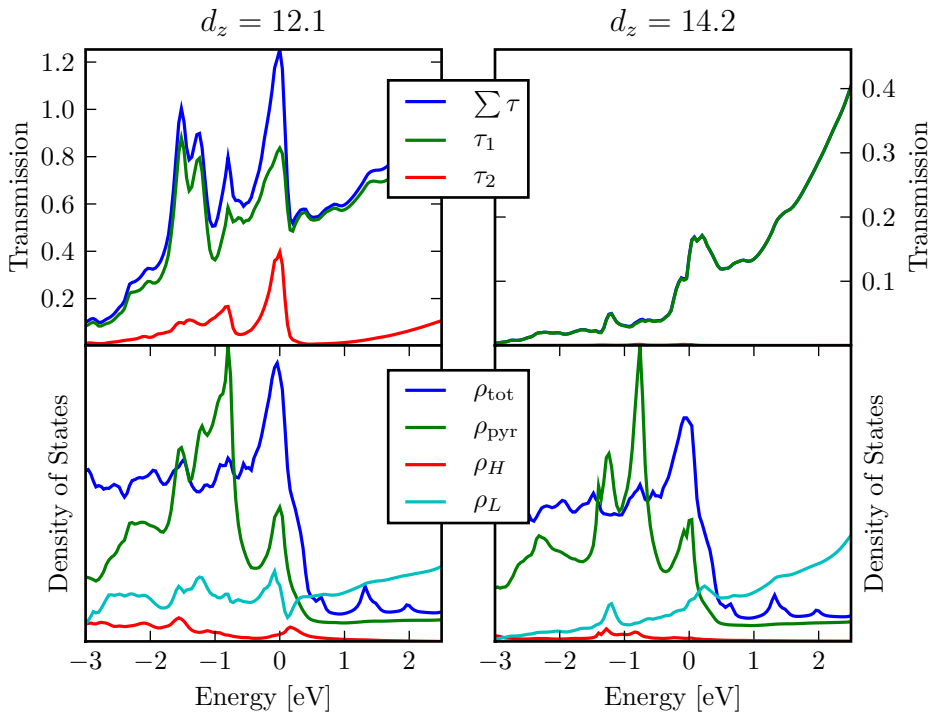


Figure 5.8: Total transmission function and the two dominant transmission eigenchannels (top) for a short (left) and long (right) electrode separation, d_z. Bottom plots show the total ρ_{tot} DOS and the projected DOS for the pyramid apex atoms ρ_{pyr} and the RMO HOMO and LUMO $\rho_{H/L}$ respectively. The pyramid and molecular PDOS have been scaled by 10 and 100 respectively to be visible on the same scale.

compressed junction due to relatively strong hybridization.

The real-space scattering states can be constructed from the Green function and lead self-energies, using the procedure by [195]. By visual inspection of, the scattering states at $\varepsilon = \epsilon_F$ (not shown), the transport is found to be dominated by direct tunneling through the tip Pt atoms for the short junctions, with only low weight on the water molecule, (which is positioned below the contact, see fig. 5.7). At larger distances, the molecule moves into the contact, and the scattering states shortcut through the oxygen atom, but apparently do not involve the bridging hydrogen atom. The direct Pt tunneling at short distances is consistent with the transmission function being dominated by the local Pt DOS of the tip.

5.2.4 Coupling to Phonons

The vibrational eigenmodes, $\{\tilde{\boldsymbol{v}}^\lambda\}$, and corresponding phonon energies, $\{\Omega^\lambda\}$, of the H_2O molecule are obtained in the Harmonic approximation by diagonalizing the mass-scaled dynamical matrix of the system using the finite difference procedure on the DFT forces described in section 2.3.4

Gas	$d_z = 14.2$	$d_z = 12.1$
-23.0	-22.5	-22.1
-10.8	-10.4	-9.9
-6.8	-7.2	-6.8
-4.9	-4.7	-4.8
1.5	3.4	3.3
2.2	5.7	5.6

Table 5.1: Comparison of gas-phase eigenvalues to renormalized molecular orbital eigen values at long, $d_z = 12.1$, and short, $d_z = 12.1$, electrode separations.

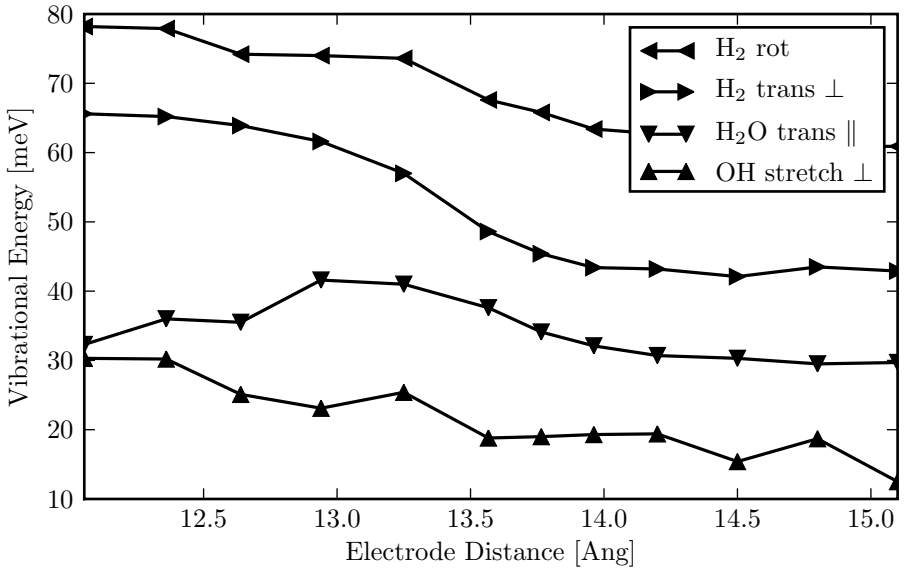


Figure 5.9: Calculated phonon energies.

Vibrations are calculated using DACAPO due to the improved accuracy over SIESTA, and to be consistent with the relaxed geometries also determined using DACAPO. Owing to the large difference in mass between the metal and molecule atoms, the dynamical matrix for the molecule can be calculated keeping all metal atoms fixed. The thusly calculated vibration eigen energies of the water molecule are shown fig. 5.9 (showing only the energetically lowest relevant modes). The phonon modes are, counting from the bottom and up: 1) a stretching mode of the OH bond transverse to transport direction. 2) A rigid translation of the entire water molecule parallel to the transport direction. 3) A frustrated translation of H_2 transverse to both transport direction and H-H bond. 4) A frustrated rotation of H_2 about the axis of the Pt-O bond.

Comparing to the experimental phonon distribution, fig. 5.3, the calculated energies do show modes at around 40 and 20 meV, although some stretching dependence is observed. To determine which modes give a signal in the experiment

(i.e. couple to the scattering states) and to determine the sign of the induced step in $\mathrm{d}I/\mathrm{d}V$, the finite bias current must be simulated including e-ph interaction. The central ingredient for this interaction is the e-ph coupling matrix $\boldsymbol{M}^\lambda$.

Intermezzo on calculating the inelastic signal

Before displaying the calculated inelastic signals for the Pt/H_2O system, I will here pause at the technicalities of calculating this, as it is far from trivial.

The chosen level of inclusion is at the first Born approximation, in which the self-energy related to the e-ph interaction, is given by (see sec. 2.5)

$$\boldsymbol{\Sigma}^{\lessgtr}_{e\text{-ph}}(\omega) = \sum_\lambda \boldsymbol{M}^\lambda \boldsymbol{G}^{\lessgtr}(\omega \pm \Omega^\lambda)\boldsymbol{M}^\lambda, \tag{5.1a}$$

$$\boldsymbol{\Sigma}^{r}_{e\text{-ph}}(\omega) = \tfrac{1}{2}[\boldsymbol{\Sigma}^{>}_{e\text{-ph}}(\omega) - \boldsymbol{\Sigma}^{<}_{e\text{-ph}}(\omega)] - \tfrac{i}{2}\mathcal{H}[\boldsymbol{\Sigma}^{>}_{e\text{-ph}}(\omega') - \boldsymbol{\Sigma}^{<}_{e\text{-ph}}(\omega')](\omega), \tag{5.1b}$$

assuming a vanishing life-time of the phonons. Here, $\mathcal{H}[f(y)](x)$ denotes the Hilbert transform of $f(x)$, and $\boldsymbol{M}^\lambda$ is the coupling matrix

$$M^\lambda_{ij} = \sum_{a\nu} \sqrt{\frac{\hbar}{2M_a\Omega^\lambda}} v^\lambda_{a\nu} \langle\phi|_i W_{a\nu}(\boldsymbol{r})|\phi_j\rangle, \tag{5.2}$$

given here in terms of the mass-scaled normalized eigen-modes $\boldsymbol{v}^\lambda$ and the gradient of the Hamiltonian $W_{a\nu}(\boldsymbol{r}) = \partial v_{\text{eff}}(\boldsymbol{r})/\partial R_{a\nu}$ (the kinetic energy obviously do not depend on $\boldsymbol{R}$).

The above expression is actually only the "Fock" diagram of the e-ph interaction. The corresponding Hartree term is neglect, as it simply leads to a static (polaron) potential shift, giving no signal at the phonon threshold voltage. Had it been included along with an electronic Hartree self-energy, it would be, at least partially, screened. In practice, both the dynamical matrix and the gradient $\boldsymbol{W}(\boldsymbol{r})$ are determined from finite differences between the self-consistent KS systems at the displaced configurations, and thus include electronic screening of the polaron shift at the DFT level.

The self-consistent Born approximation (5.1) presents two complications. One is calculation of the coupling matrix, the second is the inclusion of the dynamic self-energy approximation in the Dyson and Keldysh equations. These issues will be addressed in turn in the following.

Extraction of the coupling matrix can be done very accurately in the PAW code GPAW by basing it directly on the real-space effective potential. In PAW, the gradient of the Hamiltonian has four terms

$$W_{a\nu}(\boldsymbol{r}) = \frac{\partial \tilde{v}_{\text{eff}}(\boldsymbol{r})}{\partial R_{a\nu}} + \sum_{ij}\left[|\tilde{p}^a_i\rangle \frac{\partial \Delta H^a_{ij}}{\partial R_{a\nu}} \langle\tilde{p}^a_j| + \frac{\partial |\tilde{p}^a_i\rangle}{\partial \boldsymbol{R}_{a\nu}} \Delta H^a_{ij} \langle\tilde{p}^a_j| + |\tilde{p}^a_i\rangle \Delta H^a_{ij} \frac{\partial \langle\tilde{p}^a_j|}{\partial R_{a\nu}}\right], \tag{5.3}$$

the first being the gradient of the pseudo effective potential, and the remaining terms correcting for the PAW transform of kinetic and potential energy terms as well as the compensation of the Hartree potential, see sec. 3.1. The first two terms of are calculated by a central finite difference on $\tilde{v}_{\text{eff}}(\boldsymbol{r})$ and ΔH^a_{ij} respectively, while the last two are calculated analytically.

Interestingly, the effect of the last two terms is an order magnitude larger than the first, which is because the pseudo KS potential $\tilde{v}$ is smooth, and therefore have a small gradient in ion displacements, while the atomic corrections are sharply peaked.

Note that the gradient (5.3) is independent of the individual modes, and the ingredients for the finite-difference operations can be obtained from calculations that are anyway performed for each displaced ion for construction of the dynamical matrix. The coupling matrix then follows directly from the real-space gradient by projection onto the basis set of the equilibrium structure, see (5.2), and thereby comes at no noticeable extra computational cost besides that required to determine the dynamical matrix.

For the SIESTA calculations used in this study for the transport related part, the above method can not be applied. The output of the code only allow access to the Hamiltonian and overlap matrix resolved in the LCAO basis, for each of the displaced systems. A finite difference scheme for the gradient based directly on these matrices instead of the real-space approach, would be incorrect due to the moving basis functions. This can be corrected for by the scheme described in [103], but this also requires matrix elements inaccessible from the unmodified code (the overlap between the displaced and original basis functions). Instead, I use the following finite-difference along the individual vibration eigen modes

$$M_{ij}^{\lambda} \approx \sqrt{\frac{\hbar}{2\Omega^{\lambda}}} \frac{H_{ij}(\boldsymbol{R}_0 + \delta\tilde{\boldsymbol{v}}^{\lambda}) - H_{ij}(\boldsymbol{R}_0 - \delta\tilde{\boldsymbol{v}}^{\lambda})}{2\delta}, \tag{5.4}$$

where $\tilde{v}_{a\nu}^{\lambda} = v_{a\nu}^{\lambda}/M_a$ is the physical eigen mode (hence the missing mass factor in the prefactor). To avoid errors induced by the basis functions moving with the displaced atoms, ghost basis functions are left at the original coordinates and after obtaining the self-consistent Hamiltonian, the rows and columns corresponding to the displaced basis are removed. The method is computationally cheap if only a few modes are of interest, as only the two calculations needed for a central finite difference are required for each mode. It does however have some drawbacks in terms of accuracy. The use of ghost basis functions introduce basis set superposition errors. In addition the atoms are moved along the vibration eigen modes, which implies that some atoms move a lot, while others do not move.

Comparing to the more accurate grid-based approach in GPAW, it is found that reliable coupling matrices can be extracted by choosing δ in (5.4) such that the longest atomic displacement is about 0.1Å. A ghost basis at the original position can not be used for those atoms that are displaced less than 0.01 Å, as this leads to linear dependence in the basis set, so for these, a minor error for the moving basis must be tolerated.

This resolves the issue of determining the the coupling matrix. Note that even though only the water molecules are displaced, this has an effect also on basis functions far from contact, and in general M and hence $\Sigma_{e\text{-ph}}$ are non-zero in the entire scattering region. I have implemented the self-energies in (5.1) in the KELDYSHGF code as an addition along the same lines as the GW self-energy. In many ways the e-ph self-energy resembles the GW self-energy in that it is a dynamic correlation self-energy such that the Meir-Wingreen current formula has weight outside of the bias window, and in that it is non-local in the energy grid (due to the Hilbert transform), and therefore can not be evaluated independently at each energy. It does require a fine-spaced energy grid, but unlike the GW self-energy, it does not depend upon the entire frequency range of the electronic Green function, and in practice the energy range need only extend a few eV beyond the bias window, or the largest phonon frequency, whichever is largest. I have not made a parallel implementation though, and for the basis set sizes used here, the computations becomes very heavy numerically.

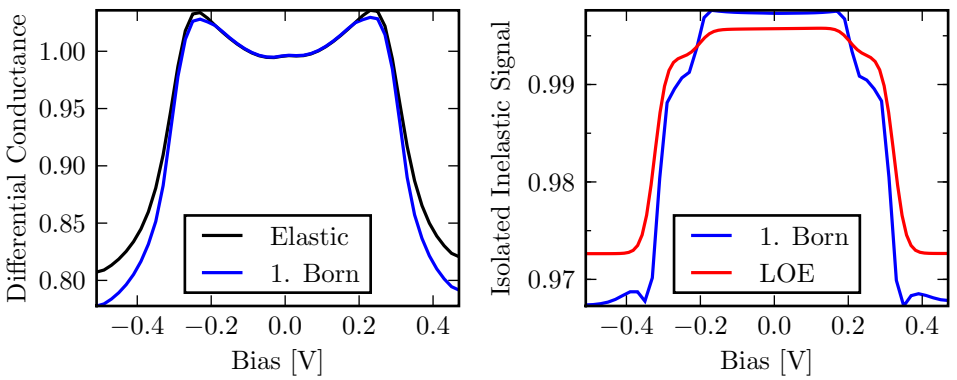


Figure 5.10: Calculated differential conductance for an H_2 molecule connected to Pt wires.

An alternative to the self-consistent use of the non-local self-energies in (5.1) exists in the form of the lowest order expansion (LOE) scheme of [196, 197]. In this scheme, the retarded electronic Green function and the lead self-energies are assumed to be energy independent, and the inelastic signal is expanded to lowest order in M. From this, the entire I/V spectrum can be determined from a relatively simple algebraic equation using only the set of phonon energies, the coupling matrix and the retarded Green function and lead self-energy evaluated at the Fermi energy. No energy convolutions or self-consistency is needed.

To compare the two methods, and test the implementation, both are applied to a one-dimensional Pt wire connecting a H_2 molecule. This should be comparable to the bulk Pt with a H_2O molecule. The comparison is shown in figure 5.10. For this system two vibrational modes exist at 205 meV and 320 meV respectively corresponding to an in-phase (center-of-mass) longitudinal mode and an out-of-phase (internal) longitudinal mode.

The left side of fig. 5.10 shows the calculated differential conductance from a non-equilibrium finite bias calculation both with phonon contributions (denoted 1. Born) and without (denoted elastic). Clearly, even the elastic signal has some structure, which is due to the narrow d-band of the metal. Thus, when increasing the bias window, one first includes the tip of the d-band peak (slightly below ϵ_F) which gives an increased conduction, then one moves off-resonance giving a reduced conduction. That the down-step of the elastic signal is right at the phonon frequency is purely accidental. The inelastic part of the signal can be isolated by plotting the difference between the two curves. This makes the 1. Born comparable with the LOE scheme which only gives the inelastic contribution. The isolated inelastic signal is shown for the LOE and 1. Born approximations on the right graph. Both phonon modes gives a conduction reduction, consistent with the almost fully transmitting channel.

From the good correspondence between the two methods despite the violation of the assumed flat DOS in the LOE scheme, it is concluded that the much simpler LOE scheme is sufficient to describe the inelastic signal and this is then the scheme applied to the bulk Pt/H_2O system.

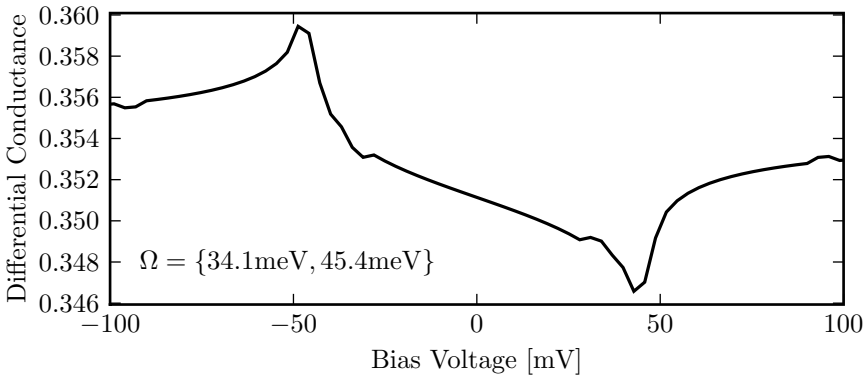


Figure 5.11: Calculated differential conductance at $d_z = 13.8$Å.

The inelastic signal for the Pt/H_2O system

Using the method described in the previous, the inelastic contribution to the differential conduction for the Pt/H_2O system is calculated at various electrode separations.

At long separations, the two modes closest to the experimental mode at 42 meV are the H_2 translation mode transverse to the junction and the translational mode of the entire molecule parallel to the junction (see fig. 5.9). The $\mathrm{d}I/\mathrm{d}V$ signal of these are shown for $d_z = 13.8$ in figure 5.11. At this distance, the molecule is positioned inside the junction, and the H_2 translation mode which the dominant signal. The signal is entirely asymmetric, and can not be associated with either a step up or down.

For short junctions, the molecule is pushed below the contact, which leads to a decreased inelastic signal strength. At $d_z = 12.1$ the signal almost completely vanish, which is consistent with the scattering states for these, which involved only direct tunneling through the Pt tips. The modes closest to the experimental mode at 42 meV are, for the short separations, the OH stretching mode and the translational mode of the entire molecule (see fig. 5.9). The $\mathrm{d}I/\mathrm{d}V$ signal of these are shown for $d_z = 12.4$ and $d_z = 12.6$ in figure 5.12. None of the figures show any signal at the OH stretching mode which is expected as it is far below the contact for short junctions where the molecule itself is pushed below the contact. For the longer separations a signal around 20 meV appears (not shown in fig. 5.11), which explains the peak at this energy in figure 5.3.

The translational mode gives rise to a conduction enhancement in both cases. This is expected for the longer junction for which the zero-bias conduction is $G(V = 0) = 0.4G_0$, but should have lead to a reduction for the short junction, for which $G(V = 0) = 0.7G_0$. The transmission probability of the dominant channel of the short junction is 0.65 which is also larger than the expected crossover of 0.5. The disagreement can not be explained by the asymmetry of the contact, as this according to [189] will lead to a crossover to a signal decrease at even lower transmissions than 0.5.

The discrepancy compared to the model may be related to the inadequacy of a single level model for this system, where the RMO HOMO and LUMO are almost

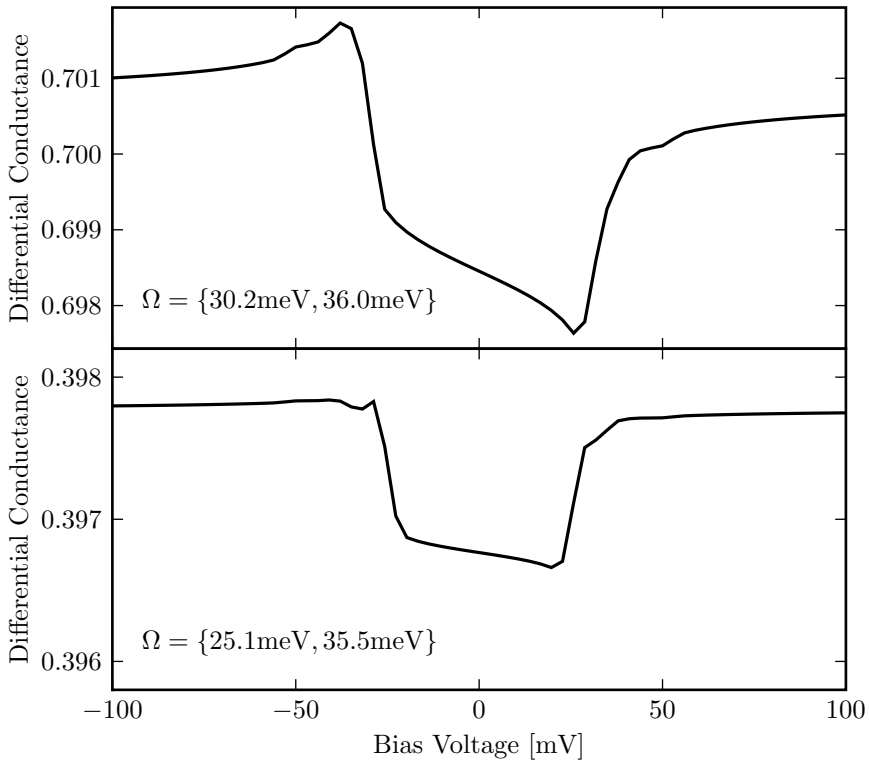


Figure 5.12: Calculated differential conductance at $d_z = 12.4$Å (top) and 12.6 Å(bottom).

equally far from the Fermi level. The discrepancy with the experimental may be explained by the elastic signal dominating the measured signal. That they see a crossover to an increase at $G = 0.65G_0$ should then be related to the e-ph induced enhancement becoming dominant at this point. This would still give an isotope substitution effect, making the step seem as a purely e-ph induces quantity.

In any case I can not with any of the systems recover high transmission, without the transport being entirely Pt mediated, and therefore unaffected by the vibrational modes of the water.

Of course it is possible that the experimental configuration should be described using more than a single water molecule in the simulation. Bridge formation involving several molecules is improbable, but water coverage on the Pt pyramid surfaces might affect the result. Water coverage of the flat Pt surface should be irrelevant.

5.3 Summary

Computational simulations of the break junction experiment of [1] on H_2O have been performed. Simulation of the experiment is complicated by the statistical basis of the experiment, and the unknown atomic junction geometry. From the large class of different possible shapes of the metal junction formed upon streching the Pt wire,

and the multitude of molecular arrangements, a representative subset have been studied, and within this, the most likely geometry has been determined.

Based on DFT, binding energies and phonon energies have been determined, and using the NEGF formalism both elstic and inelastic transmission probabilities were calculated and compared to experiment. Some feature of the experiment could be described by the simulations, but especially the inelastic transmission did not provide a consistent picture of the measured spectra.

Based on an analysis of the elastic transmission function and density of states, combined with simulations of phonon effect in an atomic wire, it is suggested that an acute dependence of the elastic transmission on the applied bias might influence the differential conductance measurements. The dependence is caused by the narrow and chemically active local *d*-band of the tip Pt atoms in the junction.

References

[1] O. Tal, M. Krieger, B. Leerink, and J. M. van Ruitenbeek. Electron-vibration interaction in single-molecule junctions: From contact to tunneling regimes. *Physical Review Letters*, 100(19):196804, 2008.

[2] O. Tal, M. Kiguchi, W. H. A. Thijssen, D. Djukic, C. Untiedt, R. H. M. Smit, and J. M. van Ruitenbeek. Molecular signature of highly conductive metal-molecule-metal junctions. *Physical Review B (Condensed Matter and Materials Physics)*, 80(8):085427, 2009.

[3] A. I. Yanson, G. Rubio Bollinger, H. E. van den Brom, N. Agrait, and J. M. van Ruitenbeek. Formation and manipulation of a metallic wire of single gold atoms. *Nature*, 395(6704):783–785, October 1998.

[4] Jesper Nygård, David Henry Cobden, and Poul Erik Lindelof. Kondo physics in carbon nanotubes. *Nature*, 408(6810):342–346, November 2000.

[5] G.E. Moore. Cramming more components onto integrated circuits. *Electronics*, 38(8), 1965.

[6] Intel corporation. http://www.intel.com/technology/silicon/high-k.htm.

[7] C. Joachim, J. K. Gimzewski, and A. Aviram. Electronics using hybrid-molecular and mono-molecular devices. *Nature*, 408(6812):541–548, November 2000.

[8] Gianaurelio Cuniberti, Giorgos Fagas, and Klaus Richter, editors. *Introducing Molecular Electronics*. Lecture Notes in Physics. Springer Berlin / Heidelberg, 2005.

[9] Arieh Aviram and Mark A. Ratner. Molecular rectifiers. *Chemical Physics Letters*, 29(2):277 – 283, 1974.

[10] Nicolás Agraït, Alfredo Levy Yeyati, and Jan M. van Ruitenbeek. Quantum properties of atomic-sized conductors. *Physics Reports*, 377(2-3):81–279, April 2003.

[11] Abraham Nitzan and Mark A. Ratner. Electron transport in molecular wire junctions. *Science*, 300(5624):1384–1389, 2003.

[12] D.M. Eigler, C.P. Lutz, and W.E. Rudge. An atomic switch realized with the scanning tunnelling microscope. *Nature*, 352(6336):600, 1991.

[13] B. Xu and N.J. Tao. Measurement of single-molecule resistance by repeated formation of molecular junctions. *Science*, 301(5637):1221, 2003.

[14] G. Rubio-Bollinger, S. R. Bahn, N. Agraït, K. W. Jacobsen, and S. Vieira. Mechanical properties and formation mechanisms of a wire of single gold atoms. *Phys. Rev. Lett.*, 87(2):026101, 2001.

[15] R. H. M. Smit, C. Untiedt, G. Rubio-Bollinger, R. C. Segers, and J. M. van Ruitenbeek. Observation of a parity oscillation in the conductance of atomic wires. *Phys. Rev. Lett.*, 91(7):076805, Aug 2003.

[16] John Moreland and J. W. Ekin. Electron tunneling experiments using nb-sn "break" junctions. *Journal of Applied Physics*, 58(10):3888–3895, 1985.

[17] C. J. Muller, J. M. van Ruitenbeek, and L. J. de Jongh. Experimental observation of the transition from weak link to tunnel junction. *Physica C: Superconductivity*, 191(3-4):485 – 504, 1992.

[18] M.A. Reed, C. Zhou, C.J. Muller, T.P. Burgin, and J.M. Tour. Conductance of a molecular junction. *Science*, 278(5336):252, 1997.

[19] J. Reichert, R. Ochs, D. Beckmann, H.B. Weber, M. Mayor, and H. L "ohneysen. Driving current through single organic molecules. *Phys. Rev. Lett.*, 88(17):176804, 2002.

[20] Hideaki Ohnishi, Yukihito Kondo, and Kunio Takayanagi. Quantized conductance through individual rows of suspended gold atoms. *Nature*, 395(6704):780–783, October 1998.

[21] N. D. Lang. Anomalous dependence of resistance on length in atomic wires. *Phys. Rev. Lett.*, 79(7):1357–1360, Aug 1997.

[22] K. S. Thygesen and K. W. Jacobsen. Four-atom period in the conductance of monatomic al wires. *Physical Review Letters*, 91(14):146801, October 2003.

[23] Mikkel Strange, Kristian S. Thygesen, James P. Sethna, and Karsten W. Jacobsen. Anomalous conductance oscillations and half-metallicity in atomic ag-o chains. *Physical Review Letters*, 101(9):096804, 2008.

[24] Jiwoong Park, Abhay N. Pasupathy, Jonas I. Goldsmith, Connie Chang, Yuval Yaish, Jason R. Petta, Marie Rinkoski, James P. Sethna, Hector D. Abruna, Paul L. McEuen, and Daniel C. Ralph. Coulomb blockade and the kondo effect in single-atom transistors. *Nature*, 417(6890):722–725, June 2002.

[25] T A Costi, A C Hewson, and V Zlatic. Transport coefficients of the anderson model via the numerical renormalization group. *Journal of Physics: Condensed Matter*, 6(13):2519–2558, 1994.

[26] D. Goldhaber-Gordon, Hadas Shtrikman, D. Mahalu, David Abusch-Magder, U. Meirav, and M. A. Kastner. Kondo effect in a single-electron transistor. *Nature*, 391(6663):156–159, January 1998.

[27] M. Brandbyge, J. L. Mozos, P. Ordejon, J. Taylor, and K. Stokbro. Density-functional method for nonequilibrium electron transport. *Physical Review B*, 65(16):165401, April 2002.

[28] Y. Q. Xue, S. Datta, and M. A. Ratner. First-principles based matrix green's function approach to molecular electronic devices: general formalism. *Chemical Physics*, 281(2-3):151–170, August 2002.

[29] J.J. Palacios, A.J. Perez-Jimenez, E. Louis, E. SanFabian, and J.A. Verges. First-principles approach to electrical transport in atomic-scale nanostructures. *Phys. Rev. B*, 66(3):35322, 2002.

[30] Yigal Meir and Ned S. Wingreen. Landauer formula for the current through an interacting electron region. *Physical Review Letters*, 68(16):2512, 1992.

[31] R. Landauer. Electrical resistance of disordered one-dimensional lattices. *Philosophical Magazine*, 21(172):863, 1970.

[32] F. Evers, F. Weigend, and M. Koentopp. Conductance of molecular wires and transport calculations based on density-functional theory. *Phys. Rev. B*, 69(23):235411, 2004.

[33] F. Evers and K. Burke. Pride, prejudice, and penury of ab initio transport calculations for single molecules. In *CRC Handbook on Molecular and Nanoelectronics, ed. S. Lyshevski*. CRC Press, 2007.

[34] M. Koentopp, C. Chang, K. Burke, and R. Car. Density functional calculations of nanoscale conductance. *J. Phys.: Condens. Matter*, 20(8):83203, 2008.

[35] V. M. García-Suárez, A. R. Rocha, S. W. Bailey, C. J. Lambert, S. Sanvito, and J. Ferrer. Conductance oscillations in zigzag platinum chains. *Phys. Rev. Lett.*, 95(25):256804, 2005.

[36] T. Frederiksen, M. Brandbyge, N. Lorente, and A. P. Jauho. Inelastic scattering and local heating in atomic gold wires. *Physical Review Letters*, 93(25):256601, December 2004.

[37] K. Stokbro, J. Taylor, M. Brandbyge, J. L. Mozos, and P. Ordejon. Theoretical study of the nonlinear conductance of di-thiol benzene coupled to au(111) surfaces via thiol and thiolate bonds. *COMPUTATIONAL MATERIALS SCIENCE*, 27(1-2):151–160, March 2003.

[38] M. Di Ventra, S.T. Pantelides, and N.D. Lang. First-principles calculation of transport properties of a molecular device. *Physical Review Letters*, 84(5):979, 2000.

[39] Su Ying Quek, Latha Venkataraman, Hyoung Joon Choi, Steven G. Louie, Mark S. Hybertsen, and J. B. Neaton. Amine-gold linked single-molecule circuits: Experiment and theory. *Nano Letters*, 7(11):3477–3482, 2007.

[40] J. Heurich, J. C. Cuevas, W. Wenzel, and G. Schön. Electrical transport through single-molecule junctions: From molecular orbitals to conduction channels. *Phys. Rev. Lett.*, 88(25):256803, Jun 2002.

[41] Fang Chen, Xiulan Li, Joshua Hihath, Zhifeng Huang, and Nongjian Tao. Effect of anchoring groups on single-molecule conductance: Comparative study of thiol-, amine-, and carboxylic-acid-terminated molecules. *Journal of the American Chemical Society*, 128(49):15874–15881, December 2006.

[42] C. Toher, A. Filippetti, S. Sanvito, and Kieron Burke. Self-interaction errors in density-functional calculations of electronic transport. *Phys. Rev. Lett.*, 95(14):146402, Sep 2005.

[43] J. B. Neaton, Mark S. Hybertsen, and Steven G. Louie. Renormalization of molecular electronic levels at metal-molecule interfaces. *Physical Review Letters*, 97(21):216405, 2006.

[44] Sergey Kubatkin, Andrey Danilov, Mattias Hjort, Jerome Cornil, Jean-Luc Bredas, Nicolai Stuhr-Hansen, Per Hedegard, and Thomas Bjornholm. Single-electron transistor of a single organic molecule with access to several redox states. *Nature*, 425(6959):698–701, October 2003.

[45] Leslie E. Ballentine. *Quantum mechanics: a modern development*. World Scientific, 2. edition, 1998.

[46] P. A. M. Dirac. Quantum mechanics of many-electron systems. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 123(792):714–733, 1929.

[47] M.L. Born and R. Oppenheimer. Zur Quantentheorie der Moleküln. *Ann. Phys.*, 84:457, 1927.

[48] Stefano Baroni, Stefano de Gironcoli, Andrea Dal Corso, and Paolo Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73(2):515–562, Jul 2001.

[49] W. Kohn. Nobel Lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.*, 71:1253–1266, 1999.

[50] J. C. Slater. *Quantum Theory of Atomic Structure, Vol. I.* McGraw-Hill, New York, 1960.

[51] M. P. Marder. *Condensed Matter Physics*. John Wiley & Sons Inc., 2000.

[52] C. Fiolhais, F. Nogueira, and M. Margues, editors. *A Primer in Density Functional Theory*, volume 620 of *Lecture Notes in Physics*. Springer, 2003.

[53] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Physical Review*, 136:B864–B871, 1964.

[54] M. Levy. Universal variational functionals of electron densities, first-oder density matrices and natural spin orbitals and solution of the ν-representability problem. *Proceedings of the National Academy of Sciences U.S.A.*, 76:6062–6065, 1979.

[55] Mel Levy. Electron densities in search of hamiltonians. *Phys. Rev. A*, 26(3):1200–1208, Sep 1982.

[56] E. H. Lieb. Density functionals for Coulomb-systems. *International Journal of Quantum Chemistry*, 24:243–277, 1983.

[57] H. English and R. English. Hohenberg-Kohn theorem and non-v-representable densities. *Physica A*, 121:253–268, 1983.

[58] B. H. Bransden and C. J. Joachain. *Quantum Mechanics*. Pearson Education Limited, 2nd edition edition, 2000. ISBN 0582-35691-1.

[59] George B. Arfken and Hans J. Weber. *Mathematical Methods for Physicists*. Acacdemic Press, 5 edition, 2001.

[60] L. H. Thomas. The calculation of atomic fields. *Proceedings of the Cambridge Philosophical Society*, 23:542–548, 1927.

[61] E. Fermi. Statistical method of investigating electrons in atoms. *Zeitschrift für Physik*, 48:73–79, 1928.

[62] P. A. M. Dirac. Exchange phenomena in the thomas atom. *Proceedings of the Cambridge Philosophical Society*, 26:376–385, 1930.

[63] N. H. March. The thomas-fermi approximation in quantum mechanics. *Advances in Physics*, 6:1–98, 1975.

[64] E. H. Lieb. Thomas-fermi and related theories of atoms and molecules. *Reviews of Modern Physics*, 53(4):603–641, 1981.

[65] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133–A1138, 1965.

[66] A. Seidl, A. Görling, P. Vogl, J. A. Majewski, and M. Levy. Generalized kohn-sham schemes and the band-gap problem. *Physical Review B*, 53:3764–3774, 1996.

[67] A. Görling and M. Levy. Hybrid schemes combining the hartree-fock methods and density-functional theory: underlying formalism and properties of correlation functional. *The Journal of Chemical Physics*, 106:2676–2680, 1997.

[68] W. Pauli.
"Uber den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren. *Z. Physik*, 31(1):765, 1925.

[69] J. P. Perdew and A. Zunger. Self-interaction correction to density-functional approximations to many-electron systems. *Physical Review B*, 23:5048–5079, 1981.

[70] D. M. Ceperly and B. J. Alder. Ground state of the electron gas by a stochastic method. *Physical Review Letters*, 45:566–569, 1980.

[71] S. J. Vosko, L. Wilk, and M. Nusair. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. *Canadian Journal of Physics*, 58:1200–1211, 1980.

[72] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Physical Review B*, 46:6671–6687, 1992.

[73] J. P. Perdew and K. Burke. Comparison shopping for a gradient-corrected density functional. *International Journal of Quantum Chemistry*, 57:309–319, 1996.

[74] J. P. Perdew, S. Kurth, A. Zupan, and P. Blaha. Accurate density functional with correct formal properties: A step beyond the generalized gradient approximation. *Physical Review Letters*, 82:2544–2547, 1999.

[75] J. P. Perdew, K. Burke, and M. Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77:3865–3868, 1996.

[76] J. P. Perdew and K. Schmidt. Jacob's ladder of density functional approximations for the exchange-correlation energy. *American Institute of Physics, Conference Proceedings*, 577:1–20, 2001.

[77] F. D. Sala and A. Görling. Efficient localized hartree-fock methods as effective exact-exchange kohn-sham methods for molecules. *Journal of Chemical Physics*, 115:5718–5732, 2001.

[78] A. M. Teale and D. J. Tozer. Exchange methods in kohn-sham theory. *Physical Chemistry Chemical Physics*, 7:2991–2998, 2005.

[79] T. Grabo, T. Kreibich, S. Kurth, and E. K. U. Gross. Orbital functionals in density functional theory: the optimized effective potential method. In V. I. Anisimov, editor, *Strong Coulomb Correlations in Electronic Structure Calculations: Beyond the Local Density Approximation*, pages 203–311, 2000.

[80] S. Kümmel and J. P. Perdew. Simple iterative construction of the optimized effective potential for orbital functionals, including exact exchange. *Physical Review Letters*, 90:043004, 2003.

[81] S. Kümmel and J. P. Perdew. Optimized effective potential made simple: Orbital functionals, orbital shifts, and the exact hohn-sham exchange potential. *Physical Review B*, 68:035103, 2003.

[82] C. Adamo and V. Barone. Towards reliable density functional methods without adjustable parameters: The PBE0 model. *Journal of Chemical Physics*, 110:6158–6170, 1999.

[83] M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos. Iterative minimization techniques for *ab initio* total-energy calculations: Molecular dynamics and conjugate gradients. *Rev. Mod. Phys.*, 64:1045–1097, Oct 1992.

[84] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6:15–50, 1996.

[85] G. Kresse and J. Furthmüller. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Physical Review B*, 54:11169–11186, 1996.

[86] The real-space PAW-DFT code GPAW is part of the CAMP Open-Source (CAMPOS) project. GPAW is freely available at `https://wiki.fysik.dtu.dk/gpaw`.

[87] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen. Real-space grid implementation of the projector augmented wave method. *Physical Review B*, 71(3):035109, Jan 2005.

[88] Atomic Simulation Environment (ASE) is part of the CAMP open-source (CAMPOS) project, and is a python based interface for atomistic calculators. ASE is freely available at https://wiki.fysik.dtu.dk/ase.

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

[90] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The SIESTA method for ab initio order-N materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002. SIESTA is available online at http://www.icmab.es/siesta/.

[91] S.R. Elliot. *The Physics and Chemistry of Solids*. John Wiley & Sons Inc., 1998.

[92] H. J. Monkhorst and J. D. Pack. Special points for Brillouin-zone integrations. *Physical Review B*, 13:5188–5192, 1976.

[93] D.J. Chadi and M.L. Cohen. Special points in the Brillouin zone. *Physical Review B*, 8(12):5747, 1973.

