Development and application of a massively parallel KKR Green function method for large scale systems

Von der Fakultät fur Mathematik, Informatik und Naturwissenschaften

der RWTH Aachen University zur Erlangung des akademischen Grades

eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Dipl.-Phys. Alexander R. Thieß

aus

Neustadt in Holstein, Deutschland

Berichter: Univ.-Prof. Dr. rer. nat. Stefan Blügel

J.-Prof. Dr. rer. nat. Riccardo Mazzarello

Tag der mündlichen Prüfung: 19.12.2011

Contents

Introduction

The impact of structural and functional materials on society is often overlooked but can in fact hardly be overestimated: In numerous examples, ranging from the improvement of steel to the invention of light emitting diodes, carbon fibers as well as cheaper and larger memories for data storage, novel materials are a key to successfully face global challenges on mobility, energy, communication and sustainability. Most strikingly visible is this influence for technologies based on electronic, optical, and magnetic materials, technologies that revolutionize computing and communication excelling mankind into the information age. With the miniaturization of devices, made possible by the invention of the transistor and the integrated circuit, enormous and still exponentially growing computing and communication capabilities are fundamentally changing how we interact, work and live. Material science and condensed matter physics are at the heart of the invention, development, design and improvement of novel materials and subsequently of novel physical phenomena and processes and are thus an excellent demonstration of the interdependence of science, technology and society.

Advances in modern material design and technology are closely linked to advances in understanding on the basis of condensed matter physics, statistical physics and quantum mechanics of the many particle problem as well as the development of powerful methods. High-performance experimental tools combined with extraordinary progress in theory and computational power provide insight on the microscopic phenomena in materials and have paved new roads towards understanding as well as raising and answering new questions. On the theory side, density functional theory takes a central position in this process. The *ab initio* description of materials from the first principles of quantum mechanics holds fundamental and highly valuable information on the interactions and interplay of electrons in solids and contributes such to the advancement of knowledge on the structural, mechanical, optical, thermal, electrical, magnetic, ferroic or transport properties in bulk solids, surfaces, thin films, heterostructures, quantum wells, clusters and molecules. The complicated task to compute material properties on the quantum mechanical level of myriad of atoms in solids became first accessible by exploiting the periodicity of crystalline solids and high symmetry of idealized systems. Density functional theory calculations exploiting the periodic boundary

conditions of such idealized structures are the basis for a deep microscopical understanding as well as modeling and predicting of new materials, their properties and functionalities.

However, real materials can deviate substantially from ideal single crystals. Real solids may crystallize in grains separated by grain boundaries. Entropy, heat treatment or mechanical impact are a source for lattice defects of different dimensionality, varying extents and distributions. Quite similar to disordered alloys and compounds, where the chemical composition, or the balance of short range and long range order determine properties unattainable by solids of one constituent, these defects are of structural and chemical nature, and most frequently they are actually defining the materials properties and functionalities such as the electronic behavior of semiconductors or oxides. A prototype example are dilute magnetic semiconductors such as gadolinium doped gallium nitride where the inclusion of gadolinium can spark a thermally stable magnetic state [1] and thereby crucially change and improve the properties of the sample. This strong influence of defects in combination with preparation techniques which allow to synthesize materials with control over defects even down to single atoms opens the fascinating possibility to use the vast phase space of defects to actively engineer material properties.

On the other hand, to understand the role of defects, defect clusters, chemical and structural inhomogeneities or disorder in a solid or nanostructure on a quantum mechanical level is a highly non-trivial endeavor. Typically defects or disorder break the translation symmetry of the periodic solid, a symmetry that made the treatment of an infinite solid possible in the first place. The properties of isolated defects, i.e. single impurities are most adequately solved within DFT by a Green-function impurity method as developed for example by Zeller and Dederichs [2], Gunnarsson *et al.* [3], Kudrnovský *et al.* [4] or Aldén *et al.* [5]. In particular, it has possibilities for dealing with the long-range Coulomb interaction for charged defects. Now, this is dealt with by adding an artificial neutralizing homogeneous back ground charge but this imperfectly represents the actual screening charge density around a defect and misses the long range $\frac{1}{r}$ tail that leads to the hydrogenic states for shallow impurities. Alloys in the limit of large concentrations are frequently investigated using the coherent potential approximation (CPA) [6–8], where it is assumed that electrons propagate in a periodic solid of a fictitious potential that has the same scattering properties as the alloy in average. The vast majority of calculations for disordered systems are carried out in a supercell approach, where the solid is modeled by large unit cells that approximate the chemical and structural nature of the alloy or disordered system in real space, which are repeated periodically. A proper treatment of the physical systems requires frequently supercells containing hundreds to many thousands of atoms. To realize statistically relevant results, calculations have to be repeated for many different structural or compositional realizations. At present only few density-functional methods are capable to routinely treat systems of considerably more than one hundred atoms in the supercell [9–14]. Beyond this limit calculations become computationally extremely demanding and can hardly be performed on conventional computers.

With the sustained increase of computer power by a factor 1000 over 10 years, high performance computing has passed the peta-flop boundary a few years ago and opens new opportunities for progress in this field. High-end scientific computing is realized by massively parallelized computers of many ten thousand central processing units (CPU), so-called cores, working in parallel connected by a complex network to achieve this performance and

are unrivaled fast compared to conventional computers [15]. Although, high-performance computing architectures are under a permanent and rapid change, the design principle of massively parallel computers will live on although details on the processors, memory access, fast co-processors, input and output units or network complexity may change. According to Amdahl's law, where the speedup of a program using multiple processors in parallel computing is limited by the time needed for the sequential fraction of the program, massively parallel computers pose unprecedented challenges on the level of parallelization of the density functional algorithm. New algorithms fulfilling the high demands on parallel scalability and low memory resources have to be designed and implemented to enable fast and efficient computation on such architectures. Maybe, the most-commonly used plane-wave basis set to represent wave functions of the electrons in a periodic solid together with a pseudopotential approach to describe the scattering of the valence electrons at the ions is not the most effective approach for this architecture.

During the last years several methods are being explored, which aim at computation of large-scale systems on parallel platforms. These are methods that work in real space such as the real-space finite difference methods [16–20] in combination with pseudopotentials. Another set of important approaches relies on the fast spatial decay of the density matrix, i.e. Kohn's nearsightedness principle [21], focusing on the twin aims of overall linear scaling and controlled accuracy. This is the basic working principle for codes as ONETEP [22], CONQUEST [23], or openMX [24] allowing for the treatment of many thousand atoms. Wang *et al.* [25] successfully follow a similar ansatz with the locally self-consistent multiple scattering method. Although high parallel computational efficiency can be achieved in these methods, they are not universally applicable since the interactions e.g. in metallic systems are not of short range nature which impedes the application of the nearsightedness principle.

The bottom line is: on the one hand an increased demand for large scale computation can be observed. On the other hand the lack of massively parallel, highly accurate general purpose methods for more than thousand atoms leaves many unresolved open questions on extended defects, interaction of defects, and complex alloys unattended. The aim of this thesis is to combine both above presented trends: an *ab initio* treatment of thousands of atoms per unit cell on massive parallel supercomputers without sacrifice of accuracy. Conceptually, this thesis can be separated into two parts. The first one addresses the implementation of the large-scale program KKR**nano**. The second covers the results obtained by the application of KKR**nano** to two materials which recently attained high interest in the scientific community. Both parts will be introduced in greater detail in the following.

The numerical framework of our considerations will be density functional theory (DFT). DFT has been developed by Hohenberg and Kohn [26] and Kohn and Sham [27] and renders possible to circumvent the complicated many-body problem in solid state physics by a mapping on effective single-particle equations. All ground-state properties of a system can then be expressed as functionals of the charge density without requiring any additional information on the full wave function. Another important advantage of DFT is that calculations are performed *ab initio* or by first-principles − in other words without any further parametrizing than the nuclear numbers and the position of the nuclei. In chapter 2 the general concept of DFT will be introduced and the important approximations to the exchange-correlation potential highlighted.

Chapter 3 deals with the KKR Green function method which can be ascribed back to work

by Korringa [28] and Kohn and Rostoker [29] and which will be the basis to conduct the DFT cycle in KKRnano. By reviewing the KKR Green function theory in its full-potential representation [30] − as we will use it − the motivation for our selection of the KKR approach will become clear. In particular two characteristics are of great advantage for large-scale and massive parallel computation. The KKR scheme is based on multiple scattering theory, where a considerable part of the computation can be performed by considering exclusively the single site scattering problem − a scenario ideally suited for a real space parallelization. The interaction of the sites or atoms is then covered by only one multiple-scattering equation, which is the Dyson equation. Its solution is at least for large systems the computationally by far most demanding part and will be in the focus when constructing the algorithm used in KKRnano.

In chapter 4 we will direct our attention to the development of the large-scale massively parallelized code KKRnano. The combination of three essential concepts is implemented in KKRnano and all of those directly affect the Dyson equation: a screened reference system [31] of repulsive spheres will be introduced with the consequence that the matrix in the Dyson equation becomes highly sparse instead of fully occupied. This sparsity will not only be crucial to match the low memory resources on supercomputers, but also will be exploited to significantly speed up the solution. For this purpose we will not anymore solve the Dyson equation directly but rather iteratively as previously proposed [32]. It will turn out that this iterative schemes can be parallelized with convincing parallel efficiency. Nevertheless, the slow convergency of the iterations can present a serious limitation. Therefore, we will come up with optimization schemes such as preconditioning of the iterative solution which will turn out to be decisive. An additional conceptual step is the optional inclusion of the nearsightedness principle [25, 32, 33]. Implementing this scheme in KKRnano allows for order N scaling of the computational effort with system size for large system of more than thousand atoms (N denotes the number of atoms in the system). Besides the issues on the improvement of the algorithms, we will introduce a flexible multi-level hybrid parallelization to KKRnano which will facilitate efficient computation on up to one hundred thousand processors.

After completing the discussing on the elementary scheme of KKRnano, important extensions will be addressed in chapter 5: i.e. the implementation of Lloyd's formula [34–36] which allows for the treatment of semiconductors and insulators as well as the inclusion of the LDA+U scheme [37] which is crucial for a proper description of strongly correlated systems. Further, since even in complex magnetic systems the Lichtenstein formula [38] enables to extract individual magnetic exchange interactions, we will show how to incorporate this formula in KKRnano.

The first application of our new approach is considering the magnetic ordering in diluted magnetic semiconductors. Ferromagnetic ordering in dilute magnetic semiconductors and accordingly the combination of magnetic and semiconducting properties fulfills all requirements to realize novel spintronic concepts [39, 40]. However, to exploit such concepts in commercially usable computing devices ferromagnetic ordering above room temperature is inescapably required. This demand triggered a long but yet unsuccessful search for magnetic semiconductors with Curie temperatures above room temperature [41–46] which where mostly based on the inclusion of magnetic transition metal impurities. In 2005 Dhar *et al.* [1] broke new ground and reported on room temperature ferromagnetic ordering in gallium

nitride which they induced by doping with the lanthanide gadolinium. Moreover, in the dilute limit of gadolinium doping colossal magnetic moments of up to $4000\mu_B$ have been observed [1]. Elaborate work has been performed both experimentally e.g. [47, 48] and theoretically e.g. [49–52] to track down the origin of this effect. However, although there is a general consensus that extrinsic defects in addition to gadolinium are responsible, it remains still unclear which kind of defects are present. In chapter 6 we will revise the theoretical description of the magnetic properties of two of the most likely occurring types of defects: nitrogen and oxygen interstitials [50] as well as gallium vacancies [51, 52]. Using KKRnano holds two important advantages above the previous studies [50–52]. First, large supercells provide statistically sound results even for low doping concentrations and secondly, by means of the Lichtenstein formula the individual exchange interactions between any pair of defects can be evaluated. Our analysis will reveal reasons why nitrogen and oxygen interstitials are unlikely causing the experimentally observed effects. On the contrary, we will show that gallium vacancies can support a ferromagnetic network already at strikingly low concentrations.

The second system which we will take under scrutiny is the phase change material germanium antimony telluride. This material is widely used in DVD's due to the ability of reversible changing the phase by optical laser pulses [53]. Although it is utilized for industrial production since 1995, the underlying microscopic mechanisms of the switching process have only recently been explained by theoretical studies [54]. The bottom line is that the ordering of the highly concentrated vacancies on the chemically disordered germanium and antimony site takes a key role. Besides this most striking characteristic of germanium antimony telluride, in 2011 Siegrist *et al.* [55] discovered disorder-induced localization which showed indications for an Anderson localization. In chapter 7 we will present a series of *ab initio* calculations of $Ge_1Sb_2Te_4$ which is one of the most important realizations of this class of compounds. We will in particular point out the significance of the local concentration of vacancies on the electronic states. Following a scheme introduced by Schubert *et al.* [56] we will trace the origin of the spatial localization by a statistical evaluation of the density of states at the Fermi level as a function of system size. For this task we will consider supercells containing up to 4000 sites which will lead to important insight on the localization phenomena. We will show that the localization arises for larger local vacancy concentrations by the formation of vacancy clusters around the Te atoms, which effects increasingly higher density of states at the Fermi level. A second observation by Siegrist *et al.* [55] is the dependency of this localization and the corresponding transport properties on the annealing temperature. We will extract the key parameters for the annealing process from our *ab initio* calculations and will perform kinetic Monte-Carlo simulations to model the annealing which leads to a suppression of larger vacancy clusters. Based on this calculations we will be able to provide a microscopic explanation for the experimentally observed temperature dependency.

Density functional theory

Density functional theory (DFT) is an elegant technique to solve the complicated many-body problems in solid state physics by reducing the complexity to effective single-particle equations. The strength of DFT is that it is operated *ab initio* without any further input than nuclei numbers and positions of atoms, which makes it straightforwardly applicable to many systems and more important gives it predictive power. Accordingly DFT is widely used as an integral tool to address problems in quantum chemistry, bio-chemistry and material science theoretically.

Over the last decades many conceptual improvements have been made in the framework of density-functional methods. Most important, subsequently the representation of the a priori unknown exchange-correlation functional has been improved from the local density via the generalized gradient approximation to more sophisticated functionals allowing for the proper description of e.g. strongly correlated systems [57, 58]. Since the computation of most of the latter approaches is in a rigid way demanding and slows down the methods by orders of magnitudes, we exclusively introduce most commonly used local functionals: LDA and GGA as well as discuss implementational issues of the LDA+U approach in chapter 5.

Further, for the systems of interest considered in this thesis the usage of the non-relativistic representation of DFT is sufficient. Therefore we restrict the following introduction to this limit and refer the reader to the literature on relativistic DFT [59, 60].

Note, that all following derivations are given in atomic unit, i.e. $\hbar = 1$; $m_e = \frac{1}{2}$ and $e^2 = 2$.

2.1 Born-Oppenheimer Approximation

The Hamiltonian describing the full interactions of the atomic nuclei and the electrons constitute a complex many-body problem, which cannot be solved numerically nor analytically for relevant systems. The well-known Born-Oppenheimer approximation is an important simplification and decouples the general Hamiltonian of the nuclei and electronic motion. Here, the assumption that electrons move significantly faster and therefore adjust adiabatically to the nuclei-positions is exploited. Accordingly, the nuclei coordinates are fixed within the Born-Oppenheimer approximation and the Hamiltonian is solved for that particular set

of atomic positions. In practice even dynamic properties as e.g. the phonon dispersion can be very successfully obtained by solving the electronic problem for several sets of atomic coordinates and evaluating the total energies and forces.

2.2 Hohenberg-Kohn Theorems

The basic idea of DFT has been developed in the 1960's by Hohenberg and Kohn [26]. They proved for a system with a non-degenerate groundstate that the potential and therefore the Hamiltonian is a unique functional of the electron density $n(\mathbf{r})$ alone. A generalized proof was given by Levy [61].

This yields the famous Hohenberg-Kohn theorem:

1. All groundstate properties of a given many-body system are unique functionals of the electron density $n(\mathbf{r})$.

2. The total energy functional $E[n]$ underlies the variational principle

$$\frac{\partial E[n]}{\partial n(\mathbf{r})} = \mu, \tag{2.1}$$

, where μ is a Lagrange-multiplicator ensuring particle conservation

$$N - \int n(\mathbf{r})d\mathbf{r} = 0 \tag{2.2}$$

and the groundstate density n_0 minimizes the energy functional:

$$E[n] \ \geq \ E[n_0] \ = \ E_0, \tag{2.3}$$

2.3 Kohn-Sham equation

The basic idea of Kohn and Sham [27] is to represent an interacting system by a non-interacting system with the same electron density to obtain a good approximation for the large contribution of the kinetic energy to the energy functional $E[n]$. Starting from the Hohenberg-Kohn theorem one has to find the energy contributions as unique functionals of the electron density $n(\mathbf{r})$. The functional $E[n]$ (2.3) then can be split into several parts and written as [62]:

$$E[n] = T_0[n] + \int \int d\mathbf{r} \ d\mathbf{r}' \ \frac{n(\mathbf{r}) \ n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[n] + \int d\mathbf{r} \ n(\mathbf{r}) \ V_{ext}(\mathbf{r}), \tag{2.4}$$

where V_{ext} is the external potential and $T_0[n]$ describes the kinetic energy of a system of non-interacting particles with the same electron density n. The second term denotes the static Coulomb interaction of the electrons in Hartree approximation and E_{xc} arises due to exchange and correlation effects of the electron density itself and covers the missing contribution to the kinetic energy as well. It is important to note that an explicit expression of the exchange correlation functional $E_{xc}[n]$ is not known and therefore has to be approximated.

Hence, developing accurate approximations to the cumbersome contribution E_{xc} is crucial for the applicability of DFT. However, for now we will proceed with the next conceptual step and leave the issue of finding approximations to $E_{xc}[n]$ for section 2.5.

For the following discussion it is intuitive to introduce a wave-function representation. Then the electron density can be expressed by

$$n(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \,, \tag{2.5}$$

where the summation extends over the N orbitals with the lowest eigenvalues. The functional for the kinetic energy T_0 reads in that basis:

$$T_0[n] = \sum_{i=1}^{N} \int \nabla \phi_i^*(\mathbf{r}) \nabla \phi_i(\mathbf{r}) d\mathbf{r}. \tag{2.6}$$

The next important step is to apply the variational principle, which is stated in the Hohenberg-Kohn theorems, to the energy functional (2.4) under the Lagrange constraint that the normalization of the wave-function ϕ_i is conserved:

$$\sum_i \epsilon_i \left(\int \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) d\mathbf{r} - 1 \right), \tag{2.7}$$

where ϵ_i are the Lagrange-parameters. The variation of the energy functional $E[n]$ then yields:

$$\left(-\nabla^2 + V_{eff}[n] \right) \phi_i = \epsilon_i \phi_i. \tag{2.8}$$

The effective potential V_{eff} covers all previously discussed contributions to the energy and reads.

$$V_{eff}[n] = \int \frac{n(\mathbf{r}')}{\mathbf{r}' - \mathbf{r}} d\mathbf{r}' + V_{ext}(\mathbf{r}) + \frac{\partial E_{xc}[n]}{\partial n}. \tag{2.9}$$

The Kohn-Sham equation (2.8) is an important simplification as it has the form of an effective single-particle Schrödinger equation and can be solved straightforwardly. Although the Kohn-Sham eigenvalues ϵ_i have been formally introduced as Lagrange parameters without any physical meaning, they are often interpreted as excitation energies and the Kohn-Sham orbitals ϕ_i as single-particle wave-functions. In that spirit ϵ_i and ϕ_i serve as practical single-particle description of the actual system.

In practice the determination of the kinetic energy T_0 via the Laplacian operator is avoided. Multiplication of ϕ_i^* to the Kohn-Sham equation (2.8) leads, as the orbitals are normalized, to:

$$T_0[n] = \sum_{i=1}^{N} \epsilon_i - \int V_{eff}'[n'] n(\mathbf{r}) d\mathbf{r}, \tag{2.10}$$

where V_{eff}' is the effective potential to a trial electron density n'. This reformulation is necessary, as the ground state density is not known per se. The energy functional then

reads:

$$E[n] = \sum_{i=1}^{N} \epsilon_i - \int V'_{eff}(\mathbf{r})n(\mathbf{r}) + \int V_{ext}(\mathbf{r})n(\mathbf{r}) + \int \int \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{xc}[n] \qquad (2.11)$$

The first term is accounted as the single-particle contribution and all other commonly referred to as double-counting terms.

It is important to point out that the solving eigenvectors ϕ_i of the Kohn-Sham equation (2.8) have to be obtained self-consistently as the effective potential is a functional of the density itself.

2.4 Spin Density Functional Theory

For our purpose of incorporating magnetic effects in our studies, DFT has to be generalized to two spin channels. Going a step beyond the Hohenberg-Kohn approach, independently Barth and Hedin [63] and Pant and Rajagopal [64] extended the method by including spin dependent electron densities or alternatively a magnetization density $m(\mathbf{r})$ and generalized the Hohenberg-Kohn theorem to the spin-polarized case. Within the latter notation the variational principle becomes:

$$E[n(\mathbf{r}), m(\mathbf{r})] \geq E[n_0(\mathbf{r}), m_0(\mathbf{r})] = E_0, \qquad (2.12)$$

where the electron and magnetization density are defined by

$$\begin{aligned} n(\mathbf{r}) &= n^{\uparrow}(\mathbf{r}) + n^{\downarrow}(\mathbf{r}) \\ m(\mathbf{r}) &= n^{\uparrow}(\mathbf{r}) - n^{\downarrow}(\mathbf{r}). \end{aligned}$$

In general the two-component Pauli-spinors or the spin-density matrix $\rho_{\alpha\beta}$ as originally proposed [63, 64] are used to derive the spin dependent Kohn-Sham equations. Analogously to the derivation of the Kohn and Sham for DFT [27], the basic equation of SDFT turns out to be an effective single-particle Pauli-Schrödinger equation

$$\left\{ -\nabla^2 + V_{eff}^{\sigma}[n,m] \right\} \phi_i^{\sigma} = \epsilon_i^{\sigma} \phi_i^{\sigma}, \qquad (2.13)$$

where the two components ϕ^{σ} are coupled to each other and optionally to an external magnetic field, which can enter the effective potential V_{eff}. It is important to note, that we will not take into account non-collinearity or relativistic effects as spin-orbit coupling which couple both spin channels. Although the effective potential is determined by both channels of the spin dependent density n and m, equation (2.13) can then be solved independently for $\sigma = \uparrow$ and $\sigma = \downarrow$.

2.5 Exchange-correlation potential

Up to this point no further approximations beyond Born-Oppenheimer have been made and the groundstate properties could therefore in principle be calculated exactly. However, no exact expression for the exchange and correlation energy functional is known. The most

important approaches to find accurate approximations to the exchange-correlation potential are the local-density-approximation (LDA) and the generalized gradient approximation (GGA).

In the LDA − or in the spin-polarized case local spin density approximation (LSDA) − the functional E_{xc}^{LDA} is assumed to be locally approximated by the exchange-correlation energy of a homogeneous electron gas of the density $n(\mathbf{r})$ [65]. This procedure is similar to the Thomas-Fermi-approximation for the kinetic energy of an inhomogeneous electron system. Integrating the locally defined function of spin dependent electron densities over space yields the total exchange-correlation energy:

$$E_{xc}^{LSD}[n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})] = \int \epsilon_x^{hom}(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})) + \epsilon_c^{hom}(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})) d\mathbf{r}. \tag{2.14}$$

While the exchange energy ϵ_x^{hom} is known exactly for the homogeneous electron liquid, developing a parametrization of the correlation energy ϵ_c^{hom} is a highly non-trivial problem. Early perturbative approaches have been replaced by expressions obtained by quantum-Monte-Carlo calculations of the homogeneous electron gas [66] which are used in the most important parametrizations of the LDA e.g. in [67].

Due to the fact that any real system has a spatially varying electron density, the LSDA approach can be improved by considering gradient corrections.

$$E_{xc}^{GGA}[n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r})] = \int f\big(n_\uparrow(\mathbf{r}), n_\downarrow(\mathbf{r}), \nabla n_\uparrow(\mathbf{r}), \nabla n_\downarrow(\mathbf{r})\big) d\mathbf{r} \tag{2.15}$$

There are many different ways to construct the function f, one of the most reliable ones has been developed by Perdew *et al.* [68]. Besides this most commonly used LDA and GGA functionals part of the additional correlation effects can be described within the LDA+U-approach[37]. We will turn back to the discussion of this LDA+U functional in chapter 5.2.

KKR-Green function method

Korringa [28] and Kohn and Rostoker [29] introduced in 1947 and 1954 the basis for the nowadays used KKR electronic structure method. With this KKR multiple scattering approach for the solution of the Schrödinger equation it was feasible to solve systems of periodically repeated potentials and it was accordingly utilized for calculations of the band structure of ideally periodic crystals. One important characteristic which can be accounted as advantage above wave-function methods is already present at this stage of development: The clear and transparent conceptual separation of the single-site and multiple-scattering problem, which appears in multiple scattering theory.

An important extension of the KKR representation has been achieved by its reformulation in Green functions for the treatment of impurities in crystals [69, 70]. This, from then on widely used concept of impurity calculations, relies on the relation of locally confined impurity potentials to the ideal periodic potential by a Dyson equation. However, this concept does not only hold for impurity calculations but in general the Green functions of two systems can be set into relation to each other with potentially appealing numerical advantages. In fact, we will not consider any localized impurity potentials but rather focus on periodically repeated unit cells. The prior aim of this section is therefore the derivation of a Dyson equation which relates the potential-free system to a periodically repeated set of scattering potentials. In order to arrive at an expression which is valid for general scattering potentials, multiple atoms per unit cell and in reciprocal space we will conduct the following steps.

Since we aim at using the KKR Green function theory within the framework of density functional theory, first general properties of Green functions and its relation to the central quantity, the charge density, will be derived. Further, the single and multiple scattering problem will be addressed and important concepts introduced. With this insight the desired Dyson equation can be derived first for sets of spherical potentials, then for potential of general shape. To complete the discussion of all sub-steps of a self-consistency cycle, which will be summarizing this chapter in 3.7, we will point at an efficient scheme for energy integration (3.6). This above sketched and below given review on the KKR Green function method is in parts conceptually closely related to previous reviews [71–73], where secondary

topics are shortened or cut and important sections for the development of KKRnano are
added or extended.

3.1 Definition and general properties of the Green function

Consider a general Hamiltonian in atomic units ($\hbar=1$, $m_e = \frac{1}{2}$, $e=-\sqrt{2}$):

$$H = -\nabla^2 + V(\mathbf{r}). \tag{3.1}$$

Let this Hamiltonian describe the time evolution of an electron system, then in wave-function
arithmetics the time-dependent Schrödinger equation reads

$$i \frac{\partial}{\partial t}\psi(t) = H\psi(t). \tag{3.2}$$

Part of the well-known solution is the time-evolution operator e^{-iHt}. In close relation to
this the retarded and advanced Green function $G^{\mathrm{R/A}}(t)$ are solutions to the equivalent
differential equation

$$\left(i \frac{\partial}{\partial t} - H\right) G^{\mathrm{R}}(t) = \delta(t). \tag{3.3}$$

For our purposes the retarded Green function, with the formal solution

$$G^{\mathrm{R}}(t) = \begin{cases} 0 & \text{for } t < 0 \\ -i\ e^{-iHt} & \text{for } t \geq 0 \end{cases} \tag{3.4}$$

is chosen as the relevant one. From this expression a Fourier-transform of the Green function
leads to the convenient representation in energy space:

$$G(E) = \int_{-\infty}^{\infty} G(t)e^{i(E+i\gamma)t}dt, \tag{3.5}$$

with the solution

$$G(E) = \lim_{\gamma \to 0} (E + i\gamma - H)^{-1}. \tag{3.6}$$

Here, the parameter γ guarantees the convergency of the Fourier-transformation for $\gamma \to 0$
and $t \to \infty$. The singularities of $G(E) = (E - H)^{-1}$ are determined by the eigenvalues of
the Hamiltonian H. For $\mathrm{Im}E \neq 0$ the Green function is analytical in the complex plane –
a property, which will be made used of, in the following.

In the spirit of density-functional theory a general formalism how the electron density can
be obtained from a general Green function $G(E)$ is desirable. In the spectral representation
in the basis of a complete set of eigenfunctions $\mid \psi_i \rangle$ and corresponding eigenvalues ε_i the
Green function is given by

$$G(E) = \sum_i \frac{\mid \psi_i \rangle \langle \psi_i \mid}{E - \varepsilon_i + i\gamma}. \tag{3.7}$$

With explicit spatial coordinates this expression translates into

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_i \frac{\psi_i(\mathbf{r})\psi_i^*(\mathbf{r}')}{E - \epsilon_i + i\gamma}, \tag{3.8}$$

which can be interpreted as an outgoing wave at $\mathbf{r}$ with source at $\mathbf{r}'$. Making use of the Dirac-integral identity

$$\lim_{\gamma \to 0^+} \frac{1}{x \pm i\gamma} = \mathcal{P}\left(\frac{1}{x}\right) \mp i\pi\delta(x), \tag{3.9}$$

where $\mathcal{P}$ denotes the Cauchy principal part of the integral. Based on equation (3.8) the spectral and spatial electron density $n(\mathbf{r}; E)$ can be set into relation with the Green function by

$$
\begin{aligned}
n(\mathbf{r}; E) &= \sum_i \mid \psi_i(\mathbf{r}) \mid^2 \delta(E - \varepsilon_i) \\
&= -\frac{1}{\pi} \operatorname{Im} G(\mathbf{r}, \mathbf{r}; E).
\end{aligned}
\tag{3.10}
$$

From this result central quantities can be deduced, such as the local density of states $n(E)$ by spatial integration over $\mathbf{r}$ and most important the electron charge density $n(\mathbf{r})$ by energy integration. In addition (3.10) reveals the property of the Green function, that only the on-site parts ($\mathbf{r} = \mathbf{r}'$) are required to obtain the full charge density.

3.2 Single-Site Scattering

For the introduction to the scattering problem it is worthwhile studying the scattering at a single spherical atomic site embedded in a constant background potential. Here, the system of constant potential serves as a reference system for which the Green function is analytically known:

$$g^0(\mathbf{r}, \mathbf{r}'; E) = -\frac{1}{4\pi} \frac{e^{i\sqrt{E}|\mathbf{r}-\mathbf{r}'|}}{\mid \mathbf{r} - \mathbf{r}' \mid}. \tag{3.11}$$

In a spherical symmetric environment an angular momentum expansion is the most convenient description. For further reading, note that the angular momentum indices l and m will be commonly abbreviated by the combined index L. In this representation an incoming plane wave with $k = \sqrt{E}$ can be expanded in real spherical harmonics $Y_L = Y_{l,m}$:

$$\psi_{\mathbf{k}}^{inc}(\mathbf{r}) = e^{i\mathbf{k}\mathbf{r}} = \sum_L 4\pi i^l j_l(\sqrt{E}r) Y_L(\mathbf{r}) Y_L(\mathbf{k}), \tag{3.12}$$

where spherical Bessel functions j_l enter. Within the angular momentum expansion the free space Green function (3.11) is given by:

$$g^0(\mathbf{r}, \mathbf{r}'; E) = \sum_L Y_L(\mathbf{r}) g_l^0(\mathbf{r}, \mathbf{r}'; E) Y_L(\mathbf{r}'), \tag{3.13}$$

with

$$g_l^0(\mathbf{r}, \mathbf{r}'; E) = -i\sqrt{E} j_l\left(\sqrt{E}r_<\right) h_l\left(\sqrt{E}r_>\right), \tag{3.14}$$

where $h_l = j_l + i n_l$ are spherical Hankel function, n_l spherical Neumann functions and $\mathbf{r}_{<(>)}$ denote the smaller (bigger) of the radii $\mathbf{r}$ and $\mathbf{r}'$. For $\lim_{r\to 0}$ the Bessel functions $j_l(r)$ have a finite value, while $n_l(r)$ and subsequently $h_l(r)$ diverge at the origin $r = 0$.