[94] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, Dec 1994.

[95] P. E. Blöchl, C. J. Först, and J. Schimpl. Projector augmented wave method: ab-initio molecular dynamics with full wave functions. *Bulletin of Materials Science*, 26:33–41, 2003.

[96] D. Vanderbilt. Soft self-consistent pseudopotentials in a generalized eigenvalue formalism. *Physical Review B*, 41(11):7892, 1990.

[97] N. Troullier and J. L. Martins. A straightforward method for generating soft transferable pseudopotentials. *Solid State Commun.*, 74:613, 1990.

[98] D. R. Hamann, M. Schlüter, and C. Chiang. Norm-conserving pseudopotentials. *Physical Review Letters*, 43:1494–1497, 1979.

[99] James W. Cooley and John W. Tukey. An algorithm for the machine calculation of complex fourier series. *Mathematics of Computation*, 19(90):297–301, April 1965.

[100] E. L. Briggs, D. J. Sullivan, and J. Bernholc. Large-scale electronic-structure calculations with multigrid acceleration. *Phys. Rev. B*, 52(8):R5471–R5474, Aug 1995.

[101] E. L. Briggs, D. J. Sullivan, and J. Bernholc. Real-space multigrid-based approach to large-scale electronic structure calculations. *Phys. Rev. B*, 54(20):14362–14375, Nov 1996.

[102] R. P. Feynman. Forces in molecules. *Physical Review*, 56:340–343, 1939.

[103] Thomas Frederiksen, Magnus Paulsson, Mads Brandbyge, and Antti-Pekka Jauho. Inelastic transport theory from first principles: Methodology and application to nanoscale devices. *Physical Review B (Condensed Matter and Materials Physics)*, 75(20):205413, 2007.

[104] K. Flensberg and H. Bruus. *Many-Body Quantum Theory in Condensed Matter Physics*. Oxford university press, New York, 2004.

[105] P. Danielewicz. Quantum-theory of nonequilibrium processes .1. *Annals of Physics*, 152(2):239–304, 1984.

[106] Antti-Pekka Jauho and Hartmut J. W. Haug. *Quantum Kinetics in Transport and Optics of Semiconductors*, volume 123 of *Springer Series in Solid-State Sciences*. Springer Berlin Heidelberg, 2nd rev. ed. edition, December 2008. ISBN: 978-3-540-73561-8.

[107] Kristian S. Thygesen and Angel Rubio. Conserving gw scheme for nonequilibrium quantum transport in molecular contacts. *Physical Review B (Condensed Matter and Materials Physics)*, 77(11):115333, 2008.

[108] F. Guinea, C. Tejedor, F. Flores, and E. Louis. Effective two-dimensional hamiltonian at surfaces. *Phys. Rev. B*, 28(8):4397–4402, Oct 1983.

[109] D. C. Langreth. In Jozef T. Devreese and V. E. van Doren, editors, *Linear and nonlinear electron transport in solids*, lectures presented at the NATO Advanced Study Institute on Linear and Nonlinear Electronic Transport in Solids. Plenum Press, State University of Antwerp (RUCA), Antwerp, Belgium, 1976.

[110] K. S. Thygesen. Electron transport through an interacting region: The case of a nonorthogonal basis set. *Physical Review B*, 73(3):035309, January 2006.

[111] K. S. Thygesen and K. W. Jacobsen. Interference and k-point sampling in the supercell approach to phase-coherent transport. *Physical Review B*, 72(3):0333401, July 2005.

[112] Lars Hedin. New method for calculating the one-particle green's function with application to the electron-gas problem. *Phys. Rev.*, 139(3A):A796–A823, Aug 1965.

[113] Wilfried G. Aulbur, Lars Jönsson, and John W. Wilkins. Quasiparticle calculations in solids. In Henry Ehrenreich and Frans Spaepen, editors, *Advances in Research and Applications*, volume 54 of *Solid State Physics*, pages 1 – 218. Academic Press, 1999. Freely available at `http://www.physics.ohio-state.edu/~wilkins/vita/gw_review.ps`.

[114] W. Lee, N. Jean, and S. Sanvito. Exploring the limits of the self-consistent born approximation for inelastic electronic transport. *Physical Review B*, 79(8):85120, 2009.

[115] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59:1758–1775, Jan 1999.

[116] N. Troullier and J. L. Martins. Efficient pseudopotentials for plane-wave calculations. *Physical Review B*, 43:1993–2006, 1991.

[117] A. Rohrbach, J. Hafner, and G. Kresse. Molecular adsorption on the surface of strongly correlated transition-metal oxides: A case study for co/nio(100). *Physical Review B*, 69(7):075413, Feb 2004.

[118] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton. Electron-energy-loss spectra and the structural stability of nickel oxide: An lsda+u study. *Phys. Rev. B*, 57(3):1505–1509, Jan 1998.

[119] F. Yin, J. Akola, P. Koskinen, M. Manninen, and R. E. Palmer. Bright beaches of nanoscale potassium islands on graphite in stm imaging. *Physical Review Letters*, 102(10):106102, 2009.

[120] W Tang, E Sanville, and G Henkelman. A grid-based bader analysis algorithm without lattice bias. *Journal of Physics: Condensed Matter*, 21(8):084204 (7pp), 2009.

[121] Bader charge analysis. The code is freely available at `http://theory.cm.utexas.edu/henkelman/research/bader/`.

[122] Michael Walter, Hannu Häkkinen, Lauri Lehtovaara, Martti Puska, Jussi Enkovaara, Carsten Rostgaard, and Jens Jørgen Mortensen. Time-dependent density-functional theory in the projector augmented-wave method. *The Journal of Chemical Physics*, 128(24):244101, 2008.

[123] Carsten Rostgaard, Karsten W. Jacobsen, and Kristian S. Thygesen. Assessment of the gw approximation for molecules. ii. ionization potentials from first principles and comparison to hybrid density functional theory. Submitted, 2009.

[124] Joachim Paier, Robin Hirschl, Martijn Marsman, and Georg Kresse. The perdew–burke–ernzerhof exchange-correlation functional applied to the g2-1 test set using a plane-wave basis set. *The Journal of Chemical Physics*, 122(23):234102, 2005.

[125] J. B. Krieger, Y. Li, and G. J. Iafrate. Systematic approximations to the optimized effective potential: application to orbital-density-functional theory. *Physical Review A*, 46:5453–5458, 1992.

[126] Xifan Wu, Annabella Selloni, and Roberto Car. Order-n implementation of exact exchange in extended insulating systems. *Physical Review B*, 79(8):085102, feb 2009.

[127] Xifan Wu, Eric J. Walter, Andrew M. Rappe, Roberto Car, and Annabella Selloni. Hybrid density functional calculations of the band gap of $Ga_xIn_{1-x}N$. arXiv, 7 2009.

[128] Manuel Guidon, Jürg Hutter, and Joost VandeVondele. Robust periodic hartree-fock exchange for large-scale simulations using gaussian basis sets. *Journal of Chemical Theory and Computation*, October 2009.

[129] Erich Runge and E. K. U. Gross. Density-functional theory for time-dependent systems. *Phys. Rev. Lett.*, 52(12):997, Mar 1984.

[130] M.E. Casida. Time-dependent density-functional response theory for molecules. In D.P. Chong, editor, *Recent Advances in Density Functional Methods, Part I*, page 155. World Scientific, Singapore, 1995.

[131] M. E. Casida. Time-dependent density functional response theory of molecular systems: Theory, computational methods, and functionals. In J. M. Seminario, editor, *Recent Developments and Applications in Modern Density-Functional Theory*, page 391. Elsevier, Amsterdam, 1996.

[132] Mark E. Casida. Propagator corrections to adiabatic time-dependent density-functional theory linear response theory. *J. Chem. Phys.*, 122(5):054111, 2005.

[133] M. Margues, C. A. Ullrich, F. Nogueira, A. Rubio, K. Burke, and E. K. U. Gross, editors. *Time-Dependent Density Functional Theory*, volume 706 of *Lecture Notes in Physics*. Springer, 2006.

[134] The Vienna *ab-initio* simulation package (VASP) is a plane-wave based PAW-DFT code. http://cms.mpi.univie.ac.at/vasp.

[135] Larry A. Curtiss, Krishnan Raghavachari, Paul C. Redfern, and John A. Pople. Assessment of gaussian-2 and density functional theories for the computation of enthalpies of formation. *The Journal of Chemical Physics*, 106(3):1063–1079, 1997. Data available from http://www.cse.anl.gov/OldCHMwebsiteContent/compmat/comptherm.htm.

[136] Larry A. Curtiss, Paul C. Redfern, and David J. Frurip. Theoretical methods for computing enthalpies of formation of gaseous compounds. In Donald B. Boyd Kenny B. Lipkowitz, editor, *Reviews in Computational Chemistry*, pages 147–211. 2007.

[137] Russell D. Johnson III. Nist standard reference database number 101. http://srdata.nist.gov/cccbdb, September 2006. Release 14.

[138] M. Marsman and G. Kresse. Relaxed core projector-augmented-wave method. *The Journal of Chemical Physics*, 125(10):104101, 2006.

[139] Adi Makmal, Rickard Armiento, Eberhard Engel, Leeor Kronik, and Stephan Kümmel. Examining the role of pseudopotentials in exact-exchange-based kohn-sham gaps. *Physical Review B (Condensed Matter and Materials Physics)*, 80(16):161204, 2009.

[140] E. Engel. Relevance of core-valence interaction for electronic structure calculations with exact exchange. *Physical Review B (Condensed Matter and Materials Physics)*, 80(16):161205, 2009.

[141] T. Koopmans. Über die zuordnung von wellenfunktionen und eigenwerten zu den einzelnen elektronen eines atoms. *Physica*, 1(1-6):104–113, 1934.

[142] Peter Fulde. *Electron correlations in molecules and solids*. Springer series in solid state sciences ; 100. Springer, Berlin, 3., enlarged ed., corr. print. edition, 2002.

[143] O. V. Gritsenko and E. J. Baerends. The analog of koopmans' theorem in spin-density functional theory. *Journal of Chemical Physics*, 117:9154–9159, 2002.

[144] Kristian S. Thygesen and Angel Rubio. Renormalization of molecular quasiparticle levels at metal-molecule interfaces: Trends across binding regimes. *Physical Review Letters*, 102(4):046802, 2009.

[145] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Ángyán. Screened hybrid density functionals applied to solids. *The Journal of Chemical Physics*, 124:154709, 2006.

[146] Jochen Heyd, Juan E. Peralta, Gustavo E. Scuseria, and Richard L. Martin. Energy band gaps and lattice parameters evaluated with the heyd-scuseria-ernzerhof screened hybrid functional. *The Journal of Chemical Physics*, 123(17):174101, 2005.

[147] Mark S. Hybertsen and Steven G. Louie. Electron correlation in semiconductors and insulators: Band gaps and quasiparticle energies. *Phys. Rev. B*, 34(8):5390–5413, Oct 1986.

[148] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. Ab initio calculation of vibrational absorption and circular dichroism spectra using density functional force fields. *The Journal of Physical Chemistry*, 98(45):11623–11627, November 1994.

[149] J. Heyd, G. E. Scuseria, and M. Ernzerhof. Hybrid functionals based on a screened coulomb potential. *The Journal of Chemical Physics*, 118:8207–8215, 2003.

[150] J. Heyd and G. E. Scuseria. Assessment and validation of a screened coulomb hybrid density functional. *The Journal of Chemical Physics*, 120:7274–7280, 2004.

[151] Jun-Qiang Sun and Rodney J. Bartlett. Second-order many-body perturbation-theory calculations in extended systems. *The Journal of Chemical Physics*, 104(21):8553–8565, 1996.

[152] G. Grosso and G. P. Parravicini. *Solid State Physics*. Cambridge University Press, 2000.

[153] C. D. Pemmaraju, T. Archer, D. Sánchez-Portal, and S. Sanvito. Atomic-orbital-based approximate self-interaction correction scheme for molecules and solids. *Physical Review B (Condensed Matter and Materials Physics)*, 75(4):045101, 2007.

[154] Giovanni Onida, Lucia Reining, and Angel Rubio. Electronic excitations: density-functional versus many-body green's-function approaches. *Rev. Mod. Phys.*, 74(2):601–659, Jun 2002.

[155] Mikkel Strange. Note on projected wannier functions. Unpublished, 2009.

[156] Xiaofeng Qian, Ju Li, Liang Qi, Cai-Zhuang Wang, Tzu-Liang Chan, Yong-Xin Yao, Kai-Ming Ho, and Sidney Yip. Quasiatomic orbitals for ab initio tight-binding analysis. *Physical Review B (Condensed Matter and Materials Physics)*, 78(24):245112, 2008.

[157] K. S. Thygesen, L. B. Hansen, and K. W. Jacobsen. Partly occupied wannier functions: Construction and applications. *Physical Review B*, 72(12):125119, September 2005.

[158] K. S. Thygesen, L. B. Hansen, and K. W. Jacobsen. Partly occupied wannier functions. *Physical Review Letters*, 94(2):026405, January 2005.

[159] A Ferretti, A Calzolari, B Bonferroni, and R Di Felice. Maximally localized wannier functions constructed from projector-augmented waves or ultrasoft pseudopotentials. *Journal of Physics: Condensed Matter*, 19(3):036215 (16pp), 2007.

[160] S. Obara and A. Saika. Efficient recursive computation of molecular integrals over cartesian gaussian functions. *The Journal of Chemical Physics*, 84(7):3963–3974, 1986.

[161] Herbert H. H. Homeier and E. Otto Steinborn. Some properties of the coupling coefficients of real spherical harmonics and their relation to gaunt coefficients. *Journal of Molecular Structure (Theochem)*, 368:31–37, 1996. Proceedings of the Second Electronic Computational Chemistry Conference.

[162] F. Aryasetiawan and O. Gunnarsson. Product-basis method for calculating dielectric matrices. *Phys. Rev. B*, 49(23):16214–16222, Jun 1994.

[163] Adrian Stan, Nils Erik Dahlen, and Robert van Leeuwen. Levels of self-consistency in the gw approximation. *The Journal of Chemical Physics*, 130(11):114105, 2009.

[164] M. Shishkin and G. Kresse. Implementation and performance of the frequency-dependent gw method within the paw framework. *Physical Review B (Condensed Matter and Materials Physics)*, 74(3):035101, 2006.

[165] M. Shishkin and G. Kresse. Self-consistent gw calculations for semiconductors and insulators. *Physical Review B (Condensed Matter and Materials Physics)*, 75(23):235102, 2007.

[166] The NEGF code KeldyshGF is freely available from the subversion server `https://svn.fysik.dtu.dk/projects/KeldyshGF/trunk/`.

[167] P. Pulay. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chemical Physics Letters*, 73:393–398, 1980.

[168] W. H. Press, B. P. Flannery, S. A. Teukolski, and W. T. Vetterling. *Numerical recipes in C*. Cambridge University Press, Cambridge, 1 edition, 1986. Older editions are freely available online.

[169] Nakhlé H. Asmar. *Partial differential equations with Fourier series and boundary value problems*. Pearson Prentice Hall, 2nd ed. edition, 2004. ISBN 0-13-148096-0.

[170] Ralph de Laer Kronig. On the theory of dispersion of x-rays. *Journal of the Optical Society of America*, 12(6):547–556, 1926.

[171] A. Stan, N. E. Dahlen, and R. van Leeuwen. Fully self-consistent gw calculations for atoms and molecules. *EPL (Europhysics Letters)*, 76(2):298–304, 2006.

[172] Jeffrey C. Grossman, Michael Rohlfing, Lubos Mitas, Steven G. Louie, and Marvin L. Cohen. High accuracy many-body calculational approaches for excitations in molecules. *Phys. Rev. Lett.*, 86(3):472–475, Jan 2001.

[173] T. A. Niehaus, M. Rohlfing, F. Della Sala, A. Di Carlo, and Th. Frauenheim. Quasiparticle energies for large molecules: A tight-binding-based green's-function approach. *Phys. Rev. A*, 71(2):022508, Feb 2005.

[174] Catalin D. Spataru, Sohrab Ismail-Beigi, Lorin X. Benedict, and Steven G. Louie. Excitonic effects and optical spectra of single-walled carbon nanotubes. *Phys. Rev. Lett.*, 92(7):077402, Feb 2004.

[175] Murilo L. Tiago and James R. Chelikowsky. Optical excitations in organic molecules, clusters, and defects studied by first-principles green's function methods. *Physical Review B (Condensed Matter and Materials Physics)*, 73(20):205334, 2006.

[176] Kristian S. Thygesen and Angel Rubio. Nonequilibrium gw approach to quantum transport in nano-scale contacts. *The Journal of Chemical Physics*, 126(9):091101, 2007.

[177] K. S. Thygesen. Impact of exchange-correlation effects on the iv characteristics of a molecular junction. *Physical Review Letters*, 100(16):166804, 2008.

[178] Latha Venkataraman, Young S. Park, Adam C. Whalley, Colin Nuckolls, Mark S. Hybertsen, and Michael L. Steigerwald. Electronics and chemistry: Varying single-molecule junction conductance using chemical substituents. *Nano Letters*, 7(2):502–506, February 2007.

[179] S.G. Lias. Ionization energy evaluation. In P.J. Linstrom and W.G. Mallard, editors, *NIST Chemistry WebBook, NIST Standard Reference Database Number 69*. National Institute of Standards and Technology, Gaithersburg MD, 20899, retrieved November 9, 2009. http://webbook.nist.gov.

[180] D. M. Newns. Self-consistent model of hydrogen chemisorption. *Phys. Rev.*, 178(3):1123–1135, Feb 1969.

[181] R. H. M. Smit, Y. Noat, C. Untiedt, N. D. Lang, M. C. van Hemert, and J. M. van Ruitenbeek. Measurement of the conductance of a hydrogen molecule. *Nature*, 419(6910):906–909, October 2002.

[182] J.C. Cuevas, J. Heurich, F. Pauly, W. Wenzel, and G. Schon. Theoretical description of the electrical conduction in atomic and molecular junctions. *Nanotechnology*, 14(8):29, 2003.

[183] Y. García, J. J. Palacios, E. SanFabían, J. A. Vergés, A. J. Pérez-Jiménez, and E. Louis. Electronic transport and vibrational modes in a small molecular bridge: H_2 in pt nanocontacts. *Phys. Rev. B*, 69(4):041402, 2004.

[184] K. S. Thygesen and K. W. Jacobsen. Conduction mechanism in a molecular hydrogen contact. *Physical Review Letters*, 94(3):036807, January 2005.

[185] D. Djukic, K. S. Thygesen, C. Untiedt, R. H. M. Smit, K. W. Jacobsen, and J. M. van Ruitenbeek. Stretching dependence of the vibration modes of a single-molecule pt-h-2-pt bridge. *Physical Review B*, 71(16):161402, April 2005.

[186] I. S. Kristensen, M. Paulsson, K. S. Thygesen, and K. W. Jacobsen. Inelastic scattering in metal-h_2-metal junctions. *Physical Review B (Condensed Matter and Materials Physics)*, 79(23):235411, 2009.

[187] D. J. Mowbray, G. Jones, and K. S. Thygesen. Influence of functional groups on charge transport in molecular junctions. *The Journal of Chemical Physics*, 128(11):111103, 2008.

[188] Nils Erik Dahlen and Robert van Leeuwen. Self-consistent solution of the dyson equation for atoms and molecules within a conserving approximation. *The Journal of Chemical Physics*, 122(16):164102, 2005.

[189] Magnus Paulsson, Thomas Frederiksen, Hiromu Ueba, Nicolás Lorente, and Mads Brandbyge. Unified description of inelastic propensity rules for electron transport through nanoscale junctions. *Physical Review Letters*, 100(22):226604, 2008.

[190] R. Landauer. Spatial variation of currents and fields due to localized scatterers in metallic conduction. *IBM Journal of Research and Development*, 1(3):223–231, 1957.

[191] D. Djukic and J. M. van Ruitenbeek. Shot noise measurements on a single molecule. *Nano Letters*, 6(4):789–793, April 2006.

[192] Ya. M. Blanter and M. Büttiker. Shot noise in mesoscopic conductors. *Physics Reports*, 336(1-2):1 – 166, 2000.

[193] Mads Brandbyge, Mads R. Sørensen, and Karsten W. Jacobsen. Conductance eigenchannels in nanocontacts. *Physical Review B*, 56(23):14956–14959, December 1997.

[194] M. Strange, K. S. Thygesen, and K. W. Jacobsen. Electron transport in a pt-co-pt nanocontact: Density functional theory calculations. *Physical Review B*, 73(12):125424, March 2006.

[195] Magnus Paulsson and Mads Brandbyge. Transmission eigenchannels from nonequilibrium green's functions. *Physical Review B (Condensed Matter and Materials Physics)*, 76(11):115117, 2007.

[196] M. Paulsson, T. Frederiksen, and M. Brandbyge. Modeling inelastic phonon scattering in atomic- and molecular-wire junctions. *Physical Review B*, 72(20):201101, November 2005.

[197] M. Paulsson, T. Frederiksen, and M. Brandbyge. Phonon scattering in nanoscale systems: Lowest order expansion of the current and power expressions. *J. Phys.: Conf. Ser.*, 35:247, 2006.

Paper I

Time-dependent density-functional theory in the projector augmented-wave method
M. Walter, H. Häkkinen, L. Lehtovaara, M. Puska, J. Enkovaara, C. Rostgaard, and J. J. Mortensen
The Journal of Chemical Physics **128**, 244101 (2008).

Time-dependent density-functional theory in the projector augmented-wave method

Michael Walter,[1,a)] Hannu Häkkinen,[1,2] Lauri Lehtovaara,[3] Martti Puska,[3] Jussi Enkovaara,[4] Carsten Rostgaard,[5] and Jens Jørgen Mortensen[5]
[1]*Department of Physics, Nanoscience Center, University of Jyväskylä, FIN-40014 Jyväskylä, Finland*
[2]*Department of Chemistry, Nanoscience Center, University of Jyväskylä, FIN-40014, Jyväskylä, Finland*
[3]*Department of Engineering Physics, Helsinki University of Technology, P.O. Box 1100, FIN-02015 TKK, Finland*
[4]*CSC Scientific Computing Ltd., FI-02101 Espoo, Finland*
[5]*CAMd, Department of Physics, Technical University of Denmark, DK-2800 Lyngby, Denmark*

(Received 9 April 2008; accepted 19 May 2008; published online 23 June 2008)

We present the implementation of the time-dependent density-functional theory both in linear-response and in time-propagation formalisms using the projector augmented-wave method in real-space grids. The two technically very different methods are compared in the linear-response regime where we found perfect agreement in the calculated photoabsorption spectra. We discuss the strengths and weaknesses of the two methods as well as their convergence properties. We demonstrate different applications of the methods by calculating excitation energies and excited state Born–Oppenheimer potential surfaces for a set of atoms and molecules with the linear-response method and by calculating nonlinear emission spectra using the time-propagation method. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2943138]

I. INTRODUCTION

The density-functional theory[1,2] (DFT) has been very successful for ground-state calculations of molecular and condensed-matter systems due to its favorable balance of cost against accuracy. Properties such as ground-state total energies, lattice constants, and equilibrium geometries are nowadays calculated routinely for systems containing up to a few hundred atoms. However, there are several scientifically and technologically interesting quantities which are related to excited states of the system and are thus beyond the realm of the standard DFT. In recent years, the time-dependent DFT (TDDFT) (Ref. 3) has become a popular tool for calculating excited-state properties such as linear and nonlinear optical responses.[4–9]

The most general realization of the TDDFT is the time-propagation scheme[5] in which the time-dependent Kohn–Sham (KS) equations are integrated over the time domain. In the linear-response regime the excitation energies can also be calculated in the frequency space by solving a matrix equation in a particle-hole basis.[4] This is the so-called linear-response scheme. The time-propagation and the linear-response scheme are complementary as they have different advantages and disadvantages. For example, the linear-response scheme provides all the excitations in a single calculation, while the time-propagation provides only the excitations corresponding to the given initial perturbation and several separate calculations may be needed. On the other hand, the time-propagation has a wider applicability as also non-linear-response phenomena, such as the high-harmonics generation in intense laser beams and general time-dependent phenomena, in which for example the ionic structure relaxes as a function of time, can be studied. Computationally, the time-propagation scales more favorably with the system size than the linear-response scheme. However, the prefactor in time-propagation is larger, so that the cross-over in efficiency is reached at relatively large systems.

Previously, there have been several implementations of the linear-response and the time-propagation formalisms using a variety of methods such as localized basis sets,[10,11] plane waves,[12–15] and real-space grids.[5,16,17] The plane-wave and the real-space implementations have used the pseudopotential approximation which has been either of the norm-conserving or ultrasoft flavor. To our knowledge, the projector augmented-wave (PAW) method[18] has not been used in time-dependent density-functional calculations previously. Here, we present implementation of both time-propagation and linear-response TDDFT in the electronic-structure program GPAW,[19,20] which uses the PAW method and uniform real-space grids.

The real-space PAW method has several advantages both in ground-state and in time-dependent calculations. First, there is a single convergence parameter, the grid spacing, which controls the accuracy of the discretization. Different boundary conditions can be handled easily and especially the ability to treat finite systems without supercells is important for TDDFT. The PAW method can be applied on the same footing to all elements, for example, it provides a reliable description of the transition metal elements and the first row elements with open *p*-shells. These are often problematic for standard pseudopotentials. Also, the PAW method reduces the number of grid points required for accurate calculations in comparison with pseudopotential calculations. Thus, the dimension of the Hamiltonian matrix is reduced and one is

[a)]Electronic mail: michael.walter@phys.jyu.fi.

also allowed to use longer time steps in time-propagation.[15] Finally, the real-space formalism allows efficient parallelization with domain-decomposition techniques.

The present paper is organized as follows. In Sec. II A we present the basic features of the PAW method. The linear-response formulation of the TDDFT within the PAW method is presented in Sec. II B and the time-propagation scheme is reviewed in Sec. II C. In Sec. III we show that the two methods give identical results in the linear-response regime by calculating the optical absorption spectra for the Na_2 and C_6H_6 molecules. Next, we focus on the linear-response scheme and calculate excitation energies for a set of divalent atoms followed by Born–Oppenheimer potentials of excited states of Na_2. The applicability of the time-propagation in the nonlinear regime is demonstrated by calculating emission spectra of the Be atom in strong laser fields. The convergence properties of the two methods are discussed in Sec. IV. Finally, we give a brief summary in Sec. V.

II. THEORY

A. Ground state

The implementation of the PAW method using a real-space grid is explained in detail in Ref. 19. We will give here just a short introduction with the main purpose of defining the quantities needed for the time-propagation and linear-response calculation. In the PAW method, a true all-electron KS wave function ψ_n can be obtained through a linear transformation from a smooth pseudo-wave-function $\tilde{\psi}_n$ via

$$\psi_n(\mathbf{r}) = \hat{\mathcal{T}}\tilde{\psi}_n, \tag{1}$$

where n denotes a combined band and spin index. Using the explicit representation of $\hat{\mathcal{T}}$, the KS wave functions can be expressed as

$$\psi_n(\mathbf{r}) = \tilde{\psi}_n(\mathbf{r}) + \sum_a \left(\psi_n^a(\mathbf{r}-\mathbf{R}_a) - \tilde{\psi}_n^a(\mathbf{r}-\mathbf{R}_a)\right), \tag{2}$$

where ψ_n^a and $\tilde{\psi}_n^a$ are the all-electron and smooth continuations of ψ_n inside the augmentation region of the atom a at position $\mathbf{R}_a$, respectively. Their difference vanishes by definition outside the augmentation region. ψ_n^a and $\tilde{\psi}_n^a$ may be expanded in terms of atom-centered all-electron wave functions ϕ^a and their smooth counterparts $\tilde{\phi}^a$, respectively, i.e.,

$$\psi_n^a(\mathbf{r}) = \sum_j P_{nj}^a \phi_j^a(\mathbf{r}), \quad \tilde{\psi}_n^a(\mathbf{r}) = \sum_j P_{nj}^a \tilde{\phi}_j^a(\mathbf{r}), \tag{3}$$

with the same coefficients $P_{nj}^a = \langle p_j | \psi_n \rangle$, where the p_j are the so called projector functions.[18,19] The main quantity of DFT, the electron density $n(\mathbf{r})$ has a similar partitioning as the wave functions (this behavior can be shown to be true for all quantities that can be expressed as expectation values of local operators[18]). Thus,

$$n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + \sum_a \left(n^a(\mathbf{r}-\mathbf{R}_a) - \tilde{n}^a(\mathbf{r}-\mathbf{R}_a)\right), \tag{4}$$

where the all-electron density inside the augmentation region

$$n^a(\mathbf{r}) = \sum_{i_1 i_2} D_{i_1 i_2}^a \phi_{i_1}^a(\mathbf{r}) \phi_{i_2}^a(\mathbf{r}) \tag{5}$$

and its smooth counterpart

$$\tilde{n}^a(\mathbf{r}) = \sum_{i_1 i_2} D_{i_1 i_2}^a \tilde{\phi}_{i_1}^a(\mathbf{r}) \tilde{\phi}_{i_2}^a(\mathbf{r}) \tag{6}$$

appear. Denoting the ground-state occupation numbers by f_n, the above atomic density matrix can be expressed as[19]

$$D_{i_1 i_2}^a = \sum_n P_{n i_1}^{a*} f_n P_{n i_2}^a. \tag{7}$$

B. Linear response

In the following we discuss the linear-response theory in the TDDFT from a practical view, rather than from its formal derivation which can be found in the original references[4,10,21] or in more recent work.[22] We follow closely the notation used by Casida,[4] who showed that in the linear-response TDDFT the calculation of excitation energies can be reduced to solving the eigenvalue equation of the following form:

$$\Omega F_I = \omega_I^2 F_I, \tag{8}$$

where $\hbar\omega_I$ is the transition energy from the ground state to the excited state I. Expanding the matrix $\mathbf{\Omega}$ in KS single particle-hole excitations leads to

$$\Omega_{ij\sigma,kq\tau} = \delta_{ik}\delta_{jq}\delta_{\sigma\tau}\varepsilon_{ij\sigma}^2 + 2\sqrt{f_{ij\sigma}\varepsilon_{ij\sigma}f_{kq\tau}\varepsilon_{kq\tau}}K_{ij\sigma,kq\tau}, \tag{9}$$

where $\varepsilon_{ij\sigma} = \varepsilon_{j\sigma} - \varepsilon_{i\sigma}$ are the energy differences and $f_{ij\sigma} = f_{i\sigma} - f_{j\sigma}$ are the occupation number differences of the KS states. The indices i, j, k, and q are band indices, whereas σ and τ denote spin projection indices. The coupling matrix can be split into two parts $K_{ij\sigma,kq\tau} = K_{ij\sigma,kq\tau}^{\mathrm{RPA}} + K_{ij\sigma,kq\tau}^{\mathrm{xc}}$. The former is the so-called random phase approximation (RPA) part,

$$K_{ij\sigma,kq\tau}^{\mathrm{RPA}} = \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{n_{ij\sigma}^*(\mathbf{r}_1) n_{kq\tau}(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} =: (n_{ij\sigma}|n_{kq\tau}), \tag{10}$$

where $n_{ij\sigma}$ is the i,j density matrix element or pair density corresponding to the spin σ. $K_{ij\sigma,kq\tau}^{\mathrm{RPA}}$ describes the effect of the linear density response via the classical Hartree energy. The second is the exchange-correlation part,

$$K_{ij\sigma,kq\tau}^{\mathrm{xc}} = \int d\mathbf{r}_1 d\mathbf{r}_2 n_{ij\sigma}^*(\mathbf{r}_1) \frac{\delta^2 E_{\mathrm{xc}}}{\delta\rho_\sigma(\mathbf{r}_1)\delta\rho_\tau(\mathbf{r}_2)} n_{kq\tau}(\mathbf{r}_2), \tag{11}$$

where ρ_σ is the spin density. $K_{ij\sigma,kq\tau}^{\mathrm{xc}}$ describes the effect of the linear density response via the exchange and correlation energy.

We discuss the forms of the coupling matrix for the two parts separately and suppress the explicit dependence on the spin projection unless it is explicitly needed. In both parts of the coupling matrix the pair density $n_{ij}(\mathbf{r}) = \psi_i^*(\mathbf{r})\psi_j(\mathbf{r})$ appears. This quantity can be partitioned in the same way as the electron density, i.e.,

$$n_{ij} = \tilde{n}_{ij} + \sum_a \left(n_{ij}^a - \tilde{n}_{ij}^a\right), \tag{12}$$

where we have dropped the dependence on the position for brevity. Inserting this expression directly into the integral in

Eq. (10) would lead to overlaps of different augmentation spheres due to the nonlocality of the Coulomb operator $|\mathbf{r}_1-\mathbf{r}_2|^{-1}$. These overlaps have to be avoided. The same problem appears already in the calculations of the Hartree energy in the ground-state problem.[18,19] It can be solved by introducing compensation charge densities $\tilde{Z}_{ij}^a$, defined to fulfill

$$\int d\mathbf{r}_2 \frac{n_{ij}^a(\mathbf{r}_2)-\tilde{n}_{ij}^a(\mathbf{r}_2)-\tilde{Z}_{ij}^a(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}=0, \tag{13}$$

for $|\mathbf{r}_1-\mathbf{R}_a|>r_c^a$, i.e., outside the augmentation sphere. The compensation charge densities can be expanded in terms of spherical harmonics Y_L,[18,19]

$$\tilde{Z}_{ij}^a(\mathbf{r})=\sum_L Q_{L,ij}^a g_l^a(r) Y_L(\hat{\mathbf{r}}), \tag{14}$$

where L stands for the combined values of angular momentum quantum numbers l and m. The choice of local functions $g_l^a(r)$ is arbitrary as long as they fulfill

$$\int dr r^{l+2} g_l^a(r)=1, \tag{15}$$

and they are sufficiently localized inside the augmentation sphere. For the particular choice of $g_l^a(r)$ in our calculations we refer to Eq. (B1) in Ref. 19. Due to Eq. (13) the coefficients $Q_{L,ij}^a$ have to be

$$Q_{L,ij}=\sum_{i_1 i_2} \Delta_{L,i_1 i_2} P_{i i_1}^a P_{j i_2}^a, \tag{16}$$

with the constants

$$\Delta_{L,i_1 i_2}=\int d\mathbf{r} r^l Y_L(\hat{\mathbf{r}})[\phi_{i_1}^a(\mathbf{r})\phi_{i_2}^a(\mathbf{r})-\tilde{\phi}_{i_1}^a(\mathbf{r})\tilde{\phi}_{i_2}^a(\mathbf{r})]. \tag{17}$$

Using the shorthand

$$\tilde{\rho}_{ij}(\mathbf{r}) := \tilde{n}_{ij}(\mathbf{r})+\sum_a \tilde{Z}_{ij}^a(\mathbf{r}-\mathbf{R}_a), \tag{18}$$

we may write the RPA part of the kernel in the following form:

$$K_{ij,kq}^{\mathrm{RPA}}=(\tilde{\rho}_{ij}|\tilde{\rho}_{kq})+\sum_a \Delta K_{ij,kq}^{\mathrm{RPA},a}, \tag{19}$$

which has the desired partitioning in a pure smooth part $(\tilde{\rho}_{ij}|\tilde{\rho}_{kq})$ and local corrections $\Delta K_{ij\sigma,kq\tau}^{\mathrm{RPA},a}$ inside the augmentation spheres. The explicit form of these corrections is given in Appendix A [Eq. (A3)].

The exchange-correlation part of the coupling matrix is evaluated in a finite-difference scheme[23,24] as

$$K_{ij\sigma,kq\tau}^{\mathrm{xc}}[n_\sigma,n_\tau]=\lim_{\varepsilon\to 0}\int d\mathbf{r} n_{ij\sigma}^*(\mathbf{r})\frac{v_{\mathrm{xc}}^\sigma[n_\sigma,n_\tau+\varepsilon n_{kq\tau}](\mathbf{r})-v_{\mathrm{xc}}^\sigma[n_\sigma,n_\tau-\varepsilon n_{kq\tau}](\mathbf{r})}{2\varepsilon}, \tag{20}$$

where we denote that $K_{ij\sigma,kq\tau}^{\mathrm{xc}}$ is a functional of the spin densities explicitly. The finite-difference scheme is quite insensitive to the actual numerical value for ε as will be shown in Sec. III. For the local density approximation (LDA) and the generalized gradient approximation for the electron exchange and correlation we can write

$$K_{ij\sigma,kq\tau}^{\mathrm{xc}}[n_\sigma,n_\tau]=\tilde{K}_{ij\sigma,kq\tau}^{\mathrm{xc}}[\tilde{n}_\sigma,\tilde{n}_\tau]+\sum_a \Delta K_{ij\sigma,kq\tau}^{\mathrm{xc},a}, \tag{21}$$

where $\tilde{K}_{ij\sigma,kq\tau}^{\mathrm{xc}}$ depends on the smooth densities and the corrections $\Delta K_{ij\sigma,kq\tau}^{\mathrm{xc},a}$ are localized inside the atomic augmentation spheres. The explicit form of these corrections is given in Appendix B [Eq. (B1)].

In optical absorption spectra not only the excitation energies but also the corresponding dipole oscillator strengths are of interest. They are dimensionless and can be written as

$$f_{I\alpha}=\frac{2m_e}{\hbar e^2}\omega_I \mu_{I\alpha}^2, \tag{22}$$

where m_e is the electron mass, e is the unit charge, and $\alpha=x,y,z$ denotes the direction of the light polarization. The dipole transition moment,

$$\boldsymbol{\mu}_I=-e\langle 0|\sum_{k=1}^N \mathbf{r}_k|I\rangle, \tag{23}$$

is defined through the many-particle ground and excited states $|0\rangle$ and $|I\rangle$, respectively. Above, N is the number of electrons with their coordinates $\mathbf{r}_k$, $k=1,\ldots,N$. In linear-response TDDFT the oscillator strength for a transition I can be obtained using the corresponding eigenvector F_I of the Ω matrix and the KS transition dipoles,

$$\boldsymbol{\mu}_{ij\sigma}=-e\langle\psi_{i\sigma}|\mathbf{r}|\psi_{j\sigma}\rangle, \tag{24}$$

between the KS states $\psi_{i\sigma}$ and $\psi_{j\sigma}$. The oscillator strengths are evaluated then as[4]

$$f_{I\alpha}=\frac{2m_e}{\hbar e^2}\left|\sum_{ij\sigma}^{f_{i\sigma}>f_{j\sigma}}(\boldsymbol{\mu}_{ij\sigma})_\alpha\sqrt{f_{ij\sigma}\varepsilon_{ij\sigma}}(F_I)_{ij\sigma}\right|^2. \tag{25}$$

In PAW the KS transition dipoles can be partitioned as

$$\boldsymbol{\mu}_{ij\sigma}=-e\langle\tilde{\psi}_{i\sigma}|\mathbf{r}|\tilde{\psi}_{j\sigma}\rangle+\sum_a\sum_{pq}P_{i\sigma p}^* P_{j\sigma q}\boldsymbol{\mu}_{pq}^a, \tag{26}$$

where the local corrections are

$$(\boldsymbol{\mu}^a_{pq})_m = -e\sqrt{4\pi}\left[\frac{\Delta^a_{1m,pq}}{\sqrt{3}} + \Delta^a_{L=0,pq}(\mathbf{R}_a)_m\right], \tag{27}$$

with the constants $\Delta^a_{L,pq}$ defined in Eq. (17).

C. Time-propagation

The scheme for propagating time-dependent KS wave functions within the ultrasoft pseudopotential or projector augmented-wave method was already described by Qian *et al.*[14] As our implementation follows closely theirs, it is reviewed only briefly here.

The all-electron time-dependent Schrödinger-type KS equation with the Hamiltonian $\hat{H}(t)$, i.e.,

$$i\hbar\frac{\partial}{\partial t}\psi_n(t) = \hat{H}(t)\psi_n(t), \tag{28}$$

is transformed to the PAW formalism as follows. First the all-electron wave function is replaced by the projector operator operating on the pseudo-wave-function $\psi_n = \hat{\mathcal{T}}\tilde{\psi}_n$. Then Eq. (28) is operated from the left by the adjoint operator $\hat{\mathcal{T}}^\dagger$, i.e.,

$$i\hbar\hat{\mathcal{T}}^\dagger\frac{\partial}{\partial t}\hat{\mathcal{T}}\tilde{\psi}_n(t) = \hat{\mathcal{T}}^\dagger\hat{H}(t)\hat{\mathcal{T}}\tilde{\psi}_n(t). \tag{29}$$

If the projector operator $\hat{\mathcal{T}}$ is independent of time, i.e., the nuclei do not move, the above equation reads as

$$i\hbar\tilde{S}\frac{\partial}{\partial t}\tilde{\psi}_n(t) = \tilde{H}(t)\tilde{\psi}_n(t), \tag{30}$$

where $\tilde{S} = \hat{\mathcal{T}}^\dagger\hat{\mathcal{T}}$ is the PAW overlap operator and $\tilde{H}(t) = \hat{\mathcal{T}}^\dagger\hat{H}(t)\hat{\mathcal{T}}$ is the time-dependent PAW Hamiltonian including the external time-dependent potential.