The next conceptual step is the description of scattering at a single potential of finite range, defined as:

$$V(\mathbf{r}) = \begin{cases} V(r) & \text{for } r < S, \\ 0 & \text{otherwise.} \end{cases} \tag{3.15}$$

Then the radial wave-functions to $R_l(r; E)$ obey the radial Schrödinger equation

$$\left[-\frac{1}{r}\frac{\partial^2}{\partial r^2}r + \frac{l(l+1)}{r^2} + V(r) - E\right] R_l(r; E) = 0. \tag{3.16}$$

The radial solution to $R_l(r; E)$ can be determined by the Lippmann-Schwinger equation, which can be solved iteratively in a Born-series.

$$R_l(r; E) = j_l(\sqrt{E}r) + \int_0^S g_l^0(r, r'; E) V(r') R_l(r'; E) r'^2 dr', \tag{3.17}$$

which can be rewritten with the definition of $g_l^0(r, r'; E)$ in (3.14) for $r > S$ as

$$R_l(r; E) = j_l(\sqrt{E}r) - i h_l(\sqrt{E}r)\sqrt{E}\int_0^S j_l(\sqrt{E}r') V(r') R_l(r'; E) r'^2 dr'. \tag{3.18}$$

The integral is equivalent to the t-matrix element

$$t_l(E) = \int_0^S j_l(\sqrt{E}r) V(r) R_l(r; E) r^2 dr, \tag{3.19}$$

which can be interpreted as the operator connecting the incoming free wave $j_l(\sqrt{E}r)$ with the waves being scattered at the potential $V(r)$. Using this definition the Lippmann-Schwinger equation for $r > S$ reads in short form

$$R_l(r; E) = j_l(\sqrt{E}r) - i\sqrt{E} t_l(E) h_l(\sqrt{E}r). \tag{3.20}$$

In the other limit of $\lim_{r\to 0}$ (3.17) reads

$$R_l(r; E) \cong j_l(\sqrt{E}r) - i\sqrt{E} j_l(\sqrt{E}r)\int_0^S h_l(\sqrt{E}r') V(r') R_l(r'; E) dr'. \tag{3.21}$$

Hence, for very small radii the regular solution $R_l(r; E)$ is determined by the Bessel function $\lim_{r\to 0} j_l(\sqrt{E}r) \cong r^l$. Accordingly, (3.21) is usually expressed by

$$R_l(r; E) \cong j_l(\sqrt{E}r)\alpha_l(E), \tag{3.22}$$

where the α-matrix enters which is defined by

$$\alpha_l(E) = 1 - i\sqrt{E} \int_0^S h_l(\sqrt{E}r')V(r')R_l(r';E)r'^2 dr'. \tag{3.23}$$

Those α-matrices can be interpreted as modulation of the free incoming waves by the scattering potential for small r. They will be of particular importance in the derivation of Lloyd's formula in section 5.1. In analogy to $R_l(r;E)$ an irregular solution of the Schrödinger equation exists, which does not describe a valid physical picture of scattering, but is important to describe the Green function. This irregular solution is defined by

$$H_l(r;E) = \tilde{H}_l(r;E) + \int_0^S g_l^0(r,r';E)V(r')H_l(r';E)r'^2 dr'. \tag{3.24}$$

In opposite to the regular solution the boundary condition of an outgoing unperturbed spherical wave has to be fulfilled, which results in

$$H_l(r;E) = h_l(\sqrt{E}r) \ \ (r \geq S). \tag{3.25}$$

By application of this requirement the choice of $\tilde{H}_l(r;E)$ can be directly determined as

$$\tilde{H}_l(r;E) = h_l(r;E)\left[1 - \sqrt{E}\int_0^S j_l(\sqrt{E}r')V(r')H_l(r';E)r'^2 dr'\right]$$

In the limit of small r the irregular solution $H_l(r;E)$ diverges:

$$H_l(r;E) \cong h_l(\sqrt{E}r)\frac{1}{\alpha_l(E)}, \tag{3.26}$$

where for $\lim_{r\to 0}$ the radial dependency is $h_l(\sqrt{E}r) \propto r^{-l-1}$. In the basis of the regular and irregular solution, we can define the Green function describing the single-site scattering at a spherical potential in analogy to (3.14) by

$$G(\mathbf{r},\mathbf{r}';E) = -i\sqrt{E}\sum_L R_l(r_<;E)H_l(r_>;E)Y_L(\mathbf{r})Y_L(\mathbf{r}') \tag{3.27}$$

3.3 Multiple-Site Scattering

In this section the scattering theory will be generalized to a set of scattering potentials. We will obtain a secular equation which connects the amplitudes of incoming and outgoing waves. This equation is the basis to evaluate the band structures of periodic crystals in the KKR methodology. Although for large systems as treated with KKRnano band-structure calculations are not our aim, the below presented concepts will be of crucial importance for a deeper understanding of the KKR representation of the Green function which will be introduced in section 3.4.

The following derivation will be conducted in the muffin-tin (MT) approximation, in which the potentials are assumed to be spherically symmetric around all scattering centers and non-overlapping. The spherically symmetric potential is restricted within the MT as

defined by the MT radius (R_{MT}). The remaining interstitial part is assumed to be constant and accordingly can be set to zero. Given a periodic structure of MT potentials which are centered at lattice sites $\mathbf{R}^n$, each outgoing wave from site $\mathbf{R}^n$ can be expanded in a basis of incoming waves at site $\mathbf{R}^{n'}$ exploiting an addition theorem for Hankel functions [74]:

$$h_L(\mathbf{r}' + \mathbf{R}^{n'} - \mathbf{R}^n; E) = \frac{i}{\sqrt{E}} \sum_{L'} g_{LL'}^{0,nn'}(E) \, j_{L'}(\mathbf{r}; E), \tag{3.28}$$

and using the abbreviations

$$
\begin{aligned}
j_L(\mathbf{r}; E) &= j_l(\sqrt{E}r) \, Y_L(\mathbf{r}) \\
h_L(\mathbf{r}; E) &= h_l(\sqrt{E}r) \, Y_L(\mathbf{r}).
\end{aligned}
\tag{3.29}
$$

The expansion coefficients, the structure constants, are defined by

$$g_{LL'}^{0,nn'}(E) = -(1 - \delta_{nn'}) \, 4\pi i \sqrt{E} \sum_{L''} i^{l-l'+l''} C_{LL'L''} \, h_{L''}(\mathbf{R}^n - \mathbf{R}^{n'}; E), \tag{3.30}$$

with the Gaunt-coefficients

$$C_{LL'L''} = \int d\Omega \, Y_L(\mathbf{r}) \, Y_{L'}(\mathbf{r}) \, Y_{L''}(\mathbf{r}), \tag{3.31}$$

which are zero for $l'' > l' + l$. This limits the summation in (3.30) to a finite series. In the above notation the free electron Green function (3.11) can be expressed as

$$
\begin{aligned}
g^0(\mathbf{r} + \mathbf{R}^n, \mathbf{r}' + \mathbf{R}^{n'}, E) &= -i\sqrt{E} \, \delta_{nn'} \sum_L j_L(\mathbf{r}_<, E) \, h_L(\mathbf{r}_>, E) \\
&\quad + \sum_{LL'} j_L(\mathbf{r}; E) \, g_{LL'}^{0,nn'}(E) \, j_{L'}(\mathbf{r}'; E)
\end{aligned}
\tag{3.32}
$$

We proceed by considering an outgoing wave ψ^{sc} in the interstitial after a scattering event at R^n which can be written as

$$\psi_{\mathbf{k}}^{\mathrm{sc},n}(\mathbf{r}) = \sum_L b_{\mathbf{k},L}^{\mathrm{sc},n} \, h_L(\mathbf{r}; E). \tag{3.33}$$

The same wave can be interpreted as incoming wave ψ^{in} at site $\mathbf{R}^{n'}$ defined by

$$\psi_{\mathbf{k}}^{\mathrm{in},n'}(\mathbf{r}') = \sum_L b_{\mathbf{k},L}^{\mathrm{in},n'} \, j_L(\mathbf{r}'; E). \tag{3.34}$$

Using this equivalence both above expression can be set into relation

$$\sum_L b_{\mathbf{k},L}^{\mathrm{in},n'} \, j_L(\mathbf{r}'; E) = \sum_L b_{\mathbf{k},L}^{\mathrm{sc},n} \, h_L(\mathbf{r}' + R^{n'} - R^n; E), \tag{3.35}$$

which reads by exploiting the identity (3.28):

$$b_{\mathbf{k},L}^{\text{in},n'} = \frac{i}{\sqrt{E}} \sum_{L'} g_{LL'}^{0,nn'} \, b_{\mathbf{k},L'}^{\text{sc},n}. \tag{3.36}$$

Due to the periodicity of the lattice the Bloch theorem[1] can be exploited which connects the scattering amplitudes of two waves at position $\mathbf{R}^n$ and $\mathbf{R}^{n'}$ by a phase factor of $e^{i\mathbf{k}(\mathbf{R}^n - \mathbf{R}^{n'})}$. Instead of the previously used scattering amplitudes b in (3.33) to (3.36) of waves scattered at single sites, now the total wave which is originated from all identical scatterers is considered. The coefficients c which describe this total incoming and scattered waves then obey

$$c_{\mathbf{k},L}^{\text{in},n'} = \frac{i}{\sqrt{E}} \sum_{L'} \sum_{n \neq n'} g_{LL'}^{0,nn'}(E) \, e^{i\mathbf{k}(\mathbf{R}^n - \mathbf{R}^{n'})} \, c_{\mathbf{k},L'}^{\text{sc},n'} = \frac{i}{\sqrt{E}} \sum_{L'} g_{LL'}^{0}(\mathbf{k}; E) \, c_{\mathbf{k},L'}^{\text{sc},n'}, \tag{3.37}$$

where the $\mathbf{k}$-dependent structure constants are given by the Fourier transform of its real space representation

$$g_{LL'}^{0}(\mathbf{k}; E) = \sum_{n \neq n'} g_{LL'}^{0,nn'}(E) \, e^{i\mathbf{k}(\mathbf{R}^n - \mathbf{R}^{n'})}. \tag{3.38}$$

It is noteworthy to point out that $g_{LL'}(\mathbf{k}; E)$ exclusively depends on the position of the scattering centers of the lattice and not on the scattering potential.

The connection of the total incoming wave to the total scattered wave can be expressed by the t-matrix elements $t_l(E)$. Under the assumption that no additional external waves are contributing to the total incoming wave this relation reads:

$$c_{\mathbf{k},L}^{\text{sc},n'} = -i\sqrt{E} \, t_l(E) \, c_{\mathbf{k},L}^{\text{in},n'}. \tag{3.39}$$

Combining (3.39) and (3.37) then results in the following system of equations

$$\sum_{L'} \left(\delta_{LL'} - g_{LL'}^{0}(\mathbf{k}; E) \, t_{l'}(E) \right) c_{\mathbf{k},L'}^{n'} = 0, \tag{3.40}$$

which eigenvalues can be obtained under the necessary and sufficient condition

$$\text{Det}\left[\delta_{LL'} - g_{LL'}^{0}(\mathbf{k}; E) \, t_{l'}(E) \right] = 0. \tag{3.41}$$

This relation is usually referred to as KKR secular equation.

3.4 KKR representation of the Green function

As discussed in the previous section, it is not necessary to solve the KKR secular equation (3.41) to obtain the electronic structure of the system. Instead, according to (3.10) the Green function of the system provides the full information on spatial distribution and energy dependency of the electronic states. In this section the KKR multiple scattering

---

[1]Here, we restrict us for the moment to the simplest case of one atom per unit cell.

representation of this Green function will be introduced for non-overlapping periodically repeated muffin-tin potentials

$$V(\mathbf{R}^n + \mathbf{r}) = V^n(r), \tag{3.42}$$

while the generalization to the more cumbersome full-potential treatment will be conducted in the next section. The Green function of this system is defined by

$$\left(-\nabla^2 + V^n(r) - E\right) G(\mathbf{R}^n + \mathbf{r}, \mathbf{R}^{n'} + \mathbf{r}; E) = -\delta_{nn'}\,\delta(\mathbf{r} - \mathbf{r}'). \tag{3.43}$$

The solving Green function can then be expressed as a sum of the general homogeneous solution for $n \neq n'$ and the inhomogeneous solution for $n = n'$, which reads in the mixed site-angular momentum representation

$$
\begin{aligned}
G(\mathbf{r} + \mathbf{R}^n, \mathbf{r}' + \mathbf{R}^{n'}, E) = &-i\sqrt{E}\,\delta_{nn'}\sum_L R_L^n(\mathbf{r}_<, E)\,H_L^n(\mathbf{r}_>, E) \\
&+ \sum_{LL'} R_L^n(\mathbf{r}; E)\,G_{LL'}^{nn'}(E)\,R_{L'}^{n'}(\mathbf{r}'; E).
\end{aligned} \tag{3.44}
$$

Here, the abbreviation $R_L^n(\mathbf{r}; E) = R_L^n(r; E)\,Y_L(\mathbf{r})$ and $H_L^n(\mathbf{r}; E) = H_L^n(r; E)\,Y_L(\mathbf{r})$ for the regular and irregular solution of the Schrödinger equation is used. The coefficients $G_{LL'}^{nn'}(E)$ are the structural Green functions, which remain to be determined. It is noteworthy that $G_{LL'}^{nn'}(E)$ carries the full information on the multiple-scattering interaction, while all remaining functions are well-defined by the single-site scattering problem. Keeping in mind that we aim at massive real space parallelization, this clear separation of single-site and multiple-scattering contributions is one of the central advantages of the KKR methodology.

For the following step it is important to recall that the structural Green functions for the potential-free case $g_{LL'}^{nn'}(E)$ have already been introduced in (3.30) and are known analytically. As first introduced by Dupree [69] the required coefficients $G_{LL'}^{nn'}(E)$ can be set into relation to $g_{LL'}^{nn'}(E)$ by the Dyson equation

$$G_{LL'}^{nn'}(E) = g_{LL'}^{0,nn'}(E) + \sum_{n'',L''} g_{LL''}^{0,nn''}(E) t_{l''}^{n''}(E) G_{L''L'}^{n''n'}(E), \tag{3.45}$$

where t-matrix $t_{l''}^{n''}$ enters, which is defined in (3.19). Expanding (3.45) on the right hand side reveals the physical background. While the first order takes into account the direct propagation from n to n', higher order terms resemble paths from n to n' via one or more scattering centers n''.

Although (3.45) is already close to the relation which will be used in KKR**nano** two important extensions have to be made: First, the periodicity of the set of scatterers can be exploited and k-dependent structural Green functions considered in similar way as introduced in (3.38). This means the Dyson equation will in practice be solved in k-space and then Fourier-transformed to obtain $G_{LL'}^{nn'}(E)$. Restricting us for the moment to one atom in the periodically repeated unit cell the t-matrix becomes independent of n and (3.45) reads

$$G_{LL'}(\mathbf{k}; E) = g_{LL'}^0(\mathbf{k}; E) + \sum_{L''} g_{LL''}^0(\mathbf{k}; E)\,t_{l''}(E)\,G_{L''L'}(\mathbf{k}; E). \tag{3.46}$$

As a second step the generalization to an arbitrary number of atoms per unit cell N has to be performed. This step is straightforward when η is introduced to cover the periodicity of the unit cell and the second index $n = 1 \ldots N$ is limited to describe the atomic sites within the cell. The generalization of the Fourier transform (3.38) then reads

$$g_{LL'}^{0,nn'}(\mathbf{k}; E) = \sum_{\eta \neq \eta'} g_{LL'}^{0,\eta n\eta'n'}(E)\, e^{i\mathbf{k}(\mathbf{R}^\eta - \mathbf{R}^{\eta'})}. \tag{3.47}$$

Accordingly the Dyson equation changes to

$$G_{LL'}^{nn'}(\mathbf{k}; E) = g_{LL'}^{0,nn'}(\mathbf{k}; E) + \sum_{n''L''} g_{LL''}^{0,nn''}(\mathbf{k}; E)\, t_{l''}^{n''}(E)\, G_{L''L'}^{n''n'}(\mathbf{k}; E), \tag{3.48}$$

where t-matrices now depend on the site n of the represented unit cell. The number of matrix elements $g_{LL''}^{0,nn''}(\mathbf{k}; E)$ grows with N^2 with number of atoms per unit cell N and accordingly the computational effort for a direct solution of (3.48) increases with $O(N^3)$. Apparent from this unfavorable scaling law, solving (3.48) with alternative schemes will be of high importance to enable large-scale calculations with thousands of atoms per unit cell. Before the implementation of more advanced techniques is discussed in chapter 4, first the algorithm will be extended to non-spherical and space-filling potentials.

3.5 Full-potential description

In case of a full-potential treatment the simple partitioning in space as used in the atomic-sphere-approximation has to be refined. In order to get a representation of the crystal in non-overlapping space-filling cells the crystal is partitioned in Wigner-Seitz cells by a Voronoi construction. Mathematically this procedure requires the introduction of step-functions on all different sites n, which have a finite value in n and are zero everywhere else

$$\Theta^n(\mathbf{r}) = \begin{cases} 1 & \text{for } \mathbf{r} \text{ in the Wigner-Seitz-cell of site } n, \\ 0 & \text{otherwise} \end{cases} \tag{3.49}$$

Those functions $\Theta^n(\mathbf{r})$ are in the following referred to as shape-functions and are used to find a continuous description of the crystal potential

$$V^n(\mathbf{r}) = V(\mathbf{r} + \mathbf{R}^n)\Theta^n(\mathbf{r}). \tag{3.50}$$

Consequently also the shape-functions are expressed in the same angular momentum expansion as the scattering events:

$$\Theta^n(\mathbf{r}) = \sum_L \Theta_L^n(\mathbf{r})Y_L(\mathbf{r}). \tag{3.51}$$

Dependent on the lattice geometry the convergency of this expansion can be rather slow. However, the integration over Gaunt coefficients ensures that at a given cutoff in l for the scattering events of $l_{\max}$ the expansion in (3.51) can be safely stopped at $4 \cdot l_{max}$ as higher contributions are zero by symmetry. It is important for practical applications that for some

classes of materials e.g. interstitial sites might have to be incorporated to guarantee a proper convergency of the lm-expansion. Although shape-functions present an important step in the practical realization of a full-potential KKR-theory implementational details are spared here. For the following derivation we indicate the spatial integration over a shape-function by denoting the upper limit of integration to the Wigner-Seitz cell by the symbol WS.

For the first conceptual step we go back to the scattering theory of a single potential within a free host. Suppose this single potential $V(\mathbf{r})$ is anisotropic and extends - for the sake of simplicity − up to finite maximal radius S. An incoming wave of wave-vector $\mathbf{k}$, which scattered at this potential then results in a wave $\psi_{\mathbf{k}}(\mathbf{r})$, which reads after expansion in spherical harmonics:

$$\psi_{\mathbf{k}}(\mathbf{r}) = \sum_{L'} 4\pi i^{l'} Y_{L'}(\mathbf{k}) R_{L'}(\mathbf{r}; E). \tag{3.52}$$

Here, $R_{L'}(\mathbf{r}; E)$ is the regular solution of the Schrödinger equation, which corresponds to an incoming spherical wave of symmetry L'. In analogy to the spherical symmetric case discussed above, the Lippmann-Schwinger equation can be generalized to $R_{L'}(\mathbf{r}; E)$ as:

$$R_{L'}(\mathbf{r}; E) = j_{l'}(\sqrt{E}r)\, Y_{L'}(\mathbf{r}) + \int_0^S g(\mathbf{r}, \mathbf{r}'; E) V(\mathbf{r}') R_{L'}(\mathbf{r}') d^3 r', \tag{3.53}$$

where the first term resembles an incoming free wave and the integration is performed over the free-electron Green functions $g(\mathbf{r}, \mathbf{r}'; E)$. To benefit most from the cite-centered coordinates of this problem by lifting the directional dependency of the wave-function $R_{L'}(\mathbf{r}; E)$ and the potential $V(\mathbf{r})$ both quantities can be expanded in spherical harmonics as well:

$$R_{L'}(\mathbf{r}; E) = \sum_L R_{LL'}(r; E) Y_L(\mathbf{r})$$
$$V(\mathbf{r}) = \sum_L V_L(r) Y_L(\mathbf{r}).$$

With this expression equation (3.53) can be rewritten as one-dimensional integral instead of a three-dimensional one:

$$R_{LL'}(r; E) = j_{l'}(\sqrt{E}r)\, \delta_{LL'} + \int_0^S g(r, r'; E) \sum_{L''} V_{LL''}(r')\, R_{L''L'}(r')\, r'^2\, dr', \tag{3.54}$$

with the following notation for the potential coefficients

$$V_{LL'}(r) = \sum_{L''} C_{LL'L''} V_{L''}(r). \tag{3.55}$$

The regular wave-function coefficients $R_{LL'}(r; E)$ are then defined by the following Schrödinger equation

$$\sum_{L''} \left[\left(-\frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{l(l+1)}{r^2} - E \right) \delta_{LL''} + V_{LL''} \right] R_{L''L'}(r; E) = 0. \tag{3.56}$$

Instead of solving this coupled differential equation an alternative procedure can be used. Taking into account that the non-spherical contribution to the potential is usually small

compared to the spherical part, first the radial solution $R_l(r; E)$ and $H_l(r; E)$ to the spherical potential are calculated. To obtain the non-spherical solution $R_{LL'}(r; E)$ by adding the non-spherical part via a Lippmann-Schwinger equation

$$R_{LL'}(r; E) = R_l^{sph}(r; E)\delta_{LL'} + \int_0^{WS} G_l^{sph}(r, r'; E) \sum_{L''} \Delta V_{LL''}(r') R_{L''L'}(r'; E)r'^2 dr', \quad (3.57)$$

where exclusively the non-spherical part of the potential is added as a perturbation

$$\Delta V_{LL''}(r') = \sum_{L'' \neq 0} C_{LL'L''} V_{L''}(r), \quad (3.58)$$

and $G_l^{sph}(r, r'; E)$ is the spherical Green function defined by

$$G(\mathbf{r}, \mathbf{r}'; E) = -i\sqrt{E} \sum_L R_l^{sph}(r_<; E) H_l^{sph}(r_>; E) \quad (3.59)$$

In the exact same fashion the non-spherical irregular solution $H_{LL'}(r; E)$ can be obtained. Within all Jülich implementations of the KKR method including KKRnano the Lippmann-Schwinger equation is solved iteratively using as starting values the spherical solution in a Born series of typically 3 to 4 steps.

For the generalization of the Dyson equation to anisotropic potentials only the spherical symmetric t-matrix $t_l^n(E)$ has to be replaced by a t-matrix $t_{LL'}(E)$ for a general potential $V^n(\mathbf{r})$:

$$t_{LL'}(E) = \int_0^{WS} j_l(\sqrt{E}r) \sum_{L''} V_{LL''} R_{L''L'}(r'; E) r'^2 dr', \quad (3.60)$$

where the potential has been expanded in spherical harmonics as discussed above. In full-potential description the Dyson equation then reads:

$$G_{LL'}^{nn'}(\mathbf{k}; E) = g_{LL'}^{0,nn'}(\mathbf{k}; E) + \sum_{n''L''L'''} g_{LL''}^{0,nn''}(\mathbf{k}; E) t_{L''L'''}^{n''}(E) G_{L'''L'}^{n''n'}(\mathbf{k}; E), \quad (3.61)$$

where n and n' run over the sites in the unit cell. The structural Green functions as such already depend on L and L' and consequently do not have to be modified as compared to the spherically symmetric formulation in (3.48). (3.61) represents the cornerstone of our investigations in the following chapter on the development of KKRnano. However, for the understanding of part of the challenges we will be facing there, it is important to review how the energy integration over the Green function is best performed in the KKR approach.

3.6 Finite-temperature complex contour integration

The charge density of a solid can be found by energy integration up to the Fermi level E_F:

$$n(\mathbf{r}) = -\frac{1}{\pi} \text{Im} \int_{-\infty}^{E_F} G(\mathbf{r}, \mathbf{r}; E) dE \quad (3.62)$$

For most applications it is a practical choice to treat the core electrons in an atomic-like fashion and the valence electron as delocalized particles. This separation, which is in fact applied in most of the existing DFT methods and in KKRnano, allows to introduce a lower boundary E_B to the energy integration. E_B usually is selected to lie well above the highest core and below the lowest valence states. The integration then extends over

$$n(\mathbf{r}) = -\frac{1}{\pi}\text{Im}\int_{E_B}^{E_\text{F}} G(\mathbf{r},\mathbf{r};E)dE. \tag{3.63}$$

Due to fact that the Green function is strongly structured close to the real axis, a straight-forward integration would result in the need for hundreds of energy points. For an efficient evaluation of the integral in (3.63) we can exploit that the Green function is analytical for Im $E \neq 0$. Hence, the integration can be performed over a arbitrarily shaped complex contour starting from E_B and ending at the Fermi energy. Of further advantage is that the Green function is the slower varying with respect to E the larger the integration points are apart from the real axis. Hence if Im E is sufficiently large, only few sampling points are required. It remains to address the starting and ending point of the integration: While E_B was selected to be well isolated from electronic states and the Green function can be safely assumed to have little structure at this energy, at the end of the contour at E_F many energy points have to be included to cover the structure of $G(\mathbf{r},\mathbf{r};E)$. An elegant way to avoid this accuracy problem is the introduction of an artificial electronic temperature. This results in an integration over a Fermi-Dirac distribution:

$$f_T(E - E_\text{F}) = \frac{1}{e^{(E-E_\text{F})/kT} + 1}. \tag{3.64}$$

The presence of $f_T(E - E_\text{F})$ has two important consequences. First, $f_T(E - E_\text{F})$ ensures for values considerably larger than E_F that the integrand is zero. Thereby the contour integration can be formally extended up to infinity. Secondly, $f_T(E - E_\text{F})$ is analytical in the complex plane except for the poles (or Matsubara energies) at $E_j = E_\text{F} \pm i\pi(2j-1)kT$, which have to be included in the integration as residues. Within this scheme the integration can be calculated by

$$n(\mathbf{r}) = -\frac{1}{\pi}\text{Im}\int_{E_B}^{\infty} dE \, f_T(E - E_\text{F}) \, G(\mathbf{r},\mathbf{r};E). \tag{3.65}$$

In practice the integration is performed on a rectangular shaped contour, stopping at values slightly above E_F and including typically 4 to 6 Matsubara energies.

A typical contour for a given distribution of electronic states is depicted in Fig. 3.1. This integration scheme which explicitly minimizes the points close to the real energy axis is of utmost importance for an efficient usage of the iterative algorithm used in KKRnano. The convergency of the iterative inversion of the Dyson equation is shown in the following sections.
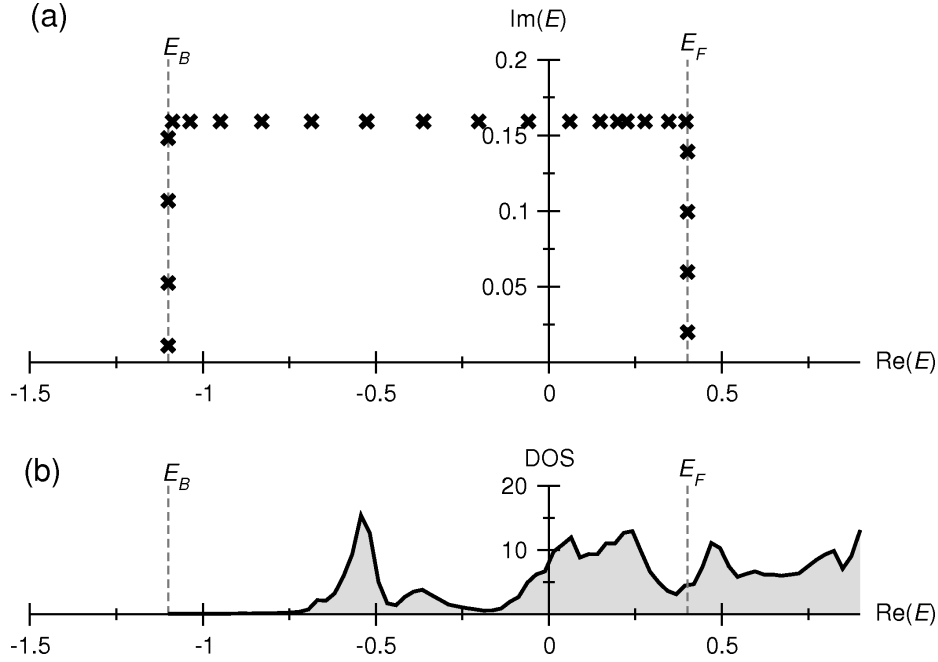
Figure 3.1.: *(a) Contour used for energy integration in the complex energy plane for a GeSbTe alloy with T=1000 K. Exemplary the density of states on one Te atom of the same sample is shown in (b). Here, the energies are given in units of Ry.*

3.7 Self-consistency cycle

After having completed the introduction of the most important equations in the KKR Green functions representation, we take a look how those are incorporated in the density-functional self-consistency cycle. The following nine steps build the basic frame of the self-consistency algorithm as utilized in existing KKR codes and implemented in KKRnano. Although in most existing methods and KKRnano a different choice of reference system is made (see section 4.1), here we restrict our considerations to the simplest case where the potential free space acts as reference system. The self-consistency cycle is then composed of the following steps:

1. Start with initial potential V^{in}

2. Calculate for each scattering site the single site properties to the potential V^{in}, i.e. $R_{LL'}$ and $t_{LL'}$ using (3.57) and (3.60), respectively.

3. Set up free space reference system $g_{LL'}^{nn'}$ according to (3.30)

4. Solve the Dyson equation (3.61) and integrate over k-space to obtain $G_{LL'}(E)$.

5. Calculate the Green function $G(\mathbf{r}, \mathbf{r}; E)$ from on-site elements $G^{nn}(E)$ by means of (3.44).

6. Perform the energy contour integration with finite electronic temperature to obtain the valence charge density by (3.65).