The linear absorption spectrum is obtained in the time-propagation scheme by applying a very weak delta-function pulse of a dipole field,[5]

$$\mathbf{E}(t) = \epsilon\mathbf{k}^{\mathbf{o}}\delta(t)\frac{\hbar}{a_0 e}, \tag{31}$$

to the system and then following the time-evolution of the dipole vector $\boldsymbol{\mu}(t)$. Above, ϵ is a unitless perturbation strength parameter, $\mathbf{k}^{\mathbf{o}}$ is a unit vector giving the polarization direction of the field, and a_0 is the Bohr radius. The delta pulse excites all possible frequencies at time zero, so that the KS wave functions change instantaneously to

$$\psi(t=0^+) = \exp\left(i\frac{\epsilon}{a_0}\mathbf{k}^{\mathbf{o}}\cdot\mathbf{r}\right)\psi(t=0^-). \tag{32}$$

Then the system is let to evolve freely.

To see the connection to the linear-response calculations, we study the effect of the delta kick in the many-body picture. If the pulse strength is weak, i.e., $\epsilon \ll 1$, the time-dependent many-body wave function after the kick is

$$|\Psi(t=0^+)\rangle = \left(1 - i\frac{\epsilon}{ea_0}\mathbf{k}^{\mathbf{o}}\cdot\hat{\boldsymbol{\mu}}\right)|0\rangle + O(\epsilon^2), \tag{33}$$

where $\hat{\boldsymbol{\mu}} = -e\Sigma_{k=1}^N\mathbf{r}_k$ is the dipole operator. When the system evolves freely it can be expanded in eigenstates $|0\rangle$ and $|I\rangle$ of the unperturbed Hamiltonian as

$$|\Psi(t)\rangle = c_0|0\rangle + \sum_I e^{-i\omega_I t}c_I|I\rangle, \tag{34}$$

with the coefficients

$$c_0 = 1 - i\frac{\epsilon}{ea_0}\mathbf{k}^{\mathbf{o}}\cdot\langle 0|\hat{\boldsymbol{\mu}}|0\rangle, \tag{35}$$

and

$$c_I = -i\frac{\epsilon}{ea_0}\mathbf{k}^{\mathbf{o}}\cdot\langle I|\hat{\boldsymbol{\mu}}|0\rangle. \tag{36}$$

The time-dependent density can be written as[25]

$$n(\mathbf{r},t) = n_0(\mathbf{r}) + \sum_I (e^{-i\omega_I t}c_I\langle 0|\hat{n}(\mathbf{r})|I\rangle + \text{c.c.}), \tag{37}$$

where $\hat{n} = \Sigma_{k=1}^N\delta(\mathbf{r}-\mathbf{r}_k)$ denotes the density operator. In the absence of magnetic fields all states can be chosen to be real resulting in the time-dependent dipole moment $\boldsymbol{\mu}(t) = -e\int d\mathbf{r}n(\mathbf{r},t)\mathbf{r}$ of the following form:

$$\boldsymbol{\mu}(t) = \boldsymbol{\mu}(0) - \frac{2\epsilon}{ea_0}\sum_I \sin(\omega_I t)(\mathbf{k}^{\mathbf{o}}\cdot\boldsymbol{\mu}_I)\boldsymbol{\mu}_I. \tag{38}$$

From this the dipole transition moment and hence the oscillator strength can be extracted via the Fourier transform. In practice, one calculates the generalization of the oscillator strength, the dipole strength tensor with respect to the polarization direction, $\mathbf{k}^o$ via[14]

$$\mathbf{S}(\omega)\mathbf{k}^o = \frac{2m_e a_0}{e\hbar\pi}\omega\frac{1}{\epsilon}\int_0^T dt\,\sin(\omega t)g(t)[\boldsymbol{\mu}(0) - \boldsymbol{\mu}(t)], \tag{39}$$

where T is the simulation time, and $g(t)$ is an envelope function being finite in the time window only. The envelope function, typically a Gaussian or an exponential decay, yields the shapes of the simulated spectral lines, Gaussians and Lorentzians, respectively, removing the effects of the finite simulation time. The dipole strength tensor is connected to the folded oscillator strength via

$$\mathbf{k}^o_\alpha\cdot\mathbf{S}(\omega)\mathbf{k}^o_\alpha = \sum_I f_{I\alpha}\tilde{g}(\omega-\omega_I), \tag{40}$$

where $\tilde{g}(\omega)$ is the normalized Fourier transform of $g(t)$ and $\mathbf{k}^o_\alpha$ is the unit vector in the direction $\alpha = x,y,z$.

In addition to the linear regime, the time-propagation can be used to interrogate the nonlinear regime of the light-matter interaction. When an atom or a molecule resides in a laser field $\mathbf{E}(t) = \mathbf{E}_0\sin(\omega t)$ of frequency ω electrons begin to oscillate with this frequency. If the field is strong enough, nonlinear terms in the polarizability of the atom begin to contribute.[26] As a result, integer multiples of the field frequency, i.e., harmonics, appear in the emission spectrum. The intensities $\mathbf{H}$ of the emitted frequencies can be calculated from the acceleration of the dipole moment,[27] i.e.,

$$\mathbf{H}(\omega) \propto \left| \int_0^T dt \exp(i\omega t) \frac{d^2}{dt^2} g(t)[\boldsymbol{\mu}(t) - \boldsymbol{\mu}(0)] \right|^2. \quad (41)$$

In the present implementation of the time-propagation, the time-dependent equations are solved using the Crank–Nicolson propagator with a predictor-corrector step[28] (this choice is not unique, for other possible propagators see Ref. 29). The predictor-corrector scheme is required to efficiently handle the nonlinearity in the Hamiltonian, i.e., to obtain a reasonable approximation for the Hamiltonian in a future time. In the predictor step, the wave functions are propagated by approximating the Hamiltonian to be constant during the time step, i.e., $\tilde{H}(t+\Delta t/2)=\tilde{H}(t)+\mathcal{O}(\Delta t)$ and then solving a linear equation for the predicted future wave functions $\tilde{\psi}_n^{\text{pred}}(t+\Delta t)$,

$$(\tilde{S} + i\tilde{H}(t)\Delta t/2\hbar)\tilde{\psi}_n^{\text{pred}}(t+\Delta t) = (\tilde{S} - i\tilde{H}(t)\Delta t/2\hbar)\tilde{\psi}_n(t) + \mathcal{O}(\Delta t^2). \quad (42)$$

The Hamiltonian in the middle of the time step is approximated as

$$\tilde{H}(t+\Delta t/2) = \tfrac{1}{2}(\tilde{H}(t) + \tilde{H}_{\text{pred}}(t+\Delta t)), \quad (43)$$

where $\tilde{H}_{\text{pred}}(t+\Delta t)$ is obtained from the predicted wave functions. In the corrector step, the improved Hamiltonian $\tilde{H}(t+\Delta t/2)$ is used to obtain the final, more accurate, propagated wave functions $\tilde{\psi}_n(t+\Delta t)$ from

$$(\tilde{S} + i\tilde{H}(t+\Delta t/2)\Delta t/2\hbar)\tilde{\psi}_n(t+\Delta t) = (\tilde{S} - i\tilde{H}(t+\Delta t/2)\Delta t/2\hbar)\tilde{\psi}_n(t) + \mathcal{O}(\Delta t^3). \quad (44)$$

The matrices in the linear equations [Eqs. (42) and (44)] are complex symmetric (not Hermitian), and we solve the equations using the biconjugate gradient stabilized method.[30]

As the Crank–Nicolson propagator is valid only for a short time step, repeated application of the propagator is required in any practical simulation. Note that no further improvement in the order of the error is obtained by repeating the corrector step with an improved approximation because the Crank–Nicolson itself is only accurate to the second order. Thus, in order to obtain more accurate results, it is more efficient to reduce the time step instead of repeating the correcting step more than once.

III. RESULTS AND DISCUSSION

In this section we will present example calculations for the linear-response and time-propagation schemes. The two computationally very different approaches are applied to the same systems and very good agreement is found in the linear-response regime. The strengths and weaknesses of both methods are discussed.

We apply consistently the LDA (Ref. 31) in all calculations. Zero Dirichlet boundary conditions are used for the finite systems studied both in the ground state as well as in the time-propagation calculations. A grid spacing of $h=0.2$ Å is used for the representation of the smooth wave functions unless otherwise specified.

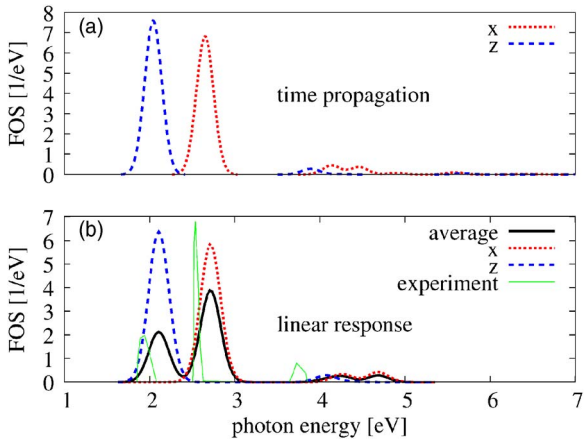


FIG. 1. (Color online) Optical absorption spectra of the Na_2 dimer represented as folded oscillator strengths [FOS's, Eq. (40)]. The results obtained (a) by the time-propagation after a delta kick and (b) by the linear-response scheme are compared. x and z denote the polarization directions of the light so that the molecule symmetry axis is aligned along the z direction. Experimental data is from Refs. 33 and 34 as quoted in Ref. 16.

Figure 1 shows a direct comparison of the absorption spectra of the Na_2 molecule at the experimental equilibrium distance of $R=3.068$ Å (Ref. 32) obtained via time-propagation after a delta kick and in the linear-response scheme. Both calculations are done using the simulation box of size $15\times15\times18$ Å^3. In the time-propagation calculation a perturbation strength of $\epsilon=1\times10^{-4}$, a grid spacing of $h=0.3$ Å, and a simulation time of 36 fs with a 1.2 as time step is used. The linear-response energy peaks are folded with Gaussians of width $\sigma=0.12$ eV corresponding to the Gaussian damping of the time-propagation. The simulated spectra agree perfectly. This proofs the correctness of the implementations of the two methods, which are technically very different. The shift of the peaks with respect to the experiment,[33,34] also seen in other calculations,[14,16,35] is probably related to the LDA.

In the next example we compare the absorption spectra of the benzene molecule obtained by the two methods. This molecule is one of the standard examples used in the literature.[14,17,36,37] The experimental spectrum shown in Fig. 2(c) consists of a strong peak at 6.9 eV and a broad feature in the range from 10 to above 25 eV. In Ref. 36 this experimental spectrum was nicely reproduced via a time-propagation scheme using a real-space grid, but in the linear-response calculation in Ref. 37 the energy of the first peak differed from the experimental value by ~0.5 eV. In our calculations the linear-response and time-propagation results are in good agreement. The time-propagation calculation results in $f_1=1.2$ for the main peak at 6.74 eV and the linear-response calculation shows $f_1=1.3$ for the main peak at 6.85 eV. The positions and strengths of the main peak coincide well with the experimental values of 6.9 eV and $f=0.9$, respectively.[36]

The differences between the spectra of the linear-response and the time-propagation schemes seen in Fig. 2 originate from the different convergence behavior (see also Sec. IV below). The time-propagation uses only occupied states, but a large unit cell has to be used in order to avoid

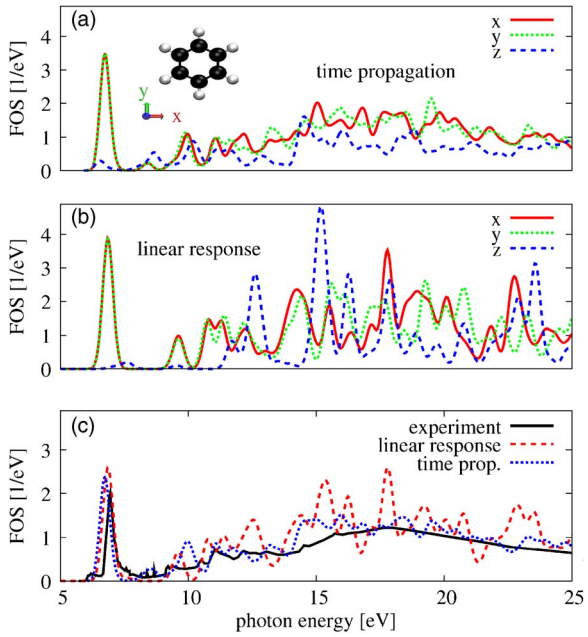


FIG. 2. (Color online) Optical absorption spectra of the benzene molecule represented as folded oscillator strengths [FOS's, Eq. (40)]. The results obtained (a) by the time-propagation after a delta kick and (b) by the linear-response scheme are shown. x, y, and z denote the polarization directions of the light as shown in the inset so that the z axis is perpendicular to the plane of the molecule. (c) The average spectra are compared with the experimental one quoted in Ref. 36. The experimental spectrum is scaled to integrate to $f=0.9$ in the energy range from 6.5 to 8.3 eV (Ref. 36).

spurious contributions from the simulation box boundaries. In contrast, the linear-response calculation has to sample unoccupied states in a range which is larger than the largest transition energy one is interested in. These unoccupied states belong already to the continuum of KS states. In practice a finite set of states can be sufficient to describe the essential features. This set is smaller if a smaller simulation box is chosen. We carry out the time-propagation with a box size of $18\times18\times13$ Å^3 and a simulation time of 24 fs with a time step of 1.2 arc sec. The linear-response calculation uses the finite box of $11\times11\times6$ Å^3 and the spectrum is folded with Gaussians of width $\sigma=0.2$ eV. Both calculations used a grid spacing of $h=0.25$ Å in this case.

A. Linear response

In the following we present results of linear-response calculations for selected divalent atoms and for Born–Oppenheimer excited state potential surfaces of the Na dimer. These systems represent standard benchmarks. We will make use of the advantage that in the linear-response calculations both singlet and triplet (s/t) excitations are directly accessible.

Table I gives the lowest $S\rightarrow P$ s/t transition energies for selected divalent atoms. The present LDA results are compared to those obtained by pseudopotential calculations and to experimental values quoted in Ref. 16. In our calculations the real-space grid spans around the atom a cubic volume with the edge length of 12 Å. Our calculated excitation energies are in reasonable agreement with experiment and conform with the pseudopotential calculations with the exception of Zn and Cd. For these atoms differences up to 0.4 eV appear. They may be related to differences in the highest occupied orbital energies $\epsilon_{\mathrm{HOMO}}^{\mathrm{LDA}}$ also listed in Table I. We note that our $\epsilon_{\mathrm{HOMO}}^{\mathrm{LDA}}$ values obtained with the PAW method for Zn and Cd are in perfect agreement with the very accurate results of Ref. 38.

Next we turn our attention to the excited states of the Na_2 dimer. Usually, only the dipole spectrum at the equilibrium distance is studied.[14,16] We want to go further and investigate the Born–Oppenheimer potential surfaces as functions of the atomic separation R. Figure 3 shows the potentials of the ground state (X) and the lowest excited singlet (A,B) and triplet (x,a,b) states according to our calculations. These are performed using a rectangular calculation volume with the edge length of 8 Å perpendicular to and of $R+8$ Å parallel to the molecules axis. The energy axis is normalized to the LDA dissociation energy of the sodium dimer, i.e., twice the energy of a spin polarized Na atom. Note that the spin-compensated LDA ground-state energy does not converge toward this limit due to the self-interaction error.[39] Therefore the ground-state potential is above zero already at $R=6$ Å and the LDA triplet potential lowers below the ground-state potential for $R>4.2$ Å. The latter effect is called the triplet instability and it results in imaginary excitation energies. For this reason the lowest triplet state does not have a minimum in contrast to all other potentials shown. The properties of the potentials are further investigated in Table II in comparison to experimental data and configuration-interaction (CI) calculations from Refs. 40–42. Our equilibrium distances R_e and the vibrational frequencies ω_e are obtained by fitting the Morse potential to the potentials in Fig. 3. The range of 2 Å$\leqslant R\leqslant$4.8 Å is used to

TABLE I. Highest occupied KS orbital energies $\epsilon_{\mathrm{HOMO}}^{\mathrm{LDA}}$ and the lowest $S\rightarrow P$ s/t (spin singlet/triplet) excitation energies for selected divalent atoms. The present ground-state or linear-response LDA results (GPAW) are compared to similar literature results. Experimental excitation values taken from Ref. 16 are also given. All values are in eV.

	$\epsilon_{\mathrm{HOMO}}^{\mathrm{LDA}}$ (eV)			$S\rightarrow P$ s/t		
Atom	Ref. 38	Ref. 16	GPAW	Ref. 16	GPAW	Expt.
Be	−5.60	−5.61	−5.60	4.94/2.45	4.82/2.41	5.28/2.72
Mg	−4.78	−4.78	−4.78	4.34/2.79	4.28/2.79	4.34/2.72
Ca	−3.86	−3.85	−3.85	3.22/1.93	3.18/1.97	2.94/1.89
Sr	−3.64	−3.59	−3.62	2.96/1.82	2.90/1.84	2.69/1.82
Zn	−6.21	−6.07	−6.21	5.71/4.27	5.89/4.41	5.79/4.05
Cd	−5.94	−5.56	−5.95	5.10/3.69	5.52/4.13	5.41/3.88

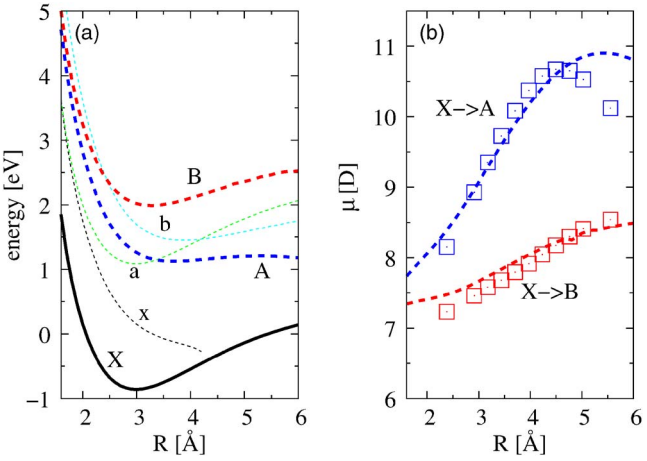


FIG. 3. (Color online) (a) Born–Oppenheimer potential curves for the Na_2 dimer in the ground state (X), in the lowest excited singlet states (A, B) and in the triplet states (x,a,b). (b) Comparison of the dipole transition moments calculated within the LDA (broken lines) with the CI results (squares) of Ref. 43. The dipole moment $\mu=|\boldsymbol{\mu}|$ is given in debyes (1 D=3.335 64 $\times 10^{-30}$ C m).

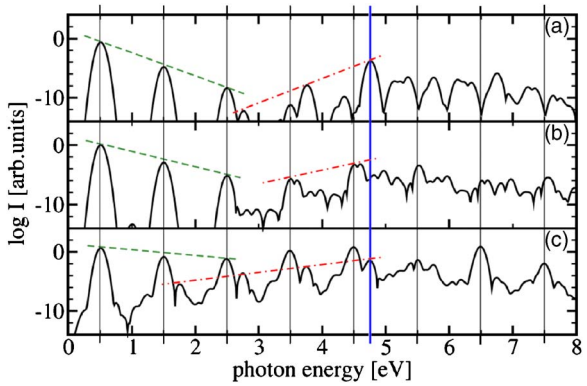


FIG. 4. (Color online) Emission spectra of a Be atom in a sinusoidal dipole field of the frequency of 0.5 eV/$\hbar$ and strengths of (a) 0.2 V/Å, (b) 0.4 V/Å, and (c) 0.8 V/Å. The thick blue vertical line at 4.82 eV denotes the frequency of the first $S\rightarrow P$ transition. The thin vertical lines denote odd harmonic frequencies. The green dashed lines are drawn to emphasize the exponential decay of the high-harmonic peak intensities as a function of the frequency in emission. The red dot-dashed lines emphasize the difference frequency mixing of the first resonance and the dipole field.

fit the X, a, b, and B potentials. The range has to be extended to 2 Å$\leqslant R\leqslant$8 Å for the A state due to the shallow potential minimum. The agreement of our excited state calculations with both experiment and CI approaches is reasonable and of similar quality as that for the ground-state calculation.

Our dipole transition moments [Eq. (23)] for the dipole allowed transitions $X\rightarrow A$ and $X\rightarrow B$ are compared in Fig. 3 with the results of the pioneering CI calculations by Stevens *et al.*[43] Our transition dipole moments calculated within the LDA are in a very good agreement with the CI results proving the accuracy of the linear-response TDDFT also for this quantity.

B. Time-propagation

Next, we present results of our time-propagation calculations in the nonlinear regime.

Figure 4 shows the calculated emission spectra of a Be atom exposed to laser fields of the frequency of 0.5 eV and strengths of E_0=0.2, 0.4, and 0.8 V/Å. Figure 5 shows similar results for the laser fields of the frequency of 1.0 eV. Only the odd harmonics are observed in the spectra as the even harmonics are forbidden due to the spherical symmetry of the atom.[44] According to Figs. 4 and 5, the effect of a nearby resonance transition[45] is clearly apparent. The high-harmonic peaks near the first dipole allowed $^1S^e\rightarrow{}^3P^o$ transition at 4.82 eV gain intensity instead of decaying exponentially as function of the frequency. We observe in Fig. 4 the difference frequency mixing[44,45] of the first resonance and the sinusoidal field, i.e., a frequency equal to the difference of the first resonance and field frequencies appears. Note that we find perfect agreement in the energy of the first dipole allowed transition in the Be atom with the result of the linear-response calculation given in Table I. Comparing the oscillator strength of this transition we find f_1=1.35 consistently in both methods and in very good agreement with the result f_1=1.375 obtained by CI calculations in Ref. 46.

IV. CONVERGENCE PROPERTIES

A. Linear response

We will now discuss the convergence properties of the linear-response calculations. Figure 6(a) shows the relative deviation of the lowest excitation energies of the Be atom for different choices of the finite-difference parameter ε in the

TABLE II. Properties of the Born–Oppenheimer potentials for the Na_2 dimer. The transition energies T_e at the experimental equilibrium distance of R =3.068 Å are given in eV, the equilibrium distances R_e in Å, and the vibration energies ω_e in cm^{-1}. The experimental data is from Ref. 32 and the theoretical data from Refs. 40–42.

	Present work			Expt.			CI calculations		
State	T_e	R_e	ω_e	T_e	R_e		T_e	R_e[a]	ω_e[a]
$X\ ^1\Sigma_g^+$		2.99	161		3.07	159		3.17	145
$x\ ^2\Sigma_u^+$	0.96	⋯	⋯	⋯	⋯	⋯	1.05[a]/0.99[b]	5.20	29
$a\ ^3\Pi_u^+$	1.86	2.98	155	⋯	⋯	⋯	1.60[a]/1.56[b]	3.21	146
$A\ ^1\Sigma_u^+$	2.13	3.75	85	1.82	3.64	117	1.86[a]/1.82[b]	3.75	115
$b\ ^3\Sigma_g^+$	2.37	3.87	97	⋯	⋯	⋯	2.34[a]/2.26[c]	3.91	101
$B\ ^3\Sigma_g^+$	2.41	3.42	126	2.52	3.42	124	2.62[a]/2.52[c]	3.63	106

[a]Reference 40.
[b]Reference 41.
[c]Reference 42.

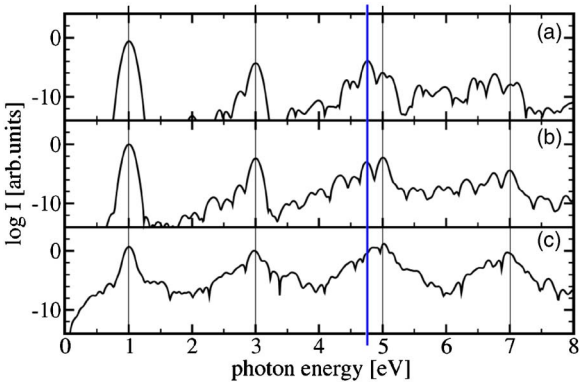


FIG. 5. (Color online) Emission spectra of a Be atom in a sinusoidal dipole field of the frequency of 1.0 eV/$\hbar$ and strengths of (a) 0.2 V/Å, (b) 0.4 V/Å, and (c) 0.8 V/Å.

calculation of the XC kernel according to Eq. (20). Too large values of ε lead out of the perturbation regime, whereas too small values of ε produce numerical errors. However, the figure shows that the results are quite insensitive to the choice of the parameter, i.e., it can be chosen in the range of $10^{-12} < \varepsilon < 0.01$ resulting in the uncertainty of less than 0.1% in the excitation energy. The effect of the size of the KS excitation basis is much more severe as shown in Fig. 6(b) for the same excited states. Here the number j of unoccupied states is varied. The three lowest unoccupied KS orbitals have the symmetry corresponding to the angular momentum $l=1$. Restricting the calculation to these states can describe only the excitations $^1S \rightarrow {}^{1/3}P^o$. Including the next $l=0$ orbital ($j>3$) enables the appearance of the $^1S \rightarrow {}^3S$ transition, but does not change the energies of the $^{1/3}P^o$ excited states. Incorporating more unoccupied orbitals of the $l=1$ symmetry

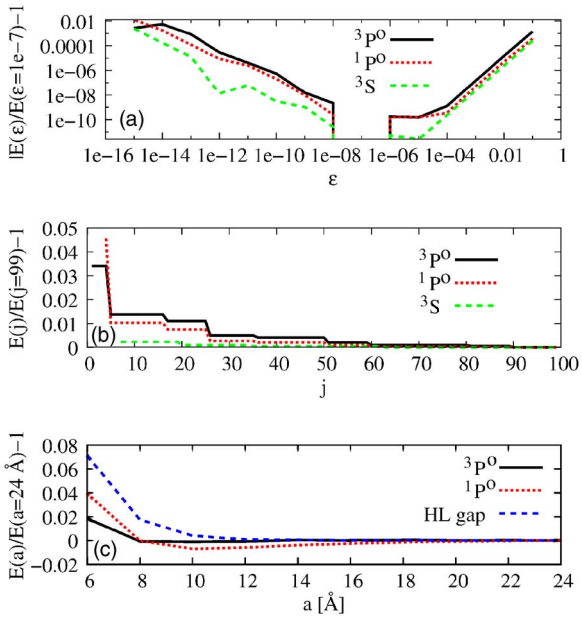


FIG. 6. (Color online) Convergence of the HOMO-LUMO gap and the excitations to the lowest ^3S, $^1P^o$, and $^3P^o$ states calculated in the linear-response scheme for the Be atom. The convergence is given as a function of (a) the finite-difference parameter ε for evaluation of the xc kernel [Eq. (20)], (b) the number j of unoccupied orbitals taken into account in the calculation, and (c) dependence on the edge length a of the cubic simulation box.

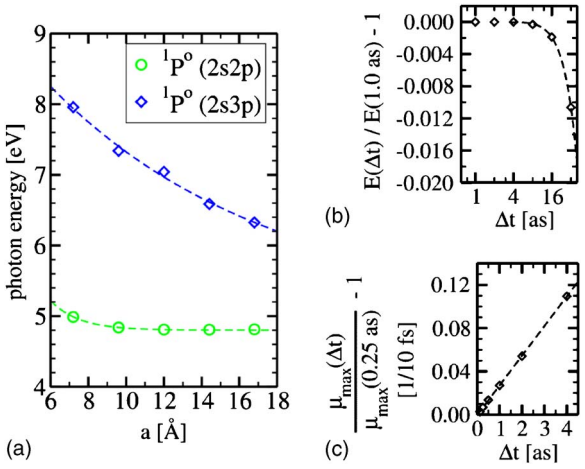


FIG. 7. (Color online) (a) Convergence of the linear-response transition energies of the Be atom as a function of the edge length a of the cubic simulation box. The dashed lines are exponential fits. (b) Convergence of the energy of the first transition as a function of the length Δt of the time step. The dashed line is a quadratic fit. (c) Convergence of the peak dipole moment as a function of the length Δt of the time step. The dashed lines is a linear fit.

produces changes of the energy of the $^{1/3}P^o$ excited state due to coupling to the $2s \rightarrow 2p$ KS transitions seen as steps in Fig. 6(b). The ^3S excited state energy is not affected due to the different symmetry of the KS transitions. Note that we have used 100 KS states to calculate the excitations of the Be atom. Such a large number of unoccupied states is not practical, but Fig. 6(b) shows also that an accuracy of more than 2% is reached already by the inclusion of a few unoccupied orbitals. Finally, we have studied the convergence as a function of the box size used for the real-space grids. The results are shown in Fig. 6(c). We use a cubic box with the edge length a, and vary a from 6 to 24 Å in steps of 2 Å. The highest occupied to lowest unoccupied orbital (HOMO-LUMO) gap shows convergence at $a \approx 10$ Å. The triplet ($^3P^o$) energy converges for even smaller box sizes, obviously due to a cancellation of errors. The slowest convergence is found for the singlet state ($^3P^o$) due to effect of the long-range RPA kernel not contributing to the triplet-state energy. However, the effect is well below 0.1 eV and therefore much lower than the accuracy of the TDDFT found above and in other calculations.[13]

B. Time-propagation

The convergence of the time-propagation method depends mainly on two factors: the simulation box size and the length of the time step. The box size should be considerably larger than that required for the ground-state calculation as excited states are more diffuse. The convergence of the first and the second transition energies of the Be atom are shown Fig. 7(a) as a function of the length a of the cubic simulation box. The second transition energy converges clearly slower than that of the first one. The latter has converged already around $a=5$ Å. Figure 7(b) shows the convergence of the first transition energy as a function of the length of the time step Δt. The transition energy behaves quadratically and it has converged around $\Delta t = 8$ as, whereas the corresponding

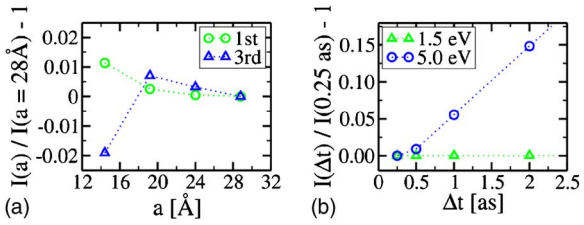


FIG. 8. (Color online) (a) Convergence of the harmonic peak intensities as a function of the edge length a of the cubic simulation box. The dotted lines are just a guide to the eye. (b) Convergence of the harmonic peak intensities as a function of the length Δt of the time step. The dotted lines are just a guide to the eye.

peak intensity in Fig. 7(c) converges only linearly in this range and time steps down to $\Delta t=1-2$ as must be used in order to obtain accurate results. The f-sum rule (or Thomas–Reiche–Kuhn sum rule),[47] $\int S_{ij}(\omega)d\omega=N\delta_{ij}$, where N is the number of electrons, is fulfilled within a few percent in the present calculations. Note that care must be taken when constructing the PAW-projectors, because, if pseudo-wavefunctions represented in the grid cannot be accurately transformed to the atomic basis by the PAW-projectors, the f-sum becomes incomplete.

Figure 8(a) shows the convergence of the intensities of the first and third harmonics as function of the edge length a of the cubic simulation box. The difference ΔI is taken with respect to the value at $a=28$ Å. Naturally, higher harmonics require a larger simulation box. For the harmonics near and beyond the first transition resonance we are not able to find converged results within our computer resources. The reason is that a part of the system is excited to the first excited state, from which it is more easily ionized by the laser field than from the ground state.

Figure 8(b) shows the convergence of the peak intensities at 1.5 and 5.0 eV as a function of the length of the time step Δt. The difference ΔI is taken with respect to the value at $\Delta t=0.25$ as. The third harmonic intensity at 1.5 eV is almost independent on Δt, whereas near (and beyond) the first transition resonance at 5.0 eV the intensity difference depends linearly on Δt.

V. CONCLUSION

We have described the implementation of the time-dependent density-functional theory in the projector augmented-wave framework, both in the time-propagation as well as within the linear-response scheme. The two approaches were compared by calculating the optical absorption spectra of Na_2 and C_6H_6 molecules in the linear regime. Good agreement of the absorption spectra was found, proving the correctness of both implementations. The strengths and weaknesses of both methods were discussed and examples of the possibilities were given. For example, the ability of the time-propagation scheme to describe nonlinear effects was demonstrated in the case of the Be atom. The convergence properties of both methods were studied in detail.

ACKNOWLEDGMENTS

This work has been supported by the Academy of Finland (Project 110013 and the Center of Excellence program) and Tekes MASI-program. The computational time was provided by CSC-the Finnish IT Center for Science. One of the authors (M.W.) thanks M. Mundt and S. Kümmel for discussions about the connection of linear-response and time-propagation schemes enabled via a bilateral AF-DAAD travel grant. The electronic structure program GPAW is developed in collaboration with CSC, CAMd/Technical University of Denmark, Department of Physics/University of Jyväskylä, Institute of Physics/Tampere University of Technology, and Laboratory of Physics, Helsinki University of Technology.

APPENDIX A: AUGMENTATION SPHERE CORRECTIONS (THE RPA PART)

This Appendix gives the explicit forms of the augmentation-sphere corrections.

Due to the use of compensation charges, all terms appearing in the augmentation-sphere corrections to the RPA part are local and can be expressed by integrals of the following type:

$$(f|g)_a := \int_a \int_a d\mathbf{r}_1 d\mathbf{r}_2 \frac{f^*(\mathbf{r}_1)g(\mathbf{r}_2)}{|\mathbf{r}_1-\mathbf{r}_2|}, \tag{A1}$$

where we have used the shorthand a for the restriction $|\mathbf{r}_{1/2}-\mathbf{R}_a|<r_c^a$. Here, r_c^a is the radius of the augmentation sphere for the atom at $\mathbf{R}_a$. Using Eq. (13) the correction $\Delta K_{ij,kq}^{\mathrm{RPA},a}$ can be written as

$$\Delta K_{ij,kq}^{\mathrm{RPA},a} = (n_{ij}^a|n_{kq}^a)_a - (\tilde{n}_{ij}^a+\tilde{Z}_{ij}^a|\tilde{n}_{kq}^a+\tilde{Z}_{kq}^a)_a. \tag{A2}$$

Inserting the explicit forms of pair densities and compensation charges leads to the expression

$$\Delta K_{ij,kq}^{\mathrm{RPA},a} = 2\sum_{i_1 i_2 i_3 i_4} P_{ii_1}^a P_{ji_2}^a P_{ki_3}^a P_{qi_4}^a C_{i_1 i_2 i_3 i_4}^a, \tag{A3}$$

with the coefficients $C_{i_1 i_2 i_3 i_4}^a$ given in Eq. (C3) of Ref. 19. These coefficients have to be calculated only once for each type of atom.

APPENDIX B: AUGMENTATION SPHERE CORRECTIONS (THE xc KERNEL)

The local corrections to the exchange-correlation kernel in the finite-difference scheme can be written as

$$\Delta K_{ij\sigma,kq\tau}^{\mathrm{xc},a} = \lim_{\varepsilon\to 0} \frac{K_{ij\sigma,kq}^{\mathrm{xc},a,+}-K_{ij\sigma,kq}^{\mathrm{xc},a,-}}{2\varepsilon}, \tag{B1}$$

with

$$K_{ij\sigma,kq\tau}^{\mathrm{xc},a,\pm} = \langle \psi_{i\sigma}^a | v_{\mathrm{xc}}[n_\sigma^a, n_\tau^a \pm \varepsilon n_{kq\tau}^a] | \psi_{j\tau}^a \rangle - \langle \tilde{\psi}_{i\sigma}^a | v_{\mathrm{xc}}[\tilde{n}_\sigma^a, \tilde{n}_\tau^a \pm \varepsilon \tilde{n}_{kq\tau}^a] | \tilde{\psi}_{j\sigma}^a \rangle. \tag{B2}$$

Using the expansion of the wave functions in Eq. (3) we obtain

$$K^{\mathrm{xc},a,\pm}_{ij\sigma,kq\tau} = \sum_{i_i i_2} P_{i\sigma i_1} P_{j\sigma i_2} I^{a,kq\tau,\pm}_{i_1 i_2}. \tag{B3}$$

Above, we have defined the integral

$$I^{a,kq\tau,\pm}_{i_1 i_2} = \int_a d\mathbf{r}[\phi_{i_1}(\mathbf{r})\phi_{i_2}(\mathbf{r})v_{\mathrm{xc}}[n_\sigma, n^{\pm}_{kq\tau}] - \tilde{\phi}_{i_1}(\mathbf{r})\tilde{\phi}_{i_2}(\mathbf{r})v_{\mathrm{xc}}[\tilde{n}_\sigma, \tilde{n}^{\pm}_{kq\tau}]], \tag{B4}$$

with the shorthands $n^{\pm}_{kq} = n^a_\tau \pm \varepsilon n^a_{kq\tau}$ and $\tilde{n}^{\pm}_{kq} = \tilde{n}^a_\tau \pm \varepsilon \tilde{n}^a_{kq\tau}$. Above, v_{xc} depends on the modified atomic density matrix compared to the $D^a_{i_1 i_2}$ in Eq. (7). A density change by $\pm\varepsilon n_{kq\tau}$ results in a change in $D^a_{i_1 i_2}$ as

$$n^a(x) \pm \varepsilon n^a_{kq}(x) = \sum \bar{D}^{a,\pm}_{i_1 i_2, kq} \phi^a_{i_1}(x)\phi^a_{i_2}(x), \tag{B5}$$

with

$$\bar{D}^{a,\pm}_{i_1 i_2, kq} = D^a_{i_1 i_2} \pm \frac{\varepsilon}{2}(P_{ki_1}P_{qi_2} + P_{ki_2}P_{qi_1}), \tag{B6}$$

where we have used a symmetric notation to point out the exchange symmetry with respect to $i_1 \leftrightarrow i_2$. The integrals [Eq. (B4)] are evaluated numerically as described in Ref. 19.