7. Calculate core states and the core contribution to the charge density.

8. Compute new potential V^{out} and total energies for each site according to the scheme introduced by Drittler [30].

9. If the difference of V^{in} and V^{out} is sufficiently small leave this cycle, otherwise mix V^{in} and V^{out} and proceed with step one.

From this view on the computational tasks, we can conclude that the KKR Green function method is due to its clear conceptual separation in single-site and only one multiple-scattering part (step 5) ideally suited for a real space parallelization. However, while the single-site part can be straightforwardly mapped on parallel processing units, the multiple scattering part is the bottleneck for large scale computation. Therefore, in the code development of KKRnano, which is presented in chapter 4 and 5, we focused to a great extent on the optimization of the treatment of this non-local part.

KKRnano

In the last decade a change of focus took place in the various fields dealing with solid state physics. While e.g. semiconductors defects play an important role since many decades, for example at surfaces often the aim was to fabricate samples with highest possible purification. Nowadays the main interest rather lying on defective structures of high geometrical complexity. This new view culminates by rather inducing defects on purpose in order to actively engineer material properties. Out of a rich field of applications, appealing examples are the fabrication of magnetic nanostructures down to the single atom [75], and the rich material properties of complex alloys [53] which are the basis for todays optical storage devices such as DVD's.

The theoretical and computational treatment of such complex structures is a challenging task. This is in particular true if high accuracy and predictive power is required which can be provided by *ab initio* density-functional methods. Most commonly used for the *ab initio* description of complex compounds or defects are three methods: first, impurity Green function methods, which embed an impurity region in an ideal host structure [2, 76]. This mathematically elegant approach allows for an efficient computation of single impurities or clusters but nevertheless runs into problems as soon as the dilute limit is left and the interaction of defects becomes crucial. Secondly, the interactions of defects can be mapped onto effective interactions as done within the coherent-potential approximation (CPA) [6, 77]. In spite of the success of the CPA as mean-field theory for alloys, certain correlation and finite size effects as well as more complex defects departing from the lattice structure cannot be treated sufficiently within either the CPA or the more advanced non-local-CPA correlation [8]. Thirdly, supercell techniques, which avoid any geometrical approximations have been extremely successful and widely used in the last two decades. The main concern about this supercell approach is to minimize finite size effects such as the spurious interaction with periodic images. In order to guarantee precise results with respect to those finite size features the supercells have to be large. This points at one of the major shortcomings of computational solid state physics. Density functional calculations with large supercells of the order of up to 1000 atoms are still the exception and performed only with few methods: Without aiming for completeness and focussing on general purpose methods, VASP [9] in
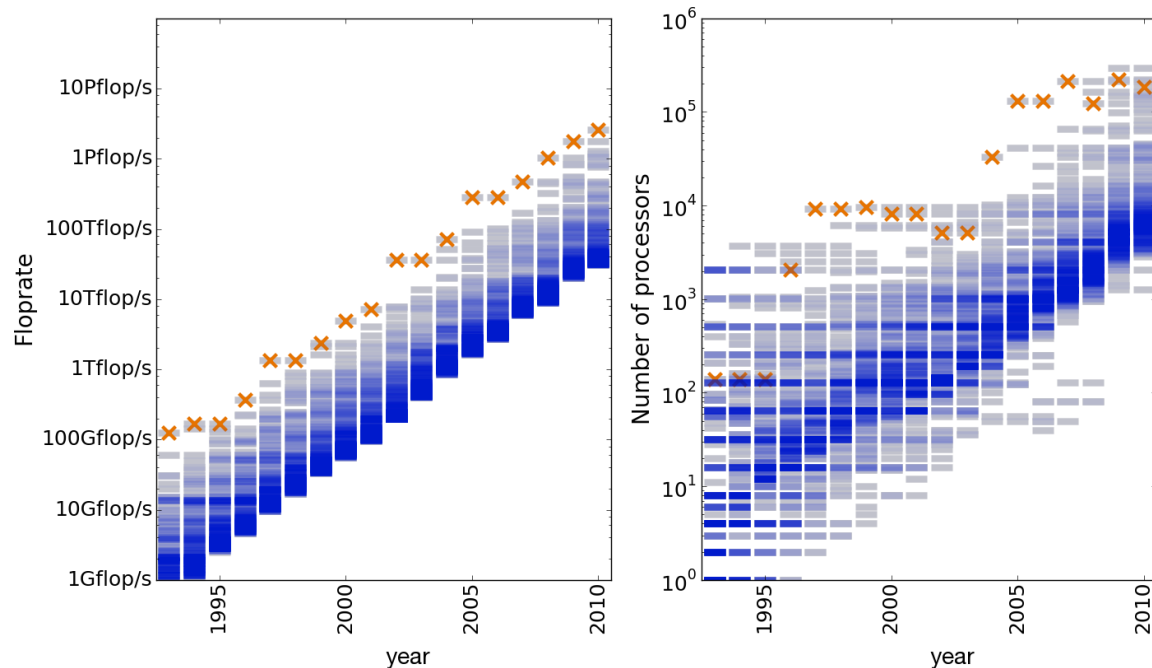
Figure 4.1.: *Benchmark floprate (left) [81] and number of processors (right) of the 500 fastest supercomputers worldwide. For each of those computers a transparent blue bar visualizes its key properties − accordingly dark blue regions correspond to multiple computers with similar floprate or number of processors. The fastest computer with respect to floprate is highlighted in yellow. All raw data are taken from [15].*

many different applications e.g. [78, 79], SIESTA [10, 80] and the LSMS method [25] have been exploited for this task. However, the vast majority of the existing electronic structure methods are still limited to a few hundreds of atoms in the supercell and often practical applications are rather restricted to less than one hundred atoms.

It is the aim of this PhD thesis to present the development of a code which enables efficient calculations of supercells containing many thousands of atoms combined with the advantages of the KKR Green function methodology. In order to achieve this goal our algorithm is especially designed to run on massively parallel computer architectures. This code will be called KKRnano from now on, which reflects the ability to compute supercells with spatial extends of more than one nanometer.

In order to enable a fast computation of large supercells we target at a high parallel performance up to many thousands of processors, which is of course motivated by the recent developments in supercomputing. The fastest computers worldwide comprise nowadays massively parallel architectures with at least tens of thousands processors. Concerning the recent development of supercomputers a couple of interesting facts can be deduced from Fig. 4.1. First, the increase in flop-rate[1] by four orders of magnitude over the last 15 years into the Petaflop-regime is to a great extent achieved by increasing the level of parallelization. E.g. in 1995 the fastest computers combined the power of a few hundred CPU's while in 2010

---

[1]Flop is used as abbreviation for floating point operation. Accordingly the flop-rate is the number of flops per second.

	Nodes	Cores	Floprate (Tflop/s)	Peak Floprate (Tflop/s)
JUGENE FZ Jülich, Germany [82]	73728	294912	826	1003
JUROPA FZ Jülich, Germany [83]	2208	17664	184	207
JAGUAR^{XT5} Oak Ridge, USA [84]	18688	224162	1759	2331
BlueGene/P Rochester, USA	4096	16384	48	56

Table 4.1.: *Hardware characteristics of the supercomputers utilized for this thesis.*

more than one hundred thousand processors are integrated. The overall rapid upgrowth of the performance of supercomputers is best pictured by the following example. The fastest supercomputer in any given year was not anymore on the Top500 list [15] after a period of only six to seven years. This observation allows us to conclude that if the power of the latest generations of supercomputers should be utilized also in the next couple of years, codes have to be flexible and parallel efficient up to hundreds of thousands of processors. One of the key strategies to reach this task is finding algorithms which minimize the communication between processes and avoiding input/output operations as much as possible.

In order to guarantee such high flexibility KKR**nano** has been tested on several supercomputers as shown in Tab. 4.1 and extensively used on both JUGENE [82] and JUROPA [83]. Besides the obvious demand for massive parallelization it is important to note, that the available memory per node or CPU can be a serious limitation. E.g. on the BlueGene/P architecture only 2048 MByte per node and accordingly 512 MByte on each of the four processors can be accessed fast without using the network. Therefore it will be of utmost importance to establish a standard of memory saving programming.

In this chapter we will discuss the crucial strategies which are used in KKR**nano**, such as introducing the screened reference system, the iterative solution of the Dyson equation and its parallelization. Further we will show how linear scaling can be achieved by truncation of the longer-range interactions.

4.1 Tight-Binding KKR-Green function method

In the traditional KKR-method the atomic geometry of interest has been set into relation to free-space structure constants $g_{LL'}^{0,nn'}$ in a Dyson equation as introduced in section 3.4. Although these free space structure constants are analytically known, their decay in real space is unfavorably slow. Therefore the matrix inversion that has to be performed in the KKR-approach leads to an $O(N^3)$ scaling algorithm with system size N.

An important improvement of the KKR method has been achieved by the development of the screened KKR formalism [31], often referred to as tight-binding KKR method. Instead of the single step from free space to the actual potential as defined by the Dyson equation (3.61)

$$G = g^0 + g^0\, t\, G \quad \Longleftrightarrow \quad (g^0)^{-1} = (G)^{-1} + t \tag{4.1}$$

an advantageous two-step procedure can be introduced. This procedure becomes possible as the Dyson-equation can connect the crystal structure constants to any kind of reference system of the same periodic structure. Here, repulsive spherical potentials are defined on each scattering center, which are placed in the same geometry as the lattice sites of the actual system as shown in Fig. 4.2(a). These potentials define a reference t-matrix t^{r}, which is set into relation to the crystal potential t via the t-matrix difference $\Delta t = t - t^{\mathrm{r}}$. With this input first the Dyson equation

$$G^{\mathrm{r}} = g^0 + g^0\, t^{\mathrm{r}}\, G^{\mathrm{r}} \quad \Longleftrightarrow \quad (g^0)^{-1} = (G^{\mathrm{r}})^{-1} + t^{\mathrm{r}} \tag{4.2}$$

connecting the free-space structure constants g with the reference structure constants G^{r} has to be solved. As a second step the reference system and crystal structure constants obey

$$G = G^{\mathrm{r}} + G^{\mathrm{r}}\, \Delta t\, G \quad \Longleftrightarrow \quad (G^{\mathrm{r}})^{-1} = (G)^{-1} + \Delta t \tag{4.3}$$

The validity of this equation can be verified by equalizing (4.1) and (4.2), using the definition of Δt.

The advantage of this alternative procedure lies in the favorable properties of the reference structure constants G^{r}. Within those previously defined potentials the G^{r} structure constants decay for the energy range of interest exponentially in space instead of the $1/r$ decay of the free-space scattering matrix g^0. For the analysis of the screening of the reference potential the partial norm of the structure constants has been defined by Zeller *et al.* [31] as

$$N_{ll'}\left(|\mathbf{R}_i - \mathbf{R}_j|\,; E\right) = \frac{|E|^{\frac{l+l'}{2}}}{(2l+1)!!\,(2l'+1)!!} \cdot \left[\sum_{mm'} \left|G_{lm,l'm'}^{\mathrm{r},nn'}\right|^2\right]^{\frac{1}{2}}, \tag{4.4}$$

where $\mathbf{R}_n$ and $\mathbf{R}_{n'}$ are the position vectors of the centered site and all other site centers, respectively. The partial norm has been calculated for an fcc test system within a finite reference cluster and the resulting spatial decay is shown in Fig. 4.2(b). This result clearly
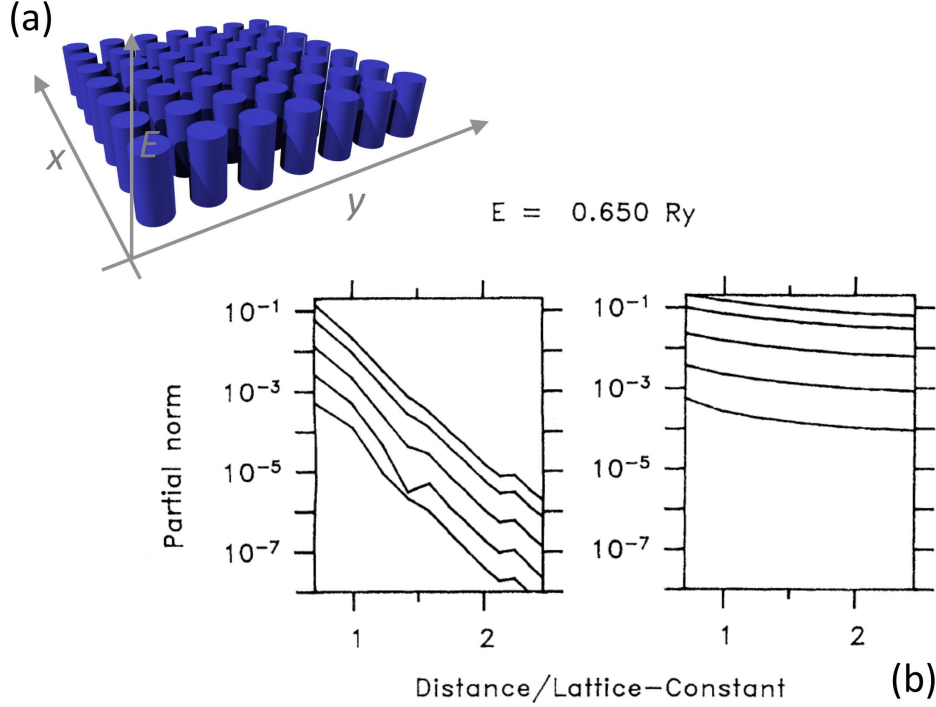
Figure 4.2.: (a) Illustration of repulsive spheres as projected on a two-dimensional plane spanned by the x- and y-axis. The third dimension corresponds to the strength of the potential. (b) Figure from [31]. Screened (left) and unscreened (right) partial norms of the reference potential structure constants G^{r} for $l = l'$ as defined in (4.4). Partial norms for an energy of 0.65 Ryd and a screening potential of 2 Ryd are shown as a function of distance in units of lattice constants. From top to bottom $l = 0, 1, 2, 3,$ and 4 are shown.

shows that G^{r} can be obtained within a relatively small cluster of a few ten to hundred scattering sites. In particular for large systems the screening of longer-range structure constants has the important effect that G^{r} becomes sparse.

The degree of sparsity is controlled by the radius R_{cls} defining the boundary of the finite reference cluster (see Fig. 4.3(a)) which contains N_{cl} sites and outside which the structure constants are assumed to be zero. As shown in Fig. 4.3(b) non-uniform lattice positions in the supercell arising e.g. due to structural relaxations require a computation of individually different reference systems for each site. In detailed notation this step reads

$$G^{\mathrm{r},nn'}_{LL'}(E) = g^{0,nn'}_{LL'}(E) + \sum_{n'',L''L'''} g^{0,nn''}_{LL''}(E)\, t^{\mathrm{r},n''}_{L''L'''}\, G^{\mathrm{r},n''n'}_{L'''L'}(E). \tag{4.5}$$

The computation of (4.5) is performed locally for each site and its individual surrounding cluster of sites and is therefore ideally suited to be distributed to independent processes. The second step is the connection of this reference system to the actual system via:

$$G^{nn'}_{LL'}(\mathbf{k}, E) = G^{\mathrm{r},nn'}_{LL'}(\mathbf{k}, E) + \sum_{n'',L''L'''} G^{\mathrm{r},nn''}_{LL''}(\mathbf{k}, E)\, \Delta t^{n''}_{L''L'''}\, G^{n''n'}_{L'''L'}(\mathbf{k}, E), \tag{4.6}$$
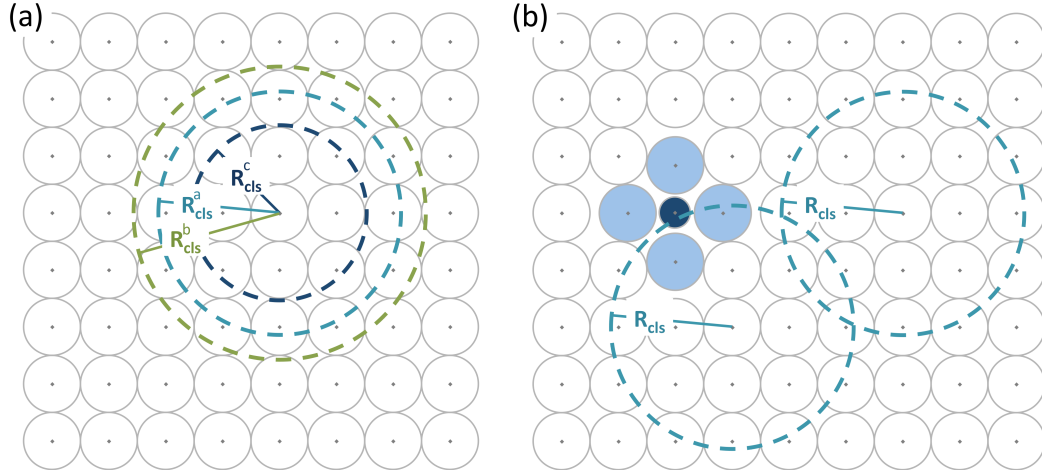
Figure 4.3.: *Schematic setup of tight-binding clusters exemplary for the (001)-plane of a simple cubic lattice. In (a) different choices of cluster defining radii $R_{cls}^b > R_{cls}^a > R_{cls}^c$ are shown resulting in 21, 13, and 9 atoms in the cluster exclusively in the shown plane. For distorted structures as visualized in (b) showing a point defect with next-nearest-neighbor relaxation different individual clusters have to be setup for each atomic center.*

where we introduced the k-dependency induced by the periodic boundary conditions of the supercell as in (3.61). It is important to underline that the sparsity of matrix G^r with the number of non-zero entries $\propto N_{cls} \cdot N$ is crucial to attain the defined goals: On the one hand for large systems of thousands of atoms a dense representation of this matrix could not even be stored in memory on supercomputers. On the other hand the number of floating point operations can be decisively reduced by exploiting the sparsity ideally.

4.2 Iterative solution of the Dyson equation

In the previous section we conducted the important step from a dense matrix equation with size $\propto N^2$ down to a sparse one with $\propto N_{\mathrm{cl}}N$ non-zero entries. While a straightforward direct inversion scales with N^3, by exploiting the sparsity this effort can be reduced using a proper scheme. Recalling that in addition a massive parallelization of this solver is inevitable for efficient computation, sparse direct inversion techniques which require intensive message passing interface (MPI) communication on the fly are not the ideal choice. It was previously shown that iterative techniques to solve the Dyson equation (4.6) can circumvent this bottleneck [32]. The iterative inversion can be an efficient scheme for sparse and parallel calculations as it takes intrinsically advantage of the sparsity and can be parallelized without any demands for communication during the inversion procedure. In the following derivation of an iterative ansatz the angular momentum and atomic site indices as well as energy and k-point dependencies will be dropped from this point on.

Starting from the Dyson equation (4.6)

$$G = G^{\mathrm{r}} + G^{\mathrm{r}}\Delta t G, \tag{4.7}$$

it is clear that G^{r} fulfills

$$G^{\mathrm{r}} = (1 - G^{\mathrm{r}}\Delta t)G. \tag{4.8}$$

With the identity

$$G^{\mathrm{r}} = -(1 - G^{\mathrm{r}}\Delta t)\Delta t^{-1} + \Delta t^{-1}. \tag{4.9}$$

it follows by multiplication with $(1 - G^{\mathrm{r}}\Delta t)^{-1}$

$$G = -(\Delta t)^{-1} + (\Delta t)^{-1}\left[(\Delta t)^{-1} - G^{\mathrm{r}}\right]^{-1}(\Delta t)^{-1}. \tag{4.10}$$

As the Δt matrices are site-diagonal all matrix-matrix multiplications in (4.10) are computationally inexpensive. The cumbersome part is computing the inverse of

$$M = (\Delta t)^{-1} - G^{\mathrm{r}}, \tag{4.11}$$

often referred to as KKR matrix or scattering path operator. In order to conduct the iterative inversion M and its inverse

$$X = M^{-1} = \left((\Delta t)^{-1} - G^{\mathrm{r}}\right)^{-1} \tag{4.12}$$

can be set into relation as

$$\Delta t M X = \Delta t. \tag{4.13}$$

Inserting the actual expression (4.11) of M leads to

$$X - \Delta t G^{\mathrm{r}} X = \Delta t \tag{4.14}$$

which can be solved using the scheme

$$X^{(i+1)} = \Delta t + \Delta t G^{\mathrm{r}} X^{(i)} \tag{4.15}$$

in an iterative cycle $X^0 \to X^1$, $X^i \to X^{i+1}$ in similar fashion as usually done for e.g. the Born iterations of the Lippmann-Schwinger equation. However, the convergency properties of such schemes are usually rather poor, motivating the choice of more sophisticated iterative methods. Most of those algorithms are formulated to solve a set of linear equations

$$A \cdot X = b. \tag{4.16}$$

In this context (4.15) can be reformulated as

$$\left(\Delta t G^{\mathrm{r}} - 1\right) X = -\Delta t, \tag{4.17}$$

which will be the basis for the iterative solution of the Dyson equation in KKRnano.

At this point the algorithm scales $\propto N^2 N_{\mathrm{cl}} N_{\mathrm{it}}$, where N_{it} stands for the number of iterations required to converge (4.17) down to a predefined accuracy. It is important to note that the linear matrix equation (4.17) decouples in N problems with a size of the solving matrix X of $N \cdot (l_{\max} + 1)^2 \times (l_{\max} + 1)^2$ each, which will be crucial for parallelization. In the following X will exemplify the solvent of one of those subproblems. For the discussion of scaling it is important to note, that we find, in agreement with [32], that N_{it} depends only weakly on the system size for $N > 1000$ and is assumed to be constant from now on. However, depending on the iterative scheme exploited and additionally on the electronic temperature T up to $N_{\mathrm{it}} \approx 1000$ iterations can be required for the energy points close to the real axis. For efficient computing it is crucial to select best suitable convergency procedures and to come up with schemes which reduce this N_{it} considerably. In the following sections we will describe the iterative scheme used in KKRnano and its optimization by e.g. finding sound starting vectors and preconditioning the matrix A.

4.2.1 Transpose Free Quasi Minimal Residual algorithm in KKRnano

In order to solve the linear equation (4.16) the transpose free quasi minimal residual (TFQMR) scheme [85] is used in KKRnano. While the derivation of the TFQMR scheme is lengthy and beyond the scope of this thesis a short survey over iterative methods will be provided which motivates our choice. Here, we will follow the book of Saad [86] to which the reader is referred for any more detailed information on TFQMR and further iterative schemes.

The most important iterative techniques are based on the projection on the Krylov subspace. I.e. an approximate solution X_m for the linear equation $A X = b$ can be expressed in terms of an arbitrary initial guess X_0 and the Krylov subspace $\mathcal{K}_m$ as

$$X_m = X_0 + \mathcal{K}_m. \tag{4.18}$$

This Krylov subspace is defined by

$$\mathcal{K}_m(A, r_0) = \mathrm{span}\{r_0, A r_0, A^2 r_0, \dots, A^{m-1} r_0\}, \tag{4.19}$$

where the residual vector $r_0 = b - A X_0$ enters. The inverse of A can then be written as

$$A^{-1} b \approx X_m = x_0 + q_{m-1}(A) r_0 \tag{4.20}$$

where q_{m-1} is a polynominal defined up the degree of $m-1$ and m is the dimension of the Krylov subspace. While the general concept is clear it remains the complicated task to determine the components defining the polynomial q_{m-1} in the Krylov basis to obtain a properly converged solution. There are two main approaches for this task:

One important class relies on the orthogonalization of the Krylov vectors in order to obtain the approximate solution. The most successful realizations are Arnoldi's method and the generalized minimal residual method (GMRES) [87]. The clear drawback is that for the orthogonalization procedure all vectors have to be stored in memory, which would be a severe limitation for the application in KKR**nano** where the size of the vectors scales with the number of atoms N.

More advanced schemes are based on the Lanczos biorthogonalization which requires to store only a few vectors. On the downside the convergency of algorithm using biorthogonalization can be less stable for several classes of problems [86]. Important implementations are the biconjugate gradient (BCG) [88] and the quasi minimal residual (QMR) method [85]. Both approaches consist of matrix-vector products with A and its transpose A^T in every iteration. Since matrix-vector multiplications represent the computationally most demanding operations of each iteration, improved schemes have been developed which are explicitly avoiding part of those operations. Thereby the conjugate gradient squared (CGS) [89] and derived from that the transpose free quasi minimal residual (TFQMR) method [90] have been developed and can consequently converge performing only half of the matrix-vector multiplications as their counterparts BCG and QMR.

In KKR**nano** we have implemented the TFQMR scheme which combines the most important advantages of low memory demands and fast computation due to its transpose free formulation. For KKR**nano** we made important modifications at the level of this algorithm, which essentially calls for a look into the used scheme which is shown in simplified form in Algorithm 4.1[2] Here, again the central linear matrix equation $AX = b$ is solved, while for the sake of simplicity the initial guess is set to $X_0 = 0$. i stands for the iteration number, the control variable is $m = \frac{i}{2}$ and all parameters and vectors in the algorithm carry an angular momentum index l underlining the separate treatment of each component. An understanding of details of the vectors and parameters used is not necessary for the upcoming discussion, but a brief overview will be given anyhow: τ_l, θ_l, c_l and η_l represent components of rotational matrices and act on the Krylov-vectors u_l and v_l which are scaled using the scalars α_l and β_l. d_l determines the update of the ith iteration on the solving vector X_l^i. The vector w_l is used in similar fashion as the residual vector in less advanced schemes than TFQMR but cannot be anymore interpreted as the residual of the actual system.

Before discussing the implementation of preconditioning, initial guess and parallelization which will alter the algorithm 4.1, firstly two important optimizations steps will be explained, which alone allow for a speedup of the TFQMR algorithm by roughly a factor of three.

Sparse Matrix-Matrix multiplication

Apparent from Alg. 4.1 the TFQMR scheme in KKR**nano** is not applied straightforwardly for each vector X but all l components are iterated simultaneously. This procedure has the important advantage that the tasks for all l can be merged before each multiplication with

---

[2]For a direct comparison the standard TFQMR scheme can be found in [85] and [86].

Algorithm 4.1 TFQMR in KKRnano

1: $\rho_{l,0} = 1$; $\eta_l = \theta_l = 0$

2: $r_{l,0} = w_{l,0} = b_l$; $d_{l,0} = u_{l,0} = v_{l,0} = 0$

3: choose $r_{l,0}^\star$ that $\left(r_{l,0}, r_{l,0}^\star \right) \neq 0$

4: **for** $m = 1$ to convergency **do**

5:     **for** $l = 1$ to $(l_{\max} + 1)^2$ **do**

6:         $\rho_{l,m} = \left(w_l, r_{l,0}^\star \right)$

7:         $\beta_l = \rho_{l,m}/\rho_{l,m-1}$

8:         $v_l = Au_l + \beta_l v_l$

9:         $u_l = w_l + \beta_l u_l$

10:     **end for**

11:     compute Au_l by sparse matrix-matrix multiplication (Algorithm 4.2)

12:     **for** $l = 1$ to $(l_{\max} + 1)^2$ **do**

13:         $v_l = Au_l + \beta_l v_l$

14:         $\alpha_l = \rho_{l,m}/\left(v_l, r_{l,0}^\star \right)$

15:         $d_l = u_l + \left(\theta_l^2/\alpha_l \right) \eta_l d_l$

16:         $w_l = w_l - \alpha_l Au_l$

17:         $\theta_l = \|w_l\|_2/\tau_l$

18:         $c_l = 1/\sqrt{1 + \theta_l^2}$

19:         $\tau_l = \tau_l \theta_l c_l$; $\eta_l = c_l^2 \alpha_l$

20:         $X_l^{i+1} = X_l^i + \eta_l d_l$

21:         $i = i + 1$

22:     **end for**

23:     **for** $l = 1$ to $(l_{\max} + 1)^2$ **do**

24:         $u_l = u_l - \alpha_l v_l$

25:         $d_l = u_l + \left(\theta_l^2/\alpha_l \right) \eta_l d_l$

26:     **end for**

27:     check if converged for all l: if yes leave the loop

28:     compute Au_l by sparse matrix-matrix multiplication (Algorithm 4.2)

29:     **for** $l = 1$ to $(l_{\max} + 1)^2$ **do**

30:         $w_l = w_l - \alpha_l Au_l$

31:         $\theta_l = \|w_l\|_2/\tau_l$

32:         $c_l = 1/\sqrt{1 + \theta_l^2}$

33:         $\tau_l = \tau_l \theta_l c_l$; $\eta_l = c_l^2 \alpha_l$

34:         $X_l^{i+1} = X_l^i + \eta_l d_l$

35:         $i = i + 1$

36:     **end for**

37:     compute AX_l^i by sparse matrix-matrix multiplication (Algorithm 4.2)

38:     **for** $l = 1$ to $(l_{\max} + 1)^2$ **do**

39:         $r_l^i = AX_l^i - b_l$

40:         if $\|r_l^i\|$ smaller than tolerance account l component as converged.

41:     **end for**

42: **end for**

---

the matrix A and split afterwards (see line 10, 27 and 36 in Alg. 4.1). This multiplication can then be conducted as matrix-matrix instead of matrix-vector multiplication, which leads to a significant increase in the flop-rate. This speedup is caused by favorable memory cache-access and of course dependent on the computing architecture. On all tested platforms we find a speedup of roughly a factor of two by applying matrix-matrix instead of matrix-vector multiplications.

Algorithm 4.2 Sparse matrix-matrix multiplication in KKRnano

1: **for** $j = 1$ to N **do**
2:      **for** $l = 1$ to $(l_{\max} + 1)^2$ **do**
3:          **if** l component is not converged **then**
4:              downfold X_l to the sparse representation X_l' which is then of size $(l_{\max} + 1)^2 \times N_{\mathrm{cl}} (l_{\max} + 1)^2$ and compatible with the sparse representation of A for the rows j.
5:          **end if**
6:      **end for**
7:      perform matrix matrix multiplication $a_j' \cdot X'$, where a_j' exploits the sparsity of A and represents the $(l_{\max} + 1)^2$ rows for site j in matrix A.
8: **end for**

Further, the matrix A is of size $N_{\mathrm{cl}} (l_{\max} + 1)^2 \times N (l_{\max} + 1)^2$, where N_{cl} is the number of sites in the screened reference system. In KKRnano this sparsity is exploited by projecting the sparse matrices onto a dense matrix of accordingly smaller size as shown in Alg. 4.2. Thereby the above algorithm is the major source for the acceleration of KKRnano due to the introduction of the screened reference system.

Adaptive determination of the residual

The residual has to be computed in every second iteration as described in Alg. 4.1 (line 38), which requires one additional sparse matrix-matrix multiplication at that time. Since those operations are by far the computationally most demanding parts, avoiding them completely has the potential to save $\frac{1}{3}$ of the overall computational time. In order exploit this potential two strategies can be used. On the one hand the convergency threshold can be based on the upper limit of the residual which is given by

$$\|r_l^m\| \leq \sqrt{m+1}\, \tau_l^m. \tag{4.21}$$

where m is the running index in Alg. 4.1, which is set into relation to the iteration number i by $2i = m$. Further, τ_l^m is computed for each iteration anyhow (see Algorithm 4.1) and therefore no demanding additional computation is required. However, selecting this upper limit of the residual as criteria for convergency leads to a notorious over-convergence of the solving vectors.