[1] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[2] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[3] E. Runge and E. K. U. Gross, Phys. Rev. Lett. **52**, 997 (1984).
[4] M. E. Casida, in *Recent Developments and Applications in Modern Density-Functional Theory*, edited by J. M. Seminario (Elsevier, Amsterdam, 1996), p. 391.
[5] K. Yabana and G. F. Bertsch, Phys. Rev. B **54**, 4484 (1996).
[6] G. Senatore and K. R. Subbaswamy, Phys. Rev. A **35**, 2440 (1987).
[7] S. J. van Gisbergen, J. G. Snijders, and E. J. Baerends, Phys. Rev. Lett. **78**, 3097 (1997).
[8] J.-I. Iwata, K. Yabana, and G. F. Bertsch, J. Chem. Phys. **115**, 8773 (2001), and references therein.
[9] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[10] R. Bauernschmitt and R. Ahlrichs, Chem. Phys. Lett. **256**, 454 (1996).
[11] A. Tsolakidis, D. Sánchez-Portal, and R. M. Martin, Phys. Rev. B **66**, 235416 (2002).
[12] N. L. Doltsinis and M. Sprik, Chem. Phys. Lett. **330**, 563 (2000).
[13] M. Moseler, H. Häkkinen, and U. Landman, Phys. Rev. Lett. **87**, 053401 (2001).
[14] X. Qian, J. Li, X. Lin, and S. Yip, Phys. Rev. B **73**, 035408 (2006).
[15] B. Walker and R. Gebauer, J. Chem. Phys. **127**, 164106 (2007).
[16] I. Vasiliev, S. Ögüt, and J. R. Chelikowsky, Phys. Rev. Lett. **82**, 1919 (1999).
[17] M. A. L. Marques, A. Castro, G. F. Bertsch, and A. Rubio, Comput. Phys. Commun. **151**, 60 (2003).
[18] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994).
[19] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005).
[20] See https://wiki.fysik.dtu.dk/gpaw.
[21] M. Casida, in *Recent Advances in Density Functional Methods, Part I*, edited by D. Chong (World Scientific, Singapore, 1995), p. 155.
[22] F. Furche, J. Chem. Phys. **114**, 5982 (2001).
[23] A. Putrino, D. Sebastiani, and M. Parrinello, J. Chem. Phys. **113**, 7102 (2000).
[24] D. Egli and S. R. Billeter, Phys. Rev. B **69**, 115106 (2004).
[25] M. Mundt and S. Kümmel, Phys. Rev. B **76**, 035413 (2007).
[26] T. Brabec and F. Krausz, Rev. Mod. Phys. **72**, 545 (2000).
[27] M. Protopapas, C. H. Keitel, and P. L. Knight, Rep. Prog. Phys. **60**, 389 (1997).
[28] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1992).
[29] A. Castro, M. A. L. Marques, and A. Rubio, J. Chem. Phys. **121**, 3425 (2004).
[30] R. Barrett, M. Berry, T. F. Chan, J. Demmel, J. Donato, J. Dongarra, V. Eijkhout, R. Pozo, C. Romine, and H. V. der Vorst, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd ed. (SIAM, Philadelphia, PA, 1994).
[31] J. P. Perdew and Y. Wang, Phys. Rev. B **46**, 12947 (1992).
[32] K. Huber and G. Herzberg, *Constants of Diatomic Molecules*, NIST Chemistry WebBook, NIST Standard Reference Database Number 69, edited by P. J. Linstrom and W. G. Mallard (National Institute of Standards and Technology, Gaithersburg MDY, 2003). See http://webbook.nist.gov; data prepared by J. W. Gallagher and R. D. Johnson III.
[33] W. R. Fredrickson and W. W. Watson, Phys. Rev. **30**, 429 (1927).
[34] S. P. Sinha, Proc. Phys. Soc., London, Sect. A **62**, 124 (1949).
[35] M. A. L. Marques, A. Castro, and A. Rubio, J. Chem. Phys. **115**, 3006 (2001).
[36] K. Yabana and G. F. Bertsch, Int. J. Quantum Chem. **75**, 55 (1999).
[37] I. Vasiliev, S. Ogut, and J. R. Chelikowsky, Phys. Rev. B **65**, 115416 (2002).
[38] S. Kotochigova, Z. Levine, E. Shirley, M. Stiles, and C. Clark, *Atomic Reference Data for Electronic Structure Calculations* (National Institute of Standards and Technology, Gaithersburg, MD, 2004). Online available: http://physics.nist.gov/PhysRefData/DFTdata/contents.html.
[39] E. J. Baerends, Phys. Rev. Lett. **87**, 133004 (2001).
[40] D. D. Konowalow, M. E. Rosenkrantz, and M. L. Olson, J. Chem. Phys. **72**, 2612 (1980).
[41] V. Bonacic-Koutecky, P. Fantucci, and J. Koutecký, J. Chem. Phys. **93**, 3802 (1990).
[42] H.-K. Chung, K. Kirby, and J. F. Babb, Phys. Rev. A **63**, 032516 (2001).
[43] W. J. Stevens, M. M. Hessel, P. J. Bertoncini, and A. C. Wahl, J. Chem. Phys. **66**, 1477 (1977).
[44] T. Brabec and F. Krausz, Phys. Rev. **127**, 1918 (1962).
[45] R. W. Boyd, *Nonlinear Optics*, 2nd ed. (Academic, New York, 2003).
[46] J. Fleming, M. R. Godefroid, K. L. Bell, A. Hibbert, N. Vaeck, J. Olsen, P. Jonsson, and C. F. Fischer, J. Phys. B **29**, 4347 (1996).
[47] S. Wang, Phys. Rev. A **60**, 262 (1999).

Paper II

Fully selfconsistent GW calculations for molecules
C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen
Accepted, Physical Review B.

Fully selfconsistent GW calculations for molecules

C. Rostgaard, K. W. Jacobsen, and K. S. Thygesen
Center for Atomic-scale Materials Design (CAMD),
Department of Physics, Technical University of Denmark, DK - 2800 Kgs. Lyngby, Denmark
(Dated: January 8, 2010)

We calculate single-particle excitation energies for a series of 33 molecules using fully selfconsistent GW, one-shot G_0W_0, Hartree-Fock (HF), and hybrid density functional theory (DFT). All calculations are performed within the projector augmented wave (PAW) method using a basis set of Wannier functions augmented by numerical atomic orbitals. The GW self-energy is calculated on the real frequency axis including its full frequency dependence and off-diagonal matrix elements. The mean absolute error of the ionization potential (IP) with respect to experiment is found to be 4.4, 2.6, 0.8, 0.4, and 0.5 eV for DFT-PBE, DFT-PBE0, HF, G_0W_0[HF], and selfconsistent GW, respectively. This shows that although electronic screening is weak in molecular systems its inclusion at the GW level reduces the error in the IP by up to 50% relative to unscreened HF. In general GW overscreens the HF energies leading to underestimation of the IPs. The best IPs are obtained from one-shot G_0W_0 calculations based on HF since this reduces the overscreening. Finally, we find that the inclusion of core-valence exchange is important and can affect the excitation energies by as much as 1 eV.

PACS numbers: 31.15.A-,33.15.Ry,31.15.V-

I. INTRODUCTION

Density functional theory (DFT)[1] with the single-particle Kohn-Sham (KS) scheme[2] is today the most widely used approach to the electronic structure problem of real materials in both solid state physics and quantum chemistry. While properties derived from total energies (or rather total energy differences) are accurately predicted by DFT, it is well known that DFT suffers from a band gap problem implying that the single-particle KS eigenvalues cannot in general be interpreted as real quasiparticle (QP) excitation energies. In particular, semilocal exchange-correlation functionals severely underestimate the fundamental gap of both insulators, semi-conductors, and molecules.[3–6]

The hybrid[7–9] and screened hybrid[10] functionals, which admix around 25% of the (screened) Fock exchange with the local DFT exchange, generally improve the description of band gaps in bulk semi-conductors and insulators[5,6]. However, the orbital energies obtained for finite systems using such functionals still underestimate the fundamental gap, $I_p - E_a$, (the difference between ionization potential and electron affinity) by up to several electron volts. In fact, for molecules the pure Hartree-Fock (HF) eigenvalues are usually closer to the true electron addition/removal energies than are the hybrid DFT eigenvalues. This is because HF is self-interaction free and because screening of the exchange interaction is a relatively weak effect in molecular systems.[4,11,13]. On the other hand, in extended systems the effect of self-interaction is less important and the long range Coulomb interaction becomes short ranged due to dynamical screening. As a consequence HF breaks down in extended systems leading to dramatically overestimated band gaps and a qualitatively incorret description of metals.[14–16]

The many-body GW approximation of Hedin[17] has been widely and succesfully used to calculate QP band structures in metals, semi-conductors, and insulators.[3,18–20] The GW approximation can be viewed as HF with a dynamically screened Coulomb interaction. The fact that the screening is determined by the system itself instead of being fixed a priori as in the screened hybrid schemes, suggests that the GW method should be applicable to a broad class of systems ranging from metals with strong screening to molecules with weak screening. With the entry of nanoscience the use of GW has been extended to low-dimensional systems and nanostructures[21–31] and more recently even nonequilibrium phenomena like quantum transport[32–36]. In view of this trend it is important to establish the performance of the GW approximation for other systems than the crystalline solids. In this work we present first-principles benchmark GW calculations for a series of small molecules. In a closely related study we compared GW and Hartree-Fock to exact diagonalization results for semi-empirical PPP models of conjugated molecules[37]. The main conclusions from the two studies regarding the qualities of the GW approximation in molecular systems, are very consistent.

Most GW calculations to date rely on one or several approximations of more technical character. These include the plasmon pole approximation, the linearized QP equation, neglect of off-diagonal matrix elements in the GW self-energy, analytic continuations from the imaginary to the real frequency axis, neglect of core states contributions to the self-energy, neglect of self-consistency. The range of validity of these approximations has been explored for solid state systems by a number of authors[38–43], however, much less is known about their applicability to molecular systems[24]. Our implementation of the GW method avoids all of these technical approximations allowing for a direct and unbiased

assessment of the GW approximation itself.

Here we report on single-shot G_0W_0 and fully self-consistent GW calculations of QP energies for a set of 33 molecules. The mean absolute error (MAE) of the calculated QP energy of the highest occupied molecular orbital (HOMO) relative to the experimental ionization potentials is 0.4 - 0.5 eV. In comparison the MAE obtained with DFT-PBE, DFT-PBE0, and Hartree-Fock is 4.4, 2.6, and 0.8 eV, respectively. Non-selfconsisistent G_0W_0 calculations starting from the HF Green function increases the HOMO energies obtained in fully self-consistent GW leading to a slight improvement with respect to the experimental IP. G_0W_0 calculations based on the PBE Green function yields non-systematic deviations from the selfconsistent GW results but similar accuracy. These findings are in good agreement with our benchmark GW results for semi-empirical models of conjugated molecules[37]. All calculations are performed within the projector augmented wave (PAW) method using a localized basis set consisting of Wannier functions augmented by numerical atomic orbitals functions. The PAW method facilitates the calculation of core-valence exchange interactions which can contribute significantly to the HF (and GW) energies. Our results show that the GW approximation yields accurate single-particle excitation energies for small molecules improving both hybrid DFT and full Hartree-Fock results.

The paper is organized as follows. In Sec. II we describe the theoretical and numerical details behind the GW calculations, including the augmented Wannier function basis set, the self-consistent solution of the Dyson equation, and the evaluation of valence-core exchange within PAW. In Sec. III we discuss and compare the results of G_0W_0, GW, HF, PBE0, and PBE calculations. We analyze the role of dynamical screening, and discuss the effect of self-consistency in the GW self-energy. We conclude in Sec. IV.

II. METHOD

A. Augmented Wannier function basis

For the GW calculations we apply a basis set consisting of projected Wannier functions (PWF) augmented by numerical atomic orbitals (NAO). The PWFs, ϕ_i, are obtained by maximizing their projections onto a set of target NAOs, Φ_{Alm}, subject to the condition that they span the set of occupied eigenstates, ψ_n. Thus we maximize the functional

$$\Omega = \sum_i \sum_{A,l,m} |\langle \phi_i | \Phi_{Alm} \rangle|^2 \qquad (1)$$

subject to the condition $\mathrm{span}\{\phi_i\} \supseteq \mathrm{span}\{\psi_n\}_{\mathrm{occ}}$ as described in Ref. 44. The target NAOs are given by $\Phi_{Alm}(\mathbf{r}) = \zeta_{Al}(r) Y_{lm}(\mathbf{r})$ where ζ_{Al} is a modified Gaussian which vanish outside a specified cut-off radius, and Y_{lm} are the spherical harmonics corresponding to the valence of atom A. The number of PWFs equals the number of target NAOs. For example we obtain one PWF for H ($l_{\max} = 0$), and four PWFs for C ($l_{\max} = 1$). The PWFs mimick the target atomic orbitals but in addition they allow for an *exact* representation of all the occupied molecular eigenstates. The latter are obtained from an accurate real-space PAW-PBE calculation[45,46].

The PWFs obtained in this way provide an exact representation of the occupied PBE eigenstates. However, this does not suffice for GW calculations because the polarizability, P, and the screened interaction, W, do not live in this subspace. Hence we augment the PWFs by additional NAOs including so-called polarization functions which have $l = l_{\max} + 1$ and/or extra radial functions (zeta functions) for the valence atomic orbitals. For more details on the definition of polarization- and higher zeta functions we refer to Ref. 46. To give an example, a double-zeta-polarized (DZP) basis consists of the PWFs augmented by one set of NAOs corresponding to $l = 0, ..., l_{\max}$ and one set of polarization orbitals. Note that the notation, SZ, SZP, DZ, DZP, etc., is normally used for pure NAO basis sets, but here we use it to denote our augmented Wannier basis set. We find that the augmented Wannier basis is significantly better for HF and GW calculations than the corresponding pure NAO basis.

The GW and HF calculations presented in Sec. III were performed using a DZP augmented Wannier basis. This gives a total of 5 basis functions per H, Li, and Na, and 13 basis functions for all other chemical elements considered. In Sec. III C we discuss convergence of the GW calculations with respect to the size of the augmented Wannier basis.

B. GW calculations

The HF and GW calculations for isolated molecules are performed using a Green function code developed for quantum transport.[47] In principle, this scheme is designed for a molecule connected to two electrodes with different chemical potentials μ_L and μ_R. However, the case of an isolated molecule can be treated as a special case by setting $\mu_L = \mu_R = \mu$ and modelling the coupling to electrodes by a constant imaginary self-energy, $\Sigma_{L/R} = i\eta$. The chemical potential μ is chosen to lie in the HOMO-LUMO gap of the molecule and the size of η, which provides an artifical broadening of the discrete levels, is reduced until the results have converged. In this limit of small η the result of the GW calculation becomes independent of the precise position of μ inside the gap.

In Ref. 47 the GW-transport scheme was described for the case of an orthogonal basis set and for a truncated, two-index Coulomb interaction. Below we generalize the relevant equations to the case of a non-orthogonal basis and a full four-index Coulomb interaction. Some central results of many-body perturbation theory in a non-

orthogonal basis can be found in Ref. 48.

The central object is the retarded Green function, G^r,

$$G^r(\varepsilon) = [(\varepsilon+i\eta)S - H_{\rm KS} + v_{\rm xc} - \Delta v_{\rm H} - \Sigma^r_{\rm xc}[G](\varepsilon)]^{-1} \quad (2)$$

In this equation all quantities are matrices in the augmented Wannier basis, e.g. $H_{{\rm KS},ij} = \langle\phi_i|\hat{H}_{\rm KS}|\phi_j\rangle$ is the KS Hamiltonian matrix and $S_{ij} = \langle\phi_i|\phi_j\rangle$ is an overlap matrix. The term $\Delta v_{\rm H}$ represents the change in the Hartree potential relative to the DFT Hartree potential already contained in $H_{\rm KS}$, see Appendix A. The local xc-potential, $v_{\rm xc}$, is subtracted to avoid double counting when adding the many-body self-energy, $\Sigma_{\rm xc}[G]$. As indicated, the latter depends on the Green function and therefore Eq. (2) must in principle be solved self-consistently in conjuction with the self-energy.

In the present study $\Sigma_{\rm xc}$ is either the bare exchange potential or the GW self-energy. To be consistent with the code used for the calculations, we present the equations for the GW self-energy on the so-called Keldysh contour. However, under the equilibrium conditions considered here the Keldysh formalism is equivalent to the ordinary time-ordered formalism.

The GW self-energy is defined by

$$\Sigma^{\rm GW}_{ij}(\tau,\tau') = i\sum_{kl} G_{kl}(\tau,\tau'^{+})W_{ik,jl}(\tau,\tau'), \quad (3)$$

where τ and τ' are times on the Keldysh contour, $\mathcal{C}$. The dynamically screened Coulomb interaction obeys the Dyson-like equation

$$W_{ij,kl}(\tau,\tau') = V_{ij,kl}\delta_{\mathcal{C}}(\tau,\tau') + \sum_{pqrs}\int_{\mathcal{C}} {\rm d}\tau_1 V_{ij,pq}P_{pq,rs}(\tau,\tau_1)W_{rs,kl}(\tau_1,\tau'), \quad (4)$$

and the polarization bubble is given by

$$P_{ij,kl}(\tau,\tau') = -iG_{ik}(\tau,\tau')G_{lj}(\tau',\tau). \quad (5)$$

In the limit of vanishing polarization, $P = 0$, W reduces to the bare Coulomb interaction

$$V_{ij,kl} = \iint \frac{{\rm d}\boldsymbol{r}\,{\rm d}\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|}\phi_i(\boldsymbol{r})\phi^*_j(\boldsymbol{r})\phi^*_k(\boldsymbol{r}')\phi_l(\boldsymbol{r}') \quad (6)$$

and the GW self-energy reduces to the exchange potential of HF theory.

From the above equations for the contour-ordered quantities, the corresponding real time components, i.e. the retarded, advanced, lesser, and greater components, can be obtained from standard conversion rules[49,50]. For completeness we give the expressions for the real time components of the GW equations in Appendix A.

The time/energy dependence of the dynamical quantities G, W, P, and Σ, is represented on a uniform grid. We switch between time and energy domains using the Fast Fourier Transform in order to avoid time consuming convolutions. A typical energy grid used for the GW calculations in this work ranges from -150 to 150 eV with a grid spacing of 0.02 eV. The code is parallelized over basis functions and energy grid points. We use a Pulay mixing scheme for updating the Green function G^r when iterating Eq. (2) to self-consistency as described in Ref. 47.

We stress that no approximation apart from the finite basis set is made in our implementation of the GW approximation. In particular the frequency dependence is treated exactly and analytic continuations from the imaginary axis are avoided since we work directly on the real frequency/time axis. The price we pay for this is the large size of the energy grid.

C. Spectral function

The single-particle excitation spectrum is contained in the spectral function

$$A(\varepsilon) = i(G^r(\varepsilon) - [G^r(\varepsilon)]^\dagger). \quad (7)$$

For a molecule $A(\varepsilon)$ shows peaks at the QP energies $\varepsilon_n = E_n(N+1) - E_0(N)$ and $\varepsilon_n = E_0(N) - E_n(N-1)$ corresponding to electron addition and removal energies, respectively. Here $E_n(N)$ denotes the energy of the nth excited state of the system with N electrons and N refers to the neutral state.

When the Green function is evaluated in a non-orthogonal basis, like the augmented Wannier basis used here, the projected density of states for orbital ϕ_i becomes

$$D_i(\varepsilon) = [SA(\varepsilon)S]_{ii} \;/\; 2\pi S_{ii}, \quad (8)$$

where matrix multiplication is implied.[48] Correspondingly, the total density of states, or quasiparticle spectrum, is given by

$$D(\varepsilon) = {\rm Tr}(A(\varepsilon)S)/2\pi. \quad (9)$$

D. Calculating Coulomb matrix elements

The calculation of all of the Coulomb matrix elements, $V_{ij,kl}$, is prohibitively costly for larger basis sets. Fortunately the matrix is to a large degree dominated by negligible elements. To systematically define the most significant Coulomb elements, we use the product basis technique of Aryasetiawan and Gunnarsson[51,52]. In this approach, the pair orbital overlap matrix

$$S_{ij,kl} = \langle n_{ij}|n_{kl}\rangle, \quad (10)$$

where $n_{ij}(\mathbf{r}) = \phi^*_i(\mathbf{r})\phi_j(\mathbf{r})$ is used to screen for the significant elements of V.

The eigenvectors of the overlap matrix Eq. (10) represents a set of "optimized pair orbitals" and the eigenvalues their norm. Optimized pair orbitals with insignificant norm must also yield a reduced contribution to the

Coulomb matrix, and are omitted in the calculation of V. We limit the basis for V to optimized pair orbitals with a norm larger than $10^{-5}a_0^{-3}$. This gives a significant reduction in the number of Coulomb elements that needs to be evaluated, and it reduces the matrix size of $P(\varepsilon)$ and $W(\varepsilon)$ correspondingly, see Appendix A.

The evaluation of the double integral in Eq. (6) is efficiently performed in real space by solving a Poisson equation using multigrid techniques[45,53].

E. Valence-core exchange

All inputs to the GW/HF calculations, i.e. the selfconsistent Kohn-Sham Hamiltonian, $H_{\rm KS}$, the xc potential $v_{\rm xc}$, the Coulomb matrix elements, $V_{ij,kl}$, are calculated using the real-space PAW[54] code GPAW[45,46].

In GPAW, the core electrons (which are treated scalar-relativistically) are frozen into the orbitals of the free atoms, and the Kohn-Sham equations are solved for the valence states only. Unlike pseudo potential schemes, these valence states are subject to the full potential of the nuclei and core electrons. This is achieved by a partitioning scheme, where quantities are divided into pseudo components augmented by atomic corrections. The operators obtained from GPAW are thus full-potential quantities, and the wave functions from which the Wannier basis functions are constructed correspond to the all-electron valence states. Ref. 53 describes how the all-electron Coulomb elements can be determined within the PAW formalism.

Since both core and all-electron valence states are available in the PAW method, we can evaluate the contribution to the valence exchange self-energy coming from the core electrons. As the density matrix is simply the identity matrix in the subspace of atomic core states, this valence-core exchange reads

$$\Sigma^{\rm core}_{x,ij} = -\sum_{k}^{\rm core} V_{ik,jk}, \qquad (11)$$

where i,j represent valence basis functions. We limit the inclusion of valence-core interactions to the exchange potential, neglecting it in the correlation. This is reasonable, because the polarization bubble, P, involving core and valence states will be small due to the large energy difference and small spatial overlap of the valence and core states. This procedure was used and validated for solids in Ref. 42. We find that the elements of $\Sigma^{\rm core}_{x,ij}$ can be significant – on average 1.2 eV for the HOMO – and are larger (more negative) for the more bound orbitals which have larger overlap with the core states. In general, the effect on the HOMO-LUMO gap is to enlarge it, on average by 0.4 eV because the more bound HOMO level is pushed further down than the less bound LUMO state. In the case of solids, the role of valence-core interaction has been investigated by a number of authors[39–42,55]. Here the effect on the QP band gap seems to be smaller than what we find for the molecular gaps. We note that most GW calculations rely on pseudopotential schemes where these valence-core interactions are not accessible. In such codes, the xc contribution from the core electrons are sometimes estimated by $\Sigma^{\rm core}_{\rm xc} \approx v_{\rm xc}[n] - v_{\rm xc}[n_{\rm val}]$ where $n_{\rm val}$ is the valence electron density, but as the local xc potential is a non-linear functional of the density, this procedure is not well justified. Instead we subtract the xc potential of the full electon density n, and add explicitly the exact exchange core contribution.

III. RESULTS

In Fig. 1 we compare the calculated HOMO energies with experimental ionization potentials for the 33 molecules listed in Table I. The geometries of the molecules, which all belong to the G2 test set, are taken from Ref. 56. The different HOMO energies correspond to: DFT-PBE[57] and DFT-PBE0[7] eigenvalues, Hartree-Fock eigenvalues, and fully selfconsistent GW. The GW energies are obtained from the peaks in the corresponding density of states Eq. (9) extrapolated to $\eta = 0$ (η gives an artificial broadening of the delta peaks).

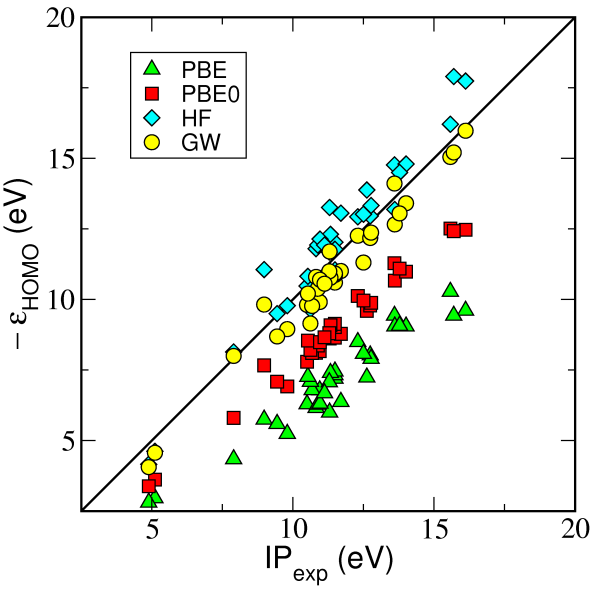


FIG. 1: (Color online) Calculated negative HOMO energy versus experimental ionization potential. Both PBE and PBE0 systematically understimates the ionization energy due to self-interaction errors while HF overestimates it slightly. The dynamical screening from the GW correlation lowers the HF energies bringing them closer to the experimental values. Numerical values are listed in Table I.

We stress the different meaning of fully selfconsistent GW and the recently introduced method of quasiparticle selfconsistent GW[58]: In fully selfconsistent GW the Green function obtained from Dyson's equation Eq. (2) with $\Sigma_{xc}[G] = \Sigma_{\rm GW}[G]$ is used to calculate the $\Sigma_{\rm GW}$ of

the next iteration. In QP-selfconsistent GW, $\Sigma_{\rm GW}$ is always evaluated using a non-interacting Green function and the self-consistency is obtained when the difference between the non-interacting GF and the interacting GF, is minimal.

Fig. 1 clearly shows that both the PBE and PBE0 eigenvalues of the HOMO severely underestimates the ionization potential. The average deviation from the experimental values are 4.35 eV and 2.55 eV, respectively. The overestimation of the single-particle eigenvalues of occupied states is a well known problem of DFT and can be ascribed to the insufficient cancellation of the self-interaction in the Hartree potential.[4,13] Part of this self-interaction is removed in PBE0. However, the fact that the HF results are significantly closer to experiments indicates that the 25% Fock exchange included in the PBE0 is not sufficient to cure the erroneous description of (occupied) molecular orbitals. On the other hand PBE0 gives good results for band gaps in semi-conductors and insulators where in contrast full Hartree-Fock does not perform well.[14–16] We conclude that the amount of Fock exchange to be used in the hybrid functionals to achieve good quasiparticle energies is highly system dependent. A similar problem is encountered with self-interaction corrected exchange-correlation functionals.[13]

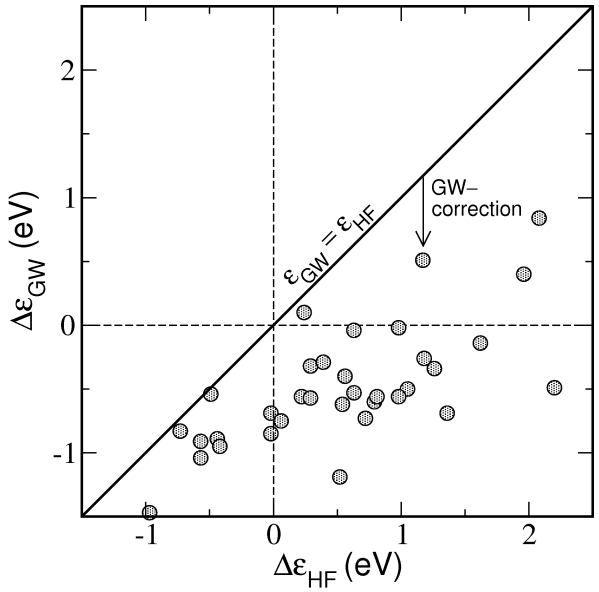


FIG. 2: The deviation of the calculated HOMO energy from the experimental ionization potential in GW and HF, respectively. The vertical displacement of points from the line $x = y$ gives the difference between the GW and HF energies and represents the effect of screening. Notice that the GW correction is always negative (corresponding to higher HOMO energy) and that it generally overcorrects the HF energies. Also notice that the GW correction is larger for molecules where HF presents the largest overestimation of the ionization potential.

As can be seen from Fig. 1, GW performs better than Hartree-Fock for the HOMO energy yielding a mean absolute error with respect to experiments of 0.5 eV compared to 0.81 eV with Hartree-Fock. As expected the difference between HF and GW is not large on an absolute scale (around 1 eV on average, see Table II) illustrating the fact that screening is weak in small molecules. On a relative scale selfconsistent GW improves the agreement with experiments by almost 30% as compared to HF.

To gain more insight into the influence of screening on the orbital energies, we compare in Fig. 2 the deviation of the HF and GW energies from $\mathrm{IP}_{\rm exp}$. The GW self-energy can be split into the bare exchange potential and an energy-dependent correlation part

$$\Sigma_{\rm GW}(\boldsymbol{r},\boldsymbol{r}';\varepsilon) = v_x(\boldsymbol{r},\boldsymbol{r}') + \Sigma_{\rm corr}(\boldsymbol{r},\boldsymbol{r}';\varepsilon) \qquad (12)$$

Accordingly the quasiparticle energy can be written as the bare HF energy and a correction due to the energy-dependent part of the GW self-energy (the dynamical screening term)

$$\varepsilon_n^{\rm QP} = \varepsilon_n^{\rm HF} + \Delta_n^{\rm GW}. \qquad (13)$$

In Fig. 2 the line $y = x$ corresponds to $\Delta_n^{\rm GW} = 0$, and the vertical displacement from the line thus represents the effect of screening on the calculated HOMO energy. We first notice that the effect of screening is to shift the HOMO level upwards in energy, i.e. to reduce the ionization potential. This can be understood by recalling that the Hartree-Fock eigenvalue represents the energy cost of removing an electron from the HOMO when orbital relaxations in the final state are neglected (Koopmans' theorem[16]). In Ref. 37 we showed, on the basis of GW and exact calculations for semi-empirical models of conjugated molecules, that $\Delta_n^{\rm GW}$ mainly describes the orbital relaxations in the final state and to a lesser extent accounts for the correlation energy of the initial and final states. This explains the negative sign of $\Delta_n^{\rm GW}$ because the inclusion of orbital relaxation in the final state lowers the energy cost of removing an electron. We note that this is different from the situation in extended, periodic systems where orbital relaxations vanish and the main effect of the GW self-energy is to account for correlations in the initial and final states.

In Table I we list the calculated HOMO energy for each of the 33 molecules. In addition to selfconsistent GW we have performed one-shot G_0W_0 calculations based on the HF and PBE Green's function, respectively. The best agreement with experiment is obtained for G_0W_0[HF]. This is because the relatively large Hartree-Fock HOMO-LUMO gap reduces the (over-)screening described by the resulting GW self-energy. There are not many GW calculations for molecules available in the litterature. Below Table I we list the few we have found. As can be seen they all compare quite well with our results given the differences in the implementation of the GW approximation.

For comparison we have included the HOMO energy predicted by second order Møller-Plesset theory (MP2) [taken from Ref. 59] with a Gaussian 6-311G** basis set. These are generally very close to our calculated HF

TABLE I: Experimental ionization potential (first column) and HOMO energy calculated using different approximations for exchange and correlation. "X-eig" refers to a single-particle eigenvalue while "X-tot" refers to a total energy difference, $E(N) - E(N-1)$. The G_0W_0(PBE) energies have been obtained from the QP equation while the GW and G_0W_0 energies are obtained from the DOS in Eq. (9). Last row shows the mean absolute error (MAE) with respect to experiments. All energies are in eV.

Molecule	Expt.[a]	PBE-eig	PBE0-eig	HF-eig	GW	G_0W_0(HF)	G_0W_0(PBE)-QP	MP2[a]	PBE-tot
LiH	7.90	4.34	5.81	8.14	8.0[b]	8.2[b]	8.0	8.20	8.02
Li_2	5.11	2.96	3.62	4.62	4.6	4.7	4.4	4.91	5.09
LiF	11.30	6.00	8.62	13.26	11.7	11.2	12.0	12.64	11.87
Na_2	4.89	2.81	3.38	4.16	4.1	4.3	4.7	4.48	4.97
NaCl	9.80	5.24	6.92	9.78	9.0	9.2	8.8	9.63	9.37
CO	14.01	9.05	10.98	14.80	13.4	14.1	13.9	15.08	13.88
CO_2	13.78	9.08	11.09	14.50	13.1	13.3	13.6	14.71	13.64
CS	11.33	7.40	9.09	12.31	10.8	11.7	11.0	12.58	11.31
C_2H_2	11.49	7.20	8.64	11.05	10.6	11.1	11.2	11.04	11.39
C_2H_4	10.68	6.79	8.11	10.11	9.8	10.4	9.6	10.18	10.67
CH_4	13.60	9.43	11.29	14.77	14.1	14.4	14.4[c]	14.82	14.10
CH_3Cl	11.29	7.08	8.80	11.68	11.0	11.4	11.1	11.90	11.10
CH_3OH	10.96	6.31	8.49	12.14	10.7	10.8	10.5	12.16	10.72
CH_3SH	9.44	5.60	7.09	9.50	8.8	9.0	8.4	9.73	9.29
Cl_2	11.49	7.32	9.02	12.03	10.9	11.3	11.5	12.37	11.22
ClF	12.77	7.90	9.88	13.33	12.4	12.4	13.0	13.63	12.48
F_2	15.70	9.43	12.42	17.90	15.2	15.2	16.2	18.20	15.39
HOCl	11.12	6.68	8.66	11.93	10.6	10.8	11.0	12.23	10.95
HCl	12.74	8.02	9.78	12.96	12.2	12.5	12.5	13.02	12.71
H_2O_2	11.70	6.38	8.78	13.06	11.0	11.1	11.1	13.00	11.18
H_2CO	10.88	6.28	8.37	11.93	10.4	10.5	10.6	11.97	10.80
HCN	13.61	9.05	10.67	13.19	12.7	13.2	12.4	13.33	13.67
HF	16.12	9.61	12.47	17.74	16.0	15.6	15.7	17.35	16.27
H_2O	12.62	7.24	9.59	13.88	12.3	12.1	11.9[d]	13.62	12.88
NH_3	10.82	6.16	8.11	11.80	10.8	11.0	10.6	11.57	11.02
N_2	15.58	10.28	12.51	16.21	15.1	15.7	15.6	16.41	15.39
N_2H_4	8.98	5.75	7.67	11.06	9.8	10.1	9.5	11.07	9.90
SH_2	10.50	6.29	7.79	10.48	9.8	10.1	9.9	10.48	10.38
SO_2	12.50	8.08	9.96	13.02	11.3	11.7	11.7	13.46	12.12
PH_3	10.95	6.79	8.17	10.38	9.9	10.3	10.0	10.50	10.39
P_2	10.62	7.09	8.21	9.65	9.2	9.8	9.0	10.09	10.37
SiH_4	12.30	8.50	10.13	12.93	12.3	12.6	12.4[e]	13.25	11.95
Si_2H_6	10.53	7.27	8.54	10.82	10.2	10.6	9.9	11.03	10.36
SiO	11.49	7.46	9.14	11.78	10.9	11.2	11.3	11.82	11.27
MAE	-	4.35	2.55	0.81	0.5	0.4	0.5	0.82	0.24

[a] From Ref. 59. The MP2 calculations use a Gaussian 6-311G** basis set.
[b] To be compared with the GW value 7.85 and the G_0W_0(HF) value 8.19 reported in Ref. 24.
[c] To be compared with the G_0W_0(LDA) value 14.3 reported in Ref. 21.
[d] To be compared with the G_0W_0(LDA) value 11.94 reported in Ref. 22.
[e] To be compared with the G_0W_0(LDA) values 12.7 and 12.66 reported in Refs. 21 and 22, respectively.

values, with a tendency to lower energies which worsens the agreement with experiment slightly as compared to HF.

We have also calculated the DFT-PBE total energy difference between the neutral and cation species, $E(N) - E(N-1)$, see last column of Table I. This procedure leads to IPs in very good agreement with the experimental values (MAE of 0.24 eV). We stress that although this method is superior to the GW method for the IP of the small molecules studied here, it can yield only the

TABLE II: Mean absolute deviation between the IPs of the 33 molecules calculated with the different methods and experiment. The mean absolute deviation with respect to experiment coincide with the last row in Table I

Method	Expt.[(a)]	PBE-eig	PBE0-eig	HF-eig	GW	G_0W_0[HF]	MP2[(a)]	PBE-tot
Expt.	0.00	4.35	2.55	0.81	0.5	0.4	0.82	0.24
PBE	4.35	0.00	1.79	4.90	3.9	4.1	4.99	4.27
PBE0	2.55	1.79	0.00	3.11	2.1	2.3	3.20	2.48
HF	0.81	4.90	3.11	0.00	1.0	0.8	0.17	0.80
GW	0.5	3.9	2.1	1.0	0.00	0.3	1.1	0.4
G_0W_0[HF]	0.4	4.1	2.3	0.8	0.3	0.00	0.9	0.3
MP2	0.82	4.99	3.20	0.17	1.1	0.9	0.00	0.84
PBE-tot	0.24	4.27	2.48	0.80	0.4	0.3	0.84	0.00

[(a)]Data taken from Ref. 59.

HOMO and LUMO levels while higher excited states are inaccessible. Moreover it applies only to isolated systems and cannot be directly used to probe QP levels of e.g. a molecule on a surface.

In Table II we provide an overview of the comparative performance of the different methods. Shown is the mean average deviation between the IPs calculated with the different methods as well as the experimental values. Note that the numbers in the experiment row/column are the same as those listed in the last row of Table II.

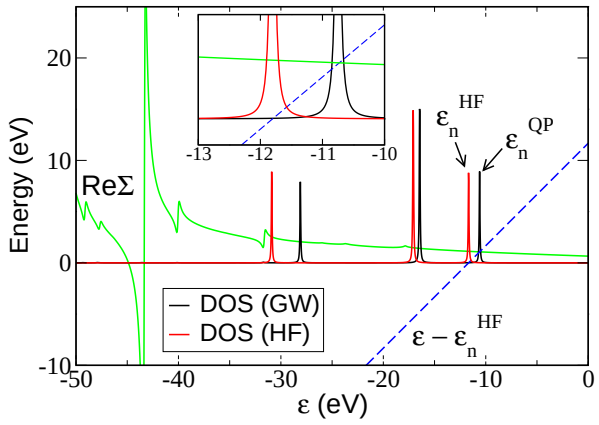


FIG. 3: (Color online) Density of states for the NH_3 molecule calculated in HF and GW, respectively. Arrows mark the level corresponding to the HOMO in the two calculations. The intersection between the line $y = \varepsilon - \varepsilon_n^{\rm HF}$ and the real part of $\langle\psi^0_{\rm HOMO}|\Sigma_{\rm corr}(\varepsilon)|\psi^0_{\rm HOMO}\rangle$ (green curve) determines the position of the GW level.

A. Linearized quasiparticle equation

In the conventional GW method the full Green function of Eq. (2) is not calculated. Rather one obtains the quasiparticle energies from the quasiparticle equation

$$\varepsilon_n^{\rm QP} = \varepsilon_n^0 + Z_n\langle\psi_n^0|\Sigma_{\rm GW}(\varepsilon_n^0) - v_{\rm xc}|\psi_n^0\rangle. \qquad (14)$$

where ψ_n^0 and ε_n^0 are eigenstates and eigenvalues of an approximate single-particle Hamiltonian (often the LDA Hamiltonian), and

$$Z_n = \left[1 - \frac{\partial\langle\psi_n^{\rm DFT}|\Sigma_{\rm GW}(\varepsilon)|\psi_n^0\rangle}{\partial\varepsilon}\Big|_{\varepsilon_n^0}\right]^{-1}. \qquad (15)$$

Moreover the GW self-energy is evaluated non-selfconsistently from the single-particle Green function, i.e. $\Sigma_{\rm GW} = iG_0W[G_0]$, with $G_0(z) = (z - H_0)^{-1}$.

The quasiparticle equation (14) relies on the assumption that off-diagonal matrix elements, $\langle\psi_n^0|\Sigma_{\rm GW}(\varepsilon_n^0) - v_{\rm xc}|\psi_m^0\rangle$, can be neglected, and that the frequency dependence of $\Sigma_{\rm GW}$ can be approximated by its first order Taylor expansion in a sufficiently large neighborhood of ε_n^0. We have found that these two assumptions are indeed fullfilled for the molecular systems studied here. More precisely, for the GW and G_0W_0(HF) self-energies, the QP energies obtained from Eq. (14) are always very close to the peaks in the density of states Eq. (9). We emphasize that this result could well be related to the rather large level spacing of small molecules, and may not hold for extended systems. An example is presented in Fig. 3 which shows the full HF and GW density of states for NH_3 together with the real part of $\langle\psi^0_{\rm HOMO}|\Sigma_{\rm corr}(\varepsilon)|\psi^0_{\rm HOMO}\rangle$. As explained in the following section this is not quite the case for the G_0W_0(PBE) calculations.

B. G_0-dependence

As stated in the previous section the GW and G_0W_0(HF) energies can be obtained either from the full spectral function or from the QP equation. In this case, returning to Table I, we see that G_0W_0(HF) yields systematically larger IPs than GW. This is easy to understand since $G_{\rm HF}$ describes a larger HOMO-LUMO gap than $G_{\rm GW}$, and therefore produces less screening. When the PBE rather than the HF Green function is used to evaluate the GW self-energy, we find that the spectral function obtained from Eq. (2) does not resemble a

simple discrete spectrum. In fact the peaks are significantly broadened by the imaginary part of $\Sigma_{\rm GW}$ and it becomes difficult to assign precise values to the QP energies. Apart from the spectral broadening, the molecular gap is significantly reduced with respect to its value in the GW and G_0W_0(HF) calculations. Both of these effects are due to the very small HOMO-LUMO gap described by $G_{\rm PBE}$ which leads to severe overscreening and QP lifetime reductions. A similar effect was observed by Ku and Eguiluz in their comparison of GW and G_0W_0(LDA) for Si and Ge crystals[40].