We gain larger control over the convergency by considering the actual computed residual $\|r_l^m\|$, but performing its calculation at significantly fewer iterations. Here, we benefit from the circumstance that the probability for convergency in the next subsequent steps can be estimated on the basis of the history of the residual norm at previous iterations.
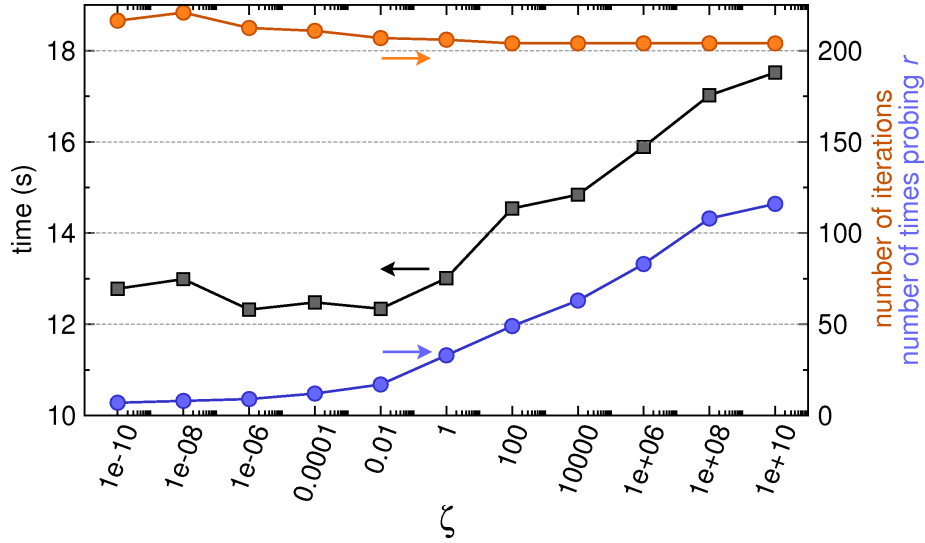
Figure 4.4.: *Performance of the TFQMR algorithm for the exemplary system Si_{432} for one atom and $l_{max} = 2$ with application of the adaptive calculation of the residual vector r and its norm $\|r\|$ controlled by ζ. Total time required on a single CPU and for all $(l_{max})^2$ components (black). For the sake of comparison the average number of TFQMR iterations performed per l component (orange) and the number of times the residual vector is probed (blue) are illustrated.*

In KKRnano the residual vector is only calculated when the following empirically set up condition is fulfilled:

$$\mathrm{mod}\,(m_l, \mu_l) = 0, \tag{4.22}$$

with

$$\mu_l = \max\left(1, \mathrm{int}\left(\log\left(\frac{\overline{\|r_l^m\|}}{\overline{t_l} \cdot \zeta}\right)\right)\right), \tag{4.23}$$

where $\overline{\|r_l\|}$ and $\overline{t_l}$ denote the average over all l components of the residual norm and the convergency tolerance, respectively. ζ is a constant parameter which is controlling the frequency of the probing of the residual norm. The logarithmic dependency has been chosen to cover the full range of $\overline{t_l}$ and $\overline{\|r_l\|}$ which can vary over order of magnitudes. Since the residual norm $\overline{\|r_l\|}$ change from one iteration to the other, μ_l is recalculated with the updated average residual norm if the condition (4.22) is fulfilled. The workflow of this adaptive calculation of the residual is best described by giving a numerical example, dropping the l-index from now on. Assume a scenario with $\overline{\|r\|} = 10^{-4}$ and $\overline{t} = 10^{-6}$. Selecting $\zeta = 1$ (4.23) leads to $\mu = 2$ and the residual is calculated for every second step according to (4.22). Choosing a smaller value e.g. $\zeta = 10^{-4}$ results in $\mu = 6$ and a less frequent evaluation of the residual in only every sixth iteration. The influence of altering ζ on the performance of the iterative scheme will be discussed in the following.

In Fig.4.4 the acceleration of the TFQMR algorithm is shown for this adaptive calculation of the residual norm. As pointed out earlier, under the assumption that all computations except for the sparse matrix-matrix operations can be neglected in the first approximation

one third of the computational time is consumed by the calculation of the residual norm. Therefore, a strict maximum of the speedup by the adaptive calculation is 1.5, which is almost reached in the test system for $\zeta \approx 10^{-4}$: Here, the execution time of KKRnano is reduced from 18 to slightly more than 12 seconds (Fig. 4.4). Leaving this minimum for smaller ζ the series of minimal residual measurements becomes too incomplete, which leads to an increase of the total number of TFQMR iterations and accordingly loss in time. In the opposite limit, if ζ is chosen too large, unnecessary calculations of the residual slow down the algorithm crucially. It remains the task to rate this trade-off and optimally find the fastest configuration.

In the practical application of KKRnano the choice of a global ζ for all different iterative inversions would result in a high degree of inefficiency. The spreading of the optimal ζ can be large comparing different energy-points and crucially depends on the quality of the initial guess and preconditioning, which will be discussed in section 4.2.2 and 4.2.3. In particular the accuracy of the starting vector will usually be increased during the self-consistency iteration, which makes a self-adjusting procedure, optimizing ζ on the fly, desirable. Therefore, in KKRnano exclusively for the first self-consistency iteration a predefined constant value is chosen for all energy-, k-points and atoms. During each of the self-consistency cycles the values of the running index m_f, m_{f-1} and m_{f-2} of the final three TFQMR steps, which where selected by the incoming ζ_{in} and the residuals are calculated and stored. Here, m is defined according to Alg. 4.1 relative to the iteration number i as $2m = i$. m_f and m_{f-2} serve as basis for a readjustment of each ζ individually as described in the flow-chart below:

Algorithm 4.3 Dynamic adjustment of ζ on the fly in KKRnano

1: **if** $m = 1$ **then**
2: $\quad \zeta = \zeta_{\text{in}} =$const.
3: **else**
4: $\quad \Delta_m = \min\left(m_f - m_{f-2}, 8\right)$
5: $\quad \zeta = \zeta_{\text{in}} \cdot 10^{\Delta_m - \Delta_{opt}}$
6: **end if**

Here $\Delta_{opt} = 5$ is chosen in KKRnano. This means, the trade-off is optimized if the residual is computed three times during the last five TFQMR iterations. In order to avoid possible overshooting of the correction additionally the adjustment Δ_m is limited to 8 (see line 4 in Alg. 4.3).

The effect of this adjustment of the performance of KKRnano is exemplified in Fig. 4.5. While for the fast converging energy points close to the bottom of the energy window ζ oscillates around the initial value of 10^{-3}, a significantly less frequent calculation of residual is required for the remaining energy points. As expected and shown in Fig. 4.5(b) this readjustment leads to a slightly enhanced total number of TFQMR iterations but effectively saves ten percent of computational time. Important for the practical calculations is that the ideal speedup is reached already after three self-consistency steps.

Overall, combining the consequent use of matrix-matrix multiplications and the adaptive calculations of the residual reduces the time spent per TFQMR iteration significantly. However, the acceleration of the convergency, i.e. the reduction of the number of iterations itself, holds huge potential to further speedup the algorithm and will be discussed in the following.
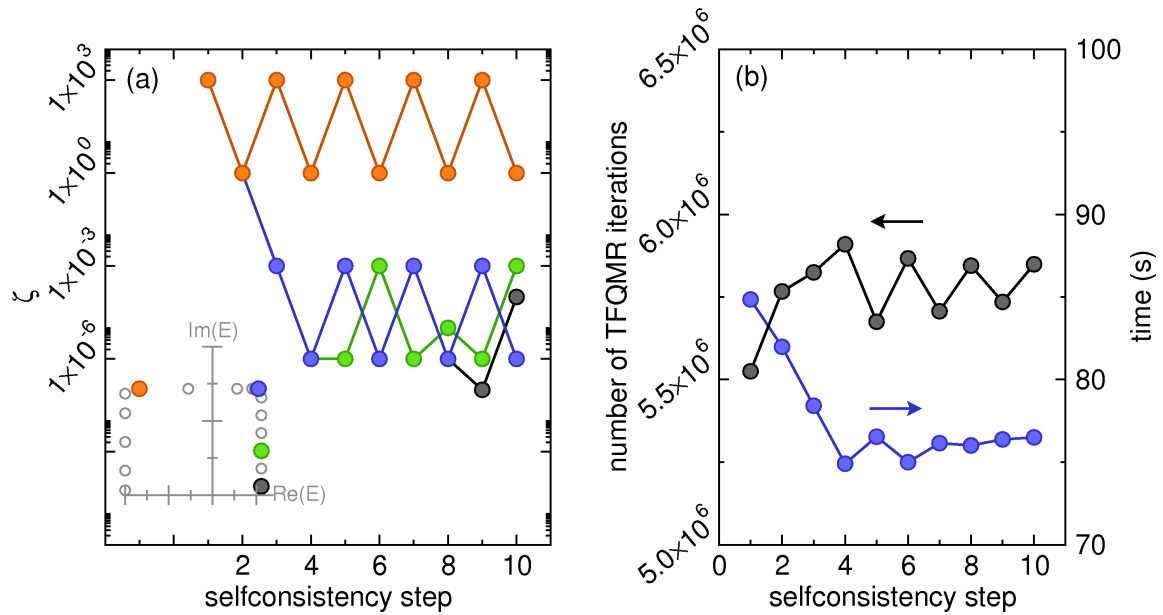
Figure 4.5.: *Adaptive scheme for the optimization of the determination of residual vectors. In (a) the self-adjusted variations of the control parameter ζ is shown for one exemplary site in Si_{432} for several energy integration points. The position of those points can be seen from the inset, where the distribution of points is schematically drawn in the corresponding color coding. In (b) the total number of TFQMR iterations and the execution time per iteration is displayed for the same self-consistency calculation as in (a). It is important to note, that here the total number of iterations include iterations for all energy points l-components and sites, which result in total numbers of 10^6 to 10^7.*

4.2.2 Initial guess

The default choice of the initial guess X_0 to the solution X of $A \cdot X = b$ is $0 + 0i$ for all entries of X_0 in the transpose-free quasi minimal (TFQMR) package [90]. Obviously, this starting point is usually far from the required solution, which motivates to introduce an − in general arbitrary − starting vector X_0, which has to be optimized. We proceed by introducing the left-over deviation $\tilde{X}$ from the solution X

$$X = X_0 + \tilde{X}. \tag{4.24}$$

Subtracting X_0 on both sides of (4.15) yields:

$$X^{(i+1)} - X_0 = \Delta t - X_0 + \Delta t G^{\mathrm{r}}(X^{(i)} - X_0) + \Delta t G^{\mathrm{r}} X_0, \tag{4.25}$$

which leads with the use of

$$\Delta t' = \Delta t - X_0 + \Delta t G^{\mathrm{r}} X_0, \tag{4.26}$$

to the effective iterative expression

$$\tilde{X}^{(i+1)} = \Delta t' + \Delta t G^{\mathrm{r}} \tilde{X}^{(i)}, \tag{4.27}$$

or accordingly written as linear equation which has to be solved instead of (4.17)

$$(\Delta t G^{\mathrm{r}} - 1)\,\tilde{X} = -\Delta t' \tag{4.28}$$

with $\tilde{X}^{(i)} = X^{(i)} - X_0$. The matrix $\Delta t'$ directly determines the quality of the starting solution X_0. Comparing (4.26) with (4.15) shows that $\Delta t'$ resembles the left-over perturbation from the initial guess to the required solution. In fact the calculation of the norm $\|\Delta t'\|$ directly corresponds to a check for quality by calculating the residual norm of the current iteration in the usual TFQMR procedure. For small $\|\Delta t'\|$, i.e. a reasonable good initial guess, the solution is expected to be computed in fewer iterations, as will be shown below. Including the initial guess in the TFQMR algorithm 4.1 requires only the computation of (4.26) which resembles b in the corresponding linear equation. At the end, X can be easily obtained by (4.24) after $\tilde{X}$ is converged. It is important to note that the convergency threshold t is usually chosen in relation to the norm of the right-hand side b in $AX = b$, e.g. $t = 10^{-5}\|b\|$. For the initial guess this choice would be highly inefficient: The better the initial guess X_0 becomes, the lower tolerances t would be required if related to the right hand side $\Delta t'$. Instead the norm of the actual right hand side Δt is used to define the tolerance in KKRnano, facilitating the best possible reduction of the number of iterations with the same accuracy as without initial guess.

After this general considerations the question has to be addressed how a well-approximating X_0 can be obtained. Below we discuss three different physically motivated approaches which can be applicable to KKRnano:

- Considering an ersatz geometry by solving a small locally defined cluster or a coherent-potential or virtual-crystal approximation with an optionally smaller angular momentum cut-off.

- The extrapolation of the solution at the previously calculated energy points.

- The result of the previous self-consistency step.

While solving an ersatz-geometry can give an excellent initial guess for a class of materials, as e.g. the solution in a local cluster for semi-conducting samples, the quality of the initial guess (IG) can be significantly worse in metallic systems. Moreover the calculation of a precise IG has the potential to be computationally cumbersome and to slow down the algorithm considerable. On the contrary, the extrapolation of the previous energy points is computationally cheap and easy to generalize. However, we find that the accuracy of such an extrapolation is rather limited. It is more advantageous to use the result of the previous self-consistency DFT cycle.
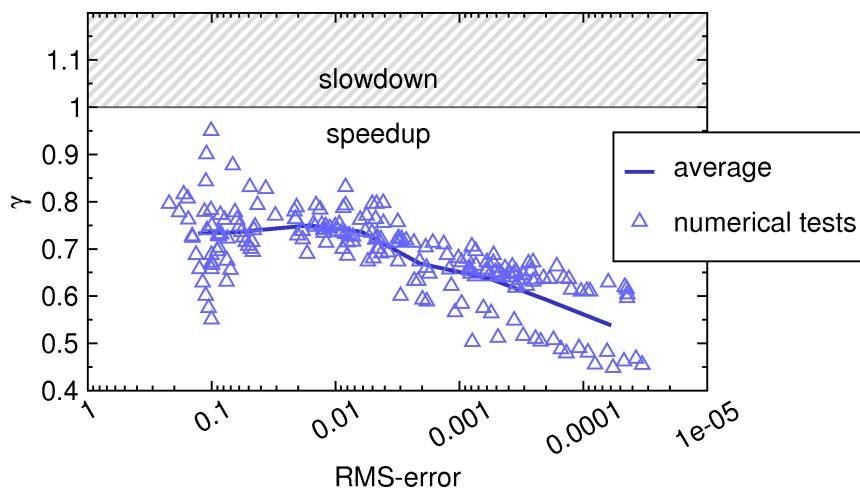
Figure 4.6.: *Reduction of required number of iterations γ by application of initial guess as described above for the system $Ge_1Sb_2Te_4$ for various intermediate system sizes ($N = [64, 512]$) as a function of convergency. The measure used to rate the quality of convergency is the RMS-error, the variance of previous and actual potential $\propto \|V^{i-1}(r) - V^i(r)\|$. The reduction ratio γ (blue triangles) is defined as the ratio of all iterations needed without and with application of the IG: $\gamma = \sum_{N,k,e,lm} N_{it}^{IG} / \sum_{N,k,e,lm} N_{it}$. For half of the plotted values for γ, block-circulant preconditioning $-$ as introduced later in section 4.2.3 $-$ is used additionally, where no difference in γ was observed in comparison to the unpreconditioned TFQMR steps. The average reduction is plotted as straight line.*

This procedure results in an IG, which exhibits usually already at the first self-consistency steps a quality of $(\|X_s\| - \|X_{s-1}\|)/\|X_s\| \approx 10^{-2}$, where s and $s-1$ label the current and previous step. Upon convergence of the self-consistency cycle the quality of the IG increases gradually without enhancing computational costs. Keeping in mind the architectures of modern supercomputers the memory limitations can be a critical bottleneck for this approach: For each atomic site the information on $X_{(s-1)}$ with size $N \cdot (l_{max}+1)^2 \times (l_{max}+1)^2$ has to be stored for each energy- and **k**-point. Therefore, in KKR**nano** exclusively contributions larger than a specified threshold are stored to represent X_0 in an array of high sparsity with single precision.