The problems encountered when attempting to solve the Dyson equation (2) using the G_0W_0(PBE) self-energy occur due to the large mismatch between $\varepsilon_n^{\rm PBE}$ and $\varepsilon_n^{\rm QP}$. On the other hand, in the QP equation, the GW self-energy is evaluated at ε_n^0 rather than $\varepsilon_n^{\rm QP}$. As a consequence the unphysical broadening and overscreening is avoided and a well defined QP energy can be obtained (last column in Table I).

To summarize, G_0 can have a very large effect on the QP spectrum when the latter is obtained via the Dyson equation (2). In particular, the use of a G_0 with a too narrow energy gap (as e.g. the $G_{\rm PBE}$) can lead to unphysical overscreening and spectral broadening. When the QP levels are obtained from the QP equation, the G_0-dependence is less pronounced because $\Sigma_{\rm GW}[G_0]$ is evaluated at ε_n^0 which is consistent with G_0.

The self-consistent GW spectrum is independent of the choice of G_0, but the number of iterations required to reach self-consistency is less when based on $G_{\rm HF}$.

C. Basis set convergence

In Figs. 4 and 5 we show the energy of the three highest occupied molecular orbitals of H_2O and CO obtained from selfconsistent GW using various sizes of the augmented Wannier basis. Clearly, the polarization functions have relatively little influence on the QP energies while the first set of additional zeta functions reduce the QP energies by up to 0.5 eV. The differences between DZP and TZDP are less than 0.15 eV for all the levels which justifies the use of DZP basis.

We have also compared the eigenvalues obtained from selfconsistent HF calculations using the DZP augmented Wannier basis to accurate HF calculations performed with the real-space code GPAW[45]. Here we obtain a MAE of 0.09 eV for the energy of the HOMO level of the 33 molecules.

IV. CONCLUSIONS

As the range of systems to which the GW method is being applied continues to expand it becomes important to establish its performance for other systems than the solids. In this work we have discussed benchmark GW calculations for molecular systems.

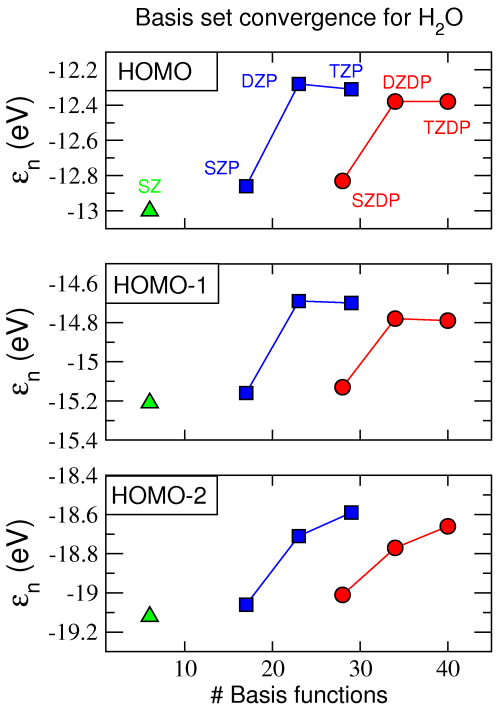


FIG. 4: (Color online) Convergence of the three highest occupied levels of H_2O obtained from GW calculations with different sizes of the augmented Wannier function basis. SZ denotes the Wannier function basis, while e.g. DZDP denotes the Wannier basis augmented by one extra radial function per valence state and two sets of polarization functions.

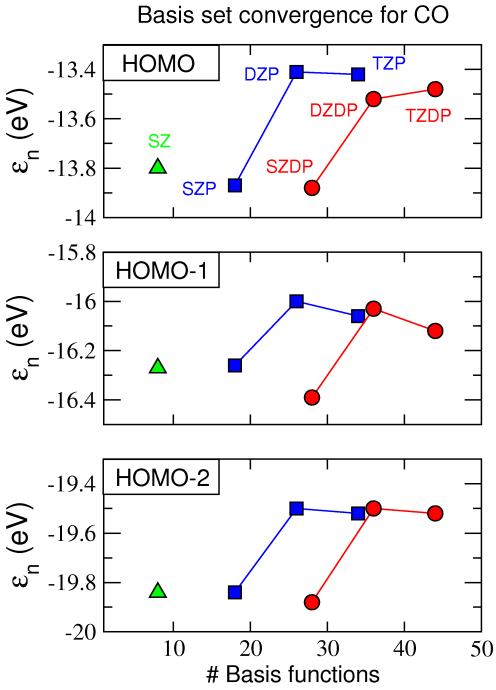


FIG. 5: (Color online) Same as Fig. 4 but for CO.

The GW calculations were performed using a novel

scheme based on the PAW method and a basis set consisting of Wannier functions augmented by numerical atomic orbitals. We found that a basis corresponding to double-zeta with polarization functions was sufficient to obtain GW energies converged to within 0.15 eV (compared to triple-zeta with double polarization functions). The GW self-energy was calculated on the real frequency axis including its full frequency dependence and off-diagonal elements. We thereby avoid all of the commonly used approximations, such as the the plasmon pole approximation, the linearized quasiparticle equation and analytical continuations from imaginary to real frequencies, and thus obtain a direct and unbiased assessment of the GW approximation itself. We found that the inclusion of valence-core exchange interactions, as facilitated by the PAW method, is important and shift the HF/GW HOMO levels by -1.2 eV on average.

The position of the HOMO for a series of 33 molecules was calculated using fully selfconsistent GW, single-shot G_0W_0, Hartree-Fock, DFT-PBE0, and DFT-PBE. Both PBE and PBE0 eigenvalues grossly overestimate the HOMO energy with a mean absolute error (MAE) with respect to the experimental ionization potentials (IP) of 4.4 and 2.5 eV, respectively. Hartree-Fock underestimates the HOMO energy but improves the agreement with experiments yielding a MAE of 0.8 eV. GW and G_0W_0 overcorrects the Hartree-Fock levels slightly leading to a small overestimation of the HOMO energy with a MAE relative to experiments of 0.4-0.5 eV. This shows that although screening is a weak effect in molecular systems its inclusion at the GW level improves the electron addition energies by 30-50% improvement relative to the unscreened Hartree-Fock. Very similar conclusions were reached comparing GW, G_0W_0 and HF to exact diagonalization for conjugated molecules described by the semi-empirical PPP model.[37]

V. ACKNOWLEDGMENTS

We thank Mikkel Strange for useful discussions and assistance with the projected Wannier functions. We acknowledge support from the Danish Center for Scientific Computing and The Lundbeck Foundation's Center for Atomic-scale Materials Design (CAMD).

APPENDIX A: THE GW SELF-ENERGY

Let U denote the rotation matrix that diagonalizes the pair orbital overlap $S_{ij,kl} = \langle n_{ij}|n_{kl}\rangle$, i.e. $U^\dagger S U = \sigma I$. The columns of U are truncated to those which have corresponding eigenvalues $\sigma_q < 10^{-5}a_0^{-3}$. We then only calculate the reduced number of Coulomb elements

$$V_{qq'} = \langle n_q|\frac{1}{|\boldsymbol{r}-\boldsymbol{r'}|}|n_{q'}\rangle, \tag{A1}$$

where $n_q(\boldsymbol{r})$ are the optimized pair orbitals

$$n_q(\boldsymbol{r}) = \sum_{ij} n_{ij}(\boldsymbol{r})U_{ij,q}/\sqrt{\sigma_q}, \tag{A2}$$

which are mutually orthonormal, i.e. $\langle n_q|n_{q'}\rangle = \delta_{qq'}$.

Determining the GW self-energy proceeds by calculating first the full polarization matrix in the time domain

$$P^<_{ij,kl}(t) = 2iG^<_{ik}(t)G^{>*}_{jl}(t), \tag{A3}$$

$$P^>_{ij,kl}(t) = P^{<*}_{ji,lk}(t). \tag{A4}$$

The factor 2 appears for spin-paired systems from summing over spin indices. This is then downfolded to the reduced representation

$$P^{\lessgtr}_{qq'} = \sum_{ij,kl}\sqrt{\sigma_q}U^*_{ij,q}P^{\lessgtr}_{ij,kl}U_{kl,q'}\sqrt{\sigma_{q'}}. \tag{A5}$$

The screened interaction can be determined from the lesser and greater polarization matrices, and the static interaction $V_{qq'}$, via the relations

$$P^r(t) = \theta(t)\left(P^>(t) - P^<(t)\right), \tag{A6}$$

$$W^r(\varepsilon) = [I - VP^r(\varepsilon)]^{-1}V, \tag{A7}$$

$$W^>(\varepsilon) = W^r(\varepsilon)P^>(\varepsilon)W^{r\dagger}(\varepsilon), \tag{A8}$$

$$W^<(\varepsilon) = W^>(\varepsilon) - W^r(\varepsilon) + W^{r\dagger}(\varepsilon), \tag{A9}$$

where all quantities are matrices in the optimized pair orbital basis and matrix multiplication is implied. We obtain the screened interaction in the original orbital basis from

$$W^{\lessgtr}_{ij,kl}(t) \approx \sum_{qq'} U_{ij,q}\sqrt{\sigma_q}W^{\lessgtr}_{qq'}(t)\sqrt{\sigma_{q'}}U^*_{kl,q'}, \tag{A10}$$

which is an approximation due to the truncation of the columns of U. Finally the GW self-energy can be determined by

$$\Sigma^{\lessgtr}_{\mathrm{GW},ij}(t) = i\sum_{kl} G^{\lessgtr}_{kl}(t)W^{\lessgtr}_{ik,jl}(t) \tag{A11}$$

$$\Sigma^r_{\mathrm{GW}}(t) = \theta(t)\left(\Sigma^>_{\mathrm{GW}}(t) - \Sigma^<_{\mathrm{GW}}(t)\right) + \delta(t)\Sigma_x. \tag{A12}$$

The exchange and Hartree potentials are given by

$$\Sigma_{x,ij} = i\sum_{kl} V_{ik,jl}G^<_{kl}(t=0) \tag{A13}$$

$$\Sigma_{\mathrm{H},ij} = -2i\sum_{kl} V_{ij,kl}G^<_{kl}(t=0) \tag{A14}$$

The Green functions are given by

$$G^r(\varepsilon) = [(\varepsilon + i\eta)S - H_{\mathrm{KS}} + v_{\mathrm{xc}} - \Delta v_{\mathrm{H}} - \Sigma^r_{\mathrm{GW}}(\varepsilon)]^{-1} \tag{A15}$$

$$G^<(\varepsilon) = -f_{\mathrm{FD}}(\varepsilon-\mu)[G^r(\varepsilon) - G^r(\varepsilon)^\dagger] \tag{A16}$$

$$G^>(\varepsilon) = (1 - f_{\mathrm{FD}}(\varepsilon-\mu))[G^r(\varepsilon) - G^r(\varepsilon)^\dagger] \tag{A17}$$

where $f_{\mathrm{FD}}(\varepsilon-\mu)$ is the Fermi-Dirac function and $\Delta v_{\mathrm{H}} = \Sigma_{\mathrm{H}}[G] - \Sigma_{\mathrm{H}}[G_{\mathrm{DFT}}]$ is the difference between the GW Hartree potential and the DFT Hartree potential. For self-consistent calculations, equation (A3)-(A17) are iterated untill convergence in G.

[1] P. Hohenberg and W. Kohn Phys. Rev. **136**, B864 (1964).
[2] W. Kohn and L. J. Sham Phys. Rev. **140**, A1133 (1965).
[3] M. S. Hybertsen and S. G. Louie, Phys. Rev. B **34**, 5390 (1986).
[4] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).
[5] J. Paier, M. Marsman, K. Hummer, G. Kresse, I. C. Gerber, and J. G. Angyan, J. Chem. Phys. **124**, 154709 (2006)
[6] J. Heyd, J. E. Peralta, G. E. Scuseria, and R. L. Martin, J. Chem. Phys. **123**, 174101 (2005).
[7] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999)
[8] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993)
[9] C. Lee, W. Yang and R. G. Parr, Phys. Rev. B **37**, 785 (1988)
[10] J. Heyd, G. E. Scuseria, and M. Ernzerhof, J. Chem. Phys. **118**, 8207 (2003)
[11] According to Janak's theorem[12], for finite systems the position of the HOMO in exact Kohn-Sham theory should coincide with the ionization potential. However, for practical exchange-correlation functionals this is far from being the case[4,13].
[12] J. F. Jarnak, Phys. Rev. B **18**, 7165 (1978).
[13] C. D. Pemmaraju, T. Archer, D. Sanchez-Portal, and S. Sanvito, Phys. Rev. B **75**, 045101 (2007)
[14] J. Q. Sun and R. J. Bartlett, J. Chem. Phys. **104**, 8553 (1996)
[15] L. Dagens and F. Perrot, Phys. Rev. B **5**, 641 (1972)
[16] *Solid State Physics*, G. Grosso and G. P. Parravicini, Cambridge University Press, Cambridge 2000
[17] L. Hedin, Phys. Rev. **139**, A796 (1965)
[18] F. Aryasetiawan and O. Gunnarson, Rep. Prog. Phys. **61**, 237 (1998)
[19] W. G. Aulbur, L. Jonsson, and J. Wilkins, in *Solid State Physics*, edited by H. Ehrenreich and F. Saepen (Academic Press, New York 2000), Vol. 54, 1.
[20] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002)
[21] J. C. Grossman, M. Rohlfing, L. Mitas, S. G. Louie, and M. L. Cohen, Phys. Rev. Lett. **86**, 472 (2001).
[22] P. H. Hahn, W. G. Schmidt, and F. Bechstedt Phys. Rev. B **72**, 245425 (2005)
[23] T. Niehaus, M. Rohlfing, F. D. Sala, A. Di Carlo, and T. Frauenhaim, Phys. Rev. A **71**, 022508 (2005)
[24] A. Stan, N. E. Dahlen, and R. van Leeuwen, Europhys. Lett. **76**, 298 (2006)
[25] C. D. Spataru, S. Ismail-Beigi, X. L. Benedict, and S. G. Louie, Phys. Rev. Lett. **92**, 077402 (2004)
[26] P. E. Trevisanutto, C. Giorgetti, L. Reining, M. Ladisa, and V. Olevano, Phys. Rev. Lett. **101**, 226405 (2008)
[27] M. L. Tiago and J. R. Chelikowsky, Phys. Rev. B **73**, 205334 (2006)
[28] J. B. Neaton, M. S. Hybertsen, and S. G. Louie Phys. Rev. Lett. **97**, 216405 (2006)
[29] K. S. Thygesen and A. Rubio, Phys. Rev. Lett. **102**, 046802 (2009)
[30] C. Freysoldt and P. Rinke and M. Scheffler, Phys. Rev. Lett. **103**, 056803 (2009)
[31] J. M. Garcia-Lastra, C. Rostgaard, A. Rubio, and K. S. Thygesen Phys. Rev. B **80**, 245427 (2009)
[32] K. S. Thygesen and A. Rubio, J. Chem. Phys. **126**, 091101 (2007).
[33] K. S. Thygesen, Phys. Rev. Lett. **100**, 166804 (2008)
[34] P. Myöhanen and A. Stan and G. Stefanucci and R. van Leeuwen, Euro. Phys. Lett. **84**, 67001 (2008)
[35] C. D. Spataru and M. S. Hybertsen and S. G. Louie and A. J. Millis, Phys. Rev. B **79**, 155110 (2009)
[36] M. P. von Friesen and C. Verdozzi and C.-O. Almbladh, Phys. Rev. Lett. **103**, 176404 (2009)
[37] K. Kaasbjerg and K. S. Thygesen (submitted)
[38] B. Holm and U. von Barth Phys. Rev. B **57**, 2108 (1998)
[39] M. Usuda, N. Hamada, T. Kotani, and M. van Schiflgaarde, Phys. Rev. B **66**, 125101 (2002)
[40] W. Ku and A. G. Eguiluz, Phys. Rev. Lett. **89**, 126401 (2002)
[41] M. van Schiflgaarde, T. Kotani, and S. V. Faleev, Phys. Rev. B **74**, 245125 (2006)
[42] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006)
[43] P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt, and M. Scheffler New J. Phys. **7** 126 (2005)
[44] X. Qian, J. Li, L. Qi, C.-Z. Wang, T.-L. Chan, Y.-X. Yao, K.-M. Ho, and S. Yip, Phys. Rev. B **78**, 245112 (2008)
[45] J. J. Mortensen, L. B. Hansen, K. W. Jacobsen, Phys. Rev. B **71**, 035109 (2005)
[46] A. H. Larsen, M. Vanin, J. J. Mortensen, K. S. Thygesen, and K. W. Jacobsen Phys. Rev. B **80**, 195112 (2009)
[47] K. S. Thygesen and A. Rubio, Phys. Rev. B **77**, 115333 (2008).
[48] K. S. Thygesen, Phys. Rev. B **73**, 035309 (2006)
[49] H. Haug and A. -P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors*, Springer (1998)
[50] R.van Leeuwen, N.E.Dahlen, G.Stefanucci, C.-O.Almbladh, and U.von Barth, Lectures Notes in Physics vol. 706 (Springer)
[51] A. Stan, N. E. Dahlen, and R. van Leeuwen, J Chem. Phys. **130**, 114105 (2009)
[52] F. Aryasetiawan and O. Gunnarsson, Phys. Rev. B **49**, 16214 (1994)
[53] M. Walter, H. Häkkinen, L. Lehtovaara, M. Puska, J. Enkovaara, C. Rostgaard, and J. J. Mortensen, J Chem. Phys. **128**, 244101 (2009)
[54] P. E. Blöchl, Phys. Rev. B **50**, 17953 (1994)
[55] E. Engel, Phys. Rev. B **80**, 161205(R) (2009)
[56] L. A. Curtiss, K. Raghavachari, P. Redfern, and J. A. Pople, J. Chem. Phys. **106**, 1063 (1997).
[57] J. P. Perdew, K. Burke, M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996)
[58] M. van Schilfgaarde, T. Kotani, and S. Faleev, Phys. Rev. Lett. **96**, 226402 (2006)
[59] NIST Computational Chemistry Comparison and Benchmark Database, NIST Standard Reference Database Number 101 Release 14, Sept 2006, Editor: Russell D. Johnson III http://srdata.nist.gov/cccbdb

Paper III

Polarization-induced renormalization of molecular levels at metallic and semiconducting surfaces
J. M. Garcia-Lastra, C. Rostgaard, A. Rubio, and K. S. Thygesen
Physical Review B **80**, 245427 (2009).

Polarization-induced renormalization of molecular levels at metallic and semiconducting surfaces

J. M. Garcia-Lastra,[1,2] C. Rostgaard,[1] A. Rubio,[2] and K. S. Thygesen[1]
[1]*Center for Atomic-Scale Materials Design (CAMD), Department of Physics, Technical University of Denmark, DK-2800 Kongens Lyngby, Denmark*
[2]*ETSF Scientific Development Centre, Depto. Física de Materiales, Centro de Física de Materiales CSIC-UPV/EHU and DIPC, Universidad del País Vasco (UPV)/EHU, Avenue Tolosa 72, 20018 San Sebastián, Spain*
(Received 28 October 2009; published 21 December 2009)

On the basis of first-principles G_0W_0 calculations we systematically study how the electronic levels of a benzene molecule are renormalized by substrate polarization when physisorbed on different metallic and semiconducting surfaces. The polarization-induced reduction in the energy gap between occupied and unoccupied molecular levels is found to scale with the substrate density of states at the Fermi level (for metals) and substrate band gap (for semiconductors). These conclusions are further supported by calculations on simple lattice models. By expressing the electron self-energy in terms of the substrate's joint density of states we relate the level shift to the surface electronic structure, thus providing a microscopic explanation of the trends in the *GW* and G_0W_0 calculations. While image charge effects are not captured by semilocal and hybrid exchange-correlation functionals, we find that error cancellations lead to remarkably good agreement between the G_0W_0 and Kohn-Sham energies for the occupied orbitals of the adsorbed molecule.

DOI: 10.1103/PhysRevB.80.245427 PACS number(s): 85.65.+h, 31.70.Dk, 71.10.−w, 73.20.−r

I. INTRODUCTION

Solid-molecule interfaces are central to a number of important areas of physics and chemistry including heterogeneous catalysis, electrochemistry, molecular- and organic electronics, and scanning tunneling spectroscopy.[1–4] Most of our current understanding of level alignment at interfaces builds on effective single-particle descriptions such as the Kohn-Sham scheme of density-functional theory (DFT).[5] Within such theories the energy levels of a molecule close to a surface are determined by hybridization, charge-transfer, and interface dipole fields—all properties of the static mean field potential defining the single-particle Hamiltonian. On the other hand, from photoemission and electron transport measurements it is well known that the dynamic polarizability of the molecule's local environment can have a large influence on the level positions.[6–10] Such polarization effects, which are induced by changes in the charge state of the molecule, are not captured by available single-particle descriptions.

Many-body perturbation theory provides a systematic method to obtain the true single-particle excitations [sometimes referred to as addition/removal energies or quasiparticle (QP) energies] from the Green function of the system. In the G_0W_0 approximation the electron self-energy is written as a product of the (noninteracting) Green function and a dynamically screened Coulomb interaction, $\Sigma = iG_0W_0$.[11,12] It is instructive to compare this to the bare exchange self-energy given by $\Sigma_x = iG_0V$, where V is the unscreened Coulomb interaction. It is well known that the Hartree-Fock (HF) eigenvalues correspond to energy differences between the N-particle groundstate and the *unrelaxed* $N \pm 1$-particle Slater determinants (Koopmans' theorem). The effect of replacing V with the screened and frequency dependent W_0 is twofold: it introduces correlations into the many-body eigenstates, and it includes the response of the other electrons to the added electron/hole, i.e., relaxation effects. For a molecule at a surface, the latter effect is particularly important as it incorporates the attractive interaction between the added electron/hole and its induced image charge, into the QP spectrum.

Recent experiments on molecular charge transport have renewed the interest for theoretical modeling of polarization-induced level renormalization. First-principles G_0W_0 calculations for a benzene molecule on graphite[13] as well as CO on NaCl/Ge(001)[14] have demonstrated significant reductions of

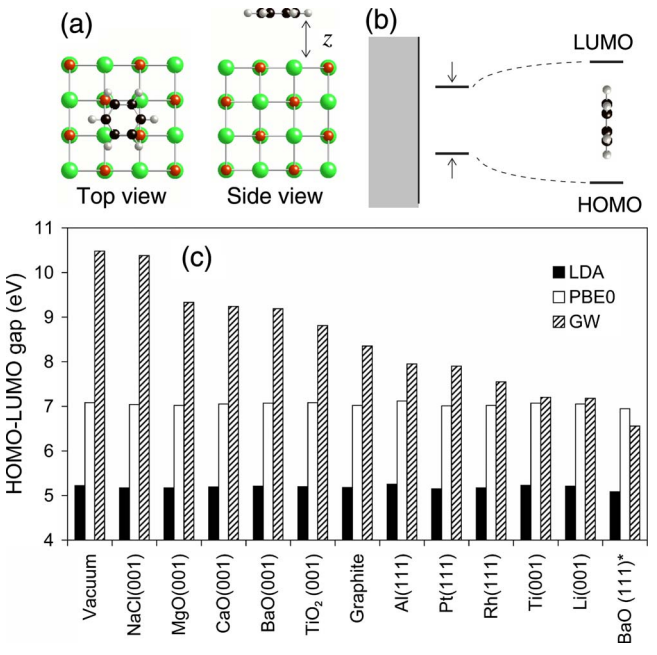


FIG. 1. (Color online) (a) Supercell used to represent benzene physisorbed on NaCl(001). (b) Reduction in a molecule's energy gap when it approaches a polarizable surface. (c) Calculated LDA, PBE0, and G_0W_0 HOMO-LUMO gap of a benzene molecule lying flat at z=4.5 Å above different surfaces. Note that BaO(111) is metallic due to surface states in the BaO band gap.

the molecular energy gap due to image charge effects. Model *GW* calculations have been used to elucidate the qualitative features of the effect across different bonding regimes.[15] Classical electrostatic models of various complexities have been developed to correct energy levels obtained from single-particle calculations.[16–19]

In this work, we present a systematic study of image charge-induced renormalization at a range of different surfaces taking both a classical and quantum many-body viewpoint. We have performed DFT calculations with local density approximation (LDA) and hybrid (PBE0) exchange-correlation functionals as well as G_0W_0 calculations for a benzene molecule weakly physisorbed on the metals Li, Al, Ti, Rh, Pt, and the semiconductors/insulators TiO_2, BaO, MgO, CaO, and NaCl. The results for the highest occupied molecular orbital (HOMO)- lowest unoccupied molecular orbital (LUMO) gap of benzene are shown in Fig. 1. While LDA and PBE0 yields a substrate independent HOMO-LUMO gap, the G_0W_0 gaps are reduced from the gas phase value by an amount which depends on the polarizability of the surface. For all systems, we find that the dependence of the QP gap on the distance to the surface can be described by a classical image charge model. However, the model parameters are sensitive to the microscopic details of the system and this limits the usefulness of the classical model in pratice. By evaluating the G_0W_0 self-energy to second order we obtain a simple analytic expression which relates the level shift to the substrate's joint density of states weighted by Coulomb interaction matrix elements. The model suggests that the HOMO-LUMO gap should scale with the substrate band gap (for semiconducting surfaces) and density of states at the Fermi level (for metallic surfaces). This trend is verified for the first-principles results and is further supported by *GW* calculations for simple lattice models. Finally, we analyze the deviation between the DFT and G_0W_0 results in more detail. We find that the occupied Kohn-Sham levels obtained with LDA (PBE0) are in very good agreement with the G_0W_0 results for benzene adsorbed on the metallic (semiconducting) surfaces, and we show that this is a result of significant error cancellation in the LDA/PBE0 approximations.

The paper is organized as follows. In Sec. II we outline the methodology used for the first-principles and model *GW* calculations. In Sec. III we investigate to what extent the first-principles G_0W_0 results can be explained by a classical image charge model. In Sec. IV we derive a simple analytical expression for the polarization-induced level shift and show that it explains the main trends in both the first-principles as well as the model calculations. At the end of the section we analyze the description of occupied and unoccupied levels separately and discuss the effect of error cancellations in the DFT results. We conclude in Sec. V

II. METHODS

A. *Ab initio* G_0W_0 calculations

To model the solid-molecule interfaces we use a slab containing four atomic layers of the substrate in the experimentally most stable phase and a benzene molecule lying flat above the surface followed by 12 Å of vacuum. The benzene molecule is not relaxed on the surface but is fixed in its gas phase structure at a distance z from the surface. An example of a supercell is shown in Fig. 1(a) for the case of benzene on NaCl(001). The number of atoms included in the supercell per atomic layer is 9 for Al, Rh, Pt, Ti; 12 for Li and TiO_2; and 16 for NaCl, MgO, CaO and BaO. This corresponds to distances between periodically repeated benzene molecules in the range 8.1 to 9.9 Å. All DFT calculations have been performed with the PWSCF code[20] which using norm-conserving pseudopotentials.[21] For exchange-correlation functionals we have used the local density approximation[22] as well as the PBE0 hybrid functional.[23,24] The Brillouin zone (BZ) was sampled on a $4\times4\times1$ k-point mesh, and the wave functions were expanded with a cut-off energy of 40 Hartree.

In the G_0W_0(LDA) method one obtains the QP energies from the linearized QP equation

$$\varepsilon_n^{\mathrm{QP}} = \varepsilon_n^{\mathrm{LDA}} + Z_n\langle\psi_n^{\mathrm{LDA}}|\Sigma_{GW}(\varepsilon_n^{\mathrm{LDA}}) - v_{\mathrm{xc}}|\psi_n^{\mathrm{LDA}}\rangle, \qquad (1)$$

where ψ_n^{LDA} and $\varepsilon_n^{\mathrm{LDA}}$ are LDA eigenstates and eigenvalues, and

$$Z_n = \left[1 - \left.\frac{\partial\langle\psi_n^{\mathrm{LDA}}|\Sigma_{GW}(\varepsilon)|\psi_n^{\mathrm{LDA}}\rangle}{\partial\varepsilon}\right|_{\varepsilon_n^{\mathrm{LDA}}}\right]^{-1}. \qquad (2)$$

The self-energy, Σ_{GW}, is evaluated non-selfconsistently from the single-particle Green function, i.e., $\Sigma_{GW}=iG_0W_0$, with $G_0(z)=(z-H^{\mathrm{LDA}})^{-1}$. It is customary to use the random phase approximation for the screened interaction, i.e., $W_0=V(1-VP)^{-1}$ with $P=-iG_0G_0$.

We have performed the G_0W_0 calculations with the YAMBO code[25] using the LDA wave functions and eigenvalues from the PWSCF calculations as input. The plasmon pole approximation has been applied with a frequency of 1 hartree (the HOMO and LUMO energies of benzene change by less than 0.05 eV when the plasmon frequency is varied between 0.5 and 2.0 hartrees). In the calculation of the self-energy we included a minimum of 200 empty states. We have checked that calculations are converged with respect to slab thickness, lateral supercell size, k-point mesh, all energy cutoffs and that we reproduce the results previously reported in Ref. 13 for benzene on graphite at $z=3.25$ Å.

B. Model *GW* calculations

In addition to the first-principles G_0W_0 calculations, we have performed (self-consistent) *GW* calculations for two lattice models representing a metal-molecule and semiconductor-molecule interface, respectively. The model Hamiltonians contain three terms

$$\hat{H} = \hat{H}_{\mathrm{sol}} + \hat{H}_{\mathrm{mol}} + \hat{U}, \qquad (3)$$

describing the solid (metal or semiconductor), the molecule, and their mutual interaction, see Fig. 2. A metallic substrate is modeled by a semi-infinite tight-binding (TB) chain (we suppress the spin for notational simplicity),

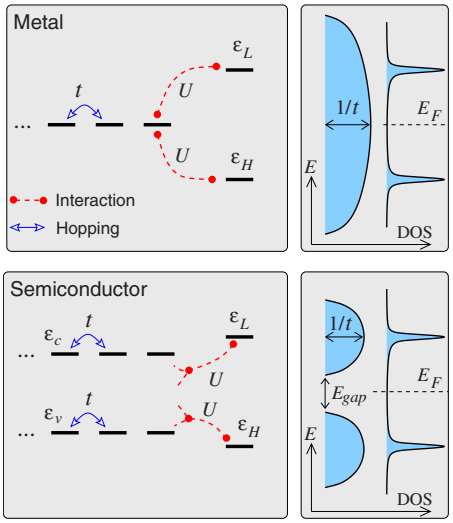


FIG. 2. (Color online) The lattice models representing a metal-molecule and semiconductor-molecule interface, respectively. We consider the weak coupling limit where no hybridization between the molecule and surface states occur. Thus the only interaction between the solid and molecule is via the nonlocal Coulomb interaction U.

$$\hat{H}_{\text{met}} = \sum_{i=-\infty}^{0} t(c_i^\dagger c_{i-1} + c_{i-1}^\dagger c_i). \tag{4}$$

A semiconducting substrate is modeled by

$$\hat{H}_{\text{sc}} = \sum_{\alpha=c,v} \sum_{i=-\infty}^{0} \varepsilon_\alpha \hat{n}_{\alpha i} + t(c_{\alpha i}^\dagger c_{\alpha i-1} + c_{\alpha i-1}^\dagger c_{\alpha i}), \tag{5}$$

where $\alpha=c,v$ refers to conduction and valence bands, respectively.

The molecule is represented by its HOMO and LUMO levels, i.e.,

$$\hat{H}_{\text{mol}} = \xi_H \hat{n}_H + \xi_L \hat{n}_L \tag{6}$$

where e.g., $\hat{n}_H = c_{H\uparrow}^\dagger c_{H\uparrow} + c_{H\downarrow}^\dagger c_{H\downarrow}$, is the number operator of the HOMO level.

Finally, the interaction between the molecule and the terminal site(s) of the substrate TB chain(s) is described by

$$\hat{U} = \begin{cases} U\hat{n}_0 \hat{N}_{\text{mol}} & \text{for metals} \\ U\sum_\sigma (c_{c0,\sigma}^\dagger c_{v0,\sigma} + c_{v0,\sigma}^\dagger c_{c0,\sigma})\hat{N}_{\text{mol}} & \text{for semicond.} \end{cases}$$

where $\hat{N}_{\text{mol}} = \hat{n}_H + \hat{n}_L$ is the number operator of the molecule. Note that since polarization of a semiconductor occurs via transitions between valence and conduction bands, only the interaction terms of the form given above contribute to the image charge effect (this will become clear in Sec. IV A).

We set $E_F=0$ corresponding to a half filled band for the metal. We choose ξ_H and ξ_L so that the molecule contains exactly two electrons (E_F in the middle of the HOMO-LUMO gap). We consider the limit of zero hybridization between the solid and molecule so that interaction between the solid and molecule occurs only via the nonlocal $\hat{U}$. The model neglects interactions within the TB chain and between the molecule and interior TB sites ($i<0$). These approximations are, however, not expected to influence the image charge physics described by the model in any qualitative way.

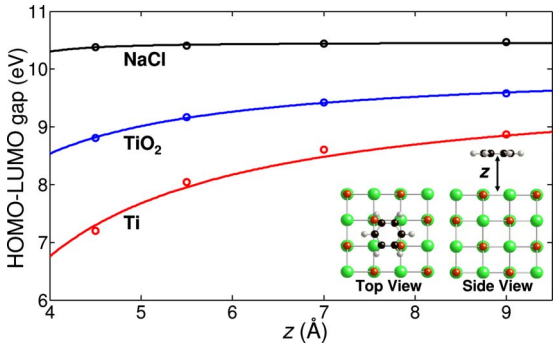


FIG. 3. (Color online) Calculated G_0W_0 energy gap of benzene on NaCl, TiO_2, and Ti surfaces (circles) as a function of the distance to the surface, and the best fit to the classical model Eq. (8) (full lines).

We obtain the Green function of the molecule from

$$G(z) = 1/(z - H_{\text{mol}} - \Sigma_{GW}[G](\varepsilon)) \tag{7}$$

where the Hartree potential due to $\hat{U}$ has been absorbed in H_{mol}. The GW self-energy is calculated fully self-consistently using a recently developed GW scheme for quantum transport.[26] The renormalized molecular QP levels are obtained as peaks in the spectral function $A_\nu(\varepsilon) = -(1/\pi)\text{Im}\, G_{\nu\nu}^r(\varepsilon)$.

III. CLASSICAL THEORY

In this section we investigate to what extent the G_0W_0 results of Fig. 1 can be described by a classical image charge model. The electrostatic energy of a point charge, q, located in vacuum at position $(0,0,z)$ above a polarizable medium filling the half-space $z<z_0$, is given by (in a.u.)

$$V = \frac{qq'}{4(z-z_0)}. \tag{8}$$

The size of the image charge is $q' = q(1-\epsilon)/(1+\epsilon)$, where ϵ is the relative dielectric constant of the medium.[27] In 1973 Lang and Kohn showed that the energy of a classical point charge above a quantum jellium surface follows Eq. (8) with $q'=-q$ (corresponding to $\epsilon=\infty$ as expected for a perfect metal), with the image plane, z_0, lying 0.5–0.9 Å outside the surface depending on the electron density.[28] More recently, *ab initio* G_0W_0 calculations have found the same asymptotic form for the potential felt by an electron outside a metallic surface.[27,29,30] From this it seems reasonable to conclude that the asymptotic position of the electronic levels of a molecule outside a surface would also follow the image potential of Eq. (8). This is, however, only true for the unoccupied levels whereas the occupied levels experience a shift in the opposite direction, i.e., the shift is upward in energy as the molecule approaches the surface. This is because the occupied

TABLE I. Position of the effective image plane, z_0, and dielectric constant, ϵ_{model}, obtained by fitting the z dependence of the HOMO-LUMO gap to Eq. (8). Last row shows the experimental optical dielectric constant of the bulk. The two values for the nonisotropic TiO_2 refers to longitudinal and transverse polarization directions. Experimental data taken from Ref. 33.

	z_0 (Å)	ϵ_{model}	$\epsilon_\infty^{\text{exp,bulk}}$
NaCl(001)	1.70	1.15	2.30
MgO(001)	1.20	2.63	2.95
CaO(001)	2.69	1.56	3.30
BaO(001)	2.74	1.77	3.83
TiO_2(001)	1.79	2.76	8.43/6.84
Al(111)	0.55	∞	∞
Pt(111)	0.60	∞	∞
Rh(111)	1.28	∞	∞
Ti(001)	1.66	∞	∞
Li(001)	1.72	∞	∞
BaO(111)	2.01	∞	3.83

levels represent the negative of the energy cost of removing an electron from the molecule. Similarly it has been found that the image potential leads to band gap narrowing at semiconductor-metal interfaces.[31,32]

To test whether the gap reductions obtained in the G_0W_0 calculations can be described by the classical image charge model we have fitted Eq. (8) to the calculated HOMO-LUMO gap for $z=4.5,5.5,7.0,9.0,\infty$ Å. In Fig. 3 we show the result of the fit for three systems (the fit is equally good for the other systems). The best-fit values for the effective image plane z_0 and the dielectric constant ϵ_{model} are given in Table I.

As can be seen ϵ_{model} is generally smaller than the experimental optical dielectric constant of the bulk, $\epsilon_\infty^{\text{exp,bulk}}$. This is expected since the latter gives the long-range response of the bulk while ϵ_{model} probes the local response at the surface. Part of the discrepancy between $\epsilon_\infty^{\text{exp}}$ and ϵ_{model} is clearly due to geometric effects. By taking the surface geometry into account, as done in Ref. 18, better estimates of ϵ_{model} can be produced from $\epsilon_\infty^{\text{exp}}$. On the other hand, electronic effects due to the local atomic structure of the surface cannot be captured by a classical model. For example, the BaO(111) surface is metallic due to surface states, and thus $\epsilon_{\text{model}} \approx \infty$ while $\epsilon_\infty^{\text{exp}}=3.83$. Similarly, impurities, defects, and surface roughness are expected to influence the local dielectric properties of the surface.

According to the classical image charge model all the molecular levels should experience the same shift (the sign of the shift being different for occupied and unoccupied levels). However, we have found that the best-fit values for z_0 and ϵ_{model} obtained by fitting the HOMO and LUMO levels separately, are in general different—most notably for the metallic surfaces. This observation, which is discussed in more detail in Sec. IV C, shows that the shape of molecular orbital also influences the size of the polarization-induced shift.

IV. MICROSCOPIC THEORY

In this section we first consider the *GW* self-energy for a molecule interacting with a surface to second-order in the electron-electron interaction. This leads to a simple microscopic model for image charge renormalization which relates the shift of molecular levels to the electronic structure of the surface, and explains general trends of the first-principles and model *GW* calculations. In the last section we consider the HOMO and LUMO levels separately and explain how error cancellations in semilocal exchange-correlation functionals can explain the surprisingly good agreement found between LDA eigenvalues and *GW* QP energies for the occupied levels of benzene on metallic surfaces.

A. Second-order expansion

In quantum many-body theory, the effect of substrate polarization on the energy levels of a molecule enters the Green function via a self-energy operator. In general, the G_0W_0 self-energy can be written symbolically as

$$\Sigma = \sum_{n=1} \Sigma^{(n)} = \sum_{n=1} iG_0V(PV)^{n-1}, \tag{9}$$

where G_0 is the Green function of the noninteracting (Kohn-Sham) Hamiltonian, and $P=-iG_0G_0$ is the polarization bubble. The first-order term, $\Sigma^{(1)}$, is simply the static exchange potential while the remaining terms account for correlations and dynamic screening. In the following we consider the second-order term, $\Sigma^{(2)}=iG_0VPV$ explicitly. This corresponds to approximating the response of the substrate by its noninteracting response, P.

For sufficiently large surface-molecule separations ($z \gtrsim 3.5$ Å) we can neglect hybridization effects, and the noninteracting eigenstates of the combined system can be taken as the eigenstates of the isolated molecule and surface. We denote these eigenstates by $\{\psi_a\}$ ("a" for adsorbate) and $\{\psi_k\}$, respectively. To see how a given electronic level, ε_a, is renormalized by polarization processes in the substrate we consider the (time-ordered) matrix element $\Sigma_{aa}^{(2)}(\omega) = \langle \psi_a | \Sigma^{(2)}(\omega) | \psi_a \rangle$, given by

$$\Sigma_{aa}^{(2)} = \sum_k^{\text{occ}} \sum_{k'}^{\text{empty}} \int iG_{0,aa}(\omega') V_{aa,kk'} P_{kk'}(\omega' - \omega) V_{k'k,aa} d\omega'. \tag{10}$$

The Feynman diagram corresponding to $\Sigma_{aa}^{(2)}$ is shown in Fig. 4(a). The polarization and Coulomb matrices are given by

$$P_{kk'}(\omega) = \frac{1}{\omega - \omega_{kk'} + i\eta} - \frac{1}{\omega + \omega_{kk'} - i\eta} \tag{11}$$

$$V_{kk',aa} = \int\int \frac{\psi_k^*(\mathbf{r})\psi_{k'}(\mathbf{r})|\psi_a(\mathbf{r}')|^2}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' \tag{12}$$

where η is a positive infinitesimal and $\omega_{kk'}=\varepsilon_{k'}-\varepsilon_k \gtrless 0$. Using that $G_{0,aa}(\omega)=1/(\omega-\varepsilon_a+\text{sgn}(\varepsilon_a-E_F)i\eta)$,[11] Eq. (10) reduces to

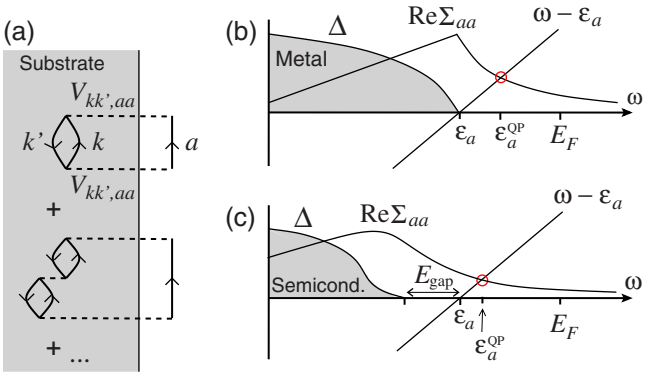


FIG. 4. (Color online) (a) Feynman diagrams representing dynamic polarization of the substrate induced by an electron propagating in the molecule. (b) and (c): generic shapes of the imaginary and real parts of the self-energy of Eq. (15) for an occupied molecular level $|a\rangle$ interacting with a metallic and semiconducting substrate assuming $V_{kk',aa}$ to be energy independent.

$$\Sigma_{aa}^{(2)}(\omega) = \frac{1}{\pi}\int \frac{\Delta(\omega')}{\omega - \omega' + \mathrm{sgn}(E_F - \varepsilon_a)i\eta} d\omega' \tag{13}$$

where we have defined the interaction strength,

$$\Delta = \pi \sum_{k}^{\text{occ}} \sum_{k'}^{\text{empty}} |V_{kk',aa}|^2 \delta(\omega_{kk'} - \mathrm{sgn}(\varepsilon_a - E_F)(\omega - \varepsilon_a)). \tag{14}$$

Note that Δ is simply the joint density of states (JDOS) of the substrate, shifted by ε_a, and weighted by the Coulomb matrix elements. The physically relevant retarded self-energy is readily obtained from Eq. (13)

$$\Sigma_{aa}^{(2),r} = \frac{\mathcal{P}}{\pi}\int \frac{\Delta(\omega')}{\omega' - \omega} d\omega' - i\Delta(\omega). \tag{15}$$

where $\mathcal{P}$ denotes the Cauchy principal value. Now, the renormalized QP energy can be obtained from the equation (neglecting off-diagonal terms)

$$\varepsilon_a^{QP} - \varepsilon_a - \mathrm{Re}\langle\psi_a|\Sigma^{(2),r}(\varepsilon_a^{QP})|\psi_a\rangle = 0 \tag{16}$$

A graphical solution to the QP equation is illustrated in Figs. 4(b) and 4(c) for the case of an occupied molecular level $\varepsilon_a < E_F$ interacting with a metal or semiconductor surface, respectively.

From Eq. (14) it follows that the image charge effect does not broaden the molecular level because Im $\Sigma^{(2)}(\varepsilon_a)=0$. We also note that the level shift is independent of the absolute value $|\varepsilon_a - E_F|$, and that the effect of changing the sign of $\varepsilon_a - E_F$ is to change the sign of the level shift. These properties are all in line with the classical theory.

In the limit where $V_{kk',aa}$ varies little with k and k', Δ is simply proportional to the shifted JDOS [the "generic" cases illustrated in Figs. 4(b) and 4(c)]. In this case the level shift is simply determined by the form of the JDOS. For a metal, the JDOS raises linearly at $\omega=0$ with a slope given by the metal's DOS at E_F. This suggests that the level shift should increase with the substrate DOS at the Fermi level. For a semiconductor, the JDOS raises smoothly at $\omega=E_{gap}$, suggesting that the level shift should decrease with E_{gap}. In the following section we investigate these relations for the model and first-principles calculations. We mention that the second-order approximation discussed above may not always provide a good description of the full *GW* self-energy. However, as we will show in the next section, it explains qualitatively the trends in G_0W_0 calculations.

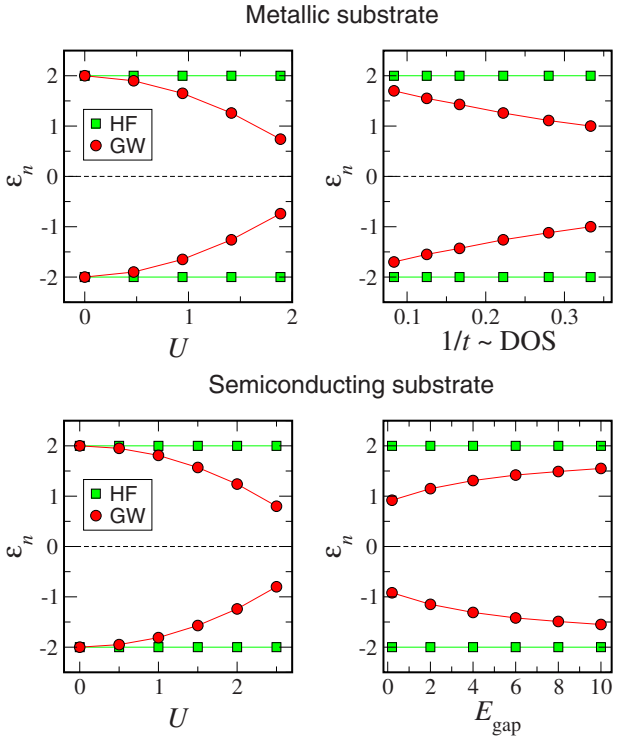


FIG. 5. (Color online) HOMO and LUMO positions obtained from the simple lattice models for a metallic substrate (upper panel) and semiconducting substrate (lower panel). In all plots we vary one parameter while keeping the remaining parameters fixed.[34]

B. Dependence of level shift on surface electronic structure

In Fig. 5 we show the HOMO and LUMO levels of the lattice models calculated with the HF and *GW* approximations. In all plots we vary one parameter of the model while keeping the remaining parameters fixed.[34]

The upper panels refer to a metallic substrate and show the dependence on the levels on the interaction strength U and the intrachain hopping parameter t. Note that the latter is inversely proportional to the projected density of states (DOS) of the terminal site evaluated at E_F. The lower panels refer to a semiconducting substrate and show the dependence of the levels on U and the substrate gap, E_{gap}. The HF eigenvalues are clearly independent of the nonlocal interaction between the molecule and substrate. This can be understood from Koopmans' theorem which states that the HF eigenvalues do not include the electronic relaxations of the substrate induced by the extra electron/hole in the molecule. In contrast the *GW* levels vary in the way predicted by the simple model discussed in the previous section: The polarization-

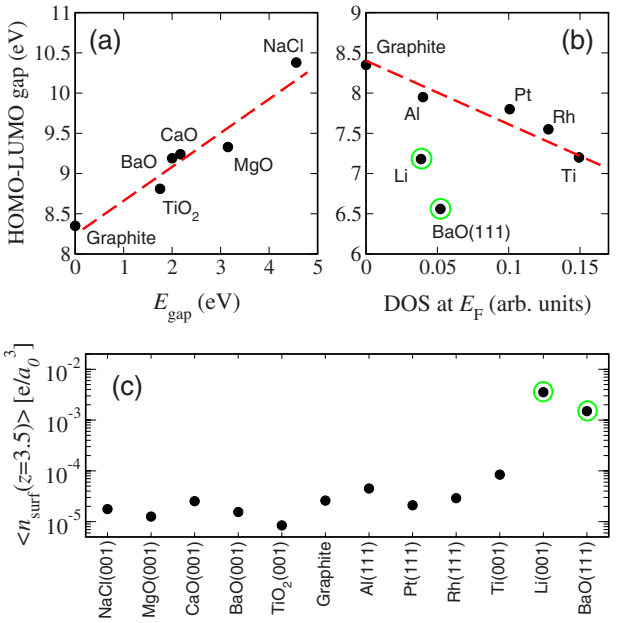


FIG. 6. (Color online) Calculated HOMO-LUMO gap of benzene at z=4.5 Å (same numbers as in Fig. 1) plotted as function of the LDA substrate band gap for the semiconductors (a) and the total DOS per volume evaluated at the Fermi level for the metals (b). Dashed lines have been added to guide the eye. (c) Average electron density in a plane lying z=3.5 Å above the clean surfaces.

induced reduction of the HOMO-LUMO gap is stronger for larger U as well as for larger substrate DOS at E_F for the metals and smaller substrate band gap for the semiconductors. A more detailed discussion of level renormalization based on the lattice model for metallic substrates, including the case of strong metal-molecule hybridization, can be found in Ref. 15.

In Figs. 6(a) and 6(b) we plot the G_0W_0 gaps from Fig. 1 versus the LDA band gap and DOS at E_F for the semiconducting and metallic substrates, respectively. For the semiconductors the reduction of the HOMO-LUMO gap clearly correlates with E_{gap}. This indicates that the interaction strength, i.e., the matrix elements $V_{kk',aa}$ of Eq. (14), do not differ too much from one surface to another. For the metals, the HOMO-LUMO gap seems to scale with the metal's DOS at E_F. However, we note that Li(001) and BaO(111) deviate from the general trend followed by the other metals. This can be explained by the larger extend of the metallic wave functions of these systems into the vacuum region, which in turn leads to larger $V_{kk',aa}$ matrix elements. Indeed, Fig. 6(c) shows the average electron density evaluated in a plane lying z=3.5 Å above the surface in the absence of the benzene molecule. The density outside the Li(001) and BaO(111) surfaces is significantly larger than for the other surfaces which on the other hand have quite similar densities.

C. DFT eigenvalues and error cancellation

In Fig. 7 we plot the energies of the HOMO and LUMO levels of benzene at z=4.5 Å. For each surface, we have shifted the LDA, PBE0, and G_0W_0 levels by the same amount so that the LDA HOMO is aligned with the HOMO in the gas phase. We note that the effect of substrate polarization is very similar for the G_0W_0 HOMO and LUMO levels which are shifted up and down, respectively, by almost the same amount. This is indeed expected from the classical image charge model. Significant deviations from this trend are, however, seen for Li(001) and BaO(111). We ascribe this to the more extended nature of the metallic states on these surfaces which reduce the validity of the point charge approximation and can introduce differences between the $V_{kk',HH}$ and $V_{kk',LL}$ matrix elements.

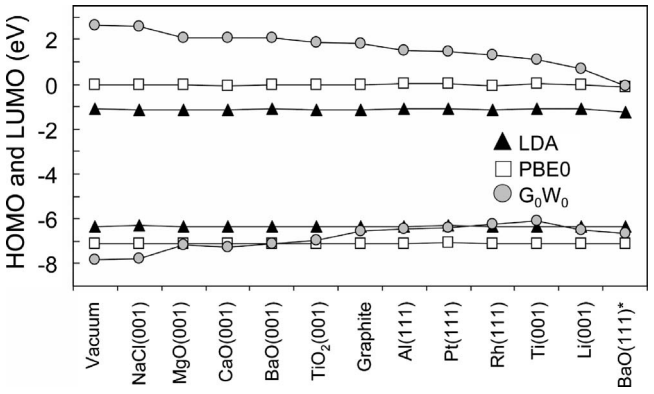


FIG. 7. LDA, PBE0, and G_0W_0 energies for the HOMO and LUMO levels of benzene at z=4.5 Å above the surfaces. The very good agreement between LDA and G_0W_0 energies for the HOMO level at the metallic surfaces is due to error cancellation in the LDA approximation.

Overall, the LDA and PBE0 eigenvalues for the HOMO are in better agreement with the G_0W_0 QP energies than is the case for the LUMO. Moreover there is a general trend that the LDA eigenvalues come closer to the G_0W_0 energies as we move from the insulating to the metallic surfaces. In fact, the LDA HOMO level is almost on top of the G_0W_0 level on the metallic surfaces. This trend is clearly a result of significant error cancellation in the LDA. Indeed, it is well known that semilocal exchange-correlation functionals overestimate (underestimate) occupied (empty) molecular levels due to self-interaction effects. At the metallic surfaces this error is compensated by the missing image charge correction. PBE0 gives better estimates for the free molecule where it opens up the LDA HOMO-LUMO gap due to partial removal of self-interaction errors. In this case, the cancellation between the missing image charge effect and the remaining self-interaction error results in very good agreement between PBE0 and G_0W_0 for the HOMO level on the semiconducting surfaces.

The cancellation between self-interaction errors and missing polarization effects will always be present in hybrid- and semilocal approximations. However, the relative size of the two contributions will in general depend on the shape of the molecule, its orientation with respect to the surface, the molecule-surface distance, and the type of substrate.

V. CONCLUSIONS

We have presented G_0W_0 calculations for a benzene molecule physisorbed on different metallic and semiconducting

surfaces. Upon physisorption the molecule's HOMO-LUMO gap is reduced from its gas phase value due to dynamic polarization of the substrate. It was shown that a classical image charge model captures the qualitative features of the effect while the magnitude of the level shift is sensitive to the detailed atomic structure of the surface. In particular the presence of metallic midgap state at the surface of a semiconductor can have a large influence on the local response of the surface. Both local and hybrid exchange-correlation potentials fail to account for the polarization effects yielding Kohn-Sham eigenvalues of physisorbed benzene which are independent of the substrate. Nevertheless we found that a cancellation between self-interaction errors and missing image charge effects in the LDA leads to a very good agreement between LDA and G_0W_0 energies for the occupied states of benzene on metallic surfaces. Similar conclusions were reached for the PBE0 energies on semiconducting substrates. Finally, we have derived a simple second-order approximation to the *GW* self-energy which expresses the polarization-induced shift of a molecular level in terms of the substrate's joint density of states weighted by Coulomb interaction matrix elements. This model was used to explain general trends in the first-principles results, namely the scaling of the benzene's HOMO-LUMO gap with the substrate density of states at E_F (for metals) and the substrate band gap (for semiconductors).

Our results clearly demonstrate the importance of nonlocal correlations for the electronic levels at solid-molecule interfaces. We expect this to have important implications for the theoretical modeling of electron transport in organic and single-molecule devices.

ACKNOWLEDGMENTS

K.S.T. and C.R. acknowledge support from the Danish Center for Scientific Computing. The Center for Atomic-scale Materials Design (CAMD) is sponsored by the Lundbeck Foundation. A.R. and J.M.G.L. acknowledge funding by the Spanish MEC (Grant No. FIS2007-65702-C02-01), "Grupos Consolidados UPV/EHU del Gobierno Vasco" (Grant No. IT-319-07), e-I3 ETSF project (Contract No. 211956) and "Red Española de Supercomputación."

[1] J. K. Nørskov, T. Bligaard, J. Rossmeisl, and C. H. Christensen, Nat. Chem. **1**, 37 (2009).
[2] K. Moth-Poulsen and T. Bjornholm, Nat. Nanotechnol. **4**, 551 (2009).
[3] A. Nitzan and M. A. Ratner, Science **300**, 1384 (2003).
[4] W. A. Hofer, A. S. Foster, and A. L. Shluger, Rev. Mod. Phys. **75**, 1287 (2003).
[5] F. Flores, J. Ortega, and H. Vazquez, Phys. Chem. Chem. Phys. **11**, 8658 (2009).
[6] P. D. Johnson and S. L. Hulbert, Phys. Rev. B **35**, 9427 (1987).
[7] S. Kubatkin,Andrey Danilov, Mattias Hjort, Jérôme Cornil, Jean-Luc Brédas, Nicolai Stuhr-Hansen, Per Hedegård, and Thomas Bjørnholm, Nature (London) **425**, 698 (2003).
[8] R. Hesper, L. H. Tjeng, and G. A. Sawatzky, Europhys. Lett. **40**, 177 (1997).
[9] J. Repp, G. Meyer, S. M. Stojković, A. Gourdon, and C. Joachim, Phys. Rev. Lett. **94**, 026803 (2005).
[10] A. Kahn, N. Koch, and W. Gao, J. Polym. Sci., B, Polym. Phys. **41**, 2529 (2003).
[11] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).
[12] G. Onida, L. Reining, and A. Rubio, Rev. Mod. Phys. **74**, 601 (2002).
[13] J. B. Neaton, Mark S. Hybertsen, and Steven G. Louie, Phys. Rev. Lett. **97**, 216405 (2006).
[14] C. Freysoldt, P. Rinke, and M. Scheffler, Phys. Rev. Lett. **103**, 056803 (2009).
[15] K. S. Thygesen and A. Rubio, Phys. Rev. Lett. **102**, 046802 (2009).
[16] S. Y. Quek, Latha Venkataraman, Hyoung Joon Choi, Steven G. Louie, Mark S. Hybertsen, and J. B. Neaton, Nano Lett. **7**, 3477 (2007).
[17] D. J. Mowbray, G. Jones, and K. S. Thygesen, J. Chem. Phys. **128**, 111103 (2008).
[18] K. Kaasbjerg and K. Flensberg, Nano Lett. **8**, 3809 (2008).
[19] R. Stadler, V. Geskin, and J. Cornil, Phys. Rev. B **79**, 113408 (2009).
[20] S. Baroni *et al.*, QUANTUM-ESPRESSO package, 2005 (http://www.quantum-espresso.org/).
[21] M. Fuchs and M. Scheffler, Comput. Phys. Commun. **119**, 67 (1999).
[22] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[23] M. Ernzerhof and G. Scuseria, J. Chem. Phys. **110**, 5029 (1999).
[24] C. Adamo and V. Barone, J. Chem. Phys. **110**, 6158 (1999).
[25] A. Marini, C. Hogan, M. Grüning, and D. Varsano, Comput. Phys. Commun. **180**, 1392 (2009).
[26] K. S. Thygesen and A. Rubio, Phys. Rev. B **77**, 115333 (2008).
[27] M. Rohlfing, N.-P. Wang, P. Krüger, and J. Pollmann, Phys. Rev. Lett. **91**, 256802 (2003).
[28] N. Lang and W. Kohn, Phys. Rev. B **7**, 3541 (1973).
[29] A. G. Eguiluz, M. Heinrichsmeier, A. Fleszar, and W. Hanke, Phys. Rev. Lett. **68**, 1359 (1992).
[30] I. D. White, R. W. Godby, M. M. Rieger, and R. J. Needs, Phys. Rev. Lett. **80**, 4265 (1998).
[31] J. C. Inkson, J. Phys. C **6**, 1350 (1973).
[32] J. P. A. Charlesworth, R. W. Godby, and R. J. Needs, Phys. Rev. Lett. **70**, 1685 (1993).
[33] www.oxmat.co.uk/crysdata/homepage.htm.
[34] For the metals: $t=5$, $U=2$, $\xi_{H/L}=\pm 2-U$. For the semiconductor: $t=2$, $U=2$, $\xi_{H/L}=\pm 2$, and $\varepsilon_{c/v}=\pm 5.5$.

Paper IV

The projector Augmented-wave Method
C. Rostgaard
arXiv:0910.1921 [cond-mat.mtrl-sci] (2009).

The Projector Augmented-wave Method

Carsten Rostgaard

October 10, 2009

Abstract

The purpose of this text is to give a self-contained description of the basic theory of the projector augmented-wave (PAW) method, as well as most of the details required to make the method work in practice. These two topics are covered in the first two sections, while the last is dedicated to examples of how to apply the PAW transformation when extracting non-standard quantities from a density-functional theory (DFT) calculation.

The formalism is based on Blöchl's original formulation of PAW [1], and the notation and extensions follow those used and implemented in the GPAW[2] code.

Contents

1 Formalism

By the requirement of orthogonality, DFT wave functions have very sharp features close to the nuclei, as all the states are non-zero in this region. Further out only the valence states are non-zero, resulting in much smoother wave functions in this interstitial region. The oscillatory behavior in the core regions, requires a very large set of plane waves, or equivalently a very fine grid, to be described correctly. One way of solving this problem is the use of pseudopotentials in which the collective system of nuclei and core electrons are described by an effective, much smoother, potential. The KS equations are then solved for the valence electrons only. The pseudopotentials are constructed such that the correct scattering potential is obtained beyond a certain radius from the core. This method reduces the number of wave functions to be calculated, since the pseudo potentials only have to be calculated and tabulated once for each type of atom, so that only calculations on the valence states are needed. It justifies the neglect of relativistic effects in the KS equations, since the valence electrons are non-relativistic (the pseudopotentials describing core states are of course constructed with full consideration of relativistic effects). The technique also removes the unwanted singular behavior of the ionic potential at the lattice points.

The drawback of the method is that all information on the full wave function close to the nuclei is lost. This can influence the calculation of certain properties, such as hyperfine parameters, and electric field gradients. Another problem is that one has no before hand knowledge of when the approximation yields reliable results.

A different approach is the augmented-plane-wave method (APW), in which space is divided into atom-centered augmentation spheres inside which the wave functions are taken as some atom-like partial waves, and a bonding region outside the spheres, where some envelope functions are defined. The partial waves and envelope functions are then matched at the boundaries of the spheres.

A more general approach is the projector augmented wave method (PAW) presented here, which offers APW as a special case[3], and the pseudopotential method as a well defined approximation[4]. The PAW method was first proposed by Blöchl in 1994[1].

1.1 The Transformation Operator

The features of the wave functions, are very different in different regions of space. In the bonding region it is smooth, but near the nuclei it displays rapid oscillations, which are very demanding on the numerical representation of the wave functions. To address this problem, we seek a linear transformation $\hat{\mathcal{T}}$ which takes us from an auxiliary smooth wave function $|\tilde{\psi}_n\rangle$ to the true all electron Kohn-Sham single particle wave function $|\psi_n\rangle$

$$|\psi_n\rangle = \hat{\mathcal{T}}|\tilde{\psi}_n\rangle, \tag{1}$$

where n is the quantum state label, containing a $\mathbf{k}$ index, a band index, and a spin index.

This transformation yields the transformed KS equations

$$\hat{\mathcal{T}}^\dagger \widehat{H} \hat{\mathcal{T}} |\tilde{\psi}_n\rangle = \epsilon_n \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}} |\tilde{\psi}_n\rangle, \tag{2}$$

which needs to be solved instead of the usual KS equation. Now we need to define $\hat{\mathcal{T}}$ in a suitable way, such that the auxiliary wave functions obtained from solving (2) becomes smooth.

Since the true wave functions are already smooth at a certain minimum distance from the core, $\hat{\mathcal{T}}$ should only modify the wave function close to the nuclei. We thus define

$$\hat{\mathcal{T}} = 1 + \sum_a \hat{\mathcal{T}}^a, \tag{3}$$

where a is an atom index, and the atom-centered transformation, $\hat{\mathcal{T}}^a$, has no effect outside a certain atom-specific augmentation region $|\mathbf{r} - \mathbf{R}^a| < r_c^a$. The cut-off radii, r_c^a should be chosen such that there is no overlap of the augmentation spheres.

Inside the augmentation spheres, we expand the true wave function in the partial waves ϕ_i^a, and for each of these partial waves, we define a corresponding auxiliary smooth partial wave $\tilde{\phi}_i^a$, and require that

$$|\phi_i^a\rangle = (1 + \hat{\mathcal{T}}^a)|\tilde{\phi}_i^a\rangle \quad \Leftrightarrow \quad \hat{\mathcal{T}}^a|\tilde{\phi}_i^a\rangle = |\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle \tag{4}$$

for all i, a. This completely defines $\hat{\mathcal{T}}$, given ϕ and $\tilde{\phi}$.

Since $\hat{\mathcal{T}}^a$ should do nothing outside the augmentation sphere, we see from (4) that we must require the partial wave and its smooth counterpart to be identical outside the augmentation sphere

$$\forall a, \quad \phi_i^a(\mathbf{r}) = \tilde{\phi}_i^a(\mathbf{r}), \text{ for } r > r_c^a$$

where $\phi_i^a(\mathbf{r}) = \langle \mathbf{r}|\phi_i^a\rangle$ and likewise for $\tilde{\phi}_i^a$.

If the smooth partial waves form a complete set inside the augmentation sphere, we can formally expand the smooth all electron wave functions as

$$|\tilde{\psi}_n\rangle = \sum_i P_{ni}^a |\tilde{\phi}_i^a\rangle, \text{ for } |\mathbf{r} - \mathbf{R}^a| < r_c^a \tag{5}$$

where P_{ni}^a are some, for now, undetermined expansion coefficients.

Since $|\phi_i^a\rangle = \hat{\mathcal{T}}|\tilde{\phi}_i^a\rangle$ we see that the expansion

$$|\psi_n\rangle = \hat{\mathcal{T}}|\tilde{\psi}_n\rangle = \sum_i P_{ni}^a |\phi_i^a\rangle, \text{ for } |\mathbf{r} - \mathbf{R}^a| < r_c^a \tag{6}$$

has identical expansion coefficients, P_{ni}^a.

As we require $\hat{\mathcal{T}}$ to be linear, the coefficients P_{ni}^a must be linear functionals of $|\tilde{\psi}_n\rangle$, i.e.

$$P_{ni}^a = \langle \tilde{p}_i^a|\tilde{\psi}_n\rangle = \int d\mathbf{r} \tilde{p}_i^{a*}(\mathbf{r} - \mathbf{R}^a)\tilde{\psi}_n(\mathbf{r}), \tag{7}$$

where $|\tilde{p}_i^a\rangle$ are some fixed functions termed smooth projector functions.

As there is no overlap between the augmentation spheres, we expect the one center expansion of the smooth all electron wave function, $|\tilde{\psi}_n^a\rangle = \sum_i |\tilde{\phi}_i^a\rangle\langle\tilde{p}_i^a|\tilde{\psi}_n\rangle$ to reduce to $|\tilde{\psi}_n\rangle$ itself inside the augmentation sphere defined by a. Thus, the smooth projector functions must satisfy

$$\sum_i |\tilde{\phi}_i^a\rangle\langle\tilde{p}_i^a| = 1 \tag{8}$$

inside each augmentation sphere. This also implies that

$$\langle \tilde{p}_{i_1}^a|\tilde{\phi}_{i_2}^a\rangle = \delta_{i_1,i_2} \text{ , for } |\mathbf{r} - \mathbf{R}^a| < r_c^a \tag{9}$$

i.e. the projector functions should be orthonormal to the smooth partial waves inside the augmentation sphere. There are no restrictions on $\tilde{p}_i^a$ outside the augmentation spheres, so for convenience we might as well define them as local functions, i.e. $\tilde{p}_i^a(\mathbf{r}) = 0$ for $r > r_c^a$.

Note that the completeness relation (8) is equivalent to the requirement that $\tilde{p}_i^a$ should produce the correct expansion coefficients of (5)-(6), while (9) is merely an implication of this restriction. Translating (8) to an explicit restriction on the projector functions is a rather involved procedure, but according to Blöchl, [1], the most general form of the projector functions is:

$$\langle \tilde{p}_i^a| = \sum_j (\{\langle f_k^a|\tilde{\phi}_l^a\rangle\})_{ij}^{-1}\langle f_j^a|, \tag{10}$$

where $|f_j^a\rangle$ is any set of linearly independent functions. The projector functions will be localized if the functions $|f_j^a\rangle$ are localized.

Using the completeness relation (8), we see that

$$\hat{\mathcal{T}}^a = \sum_i \hat{\mathcal{T}}^a|\tilde{\phi}_i^a\rangle\langle\tilde{p}_i^a| = \sum_i \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle\right)\langle\tilde{p}_i^a|,$$

where the first equality is true in all of space, since (8) holds inside the augmentation spheres and outside $\hat{\mathcal{T}}^a$ is zero, so anything can be multiplied with it. The second equality is due to (4) (remember that $|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle = 0$ outside the augmentation sphere). Thus we conclude that

$$\hat{\mathcal{T}} = 1 + \sum_a \sum_i \left(|\phi_i^a\rangle - |\tilde{\phi}_i^a\rangle\right)\langle \tilde{p}_i^a|. \tag{11}$$

To summarize, we obtain the all electron KS wave function $\psi_n(\mathbf{r}) = \langle \mathbf{r}|\psi_n\rangle$ from the transformation

$$\psi_n(\mathbf{r}) = \tilde{\psi}_n(\mathbf{r}) + \sum_a \sum_i \left(\phi_i^a(\mathbf{r}) - \tilde{\phi}_i^a(\mathbf{r})\right)\langle \tilde{p}_i^a|\tilde{\psi}_n\rangle, \tag{12}$$

where the smooth (and thereby numerically convenient) auxiliary wave function $\tilde{\psi}_n(\mathbf{r})$ is obtained by solving the eigenvalue equation (2).

The transformation (12) is expressed in terms of the three components: a) the partial waves $\phi_i^a(\mathbf{r})$, b) the smooth partial waves $\tilde{\phi}_i^a(\mathbf{r})$, and c) the smooth projector functions $\tilde{p}_i^a(\mathbf{r})$.

The restriction on the choice of these sets of functions are: a) Since the partial- and smooth partial wave functions are used to expand the all electron wave functions, i.e. are used as atom-specific basis sets, these must be complete (inside the augmentation spheres). b) the smooth projector functions must satisfy (8), i.e. be constructed according to (10). All remaining degrees of freedom are used to make the expansions converge as fast as possible, and to make the functions termed 'smooth', as smooth as possible. For a specific choice of these sets of functions, see section 2.2. As the partial- and smooth partial waves are merely used as basis sets they can be chosen as real functions (any imaginary parts of the functions they expand, are then introduced through the complex expansion coefficients P_{ni}^a). In the remainder of this document ϕ and $\tilde{\phi}$ will be assumed real.

Note that the sets of functions needed to define the transformation are system independent, and as such they can conveniently be pre-calculated and tabulated for each element of the periodic table.

For future convenience, we also define the one center expansions

$$\psi_n^a(\mathbf{r}) = \sum_i \phi_i^a(\mathbf{r})\langle \tilde{p}_i^a|\tilde{\psi}_n\rangle, \tag{13a}$$

$$\tilde{\psi}_n^a(\mathbf{r}) = \sum_i \tilde{\phi}_i^a(\mathbf{r})\langle \tilde{p}_i^a|\tilde{\psi}_n\rangle. \tag{13b}$$

In terms of these, the all electron KS wave function is

$$\psi_n(\mathbf{r}) = \tilde{\psi}_n(\mathbf{r}) + \sum_a \left(\psi_n^a(\mathbf{r} - \mathbf{R}^a) - \tilde{\psi}_n^a(\mathbf{r} - \mathbf{R}^a)\right). \tag{14}$$

So what have we achieved by this transformation? The trouble of the original KS wave functions, was that they displayed rapid oscillations in some parts of space, and smooth behavior in other parts of space. By the decomposition (12) we have separated the original wave functions into auxiliary wave functions which are smooth everywhere and a contribution which contains rapid oscillations, but only contributes in certain, small, areas of space. This decomposition is shown on the front page for the hydrogen molecule. Having separated the different types of waves, these can be treated individually. The localized atom centered parts, are indicated by a superscript a, and can efficiently be represented on atom centered radial grids. Smooth functions are indicated by a tilde $\tilde{}$. The delocalized parts (no superscript a) are all smooth, and can thus be represented on coarse Fourier- or real space grids.

1.2 The Frozen Core Approximation

In the frozen core approximation, it is assumed that the core states are naturally localized within the augmentation spheres, and that the core states of the isolated atoms are not changed by the formation of molecules or solids. Thus the core KS states are identical to the atomic core states

$$|\psi_n^c\rangle = |\phi_\alpha^{a,\text{core}}\rangle,$$

where the index n on the left hand site refers to both a specific atom, a, and an atomic state, α.

In this approximation only valence states are included in the expansions of $|\psi_n\rangle$, (6), and $|\tilde{\psi}_n\rangle$, (5).

Figure 1, shows the atomic states of Platinum in its ground state, obtained with an atomic DFT program at an LDA level, using spherical averaging, i.e. a spin-compensated calculation, assuming the degenerate occupation 9/10 of all 5d states, and both of the 6s states half filled. It is seen that at the typical length of atomic interaction (the indicated cut-off $r_c = 2.5$ Bohr is approximately half the inter-atomic distance in bulk Pt), only the 5d and 6s states are non-zero.

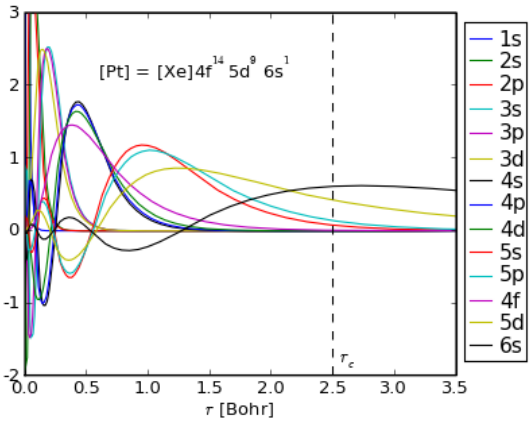


Figure 1: *The core states of Platinum*

1.3 Expectation Values

The expectation value of an operator $\widehat{O}$ is, within the frozen core approximation, given by

$$\langle \widehat{O} \rangle = \sum_n^{\text{val}} f_n \langle \psi_n | \widehat{O} | \psi_n \rangle + \sum_a \sum_\alpha^{\text{core}} \langle \phi_\alpha^{a,\text{core}} | \widehat{O} | \phi_\alpha^{a,\text{core}} \rangle. \tag{15}$$

Using that $\langle \psi_n | \widehat{O} | \psi_n \rangle = \langle \tilde{\psi}_n | \hat{\mathcal{T}}^\dagger \widehat{O} \hat{\mathcal{T}} | \tilde{\psi}_n \rangle$, and skipping the state index for notational convenience, we see that

$$\begin{aligned}
\langle \psi | \widehat{O} | \psi \rangle &= \langle \tilde{\psi} + \sum_a (\psi^a - \tilde{\psi}^a) | \widehat{O} | \tilde{\psi} + \sum_a (\psi^a - \tilde{\psi}^a) \rangle \\
&= \langle \tilde{\psi} | \widehat{O} | \tilde{\psi} \rangle + \sum_{aa'} \langle \psi^a - \tilde{\psi}^a | \widehat{O} | \psi^{a'} - \tilde{\psi}^{a'} \rangle + \sum_a \Big(\langle \tilde{\psi} | \widehat{O} | \psi^a - \tilde{\psi}^a \rangle + \langle \psi^a - \tilde{\psi}^a | \widehat{O} | \tilde{\psi} \rangle \Big) \\
&= \langle \tilde{\psi} | \widehat{O} | \tilde{\psi} \rangle + \sum_a \Big(\langle \psi^a | \widehat{O} | \psi^a \rangle - \langle \tilde{\psi}^a | \widehat{O} | \tilde{\psi}^a \rangle \Big) \\
&\quad + \sum_a \Big(\langle \psi^a - \tilde{\psi}^a | \widehat{O} | \tilde{\psi} - \tilde{\psi}^a \rangle + \langle \tilde{\psi} - \tilde{\psi}^a | \widehat{O} | \psi^a - \tilde{\psi}^a \rangle \Big) \\
&\quad + \sum_{a \neq a'} \langle \psi^a - \tilde{\psi}^a | \widehat{O} | \psi^{a'} - \tilde{\psi}^{a'} \rangle.
\end{aligned} \tag{16}$$

For local operators[1] the last two lines does not contribute. The first line, because $|\psi^a - \tilde{\psi}^a\rangle$ is only non-zero inside the spheres, while $|\tilde{\psi} - \tilde{\psi}^a\rangle$ is only non-zero outside the spheres. The second line simply because $|\psi^a - \tilde{\psi}^a\rangle$ is zero outside the spheres, so two such states centered on different nuclei have no overlap (provided that the augmentation spheres do not overlap).

Reintroducing the partial waves in the one-center expansions, we see that

$$\sum_n^{\text{val}} f_n \langle \psi_n^a | \widehat{O} | \psi_n^a \rangle = \sum_n^{\text{val}} f_n \sum_{i_1 i_2} \langle \phi_{i_1}^a P_{ni_1}^a | \widehat{O} | \phi_{i_2}^a P_{ni_2}^a \rangle = \sum_{i_1 i_2} \langle \phi_{i_1}^a | \widehat{O} | \phi_{i_2}^a \rangle \sum_n^{\text{val}} f_n P_{ni_1}^{a*} P_{ni_2}^a, \tag{17}$$

and likewise for the smooth waves.

Introducing the Hermitian one-center density matrix

$$D_{i_1 i_2}^a = \sum_n f_n P_{ni_1}^{a*} P_{ni_2}^a = \sum_n f_n \langle \tilde{\psi}_n | \tilde{p}_{i_1}^a \rangle \langle \tilde{p}_{i_2}^a | \tilde{\psi}_n \rangle. \tag{18}$$

[1] Local operator $\widehat{O}$: An operator which does not correlate separate parts of space, i.e. $\langle \mathbf{r} | \widehat{O} | \mathbf{r}' \rangle = 0$ if $\mathbf{r} \neq \mathbf{r}'$.

We conclude that for any local operator $\widehat{O}$, the expectation value is

$$\langle \widehat{O} \rangle = \sum_n^{\text{val}} f_n \langle \tilde{\psi}_n | \widehat{O} | \tilde{\psi}_n \rangle + \sum_a \sum_{i_1 i_2} \left(\langle \phi^a_{i_1} | \widehat{O} | \phi^a_{i_2} \rangle - \langle \tilde{\phi}^a_{i_1} | \widehat{O} | \tilde{\phi}^a_{i_2} \rangle \right) D^a_{i_1 i_2} + \sum_a \sum_\alpha^{\text{core}} \langle \phi^{a,\text{core}}_\alpha | \widehat{O} | \phi^{a,\text{core}}_\alpha \rangle. \tag{19}$$

1.4 Densities

The electron density is obviously a very important quantity in DFT, as all observables in principle are calculated as functionals of the density. In reality the kinetic energy is calculated as a functional of the orbitals, and some specific exchange-correlation functionals also rely on KS-orbitals rather then the density for their evaluation, but these are still *implicit* functionals of the density.

To obtain the electron density we need to determine the expectation value of the real-space projection operator $|\mathbf{r}\rangle\langle\mathbf{r}|$

$$n(\mathbf{r}) = \sum_n f_n \langle \psi_n | \mathbf{r} \rangle \langle \mathbf{r} | \psi_n \rangle = \sum_n f_n |\psi_n(\mathbf{r})|^2, \tag{20}$$

where f_n are the occupation numbers.

As the real-space projection operator is obviously a local operator, we can use the results (19) of the previous section, and immediately arrive at

$$n(\mathbf{r}) = \sum_n^{\text{val}} f_n |\tilde{\psi}_n|^2 + \sum_a \sum_{i_1 i_2} \left(\phi^a_{i_1} \phi^a_{i_2} - \tilde{\phi}^a_{i_1} \tilde{\phi}^a_{i_2} \right) D^a_{i_1 i_2} + \sum_a \sum_\alpha^{\text{core}} |\phi^{a,\text{core}}_\alpha|^2. \tag{21}$$

To ensure that (21) reproduce the correct density even though some of the core states are not strictly localized within the augmentation spheres, a smooth core density, $\tilde{n}_c(\mathbf{r})$, is usually constructed, which is identical to the core density outside the augmentation sphere, and a smooth continuation inside. Thus the density is typically evaluated as

$$n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + \sum_a \left(n^a(\mathbf{r}) - \tilde{n}^a(\mathbf{r}) \right), \tag{22}$$

where

$$\tilde{n}(\mathbf{r}) = \sum_n^{\text{val}} f_n |\tilde{\psi}_n(\mathbf{r})|^2 + \tilde{n}_c(\mathbf{r}) \tag{23a}$$

$$n^a(\mathbf{r}) = \sum_{i_1 i_2} D^a_{i_1 i_2} \phi^a_{i_1}(\mathbf{r}) \phi^a_{i_2}(\mathbf{r}) + n^a_c(\mathbf{r}) \tag{23b}$$

$$\tilde{n}^a(\mathbf{r}) = \sum_{i_1 i_2} D^a_{i_1 i_2} \tilde{\phi}^a_{i_1}(\mathbf{r}) \tilde{\phi}^a_{i_2}(\mathbf{r}) + \tilde{n}^a_c(\mathbf{r}) \tag{23c}$$

1.5 Total Energies

The total energy of the electronic system is given by:

$$E[n] = T_s[n] + U_H[n] + V_{ext}[n] + E_{xc}[n]. \tag{24}$$

In this section, the usual energy expression above, is sought re-expressed in terms of the PAW quantities: the smooth waves and the auxiliary partial waves.

For the local and semi-local functionals, we can utilize (19), while the nonlocal parts needs more careful consideration.

1.5.1 The Semi-local Contributions

The kinetic energy functional $T_s = \sum_n f_n \langle\psi_n| - \frac{1}{2}\nabla^2|\psi_n\rangle$ is obviously a (semi-) local functional, so we can apply (19) and immediately arrive at:

$$\begin{aligned} T_s[\{\psi_n\}] &= \sum_n f_n \langle\psi_n| - \tfrac{1}{2}\nabla^2|\psi_n\rangle \\ &= \sum_n^{\text{val}} f_n \langle\tilde{\psi}_n| - \tfrac{1}{2}\nabla^2|\tilde{\psi}_n\rangle + \sum_a \left(T_c^a + \Delta T^a_{i_1 i_2} D^a_{i_1 i_2}\right), \end{aligned} \tag{25}$$

where

$$T_c^a = \sum_\alpha^{\text{core}} \langle\phi_\alpha^{a,\text{core}}| - \tfrac{1}{2}\nabla^2|\phi_\alpha^{a,\text{core}}\rangle \quad \text{and} \quad \Delta T^a_{i_1 i_2} = \langle\phi^a_{i_1}| - \tfrac{1}{2}\nabla^2|\phi^a_{i_2}\rangle - \langle\tilde{\phi}^a_{i_1}| - \tfrac{1}{2}\nabla^2|\tilde{\phi}^a_{i_2}\rangle. \tag{26}$$

For LDA and GGA type exchange-correlation functionals, E_{xc} is likewise, per definition, a semi-local functional, such that it can be expressed as

$$E_{xc}[n] = E_{xc}[\tilde{n}] + \sum_a \left(E_{xc}[n^a] - E_{xc}[\tilde{n}^a]\right). \tag{27}$$

By virtue of (23b)-(23c) we can consider the atomic corrections as functionals of the density matrix defined in (18), i.e.

$$E_{xc}[n] = E_{xc}[\tilde{n}] + \sum_a \Delta E^a_{xc}[\{D^a_{i_1 i_2}\}], \tag{28}$$

where

$$\Delta E^a_{xc}[\{D^a_{i_1 i_2}\}] = E_{xc}[n^a] - E_{xc}[\tilde{n}^a]. \tag{29}$$

1.5.2 The Nonlocal Contributions

The Hartree term is both nonlinear and nonlocal, so more care needs to be taken when introducing the PAW transformation for this expression.

In the following we will assume that there is no 'true' external field, such that $V_{\text{ext}}[n]$ is only due to the static nuclei, i.e. it is a sum of the classical interaction of the electron density with the static ionic potential, and the electrostatic energy of the nuclei.

We define the total classical electrostatic energy functional as

$$E_C[n] = U_H[n] + V_{\text{ext}}[n] = \frac{1}{2}((n)) + (n|\textstyle\sum_a Z^a) + \frac{1}{2}\sum_{a\neq a'}(Z^a|Z^{a'}), \tag{30}$$

where the notation (f|g) indicates the Coulomb integral

$$(f|g) = \iint d\mathbf{r} d\mathbf{r}' \frac{f^*(\mathbf{r}) g(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \tag{31}$$

and I have introduced the short hand notation $((f)) = (f|f)$. In (30), $Z^a(\mathbf{r})$ is the charge density of the nucleus at atomic site a, which in the classical point approximation is given by

$$Z^a(\mathbf{r}) = -\mathcal{Z}^a \delta(\mathbf{r} - \mathbf{R}^a) \tag{32}$$

with $\mathcal{Z}^a$ being the atomic number of the nuclei. As the Hartree energy of a density with non-zero total charge is numerically inconvenient, we introduce the charge neutral total density

$$\rho(\mathbf{r}) = n(\mathbf{r}) + \sum_a Z^a(\mathbf{r}) \quad (= n_{\text{electrons}} + n_{\text{nuclei}}). \tag{33}$$

In terms of this, the coulombic energy of the system can be expressed by

$$E_C[n] = U'_H[\rho] = \frac{1}{2}((n + \textstyle\sum_a Z^a))' \tag{34}$$

where the prime indicates that one should remember the self-interaction error of the nuclei introduced in the Hartree energy of the total density. This correction is obviously ill defined, and different schemes exist for making this correction. As it turns out, this correction is handled very naturally in the PAW formalism.

For now, we will focus on the term $((\rho)) = ((n + \sum_a Z^a))$. If one where to directly include the expansion of $n(\mathbf{r})$ according to (22), one would get:

$$\begin{aligned}((n + \textstyle\sum_a Z^a)) &= ((\tilde{n} + \textstyle\sum_a n^a - \tilde{n}^a + Z^a)) \\ &= ((\tilde{n})) + \sum_{aa'}(n^a - \tilde{n}^a + Z^a | n^{a'} - \tilde{n}^{a'} + Z^{a'}) + 2\sum_a (\tilde{n} | n^a - \tilde{n}^a + Z^a),\end{aligned}$$

where in the last expression, the first term is the Hartree energy of the smooth electron density, which is numerically problematic because of the nonzero total charge. The second term contains a double summation over all nuclei, which would scale badly with system size, and the last term involves integrations of densities represented on incompatible grids (remember that the one-center densities are represented on radial grids to capture the oscillatory behavior near the nuclei)[2]. This is clearly not a feasible procedure. To correct these problem we add and subtract some atom centered compensation charges $\tilde{Z}^a$:

$$\begin{aligned}((n + \textstyle\sum_a \tilde{Z}^a + \sum_a \left[Z^a - \tilde{Z}^a\right])) = ((\tilde{n} + \textstyle\sum_a \tilde{Z}^a)) + \sum_{aa'}(n^a - \tilde{n}^a + Z^a - \tilde{Z}^a | n^{a'} - \tilde{n}^{a'} + Z^{a'} - \tilde{Z}^a) \\ + 2\sum_a (\tilde{n} + \textstyle\sum_{a'} \tilde{Z}^{a'} | n^a - \tilde{n}^a + Z^a - \tilde{Z}^a).\end{aligned}$$

If we define $\tilde{Z}^a(\mathbf{r})$ in such a way that $n^a(\mathbf{r}) - \tilde{n}^a(\mathbf{r}) + Z^a(\mathbf{r}) - \tilde{Z}^a(\mathbf{r})$ has no multipole moments, i.e.

$$\int d\mathbf{r} r^l Y_L(\widehat{\mathbf{r} - \mathbf{R}^a})(n^a - \tilde{n}^a + Z^a - \tilde{Z}^a) = 0 \tag{35}$$

for all a, the potentials of these densities are zero outside their respective augmentation spheres ($L = (l, m)$ is a collective angular- and magnetic quantum number). Exploiting this feature, the Coulomb integral reduce to

$$\begin{aligned}((n + \textstyle\sum_a Z^a)) &= ((\tilde{n} + \textstyle\sum_a \tilde{Z}^a)) + \textstyle\sum_a((n^a - \tilde{n}^a + Z^a - \tilde{Z}^a)) + 2\textstyle\sum_a(\tilde{n}^a + \tilde{Z}^a | n^a - \tilde{n}^a + Z^a - \tilde{Z}^a) \\ &= ((\tilde{n} + \textstyle\sum_a \tilde{Z}^a)) + \textstyle\sum_a \left(((n^a + Z^a)) - ((\tilde{n}^a + \tilde{Z}^a))\right)\end{aligned}$$

where it has been used that inside the augmentation spheres $\tilde{n} = \tilde{n}^a$. In this expression, we have circumvented all of the previous problems. None of the terms correlates functions on different grids, there is only a single summation over the atomic sites, and furthermore the only thing that has to be evaluated in the full space is the Hartree energy of $\tilde{n}(\mathbf{r}) + \sum_a \tilde{Z}^a(\mathbf{r})$ which is charge neutral (see eq. (42)).

Inserting the final expression in (30), we see that

$$\begin{aligned}E_C[n] &= \frac{1}{2}((\tilde{n} + \textstyle\sum_a \tilde{Z}^a)) + \frac{1}{2}\sum_a \left(((n^a + Z^a))' - ((\tilde{n}^a + \tilde{Z}^a))\right) \\ &= U_H[\tilde{\rho}] + \frac{1}{2}\sum_a \left(((n^a)) + 2(n^a | Z^a) - ((\tilde{n}^a + \tilde{Z}^a))\right)\end{aligned} \tag{36}$$

[2] One could separate the terms in other ways, but it is impossible to separate the smooth and the localized terms completely.

where we have introduced the smooth total density

$$\tilde{\rho}(\mathbf{r}) = \tilde{n} + \sum_a \tilde{Z}^a(\mathbf{r}). \tag{37}$$

Note that the problem with the self interaction error of the nuclei could easily be resolved once it was moved to the atom centered part, as handling charged densities is not a problem on radial grids.

To obtain an explicit expression for the compensation charges, we make a multipole expansion of $\tilde{Z}^a(\mathbf{r})$

$$\tilde{Z}^a = \sum_L Q^a_L \; \tilde{g}^a_L(\mathbf{r}), \tag{38}$$

where $\tilde{g}^a_L(\mathbf{r})$ is any smooth function localized within $|\mathbf{r} - \mathbf{R}^a| < r^a_c$, satisfying

$$\int d\mathbf{r} r^l Y_L(\widehat{\mathbf{r} - \mathbf{R}^a}) \tilde{g}^a_{L'}(\mathbf{r}) = \delta_{LL'}. \tag{39}$$

Plugging the expansion (38) into equations (35), we see that the expansion coefficients Q^a_L from must be chosen according to

$$Q^a_L = \int d\mathbf{r} r^l Y_L(\hat{\mathbf{r}}) \left(n^a(\mathbf{r}) - \tilde{n}^a(\mathbf{r}) + Z^a(\mathbf{r})\right) = \Delta^a \delta_{l,0} + \sum_{i_1 i_2} \Delta^a_{L i_1 i_2} D^a_{i_1 i_2} \tag{40}$$

where

$$\Delta^a = \int dr \left(n^a_c(r) - \tilde{n}^a_c(r)\right) - \mathcal{Z}^a/\sqrt{4\pi} \tag{41a}$$

$$\Delta^a_{L i_1 i_2} = \int d\mathbf{r} r^l Y_L(\hat{\mathbf{r}}) [\phi^a_{i_1}(\mathbf{r})\phi^a_{i_2}(\mathbf{r}) - \tilde{\phi}^a_{i_1}(\mathbf{r})\tilde{\phi}^a_{i_2}(\mathbf{r})] \tag{41b}$$

and it has been used that the core densities are spherical $n^a_c(\mathbf{r}) = n^a_c(r) Y_{00}(\hat{\mathbf{r}})$ (we consider only closed shell frozen cores). This completely defines the compensation charges $\tilde{Z}^a(\mathbf{r})$.

Note that the special case $l = 0$ of (35), implies that

$$\begin{aligned}
&\int d\mathbf{r} \left[n^a - \tilde{n}^a + Z^a - \tilde{Z}^a\right] = 0 \\
&\qquad \Downarrow \\
&\int d\mathbf{r} \left[\tilde{n}(\mathbf{r}) + \sum_a \tilde{Z}^a(\mathbf{r})\right] = \int d\mathbf{r} \left[n(\mathbf{r}) + \sum_a Z^a(\mathbf{r})\right] \\
&\qquad \Updownarrow \\
&\int d\mathbf{r} \; \tilde{\rho}(\mathbf{r}) = \int d\mathbf{r} \; \rho(\mathbf{r}) = 0
\end{aligned} \tag{42}$$

i.e. that the smooth total density has zero total charge, making the evaluation of the Hartree energy numerically convenient.

In summary, we conclude that the classical coulomb interaction energy which in the KS formalism was given by $E_C[n] = U'_H[\rho]$, in the PAW formalism becomes a pure Hartree energy (no self-interaction correction) of the smooth total density $\tilde{\rho}$ plus some one-center corrections

$$E_C[n] = U_H[\tilde{\rho}] + \sum_a \Delta E^a_C[\{D^a_{i_1 i_2}\}] \tag{43}$$

where the corrections

$$\begin{aligned}\Delta E^a_C[\{D^a_{i_1i_2}\}] &= \tfrac{1}{2}((n^a)) + (n^a|Z^a) - \tfrac{1}{2}((\tilde{n}^a)) - (\tilde{n}^a|\tilde{Z}^a) - \tfrac{1}{2}((\tilde{Z}^a)) \\ &= \tfrac{1}{2}\left[((n^a_c)) - ((\tilde{n}^a_c))\right] - \mathcal{Z}^a \int d\mathbf{r} \frac{n^a_c(r)}{r} - \sum_L Q^a_L(\tilde{n}^a_c|\tilde{g}^a_L) \\ &\quad + \sum_{i_1i_2} D^{a*}_{i_1i_2}\left[(\phi^a_{i_1}\phi^a_{i_2}|n^a_c) - (\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{n}^a_c) - \mathcal{Z}^a\int d\mathbf{r}\frac{\phi^a_{i_1}(\mathbf{r})\phi^a_{i_2}(\mathbf{r})}{r} - \sum_L Q^a_L(\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{g}^a_L)\right] \\ &\quad + \frac{1}{2}\sum_{i_1i_2i_3i_4} D^{a*}_{i_1i_2}\left[(\phi^a_{i_1}\phi^a_{i_2}|\phi^a_{i_3}\phi^a_{i_4}) - (\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{\phi}^a_{i_3}\tilde{\phi}^a_{i_4})\right] D^a_{i_3i_4} - \frac{1}{2}\sum_{LL'} Q^a_L Q^a_{L'}(\tilde{g}^a_L|\tilde{g}^a_{L'})\end{aligned}$$

Using that the potential of a spherical harmonic (times some radial function) is itself a spheical harmonic of the same angular momentum, we see that $(\tilde{g}^a_L|\tilde{g}^a_{L'}) \propto \delta_{LL'}$ and $(\tilde{n}^a_c|\tilde{g}^a_L) \propto \delta_{L0}$. Noting that Q^a_L by virtue of (40) is a functional of the density matrix, and inserting this, we get

$$\Delta E^a_C = \Delta C^a + \sum_{i_1i_2}\Delta C^a_{i_1i_2} D^a_{i_1i_2} + \sum_{i_1i_2i_3i_4} D^{a*}_{i_1i_2}\Delta C^a_{i_1i_2i_3i_4} D^a_{i_3i_4} \tag{44}$$

where

$$\Delta C^a = \tfrac{1}{2}\left[((n^a_c)) - ((\tilde{n}^a_c)) - (\Delta^a)^2((\tilde{g}^a_{00}))\right] - \Delta^a(\tilde{n}^a_c|\tilde{g}^a_{00}) - \sqrt{4\pi}\mathcal{Z}^a\int dr\frac{n^a_c(r)}{r} \tag{45}$$

$$\begin{aligned}\Delta C^a_{i_1i_2} =& (\phi^a_{i_1}\phi^a_{i_2}|n^a_c) - (\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{n}^a_c) - \mathcal{Z}^a\int d\mathbf{r}\frac{\phi^a_{i_1}(\mathbf{r})\phi^a_{i_2}(\mathbf{r})}{r} \\ &- \Delta^a(\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{g}^a_{00}) - \Delta^a_{00,i_1i_2}\Big(\Delta^a(\tilde{n}^a_c|\tilde{g}^a_{00}) + ((\tilde{g}^a_{00}))\Big)\end{aligned} \tag{46}$$

$$\begin{aligned}\Delta C^a_{i_1i_2i_3i_4} =& \tfrac{1}{2}\left[(\phi^a_{i_1}\phi^a_{i_2}|\phi^a_{i_3}\phi^a_{i_4}) - (\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{\phi}^a_{i_3}\tilde{\phi}^a_{i_4})\right] \\ &- \sum_L\left[\tfrac{1}{2}\Delta^a_{Li_1i_2}(\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{g}^a_L) + \tfrac{1}{2}\Delta^a_{Li_3i_4}(\tilde{\phi}^a_{i_3}\tilde{\phi}^a_{i_4}|\tilde{g}^a_L) + \Delta^a_{Li_1i_2}((\tilde{g}^a_L))\Delta^a_{Li_3i_4}\right]\end{aligned} \tag{47}$$

Note that all integrals can be limited to the inside of the augmentation sphere. For example $(\phi^a_{i_1}\phi^a_{i_2}|n^a_c)$ has contributions outside the augmentation sphere, but these are exactly canceled by the contributions outside the spheres of $(\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}|\tilde{n}^a_c)$, in which region the two expressions are identical.

The $\Delta C^a_{i_1i_2i_3i_4}$ tensor has been written in a symmetric form, such that it is invariant under the following symmetry operations:

$$i_1 \leftrightarrow i_2 \qquad i_3 \leftrightarrow i_4 \qquad i_1i_2 \leftrightarrow i_3i_4 \tag{48}$$

1.5.3 Summary

Summing up all the energy contributions, we see that the Kohn-Sham total energy

$$E[n] = T_s[\{\psi_n\}] + U'_H[\rho] + E_{xc}[n]$$

can be separated into a part calculated on smooth functions, $\tilde{E}$, and some atomic corrections, ΔE^a, involving quantities localized around the nuclei only.

$$E = \tilde{E} + \sum_a \Delta E^a \tag{49}$$

where the smooth part

$$\tilde{E} = T_s[\{\tilde{\psi}_n\}] + U_H[\tilde{\rho}] + E_{xc}[\tilde{n}] \tag{50}$$

is the usual energy functional, but evaluated on the smooth functions $\tilde{n}$ and $\tilde{\rho}$ instead of n and ρ, and with the soft compensation charges $\tilde{Z}^a$ instead of the nuclei charges $Z^a(\mathbf{r})$. The corrections are given by

$$\Delta E^a = \Delta T_c^a + \Delta C^a + \sum_{i_1 i_2} \left(\Delta T^a_{i_1 i_2} + \Delta C^a_{i_1 i_2}\right) + \sum_{i_1 i_2 i_3 i_4} D^{a*}_{i_1 i_2} \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_3 i_4} + \Delta E^a_{xc}[\{D^a_{i_1 i_2}\}] \quad (51)$$

where T_c^a, $\Delta T^a_{i_1 i_2}$, $\Delta C^a_{i_1 i_2}$, and $\Delta C^a_{i_1 i_2 i_3 i_4}$ are system independent tensors that can be pre-calculated and stored for each specie in the periodic table of elements.

Both the Hamiltonian and the forces can be derived from the total energy functional (49) as will be shown in the following two sections.

1.6 The Transformed Kohn-Sham Equation

The variational quantity in the PAW formalism is the smooth wave function $\tilde{\psi}_n$. From this, all other quantities, such as the density matrix, the soft compensation charges, the transformation operator, etc. are determined by various projections of $\tilde{\psi}_n$ onto the projector functions, and expansions in our chosen basis functions, the partial and smooth partial waves. To obtain the smooth wave functions, we need to solve the eigenvalue equation

$$\widehat{\widetilde{H}} \tilde{\psi}_n(\mathbf{r}) = \epsilon_n \hat{S} \tilde{\psi}_n(\mathbf{r}), \quad (52)$$

where the overlap operator $\hat{S} = \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}}$ and $\widehat{\widetilde{H}} = \hat{\mathcal{T}}^\dagger \widehat{H} \hat{\mathcal{T}}$ is the transformed Hamiltonian.

1.6.1 Orthogonality

In the original form, the eigen states of the KS equation where orthogonal, i.e. $\langle \psi_n | \psi_m \rangle = \delta_{nm}$ while in the transformed version

$$\langle \tilde{\psi}_n | \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}} | \tilde{\psi}_m \rangle = \delta_{nm} \quad (53)$$

i.e. the smooth wave function are only orthogonal with respect to the weight $\hat{S}$.

The explicit form of the overlap operator is

$$\begin{aligned}
\hat{S} &= \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}} \\
&= \left(1 + \textstyle\sum_a \hat{\mathcal{T}}^a\right)^\dagger \left(1 + \textstyle\sum_a \hat{\mathcal{T}}^a\right) \\
&= 1 + \sum_a \left(\hat{\mathcal{T}}^{a\dagger} + \hat{\mathcal{T}}^a + \hat{\mathcal{T}}^{a\dagger} \hat{\mathcal{T}}^a\right) \\
&= 1 + \sum_a \Big[\sum_{i_1} |\tilde{p}^a_{i_1}\rangle (\langle \phi^a_{i_1}| - \langle \tilde{\phi}^a_{i_1}|) \sum_{i_2} |\tilde{\phi}^a_{i_2}\rangle \langle \tilde{p}^a_{i_2}| + \sum_{i_2} |\tilde{\phi}^a_{i_2}\rangle \langle \tilde{p}^a_{i_2}| \sum_{i_1} (|\phi^a_{i_1}\rangle - |\tilde{\phi}^a_{i_1}\rangle) \langle \tilde{p}^a_{i_1}| \\
&\qquad + \sum_{i_1} |\tilde{p}^a_{i_1}\rangle (\langle \phi^a_{i_1}| - \langle \tilde{\phi}^a_{i_1}|) \sum_{i_2} (|\phi^a_{i_2}\rangle - |\tilde{\phi}^a_{i_2}\rangle) \langle \tilde{p}^a_{i_2}| \Big] \\
&= 1 + \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle (\langle \phi^a_{i_1} | \phi^a_{i_2}\rangle - \langle \tilde{\phi}^a_{i_1} | \tilde{\phi}^a_{i_2}\rangle) \langle \tilde{p}^a_{i_2}| \\
&= 1 + \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle \sqrt{4\pi} \Delta^a_{00, i_1 i_2} \langle \tilde{p}^a_{i_2}|
\end{aligned} \quad (54)$$

The orthogonality condition (53) must be kept in mind when applying numerical schemes for solving (52). For example plane waves are no longer orthogonal, in the sense that $\langle \mathbf{G} | \hat{S} | \mathbf{G}' \rangle \neq \delta_{\mathbf{G},\mathbf{G}'}$.

1.6.2 The Hamiltonian

To determine the transformed Hamiltonian, one could apply the transformation $\widehat{\widetilde{H}} = \hat{\mathcal{T}}^\dagger \widehat{H} \hat{\mathcal{T}}$ directly, which would be straight forward for the local parts of $\widehat{H}$, but to take advantage of the

trick used to determine the total energy of the nonlocal term ($E_C[n]$), we make use of the relation

$$\frac{\delta E}{\delta \tilde{\psi}_n^*(\mathbf{r})} = f_n \widehat{\widetilde{H}} \tilde{\psi}_n(\mathbf{r}). \tag{55}$$

Using this, we get

$$\begin{aligned}
\frac{\delta E}{\delta \tilde{\psi}_n^*(\mathbf{r})} &= \frac{\delta}{\delta \tilde{\psi}_n^*(\mathbf{r})} \left[T_s[\{\tilde{\psi}_n\}] + E_{xc}[\tilde{n}] + U_H[\tilde{\rho}] + \Delta E^a[\{D^a_{i_1 i_2}\}] \right] \\
&= \frac{\delta T_s[\{\tilde{\psi}_n\}]}{\delta \tilde{\psi}_n^*(\mathbf{r})} \\
&\quad + \int d\mathbf{r}' \left[\frac{\delta E_{xc}[\tilde{n}]}{\delta \tilde{n}(\mathbf{r}')} + \frac{\delta U_H[\tilde{\rho}]}{\delta \tilde{n}(\mathbf{r}')} \right] \frac{\delta \tilde{n}(\mathbf{r}')}{\delta \tilde{\psi}_n^*(\mathbf{r})} \\
&\quad + \sum_a \sum_{i_1 i_2} \left[\int d\mathbf{r}' \frac{\delta U_H[\tilde{n} + \sum_a \tilde{Z}^a]}{\delta \tilde{Z}^a(\mathbf{r}')} \frac{\delta \tilde{Z}^a(\mathbf{r}')}{\delta D^a_{i_1 i_2}} + \frac{\delta \Delta E^a}{\delta D^a_{i_1 i_2}} \right] \frac{\delta D^a_{i_1 i_2}}{\delta \tilde{\psi}_n^*(\mathbf{r})} \\
&= f_n(-\tfrac{1}{2}\nabla^2)\psi_n(\mathbf{r}) \\
&\quad + \int d\mathbf{r}' \left[v_{xc}[\tilde{n}](\mathbf{r}') + u_H[\tilde{\rho}](\mathbf{r}') \right] f_n \delta(\mathbf{r} - \mathbf{r}') \tilde{\psi}_n(\mathbf{r}') \\
&\quad + \sum_a \sum_{i_1 i_2} \left[\int d\mathbf{r}' u_H[\tilde{n} + \textstyle\sum_a \tilde{Z}^a](\mathbf{r}') \sum_L \Delta^a_{L i_1 i_2} \tilde{g}^a_L(\mathbf{r}') + \frac{\delta \Delta E^a}{\delta D^a_{i_1 i_2}} \right] f_n \tilde{p}^a_{i_1}(\mathbf{r}) P^a_{n i_2}
\end{aligned}$$

where $v_{xc}[n](\mathbf{r}) = \frac{\delta E_{xc}[n]}{\delta n(\mathbf{r})}$ is the usual local (LDA) or semilocal (GGA) exchange correlation potential, and $u_H[n](\mathbf{r}) = \frac{\delta U_H[n]}{\delta n(\mathbf{r})} = \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}$ is the usual Hartree potential.

From these results, we can write down the transformed Hamiltonian as

$$\widehat{\widetilde{H}} = -\tfrac{1}{2}\nabla^2 + u_H[\tilde{\rho}] + v_{xc}[\tilde{n}] + \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle \Delta H^a_{i_1 i_2} \langle \tilde{p}^a_{i_2}|, \tag{56}$$

where the nonlocal part of the Hamiltonian is given in terms of the tensor

$$\begin{aligned}
\Delta H^a_{i_1 i_2} &= \sum_L \Delta^a_{L i_1 i_2} \int d\mathbf{r} u_H[\tilde{\rho}](\mathbf{r}) \tilde{g}^a_L(\mathbf{r}) + \frac{\delta \Delta E^a}{\delta D^a_{i_1 i_2}} \\
&= \sum_L \Delta^a_{L i_1 i_2} \int d\mathbf{r} u_H[\tilde{\rho}](\mathbf{r}) \tilde{g}^a_L(\mathbf{r}) + \Delta T^a_{i_1 i_2} + \Delta C^a_{i_1 i_2} + 2 \sum_{i_3 i_4} \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_3 i_4} + \frac{\delta \Delta E_{xc}}{\delta D^a_{i_1 i_2}}.
\end{aligned} \tag{57}$$

Note that to justify taking the derivative with respect to D only, and not its complex conjugate, the symmetry properties (48) has been used to get $D^{a*}_{i_1 i_2} \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_3 i_4} = D^a_{i_1 i_2} \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_3 i_4}$.

1.7 Forces in PAW

In the ground state, the forces on each nuclei can be calculated directly from

$$\begin{aligned}
\mathbf{F}^a &= -\frac{dE}{d\mathbf{R}^a} \\
&= -\frac{\partial E}{\partial \mathbf{R}^a} - \sum_n \left\{ \frac{\partial E}{\partial |\tilde{\psi}_n\rangle} \frac{d|\tilde{\psi}_n\rangle}{d\mathbf{R}^a} + h.c. \right\} \\
&= -\frac{\partial E}{\partial \mathbf{R}^a} - \sum_n f_n \epsilon_n \left\{ \langle \tilde{\psi}_n | \hat{S} \frac{d|\tilde{\psi}_n\rangle}{d\mathbf{R}^a} + h.c. \right\} \\
&= -\frac{\partial E}{\partial \mathbf{R}^a} + \sum_n f_n \epsilon_n \langle \tilde{\psi}_n | \frac{d\hat{S}}{d\mathbf{R}^a} |\tilde{\psi}_n\rangle
\end{aligned} \tag{58}$$

where *h.c.* denotes the hermitian conjugate. To get to the second line, the chain rule has been applied. The third line follows from the relation

$$\frac{\partial E}{\partial \langle \tilde{\psi}_n|} = f_n \widehat{\tilde{H}} |\tilde{\psi}_n\rangle = f_n \epsilon_n \hat{S} |\tilde{\psi}_n\rangle. \tag{59}$$

The last line of (58) is obtained from the following manipulation of the orthogonality condition (53)

$$\begin{aligned} & \delta_{nm} = \langle \tilde{\psi}_n | \hat{S} | \tilde{\psi}_m \rangle \\ \Rightarrow \quad & 0 = \frac{d}{d\mathbf{R}^a} \langle \tilde{\psi}_n | \hat{S} | \tilde{\psi}_m \rangle = \frac{d\langle \tilde{\psi}_n|}{d\mathbf{R}^a} \hat{S} |\tilde{\psi}_m\rangle + \langle \tilde{\psi}_n | \frac{d\hat{S}}{d\mathbf{R}^a} |\tilde{\psi}_m\rangle + \langle \tilde{\psi}_n | \hat{S} \frac{d|\tilde{\psi}_m\rangle}{d\mathbf{R}^a} \\ \Leftrightarrow \quad & \frac{d\langle \tilde{\psi}_n|}{d\mathbf{R}^a} \hat{S} |\tilde{\psi}_m\rangle + h.c. = -\langle \tilde{\psi}_n | \frac{d\hat{S}}{d\mathbf{R}^a} |\tilde{\psi}_m\rangle \end{aligned} \tag{60}$$

From the expression for the overlap operator (54), it follows that

$$\frac{d\hat{S}}{d\mathbf{R}^a} = \sum_{i_1 i_2} \Delta S^a_{i_1 i_2} \left(\frac{d|\tilde{p}^a_{i_1}\rangle}{d\mathbf{R}^a} \langle \tilde{p}^a_{i_2}| + h.c. \right) \tag{61}$$

which, when inserted in (58), gives the force expression

$$\mathbf{F}^a = -\frac{\partial E}{\partial \mathbf{R}^a} + \sum_n f_n \epsilon_n \sum_{i_1 i_2} \Delta S^a_{i_1 i_2} \left(P^{a*}_{ni_1} \langle \frac{d\tilde{p}^a_{i_2}}{d\mathbf{R}^a} |\tilde{\psi}_n\rangle + \langle \tilde{\psi}_n | \frac{d\tilde{p}^a_{i_1}}{d\mathbf{R}^a} \rangle P^a_{ni_2} \right) \tag{62}$$

In the case of standard xc approximations, the dependence of the total energy on the nuclei coordinates is

$$\begin{aligned} \frac{\partial E}{\partial \mathbf{R}^a} &= \int d\mathbf{r}' \frac{\delta E}{\delta \tilde{n}(\mathbf{r}')} \frac{\partial \tilde{n}(\mathbf{r}')}{\partial \mathbf{R}^a} + \sum_{i_1 i_2} \frac{\partial E}{\partial D^a_{i_1 i_2}} \frac{\partial D^a_{i_1 i_2}}{\partial \mathbf{R}^a} + \int d\mathbf{r}' \sum_L \frac{\delta E}{\delta \tilde{g}^a_L(\mathbf{r}')} \frac{\partial \tilde{g}^a_L(\mathbf{r}')}{\partial \mathbf{R}^a} \\ &= \int d\mathbf{r}' \tilde{v}_{\text{eff}}(\mathbf{r}') \frac{\partial \tilde{n}^a_c(\mathbf{r}')}{\partial \mathbf{R}^a} + \sum_{i_1 i_2} \Delta H^a_{i_1 i_2} \frac{\partial D^a_{i_1 i_2}}{\partial \mathbf{R}^a} + \int d\mathbf{r}' \sum_L u_H(\mathbf{r}') Q_L \frac{\partial \tilde{g}^a_L(\mathbf{r}')}{\partial \mathbf{R}^a} \end{aligned} \tag{63}$$

giving the force expression

$$\begin{aligned} \mathbf{F}^a = & -\int d\mathbf{r}' \left\{ \tilde{v}_{\text{eff}}(\mathbf{r}') \frac{\partial \tilde{n}^a_c(\mathbf{r}')}{\partial \mathbf{R}^a} + u_H(\mathbf{r}') \sum_L Q_L \frac{\partial \tilde{g}^a_L(\mathbf{r}')}{\partial \mathbf{R}^a} \right\} \\ & - \sum_n f_n \sum_{i_1 i_2} \left\{ \Delta H^a_{i_1 i_2} - \epsilon_n \Delta S^a_{i_1 i_2} \right\} \left(P^{a*}_{ni_1} \langle \frac{d\tilde{p}^a_{i_2}}{d\mathbf{R}^a} |\tilde{\psi}_n\rangle + \langle \tilde{\psi}_n | \frac{d\tilde{p}^a_{i_1}}{d\mathbf{R}^a} \rangle P^a_{ni_2} \right) \end{aligned} \tag{64}$$

1.8 Summary

The PAW KS equation to be solved is

$$\widehat{\tilde{H}} |\tilde{\psi}_n\rangle = \epsilon_n \hat{S} |\tilde{\psi}_n\rangle \tag{65}$$

with $\hat{S}$, and $\widehat{\tilde{H}}$ given by

$$\hat{S} = \hat{1} + \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle \sqrt{4\pi} \Delta^a_{00,i_1 i_2} \langle \tilde{p}^a_{i_2}| \tag{66a}$$

$$\widehat{\tilde{H}} = -\tfrac{1}{2}\nabla^2 + u_H[\tilde{\rho}](\mathbf{r}) + v_{xc}[\tilde{n}](\mathbf{r}) + \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle \Delta H^a_{i_1 i_2} \langle \tilde{p}^a_{i_2}| \tag{66b}$$

where

$$\Delta H^a_{i_1 i_2} = \sum_L \Delta^a_{L i_1 i_2} \int d\mathbf{r} u_H[\tilde{\rho}](\mathbf{r}) \tilde{g}^a_L(\mathbf{r}) + \Delta T^a_{i_1 i_2} + \Delta C^a_{i_1 i_2} + 2 \sum_{i_3 i_4} \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_3 i_4} + \frac{\delta \Delta E_{xc}}{\delta D^a_{i_1 i_2}} \quad (67)$$

The total energy can then be evaluated by

$$E = T_s[\{\tilde{\psi}_n\}] + U_H[\tilde{\rho}] + E_{xc}[\tilde{n}] + \sum_a \Delta E^a \quad (68)$$

with ΔE^a given by

$$\Delta E^a = T^a_c + \sum_{i_1 i_2} \left(\Delta T^a_{i_1 i_2} + \Delta C^a_{i_1 i_2} \right) D^a_{i_1 i_2} + \sum_{i_1 i_2 i_3 i_4} D^{a*}_{i_1 i_2} \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_3 i_4} + \Delta E^a_{xc}(\{D^a_{i_1 i_2}\}) \quad (69)$$

Having solved the eigenvalue problem (65), the eigenvalues are known. This can be used to determine, for example, the kinetic energy of the pseudo wave functions, $T_s[\tilde{n}]$, without doing the explicit (and computationally costly) computation. This can be seen by operating with $\sum_n f_n \langle \tilde{\psi}_n |$ on eq. (66b) to get:

$$T_s[\{\tilde{\psi}_n\}] = \sum_n f_n \epsilon_n - \int d\mathbf{r} [\tilde{n}(\mathbf{r}) - \tilde{n}_c(\mathbf{r})] \left[u_H[\tilde{\rho}](\mathbf{r}) + v_{xc}[\tilde{n}](\mathbf{r}) \right] - \sum_a \sum_{i_1 i_2} \Delta H^a_{i_1 i_2} D^a_{i_1 i_2} \quad (70)$$

2 Implementing PAW

For an implementation of PAW, one must specify a large number of data for each chemical element. This constitutes a data set which uniquely determines how the on-site PAW transformation works, at the site of the specific atom. For the generation of such data sets, one needs an atomic DFT program, with which basis sets can be generated. How to perform DFT calculations efficiently on an isolated atom will be discussed in the first section of this chapter, and the actual choice of data set parameters will be discussed in the second. The atomic DFT program plays the additional role of a small test program, against which implementations in the full PAW program can be tested.

2.1 Atoms

If we consider the Kohn-Sham equation for an isolated atom, (described by a non spin-dependent Hamiltonian), it is well known that the eigenstates can be represented by the product

$$\phi_{i\sigma_i}(\mathbf{r}\sigma) = R_j(r) Y_L(\hat{\mathbf{r}}) \chi_{\sigma_i}(\sigma) \quad (71)$$

where R_j are real radial function, and Y_L are the (complex valued) spherical harmonics. $i = (n, l, m)$, $j = (n, l)$, and $L = (l, m)$.

Assuming identical filling of all atomic orbitals, i.e. $f_{i\sigma} = f_j$, the density becomes

$$n(\mathbf{r}) = \sum_i \sum_{\sigma_i} f_j |\phi_{i\sigma_i}(\mathbf{r}\sigma)|^2 = \sum_j 2 \frac{2l+1}{4\pi} f_j |R_j(r)|^2 \quad (72)$$

where the first factor of 2 comes from the sum over spin, and the second factor from the sum over the magnetic quantum number using that

$$\sum_m |Y_{lm}|^2 = \frac{2l+1}{4\pi} \quad (73)$$

The identical filling of degenerate states is exact for closed shell systems, and corresponds to a spherical averaging of the density for open shell systems.

Determining potentials in a spherical coordinate system is usually done by exploiting the expansion of the Coulomb kernel

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \sum_L \frac{4\pi}{2l+1}\frac{r_<^l}{r_>^{l+1}} Y_L^*(\hat{\mathbf{r}})Y_L(\hat{\mathbf{r}}') \tag{74}$$

with $r_< = \min(r, r')$ and $r_> = \max(r, r')$. Using this it is seen that for any density with a known angular dependence, e.g. the density $R(r)Y_L(\hat{\mathbf{r}})$, the potential can be determined by

$$\begin{aligned} v[R(r)Y_L(\hat{\mathbf{r}})](\mathbf{r}) &= \int d\mathbf{r}' \frac{R(r')Y_L(\hat{\mathbf{r}}')}{|\mathbf{r}-\mathbf{r}'|} \\ &= \frac{4\pi}{2l+1} Y_L(\hat{\mathbf{r}}) \int_0^\infty r'^2 dr' R(r') \frac{r_<^l}{r_>^{l+1}} \\ &= \frac{4\pi}{2l+1} Y_L(\hat{\mathbf{r}}) \left[\int_0^r dr' R(r') r' \Big(\frac{r'}{r}\Big)^{l+1} + \int_r^\infty dr' R(r') r' \Big(\frac{r}{r'}\Big)^{l} \right] \end{aligned} \tag{75}$$

if the angular dependence is not a spherical harmonic, one can always do a multipole expansion, and use the above expression on the individual terms.

In the case of a radial density $n(\mathbf{r}) = n(r)$, the Hartree potential becomes

$$u_H(r) = \frac{4\pi}{r}\int_0^r dr' n(r') r'^2 + 4\pi \int_r^\infty dr' n(r') r' \tag{76}$$

A purely radial dependent density also implies that the xc-potential is a radial function. Using this, the entire KS equation can be reduced to a 1D problem in r, while the angular part is treated analytically.

2.1.1 The Radial Kohn-Sham Equation

For a spherical KS potential, and using that Y_L are eigenstates of the Laplacian, the KS equation can be reduced to the simpler one-dimensional second order eigenvalue problem

$$\left[-\frac{1}{2}\frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} + \frac{l(l+1)}{2r^2} + v_s(r) \right] R_j(r) = \epsilon_j R_j(r) \tag{77}$$

If we introduce the radial wave function $u_j(r)$ defined by

$$rR_j(r) = u_j(r) \tag{78}$$

the KS equation can be written as

$$u_j''(r) + \left(2\epsilon_j - 2v_s(r) - \frac{l(l+1)}{r^2} \right) u_j(r) = 0 \tag{79}$$

which is easily integrated using standard techniques. See e.g. [5, chapter 6].

2.2 The Atomic Data Set of PAW

The very large degree of freedom when choosing the functions defining the PAW transformation means that the choice varies a great deal between different implementations. In any actual implementation expansions are obviously finite, and many numerical considerations must be made when choosing these basis sets, to ensure fast and reliable convergence. This section provides an overview of the information needed for uniquely defining the PAW transformation, and the level of freedom when choosing the parameters.

The Partial Waves

The basis functions, ϕ_i^a, for the expansion of $|\psi_n\rangle$ should be chosen to ensure a fast convergence to the KS wave function. For this reason we choose the partial waves as the eigenstates of the Schrödinger equation for the isolated spin-saturated atoms. Thus the index i is a combination of main-, angular-, and magnetic quantum number, (n, l, m). And the explicit form is

$$\phi_i^a(\mathbf{r}) = \phi_{nl}^a(r) Y_{lm}(\hat{\mathbf{r}})$$

where $\phi_{nl}^a(r)$ are the solutions of the radial KS Schrödinger equation (77), and Y_{lm} are the spherical harmonics. For convenience we choose $\phi_i^a(\mathbf{r})$ to be real, i.e. we use the real spherical harmonics instead of the complex valued. This choice of partial waves implies that the smooth partial waves and the smooth projector functions can also be chosen real, and as products of some radial functions and the same real spherical harmonic.

Note that including unbound states of the radial KS equation in the partial waves is not a problem, as the diverging tail is exactly canceled by the smooth partial waves. In practice we only integrate the KS equation outward from the origin to the cutoff radius for unbound states, thus making the energies free parameters. In principle the same could be done for the bound states, but in GPAW, the energies of bound states are fixed by making the inward integration for these states and doing the usual matching (see e.g. [5, chapter 6]), i.e. the energies are chosen as the eigen energies of the system.

The Smooth Partial Waves

The smooth partial waves $\tilde{\psi}_i^a(\mathbf{r})$ are per construction identical to the partial waves outside the augmentation sphere. Inside the spheres, we can choose them as any smooth continuation. Presently GPAW uses simple 6'th order polynomials of even powers only (as odd powers in r results in a kink in the functions at the origin, i.e. that the first derivatives are not defined at this point), where the coefficients are used to match the partial waves smoothly at $r = r_c$. Other codes uses Bessel functions[4] or Gaussians.

The Smooth Projector Functions

The smooth projector functions are a bit more tricky. Making them orthonormal to $\tilde{\phi}_i^a(\mathbf{r})$ is a simple task of applying an orthonormalization procedure. This is the only formal requirement, but in any actual implementation all expansions are necessarily finite, and we therefore want them to converge as fast as possible, so only a few terms needs to be evaluated.

A popular choice is to determine the smooth projector functions according to

$$|\tilde{p}_i^a\rangle = \left(-\tfrac{1}{2}\nabla^2 + \tilde{v}_s - \epsilon_i\right)|\tilde{\phi}_i^a\rangle \tag{80}$$

or equivalently

$$\tilde{p}_j^a(r) = \left[-\frac{1}{2}\frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} + \frac{l(l+1)}{2r^2} + \tilde{v}_s(r) - \epsilon_j\right]\tilde{\phi}_j^a(r) \tag{81}$$

where $\tilde{v}_s(r)$ is the smooth KS potential $u_H[\tilde{\rho}](r) + v_{xc}[\tilde{n}](r)$. And then enforce the complementary orthogonality condition $\langle\tilde{p}_j^a|\tilde{\phi}_{j'}^a\rangle = \delta_{j,j'}$ inside the augmentation sphere, e.g. by a standard Gram-Schmidt procedure. Using this procedure ensures that the reference atom is described correctly despite the finite number of projectors.

The Smooth Compensation Charge Expansion Functions

The smooth compensation charges $\tilde{g}_L^a(\mathbf{r})$, are products of spherical harmonics, and radial functions $\tilde{g}_l^a(r)$ satisfying that

$$\int d\mathbf{r} r^l Y_L(\hat{\mathbf{r}})\tilde{g}_{L'}^a(\mathbf{r}) = \delta_{LL'} \tag{82}$$

In GPAW the radial functions are chosen as generalized Gaussian according to (here shown for $\mathbf{R}^a = 0$):

$$\tilde{g}^a_L(\mathbf{r}) = \tilde{g}^a_l(r)Y_L(\hat{\mathbf{r}}) \quad , \quad \tilde{g}^a_l(r) = \frac{1}{\sqrt{4\pi}}\frac{l!}{(2l+1)!}(4\alpha^a)^{l+3/2}r^l e^{-\alpha^a r^2} \tag{83}$$

where the atom-dependent decay factor α is chosen such that the charges are localized within the augmentation sphere.

The Core- and Smooth Core Densities

The core density follows directly from the all electron partial waves by

$$n_c(r) = \sum_i^{\text{core}} |\phi_i(\mathbf{r})|^2 = \sum_j^{\text{core}} 2(2l+1)|\phi_j(r)|^2/4\pi \tag{84}$$

The smooth core densities $\tilde{n}^a_c(\mathbf{r})$ are like the smooth partial waves expanded in a few (two or three) Bessel functions, Gaussians, polynomials or otherwise, fitted such that it matches the true core density smoothly at the cut-off radius.

The Localized Potential

An additional freedom in PAW is that for any operator $\widehat{L}$, localized within the augmentation spheres, we can exploit the identity (8)

$$\sum_i |\tilde{\phi}^a_i\rangle\langle\tilde{p}^a_i| = 1 \tag{85}$$

valid within the spheres, to get

$$\widehat{L} = \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle\langle\tilde{\phi}^a_{i_1}|\widehat{L}|\tilde{\phi}^a_{i_2}\rangle\langle\tilde{p}^a_{i_2}|$$

so for any potential $\bar{v}(\mathbf{r}) = \sum_a \bar{v}^a(\mathbf{r}-\mathbf{R}^a)$ localized within the augmentation spheres (i.e. $\bar{v}^a(\mathbf{r}) = 0$ for $r > r^a_c$) we get the identity

$$0 = \int d\mathbf{r}\tilde{n}(\mathbf{r})\sum_a \bar{v}^a(\mathbf{r}) - \sum_a \int d\mathbf{r}\tilde{n}^a\bar{v}^a(\mathbf{r})$$

This expression can be used as an 'intelligent zero'. Using this, we can make the replacement of the smooth potential

$$\tilde{v}_s(\mathbf{r}) = u_H[\tilde{\rho}](\mathbf{r}) + v_{xc}[\tilde{n}](\mathbf{r}) \rightarrow \tilde{v}_s(\mathbf{r}) = u_H[\tilde{\rho}](\mathbf{r}) + v_{xc}[\tilde{n}](\mathbf{r}) + \bar{v}(\mathbf{r}) \tag{86}$$

if we at the same time add

$$B^a + \sum_{i_1 i_2} \Delta B^a_{i_1 i_2} D^a_{i_1 i_2} \tag{87}$$

to the energy corrections ΔE^a, where

$$B^a = -\int d\mathbf{r}\tilde{n}^a_c\bar{v}^a(\mathbf{r}) \quad \text{and} \quad \Delta B^a_{i_1 i_2} = -\int d\mathbf{r}\tilde{\phi}^a_{i_1}\tilde{\phi}^a_{i_2}\bar{v}^a(\mathbf{r}) \tag{88}$$

This also implies that $B^a_{i_1 i_2}$ should be added to $\Delta H^a_{i_1 i_2}$.

The advantage of doing this is that the Hartree potential and the xc-potential might not be optimally smooth close to the nuclei, but if we define the localized potential properly, the sum of the three potentials might still be smooth. Thus one can initially evaluate $u_H[\tilde{\rho}](\mathbf{r})$ and $v_{xc}[\tilde{n}](\mathbf{r})$ on an extra fine grid, add $\bar{v}(\mathbf{r})$ and then restrict the total potential to the coarse grid again before solving the KS equation.

The typical way of constructing the localized potentials $\bar{v}^a$ is by expanding it in some basis, and then choosing the coefficients such that the potential $u_H[\tilde{\rho}](\mathbf{r}) + v_{xc}[\tilde{n}](\mathbf{r}) + \bar{v}(\mathbf{r})$ is optimally smooth at the core for the reference system.

Inclusion of $\bar{v}^a(\mathbf{r})$ changes the forces on each atom only through the redefinitions of $\tilde{v}_s(\mathbf{r})$ and $\Delta H^a_{i_1 i_2}$.

Summary

When constructing a data set for a specific atom, one must specify the following quantities, all defined within the augmentation spheres only:

1. ϕ_i^a from radial KS equation
2. $\tilde{\phi}_i^a$ by appropriate smooth continuation of ϕ_i^a
3. $\tilde{p}_i^a$ from equation (80)
4. $\tilde{g}_L^a$ localized within $r < r_c$, and satisfying $\int d\mathbf{r} r^{l'} \tilde{g}_L^a(\mathbf{r}) Y_{L'}(\widehat{\mathbf{r} - \mathbf{R}^a}) = \delta_{LL'}$
5. n_c^a follows from ϕ_i^a by (84)
6. $\tilde{n}_c^a$ by appropriate smooth continuation of n_c^a
7. $\bar{v}^a$ localized within $r < r_c^a$, otherwise freely chosen to make $\tilde{v}_s$ optimally smooth for the reference system

The adjustable parameters besides the shape of $\tilde{\phi}^a$, $\tilde{g}_L^a$, $\bar{v}^a$, and $\tilde{n}_c^a$ are

1. Cut-off radii r_c^a (which can also depend on i)
2. Frozen core states
3. Number of basis set functions (range of index i on ϕ_i^a, $\tilde{\phi}_i^a$, and $\tilde{p}_i^a$)
4. Energies of unbound partial waves

Choosing these parameters in such a way that the basis is sufficient for the description of all possible environments for the specific chemical element, while still ensuring a smooth pseudo wave function is a delicate procedure, although the optimal parameter choice is more stable than for e.g. norm conserving or ultra soft pseudopotentials.

Once the quantities above have been constructed, all other ingredients of the PAW transformation follows, such as Δ^a, $\Delta^a_{Lii'}$, T_c^a, $\Delta T^a_{i_1 i_2}$, ΔC^a, $\Delta C^a_{i_1 i_2}$, $\Delta C^a_{i_1 i_2 i_3 i_4}$, $\Delta \bar{v}^a$, and $\Delta \bar{v}^a_{i_1 i_2}$. The first two are needed for the construction of the compensation charges and the overlap operator, and the rest for determining the Hamiltonian, and for evaluating the total energy.

3 Non-standard Quantities

The preceding sections have described the details of making a standard DFT scheme work within the PAW formalism. This section will focus on what the PAW transform does to quantities needed for post-processing or expansions to DFT.

It is a big advantage of the PAW method, that it is formally exactly equivalent to all-electron methods (with a frozen core) but is computationally comparable to doing pseudopotential calculations. In pseudopotential approaches, projecting out the core region is handled by a static projection kernel, while in PAW this projection kernel is dynamically updated during the SCF-cycle via an expansion of the core region in a local atomic basis set. This has the drawback for the end user, the equations for all quantities most be modified to account the dual basis set description.

3.1 The External Potential in PAW

As an example of the principle in accommodating expressions to the PAW formalism, we will here consider the application of an external potential in DFT.

Without the PAW transformation, this addition is trivial, as the desired potential, $v_{\text{ext}}(\mathbf{r})$, should simply be added to the effective KS potential, and the total energy adjusted by the energy associated with the external potential $E_{\text{ext}} = \int d\mathbf{r} n(\mathbf{r}) v_{\text{ext}}(\mathbf{r})$.

In PAW, the density decomposes into pseudo and atomic parts, so that

$$E_{\text{ext}} = \int d\mathbf{r}\tilde{n}(\mathbf{r})v_{\text{ext}}(\mathbf{r}) + \sum_a \int d\mathbf{r}\left[n^a(\mathbf{r}) - \tilde{n}^a(\mathbf{r})\right] v_{\text{ext}}(\mathbf{r}).$$

Implying that both a pseudo energy contribution $\tilde{E}_{\text{ext}} = \int d\mathbf{r}\tilde{n}(\mathbf{r})v_{\text{ext}}(\mathbf{r})$ and atomic corrections $\Delta E^a_{\text{ext}} = \int d\mathbf{r}\left[n^a(\mathbf{r}) - \tilde{n}^a(\mathbf{r})\right] v_{\text{ext}}(\mathbf{r})$ should be added to the total energy.

In PAW, the Hamiltonian has the structure:

$$H = \frac{1}{f_n|\psi_n\rangle}\frac{\partial E}{\partial\langle\psi_n|} = \tilde{H} + \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle \Delta H^a_{i_1 i_2} \langle\tilde{p}^a_{i_2}|$$

In our case, the extra contributions due to the external potential are:

$$\tilde{H}_{\text{ext}}(\mathbf{r}) = v_{\text{ext}}(\mathbf{r})$$

and

$$\Delta H^{a,\text{ext}}_{i_1 i_2} = \int d\mathbf{r} v_{\text{ext}}(\mathbf{r}) \left\{\phi^a_{i1}(\mathbf{r})\phi^a_{i2}(\mathbf{r}) - \tilde{\phi}^a_{i1}(\mathbf{r})\tilde{\phi}^a_{i2}(\mathbf{r})\right\} \tag{89}$$

Thus we can write the atomic energy contribution as:

$$\begin{aligned}\Delta E^a_{\text{ext}} &= \int d\mathbf{r} v_{\text{ext}}(\mathbf{r}) \left[n^a_c(\mathbf{r}) - \tilde{n}^a_c(\mathbf{r}) + \sum_{i_1 i_2} D^a_{i_1 i_2} \left\{\phi^a_{i1}(\mathbf{r})\phi^a_{i2}(\mathbf{r}) - \tilde{\phi}^a_{i1}(\mathbf{r})\tilde{\phi}^a_{i2}(\mathbf{r})\right\}\right] \\ &= \int d\mathbf{r} v_{\text{ext}}(\mathbf{r}) \left[n^a_c(\mathbf{r}) - \tilde{n}^a_c(\mathbf{r})\right] + \sum_{i_1 i_2} D^a_{i_1 i_2} \Delta H^{a,\text{ext}}_{i_1 i_2}\end{aligned}$$

To evaluate the first term in the last line, the external potential should be expanded in some radial function at each nuclei e.g. the gaussians $\tilde{g}^a_L$, as the integral of these with the core densities is already precalculated.

For example, a zero-order (monopole) expansion, equivalent to the assumption

$$v_{\text{ext}}(\mathbf{r}) \approx v_{\text{ext}}(\mathbf{R}^a) \text{ , for } |\mathbf{r} - \mathbf{R}^a| < r^a_c$$

Leads to the expression:

$$\begin{aligned}\Delta E^a_{\text{ext}} &= v_{\text{ext}}(\mathbf{R}^a)(\sqrt{4\pi}Q^a_{00} + \mathcal{Z}^a) \\ \Delta H^{a,\text{ext}}_{i_1 i_2} &= v_{\text{ext}}(\mathbf{R}^a)\sqrt{4\pi}\Delta^a_{00,i_1 i_2}\end{aligned}$$

Linear external potentials (corresponding to a homogeneous applied electric field) can be handled exactly by doing an expansion to first order. This has been used in GPAW in e.g. the paper [6].

3.2 All-electron Density

During the self-consistency cycle of DFT, the all-electron quantities are at all times available in principle. In practise, they are never handled directly, but rather in the composite basis representation of a global pseudo description augmented by local atomic basis functions. For some post processing purposes it can however be desirable to reconstruct all-electron quantities on a single regular grid.

As an example, consider the all-electron density, which can formally be reconstructed by

$$n(\mathbf{r}) = \tilde{n}(\mathbf{r}) + \sum_a \left[n^a_c(\mathbf{r}) + \sum_{i_1 i_2} D^a_{i_1 i_2} \left(\phi^a_{i_1}(\mathbf{r})\phi^a_{i_2}(\mathbf{r}) - \tilde{\phi}^a_{i_1}(\mathbf{r})\tilde{\phi}^a_{i_2}(\mathbf{r})\right)\right].$$

Transferring this to a uniform grid will of coarse re-introduce the problem of describing sharply peaked atomic orbitals on a uniform grid, but as it is only needed for post processing, and not in

the self-consistency, it can be afforded to interpolating the pseudo density to an extra fine grid, before adding the summed atomic corrections from the radial grid.

One common use of the all-electron density is for the application of Bader analysis[7]. The advantage of applying this to the all-electron density instead of the pseudo density, is that it can be proved that the total electron density only has maxima's at the nuclei, such that there will only be one Bader volume associated with each atom. This does not hold for the pseudo density, which can result in detached Bader volumes. In addition, the dividing surfaces found if applied to the pseudo density will be wrong if these intersect the augmentation sphere.

In practice, the reconstructed total density will not integrate correctly due to the inaccurate description of a uniform grid in the core regions of especially heavy elements. But as the numerically exact value of the integral over the atomic corrections are known from the radial grid description ($= 4\pi \sum_{ij} D^a_{ij} \Delta^a_{L,ij}$), this deficiency can easily be remedied. As long as the density is correctly described at the dividing surfaces, these will still be determined correctly.

3.3 Wannier Orbitals

When constructing Wannier functions, the only quantities that needs to be supplied by the DFT calculator are the integrals $Z^{\mathbf{G}}_{n_1 n_2} = \langle \psi_{n_1} | e^{-\mathbf{G}\cdot\mathbf{r}} | \psi_{n_2} \rangle$, where $\mathbf{G}$ is one of at most 6 possible (3 in an orthorhombic cell) vectors connecting nearest neighbor cells in the reciprocal lattice.[8, 9]

When introducing the PAW transformation, this quantity can be expressed as

$$Z^{\mathbf{G}}_{n_1 n_2} = \langle \tilde{\psi}_{n_1} | e^{-\mathbf{G}\cdot\mathbf{r}} | \tilde{\psi}_{n_2} \rangle + \sum_a \sum_{i_1 i_2} P^{a*}_{n_1 i_1} P^a_{n_2 i_2} \left(\langle \phi^a_{i_1} | e^{-\mathbf{G}\cdot\mathbf{r}} | \phi^a_{i_2} \rangle - \langle \tilde{\phi}^a_{i_1} | e^{-\mathbf{G}\cdot\mathbf{r}} | \tilde{\phi}^a_{i_2} \rangle \right).$$

Even for small systems, the phase of the exponential of the last integral does not vary significantly over the region of space, where $\tilde{p}^a_i$ is non-zero. The integral in the last term can therefore safely be approximated by

$$e^{-\mathbf{G}\cdot\mathbf{R}^a} \sum_{i_1 i_2} P^{a*}_{n_1 i_1} P^a_{n_2 i_2} \sqrt{4\pi} \Delta^a_{00, i_1 i_2}.$$

3.4 Local Properties

This section describes quantities that can somehow be related to a specific atom. As the PAW transform utilizes an inherent partitioning of space into atomic regions, such quantities are usually readily extractable from already determined atomic attributes, such as the density matrices or the projector overlaps P^a_{ni}, which are by construction simultaneous expansion coefficients of both the pseudo and the all-electron wave functions inside the augmentation spheres.

3.4.1 Projected Density of States

The projection of the all electron wave functions onto the all electron partial waves, (i.e. the all electron wave functions of the isolated atoms) ϕ^a_i, is within the PAW formalism given by

$$\langle \phi^a_i | \psi_n \rangle = \langle \tilde{\phi}^a_i | \tilde{\psi}_n \rangle + \sum_{a'} \sum_{i_1 i_2} \langle \tilde{\phi}^a_i | \tilde{p}^{a'}_{i_1} \rangle \left(\langle \phi^{a'}_{i_1} | \phi^{a'}_{i_2} \rangle - \langle \tilde{\phi}^{a'}_{i_1} | \tilde{\phi}^{a'}_{i_2} \rangle \right) \langle \tilde{p}^{a'}_{i_2} | \tilde{\psi}_n \rangle \tag{90}$$

Using that projectors and pseudo partial waves form a complete basis within the augmentation spheres, this can be re-expressed as

$$\langle \phi^a_i | \psi_n \rangle = P^a_{ni} + \sum_{a' \neq a} \sum_{i_1 i_2} \langle \tilde{\phi}^a_i | \tilde{p}^{a'}_{i_1} \rangle \Delta S^{a'}_{i_1 i_2} P^{a'}_{n i_2} \tag{91}$$

if the chosen orbital index 'i' correspond to a bound state, the overlaps $\langle \tilde{\phi}^a_i | \tilde{p}^{a'}_{i_1} \rangle$, $a' \neq a$ will be small, and we see that we can approximate

$$\langle \phi^a_i | \psi_n \rangle \approx \langle \tilde{p}^a_i | \tilde{\psi}_n \rangle \tag{92}$$

The coefficients $P^a_{ni} = \langle \tilde{p}^a_i | \tilde{\psi}_n \rangle$, can thus be used as a qualitative measure of the local character of the true all electron wave functions. As the coefficients are already calculated and used in the SCF cycle, it involves no extra computational cost to determine quantities related directly to these.

These can be used to define an atomic orbital projected density of states

$$n_i(\varepsilon) = \sum_n \delta(\varepsilon - \epsilon_n) \left| P^a_{ni} \right|^2 . \tag{93}$$

3.4.2 Local Magnetic Moments

As the projection coefficients are simultaneous expansion coefficients of the pseudo and the all-electron wave functions inside the augmentation spheres, it can be seen that inside these, the all-electron density is given by (for a complete set of partial waves)

$$n(\mathbf{r}) = \sum_{i_1 i_2} D^a_{i_1 i_2} \phi^a_{i_1}(\mathbf{r}) \phi^a_{i_2}(\mathbf{r}) + n^a_c(\mathbf{r}) \ , \quad |\mathbf{r} - \mathbf{R}^a| < r^a_c. \tag{94}$$

This can be used to assign a local magnetic moment to each atom according to

$$M^a = \sum_{i_1 i_2} \Delta N^a_{i_1 i_2} \left[D^a_{i_1 i_2}(\uparrow) - D^a_{i_1 i_2}(\downarrow) \right],$$

where ΔN is an integration over products of AE waves truncated to the interior of the augmentation sphere

$$\Delta N^a_{i_1 i_2} = \int_{\mathbf{r} \in \Omega^a} d\mathbf{r} \phi^a_{i_1}(\mathbf{r}) \phi^a_{i_2}(\mathbf{r}).$$

Note that this will not add up to the total magnetic moment $\int d\mathbf{r}(n_\uparrow(\mathbf{r}) - n_\downarrow(\mathbf{r}))$, due to the interstitial space between augmentation spheres, and must be scaled if this is desired.

3.4.3 LDA + U

The atom projected density matrix $D^a_{i_1 i_2}$ can also be used to do LDA + U calculations. The GPAW implementation follows the LDA + U implementation in VASP[10], which is based on the particular branch of LDA + U suggested by Dudarev *et al.*[11], where you set the effective (U-J) parameter. The key notion is that from (94) one can define an (valence-) orbital density matrix

$$\hat{\rho}^a_{i_1 i_2} = |\phi^a_{i_1}\rangle D^a_{i_1 i_2} \langle \phi^a_{i_2}|.$$

Thus doing LDA + U is a simple matter of picking out the d-type elements of D^a, and adding to the total energy the contribution

$$\sum_a \sum_{i_1 i_2}^{\text{d type}} \frac{U}{2} \operatorname{Tr} \left(D^a_{i_1 i_2} - \sum_{i_3} D^a_{i_1 i_3} D^a_{i_3 i_2} \right) \tag{95}$$

and adding the gradient of this to the Hamiltonian

$$\sum_a \sum_{i_1 i_2}^{\text{d type}} |\tilde{p}^a_{i_1}\rangle \frac{U}{2} \left(\delta_{i_1 i_2} - 2 D^a_{i_1 i_2} \right) \langle \tilde{p}^a_{i_1}| \tag{96}$$

3.5 Coulomb Integrals

When trying to describe electron interactions beyond the level of standard (semi-) local density approximations, one will often need Coulomb matrix elements of the type

$$K_{nn',mm'} = (n_{nn'}|n_{mm'}) := \iint \frac{d\mathbf{r} d\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|} n^*_{nn'}(\mathbf{r}) n_{mm'}(\mathbf{r}'), \tag{97}$$

where the orbital pair density $n_{nn'}(\mathbf{r}) = \psi_n^*(\mathbf{r})\psi_{n'}(\mathbf{r})$.

Such elements are needed in some formulations of vdW functionals (although not the one implemented in GPAW), in linear-response TDDFT (see e.g. [12]) where only pair densities corresponding to electron-hole pairs are needed, in exact exchange or hybrid functionals (see next section) where only elements of the form $K_{nn',nn'}$ where both indices correspond to occupied states, are needed, and for GW calculations (see e.g. [13]), where all elements are needed.

Introducing the PAW transformation in (97), the pair densities partition according to

$$n_{nn'}(\mathbf{r}) = \tilde{n}_{nn'}(\mathbf{r}) + \sum_a \left(n^a_{nn'}(\mathbf{r}) - \tilde{n}^a_{nn'}(\mathbf{r})\right) \tag{98}$$

with the obvious definitions

$$\tilde{n}_{nn'} = \tilde{\psi}^*_n \tilde{\psi}_{n'} \qquad n^a_{nn'} = \sum_{i_1 i_2} P^{a*}_{ni_1} P^a_{n'i_2} \phi^{a*}_{i_1} \phi^{a*}_{i_2} \qquad \tilde{n}^a_{nn'} = \sum_{i_1 i_2} P^{a*}_{ni_1} P^a_{n'i_2} \tilde{\phi}^{a*}_{i_1} \tilde{\phi}^{a*}_{i_2}. \tag{99}$$

Exactly like with the Hartree potential, direct insertion of this in (97) would, due to the non-local nature of the Coulomb kernel, lead to undesired cross terms between different augmentation spheres. As before, such terms can be avoided by introducing some compensation charges, $\tilde{Z}^a_{nn'}$, chosen such that the potential of $n^a_{nn'} - \tilde{n}^a_{nn'} - \tilde{Z}^a_{nn'}$ are zero outside their respective augmentation spheres. This is achieved by doing a multipole expansion and requiring the expansion coefficients to be zero, and entails a compensation of the form

$$\tilde{Z}^a_{nn'}(\mathbf{r}) = \sum_L Q^a_{L,nn'} \tilde{g}^a_L(\mathbf{r}), \quad Q^a_{L,nn'} = \sum_{i_1 i_2} \Delta^a_{L,i_1 i_2} P^{a*}_{ni_1} P^a_{n'i_2} \tag{100}$$

(the constants $\Delta^a_{L,i_1 i_2}$ are identical to those in (41b)).

Introduction of such compensation charges makes it possible to obtain the clean partitioning

$$K_{nn',mm'} = (\tilde{\rho}_{nn'}|\tilde{\rho}_{mm'}) + 2\sum_a \sum_{i_1 i_2 i_3 i_4} P^a_{mi_1} P^{a*}_{ni_2} \Delta C^a_{i_1 i_2 i_3 i_4} P^{a*}_{n'i_3} P^a_{m'i_4}. \tag{101}$$

Here the last term is a trivial functional of the expansion coefficients P^a_{ni} involving only the constants $\Delta C^a_{i_1 i_2 i_3 i_4}$ already precalculated for the atomic corrections to the Coulomb energy (47). The only computationally demanding term relates to the Coulomb matrix element of the smooth compensated pair densities $\tilde{\rho}_{ij} = \tilde{n}_{ij} + \sum_a \tilde{Z}^a_{ij}$, which are expressible on coarse grids.

The formally exact partitioning (101) makes it possible, at moderate computational effort, to obtain Coulomb matrix elements in a representation approaching the infinite basis set limit. In standard implementations, such elements are usually only available in atomic basis sets, where the convergence of the basis is problematic. At the same time, all information on the nodal structure of the all-electron wave functions in the core region is retained, which is important due the non-local probing of the Coulomb operator. In standard pseudopotential schemes, this information is lost, leading to an uncontrolled approximation to $K_{nn',mm'}$.

As a technical issue, we note that integration over the the Coulomb kernel $1/|\mathbf{r} - \mathbf{r}'|$ is done by solving the associated Poisson equation, as for the Hartree potential, whereby the calculation of each element can be efficiently parallelized using domain decomposition. The integral $\int d\mathbf{r}\tilde{\rho}_{nn'}(\mathbf{r}) = \delta_{nn'}$ shows that the compensated pair densities $\tilde{\rho}_{nn}$ have a non-zero total charge, which is problematic for the determination of the associated potential. For periodic systems, charge neutrality is enforced by subtracting a homogeneous background charge, and the error so introduced is removed to leading order ($V^{-1/3}$ where V the the volume of the simulation box) by adding the potential of a missing probe charge in an otherwise periodically repeated array of probe charges embedded in a compensating homogeneous background charge. This can be determined using the standard Ewald technique, and corresponds to a rigid shift of the potential. For non-periodic systems, all charge is localized in the box, and the Poisson equation can be solved by adjusting the boundary values according to a multipole expansion of the pair density with respect to the center of the simulation box. A monopole correction is correct to the same order as the correction for periodic cells, but the prefactor on the error is much smaller, and leads to converged potentials even for small cells.

3.5.1 Exact Exchange

The EXX energy functional is given by

$$E_{\text{xx}} == -\frac{1}{2}\sum_{nm} f_n f_m \delta_{\sigma_n,\sigma_m} K_{nm,nm}. \tag{102}$$

Terms where n and m both refer to valence states transform in PAW as in equation (101). Terms where either index refers to a core orbital can be reduce to trivial functionals of P^a_{ni}, resulting in (see e.g. [14])

$$\begin{aligned} E_{\text{xx}} = &-\frac{1}{2}\sum_{nm}^{\text{val}} f_n f_m \delta_{\sigma_n,\sigma_m} (\tilde{\rho}_{nm}|\tilde{\rho}_{nm}) \\ &- \sum_a \left[\sum_\sigma \sum_{i_1 i_2 i_3 i_4} D^{a*}_{i_1 i_3}(\sigma) \Delta C^a_{i_1 i_2 i_3 i_4} D^a_{i_2 i_4}(\sigma) + \sum_{i_1 i_2} D^a_{i_1 i_2} X^a_{i_1 i_2} + E^{a,\text{c-c}}_{\text{xx}} \right]. \end{aligned} \tag{103}$$

The term involving the ΔC^a tensor is the PAW correction for the valence-valence interaction, and is similar to the correction in the equivalent expression for the Hartree energy, except that the order of the indices on the density matrices are interchanged. The term involving the X^a tensor represents the valence-core exchange interaction energy. $E^{a,\text{c-c}}_{\text{xx}}$ is simply the (constant) exchange energy of the core electrons.

The system independent Hermitian tensor $X^a_{i_1 i_2}$ is given by:

$$\begin{aligned} X^a_{i_1 i_2} &= \frac{1}{2}\sum_{\alpha}^{\text{core}} \iint d\mathbf{r} d\mathbf{r}' \frac{\phi^a_{i_1}(\mathbf{r})\phi^{a,\text{core}}_\alpha(\mathbf{r})\phi^a_{i_2}(\mathbf{r}')\phi^{a,\text{core}}_\alpha(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} \\ &= \sum_{j_c}^{\text{core}} \sum_l \frac{4\pi}{2l+1} \left(\sum_{m m_c} G^L_{L_1 L_c} G^L_{L_2 L_c} \right) \iint dr dr' \frac{r^l_<}{r^{l+1}_>} u^a_{j_1}(r) u^a_{j_c}(r) u^a_{j_c}(r') u^a_{j_2}(r'). \end{aligned} \tag{104}$$

Although the valence-core interaction is computationally trivial to include, it is not unimportant, giving rise to shifts in the valence eigenvalues of up to 1eV (though only a few kcal/mol in atomization energies), and we note that this contribution is unavailable in pseudopotential schemes. The core-core exchange is simply a reference energy, and will not affect self-consistency or energy differences.

For the iterative minimization schemes used in real-space and plane wave codes, the explicit form of the non-local Fock operator $v^{\text{NL}}(\mathbf{r},\mathbf{r}')$ is never needed, and would indeed be impossible to represent on any realistic grid. Instead only the action of the operator on a state is needed. As with the Hamiltonian operator, the action on the pseudo waves is derived via the relation $f_n \hat{v}^{\text{NL}}|\tilde{\psi}_n\rangle = \partial E_{\text{xx}}/\partial\langle\tilde{\psi}_n|$. Referring to [14] for a derivation, we merely state the result

$$\begin{aligned} \hat{v}^{\text{NL}}|\tilde{\psi}_n\rangle = &\sum_m f_m \tilde{v}_{nm}(\mathbf{r})|\tilde{\psi}_m\rangle \\ &+ \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle \left[\sum_m v^a_{nm,i_1 i_2} P^a_{m i_2} - X^a_{i_1 i_2} P^a_{n i_2} - 2\left(\sum_{i_3 i_4} C^a_{i_1 i_3 i_2 i_4} D^a_{i_3 i_4} \right) P^a_{n i_2} \right] \end{aligned} \tag{105}$$

where $\tilde{v}_{nm}$ is the solution of $\nabla^2 \tilde{v}_{nm}(\mathbf{r}) = -4\pi\tilde{\rho}_{nm}(\mathbf{r})$, and $v^a_{nm,i_1 i_2} = \sum_L \Delta^a_{L i_1 i_2} \int d\mathbf{r} \tilde{g}^a_L(\mathbf{r}) \tilde{v}_{nm}(\mathbf{r})$.

Again the computationally demanding first term is related to smooth pseudo quantities only, which can be accurately represented on coarse grids, making it possible to do basis set converged self-consistent EXX calculations at a relatively modest cost. Applying the Fock operator is however still expensive, as a Poisson equation must be solved for all pairs of orbitals.

As a technical consideration, note that the effect of the atomic corrections due to valence-valence, valence-core, and core-core exchange interactions can simply be incorporated into the standard equations by redefining equations (47), (46), and (45) respectively, which will also take care of

the last two terms in the Fock operator above. The introduction of the pair orbital compensation charges does however lead to a non-trivial correction to the Fock operato; the term proportional to v^a_{nm,i_1i_2}. This term also leads to a distinct contribution when calculating the kinetic energy via the eigenvalues as done in equation (70). The additional term (besides those related to redefining (45)–(47))

$$\sum_{nm} f_n \left[f_m \delta_{\sigma_n,\sigma_m} \int d\mathbf{r} \tilde{v}_{nm}(\mathbf{r}) \tilde{\psi}^*_n(\mathbf{r}) \tilde{\psi}_m(\mathbf{r}) - \sum_a \sum_{i_1 i_2} P^a_{ni_1} P^a_{mi_2} v^a_{nm,i_1 i_2} \right], \tag{106}$$

should be added to the right hand side of (70) on inclusion of exact exchange.

In a similar fashion, the compensation charges leads to an additional force contribution in equation (64) given by

$$\begin{aligned} \mathbf{F}^a_{xx} = \sum_{nm} f_n f_{nm} \delta_{\sigma_n \sigma_m} \Bigg\{ & \int d\mathbf{r}' \tilde{v}_{nm}(\mathbf{r}') \sum_{i_1 i_2} P^{a*}_{ni_1} P^a_{mi_2} \sum_L \Delta_{Li_1i_2} \frac{\partial \tilde{g}^a_L(\mathbf{r}')}{\partial \mathbf{R}^a} \\ & + \sum_{i_1 i_2} v^a_{n_1 n_2 i_1 i_2} \left(P^{a*}_{ni_1} \langle \frac{d\tilde{p}^a_{i_2}}{d\mathbf{R}^a} | \tilde{\psi}_m \rangle + \langle \tilde{\psi}_n | \frac{d\tilde{p}^a_{i_1}}{d\mathbf{R}^a} \rangle P^a_{mi_2} \right) \Bigg\}. \end{aligned} \tag{107}$$

3.5.2 Optimized Effective Potential

The optimized effective potential (OEP) method, is a way of converting the non-local Fock operator $\hat{v}^{\mathrm{NL}}_x$ into a local form $\hat{v}^{\mathrm{L}}_x = v^{\mathrm{L}}_x(\mathbf{r})$.

One way to derive the OEP equations in standard KS-DFT, is to use perturbation theory along the adiabatic connection (Görling-Levy perturbation theory [15]).

On converting the OEP equation to the PAW formalism, it should be remembered that local potentials in PAW transform to a local pseudo part plus non-local atomic corrections. Hence we want to arrive at a potential of the form

$$\hat{v}^{\mathrm{L}}_x = \tilde{v}^{\mathrm{L}}_x(\mathbf{r}) + \sum_a \sum_{i_1 i_2} |\tilde{p}^a_{i_1}\rangle \Delta v^a_{i_1 i_2} \langle \tilde{p}^a_{i_2}|, \tag{108}$$

where both the pseudo part $\tilde{v}^{\mathrm{L}}_x$ as well as the coefficients $\Delta v^a_{i_1 i_2}$ should be determined.

The derivation is more or less straight forward, if one remembers the the PAW KS equation is a generalized eigenvalue problem, that the variational quantity is the pseudo orbitals, and that the first order shift in the density has both a pseudo and an atomic part. The result is

$$\sum_n f_n \tilde{\psi}^*_n(\mathbf{r}) \sum_{m \neq n} \tilde{\psi}_m(\mathbf{r}) \frac{\langle \tilde{\psi}_m | \hat{v}^{\mathrm{NL}}_x - \hat{v}^{\mathrm{L}}_x | \tilde{\psi}_n \rangle}{\epsilon_n - \epsilon_m} + c.c. = 0 \tag{109a}$$

$$\sum_n f_n P^{a*}_{ni_1} \sum_{m \neq n} P^a_{mi_2} \frac{\langle \tilde{\psi}_m | \hat{v}^{\mathrm{NL}}_x - \hat{v}^{\mathrm{L}}_x | \tilde{\psi}_n \rangle}{\epsilon_n - \epsilon_m} + c.c. = 0 \tag{109b}$$

where $\hat{v}^{\mathrm{NL}}_x$ is the non-local exchange operator of equation (105) and $\hat{v}^{\mathrm{L}}_x$ is the local version in (108).

These can be solved iteratively starting from a local density-function approximation to the exchange potential in the spirit of [16].

It might seem that OEP is just extra work on top of the already expensive non-local operator, but it can in some cases be faster, as the number of SCF iterations in the KS cycle are greatly reduced.

References

[1] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, Dec 1994.

[2] The open-source project gpaw. freely available at https://wiki.fysik.dtu.dk/gpaw.

[3] P. E. Blöchl, C. J. Först, and J. Schimpl. Projector augmented wave method: ab-initio molecular dynamics with full wave functions. *Bulletin of Materials Science*, 26:33–41, 2003.

[4] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59:1758–1775, 1999.

[5] C. Fiolhais, F. Nogueira, and M. Margues, editors. *A Primer in Density Functional Theory*, volume 620 of *Lecture Notes in Physics*. Springer, 2003.

[6] F. Yin, J. Akola, P. Koskinen, M. Manninen, and R. E. Palmer. Bright beaches of nanoscale potassium islands on graphite in stm imaging. *Physical Review Letters*, 102(10):106102, 2009.

[7] W Tang, E Sanville, and G Henkelman. A grid-based bader analysis algorithm without lattice bias. *Journal of Physics: Condensed Matter*, 21(8):084204 (7pp), 2009.

[8] K. S. Thygesen, L. B. Hansen, and K. W. Jacobsen. Partly occupied wannier functions: Construction and applications. *Physical Review B*, 72(12):125119, September 2005.

[9] A Ferretti, A Calzolari, B Bonferroni, and R Di Felice. Maximally localized wannier functions constructed from projector-augmented waves or ultrasoft pseudopotentials. *Journal of Physics: Condensed Matter*, 19(3):036215 (16pp), 2007.

[10] A. Rohrbach, J. Hafner, and G. Kresse. Molecular adsorption on the surface of strongly correlated transition-metal oxides: A case study for co/nio(100). *Physical Review B*, 69(7):075413, Feb 2004.

[11] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton. Electron-energy-loss spectra and the structural stability of nickel oxide: An lsda+u study. *Physical Review B*, 57(3):1505–1509, Jan 1998.

[12] Michael Walter, Hannu Häkkinen, Lauri Lehtovaara, Martti Puska, Jussi Enkovaara, Carsten Rostgaard, and Jens Jørgen Mortensen. Time-dependent density-functional theory in the projector augmented-wave method. *The Journal of Chemical Physics*, 128(24):244101, 2008.

[13] Carsten Rostgaard, Karsten W. Jacobsen, and Kristian S. Thygesen. Assessment of the gw approximation for molecules. ii. ionization potentials from first principles and comparison to hybrid density functional theory. Submitted, 2009.

[14] Joachim Paier, Robin Hirschl, Martijn Marsman, and Georg Kresse. The perdew–burke–ernzerhof exchange-correlation functional applied to the g2-1 test set using a plane-wave basis set. *The Journal of Chemical Physics*, 122(23):234102, 2005.

[15] A. Görling and M. Levy. Exact kohn-sham scheme based on perturbation theory. *Physical Review A*, 50:196–204, 1994.

[16] S. Kümmel and J. P. Perdew. Simple iterative construction of the optimized effective potential for orbital functionals, including exact exchange. *Physical Review Letters*, 90:043004, 2003.