The overall performed TFQMR iterations (Fig. 4.6) exhibit for all attempts a significant reduction by a factor of 0.45 to 0.95 by application of the IG. As intuitively expected the increase in quality of the IG for almost converged self-consistency iterations leads to a reliable incremental reduction of required TFQMR steps. Given the fact that only little computational effort has to be made to apply the initial guess, the reduction of the number of iterations directly translates into a significant acceleration of the algorithm.

4.2.3 Preconditioning

On top of the usage of the initial guess, the number of TFQMR iterations can be reduced by application of preconditioning schemes [91, 92]. Here, the main challenge is to find an approximation to the matrix A in (4.16), which can be easily inverted. With the help of such an approximate matrix

$$P = P_1 \cdot P_2 \tag{4.29}$$

a modified linear matrix-equation

$$A' \cdot Y = b' \tag{4.30}$$

can be deduced, where $A' = P_1^{-1} A P_2^{-1}$, $b' = P_1^{-1}(b - AX_0)$ and $Y = P_2(X - X_0)$. X_0 denotes an optional initial guess to the solution X of $AX = b$. The residual vector for the preconditioned system then reads $r'_n = b' - A'Y_n$. When a sufficiently small residual vector is obtained the solution of the original system can be calculated as:

$$X_n = X_0 + P_2^{-1} Y_n. \tag{4.31}$$

The residual of the original system translates as:

$$r_n = P_1 r'_n \tag{4.32}$$

For our purposes we limit the preconditioning to right preconditioning by setting $P_1 = 1$, which then leaves the required size of the minimal residual unchanged from the one for the original system. If the initial guess X_0 is zero, the additional computation consists of two steps:

- P_2^{-1} is applied to X before every matrix-vector multiplication AX in algorithm 4.1 in line 10, 27 and 36.

- the solution of the original system X is obtained from Y by (4.31) after convergency.

The remaining and most challenging task is to find an easily invertible matrix P_2 approximating A. One approach is to obtain P_2^{-1} by applying a sparse incomplete LU decomposition of A which would be functional also for cells with large relaxations or amorphous systems. However, since we aim to develop a highly parallelized code and ILU preconditionier are difficult to parallelize efficiently, we restrict to systems with structural relaxations well below 10% of the lattice constant. Then, we can exploit the fact that in such lattices G^{r} and partly also Δt are roughly periodic on a smaller length scale than the size of the actual supercell. This idea is the basis for preconditioning by a block-circulant matrix which was recently introduced by Bolten *et al.* [93]. We will show that this scheme is optimally suited to efficiently obtain a high quality preconditioning matrix in KKRnano.

Right-preconditioning by block-circulant

In all cubic or rectangular lattices M_{bl}^x, M_{bl}^y, M_{bl}^z subblocks in x, y and z spatial direction which consist of N_{bl} atoms each can be defined such that those $M_{\text{bl}} = M_{\text{bl}}^x \cdot M_{\text{bl}}^y \cdot M_{\text{bl}}^z$ blocks build a new basis for the supercell. The matrix $A = (\Delta t G^{\text{r}} - I)$ can then be composed out of $M_{\text{bl}} \times M_{\text{bl}}$-submatrices and reads in full representation as:

$$A_{LL'}^{nn'} = \begin{pmatrix} (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{11} & (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{12} & \cdots & (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{1M_{\text{bl}}} \\ (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{21} & (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{22} & \cdots & (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{2M_{\text{bl}}} \\ \vdots & \vdots & \ddots & \vdots \\ (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{M_{\text{bl}}1} & (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{M_{\text{bl}}2} & \cdots & (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{M_{\text{bl}}M_{\text{bl}}} \end{pmatrix}, \tag{4.33}$$

$(a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{ij}$ are of dimension $N_{\text{bl}}(l_{\max} + 1)^2 \times N_{\text{bl}}(l_{\max} + 1)^2$ and n_{bl} is accordingly running from 1 to N_{bl}. In this representation, the diagonal submatrices ($i = j$) cover the intrablock-interactions while the interblock-interactions is accounted for on the off-diagonal submatrices ($i \neq j$).

Although chemical or geometrical disorder on the lattice leads to a clear distinction between individual subblocks, we assume that average subblocks are suitable to describe the entire lattice in coarse approximation. Dropping the indices of the submatrices $a_{ij} = (a_{LL'}^{n_{\text{bl}}n_{\text{bl}}'})_{ij}$ all submatrices and can be down-folded to set of averaged submatrices $\overline{a}_i$. The mean submatrix carrying the intrablock interaction then reads:

$$\overline{a}_1 = \frac{1}{N_b} \sum_{j}^{N_b} a_{jj}. \tag{4.34}$$

For a generalization to off-diagonal blocks of (4.34) and thereby to interblock interactions it is convenient to use not the row index i of submatrices directly, but instead a local relative index $i_l(j)$: From here on each $i_l(j)$ marks a subblock of specific relative geometrical position to the central subblock $i = j$ of column j (For an illustration of this definition see Fig. 4.7(c)). This relative geometrical position of the block with respect to the diagonal block is given by $\boldsymbol{\Delta}_{i_l(j)} = (\Delta_{i_l(j)}^x, \Delta_{i_l(j)}^y, \Delta_{i_l(j)}^z)$. E.g. in Fig. 4.7(c) the block neighboring the diagonal block in x direction would be addressed by $\boldsymbol{\Delta}_2 = (1, 0, 0)$. Utilizing this notation we can generalize (4.34) to:

$$\overline{a}_{i_l(j)} = \frac{1}{M_{\text{bl}}} \sum_{j}^{M_{\text{bl}}} a_{i_l(j)j}, \tag{4.35}$$

which is for the intrablock interaction $i_l(j) = 1$ equivalent to (4.34). This averaging operation (4.35) is schematically visualized in Fig. 4.7(d) and (e). With this set of averaged
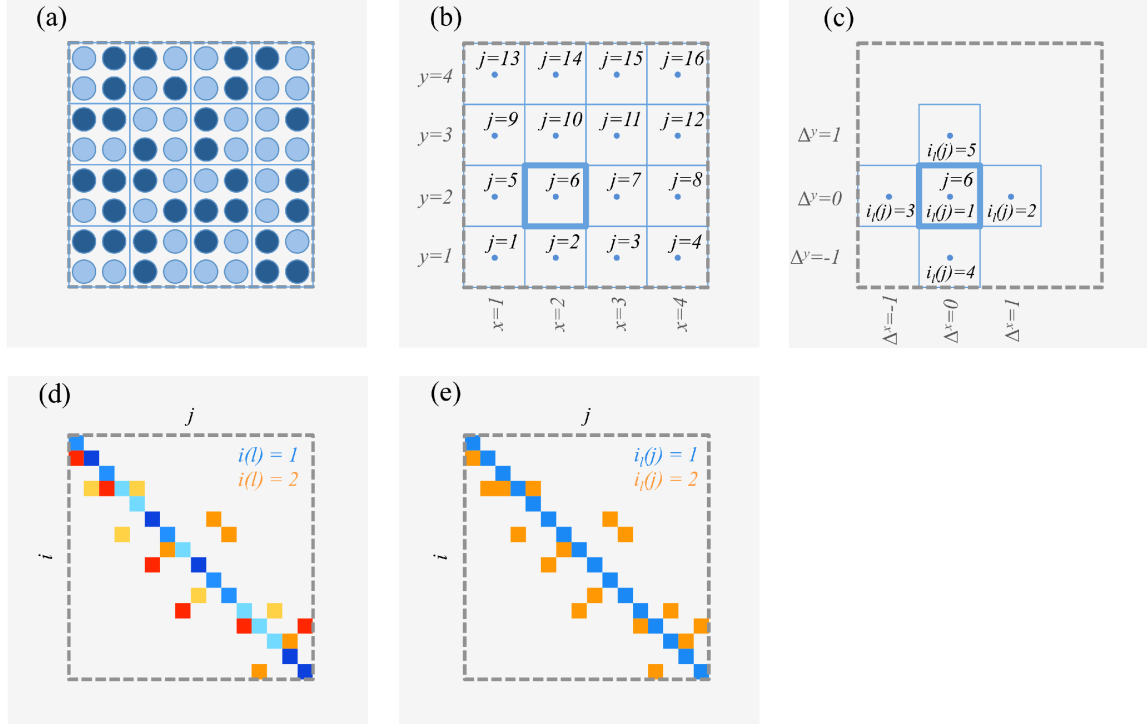
Figure 4.7.: *In all five panels, a two-dimensional supercell (boundary marked with gray dashed lines) is drawn and the five different conceptual steps needed for the block-circulant preconditioning scheme are illustrated. (a) shows the supercell on the level of atoms, which is here exemplified by a two-dimensional disordered lattice of two arbitrary types of atoms (light and dark blue) with overall $N = 64$ atoms. In addition, the supercell is partitioned into $M_{\mathrm{bl}}=16$ subblocks, where each subblock contains $N_{\mathrm{bl}} = 4$. The borders of the subblock are indicated by blue lines. Then overall $M^x_{\mathrm{bl}} = 4$ and $M^y_{\mathrm{bl}} = 4$ subblocks in x and y direction are required to represent the full supercell. (b) illustrates how those subblocks are labeled over the entire supercell by the index j from $j = 1$ to $j = M_{\mathrm{bl}} = M^x_{\mathrm{bl}} \cdot M^y_{\mathrm{bl}}$. In addition it is shown how this index j is related to the position (x_j, y_j) in real space. In (c) the interblock interactions of subblock $j = 6$, which is highlighted by thick blue lines in (b) and (c), are depicted in the space of the row index $i_l(j)$. For the sake of simplicity, index $i_l(j)$ is running exclusively over nearest neighboring subblocks. Further the relative geometrical position of the interacting blocks $\Delta^x_{i_l(j)}$ and $\Delta^y_{i_l(j)}$ are specified for this simplified example. (d) schematically shows the full interaction matrix of the supercell from all $M_{\mathrm{bl}} = 16$ blocks amongst each other, highlighting two selected types of interaction $i_l(j) = 1$ (orange colors) and $i_l(j) = 2$ (blue colors), where the variations in color represent variations in the individual interactions. In direct contrast (e) depicts the consequence of averaging the full interaction matrix to effective interactions by means of (3.35), which are accordingly represented by uniform colors. Note that (d) and (e) are schematic illustrations being not one-to-one related to (a-c).*

blocks we can proceed representing the full matrix $A_{LL'}^{nn'}$ by a block-circulant matrix:

$$
A_{LL'}^{nn'} \approx \begin{pmatrix}
\overline{a}_1 & 0 & \ldots & \ldots & & \ldots & \overline{a}_{M_{\mathrm{bl}}'} & \overline{a}_{M_{\mathrm{bl}}'-1} & \ldots & \overline{a}_3 & \overline{a}_2 \\
\overline{a}_2 & \overline{a}_1 & \ldots & \ldots & & \ldots & 0 & \overline{a}_{M_{\mathrm{bl}}'} & \ldots & \overline{a}_4 & \overline{a}_3 \\
\vdots & \vdots & \ddots & & & & \vdots & \vdots & \ddots & \vdots & \vdots \\
\vdots & \vdots & & \overline{a}_1 & 0 & 0 & 0 & \ldots & & \overline{a}_{M_b'} & \overline{a}_{M_{\mathrm{bl}}'-1} \\
\vdots & \vdots & & \overline{a}_2 & \overline{a}_1 & 0 & 0 & \ldots & & 0 & \overline{a}_{M_{\mathrm{bl}}'} \\
\overline{a}_{M_{\mathrm{bl}}'} & \overline{a}_{M_{\mathrm{bl}}'-1} & \ldots & \overline{a}_3 & \overline{a}_2 & \overline{a}_1 & 0 & & & \vdots & \vdots \\
0 & \overline{a}_{M_{\mathrm{bl}}'} & \ldots & \overline{a}_4 & \overline{a}_3 & \overline{a}_2 & \overline{a}_1 & & & \vdots & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots & & & \ddots & & \vdots & \vdots \\
0 & 0 & \ldots & \overline{a}_{M_{\mathrm{bl}}'} & \overline{a}_{M_{\mathrm{bl}}'-1} & \ldots & \ldots & & & \overline{a}_1 & 0 \\
0 & 0 & \ldots & 0 & \overline{a}_{M_{\mathrm{bl}}'} & \ldots & \ldots & & & \overline{a}_2 & \overline{a}_1
\end{pmatrix} .
$$

Here, we restrict the algorithm to M_{bl}'-submatrices, which introduces a fraction of zero-matrices to A. M_{bl}' is in practice chosen such that next-nearest-neighbor subblocks are included, i.e. usually it holds $M_{\mathrm{bl}}' < M_{\mathrm{bl}}$. This cut-off is justified by the fact that, as a consequence of the use of screened reference potentials, blocks being geometrically far from the centered diagonal block can be neglected.

The block-circulant matrix (4.2.3) will be used as preconditioning matrix P_2. For a fast computation of the inverse of P_2 an important property of circulant matrices is exploited: Given a Fourier transform defined as

$$
\overline{\alpha}_j = \sum_i \overline{a}_i \, e^{-2\pi i \mathbf{\Delta}_{i_l(j)} \mathbf{k}_j}, \tag{4.36}
$$

where

$$
\mathbf{k}_j = (k_x, k_y, k_z) = \left(\frac{x_j - 1}{M_{\mathrm{bl}}^x}, \frac{y_j - 1}{M_{\mathrm{bl}}^y}, \frac{z_j - 1}{M_{\mathrm{bl}}^z} \right), \tag{4.37}
$$

an illustration of the definition of the spatial indices x_j, y_j and z_j can be found in Fig. 4.7(b) and the blocks $\overline{\alpha}_j$ are of size $N_{\mathrm{bl}}(l_{\max}+1)^2 \times N_{\mathrm{bl}}(l_{\max}+1)^2$. This Fourier transform (4.36) of a circulant or block-circulant matrix P_2 creates a block-diagonal representation of P_2, $(P_2)_{\mathbf{k}}$, in reciprocal space [93]. The submatrices of $(P_2)_{\mathbf{k}}$, $\overline{\alpha}_j$, can now be block-wise inverted by means of LU decomposition. These are fast operations due to the small block-sizes, e.g. for $N_{\mathrm{bl}} = 10$ atoms and $l_{\max} = 3$ the blocks have a size of 160×160. The required matrix for preconditioning P_2^{-1} is then constructed out of

$$
(P_2)_{\mathbf{k}}^{-1} = \begin{pmatrix}
\overline{\alpha}_1^{-1} & 0 & \ldots & 0 \\
0 & \overline{\alpha}_2^{-1} & & \vdots \\
\vdots & & \ddots & 0 \\
0 & \ldots & 0 & \overline{\alpha}_{M_{\mathrm{bl}}}^{-1}
\end{pmatrix} . \tag{4.38}
$$

Multiplication of $(P_2)_{\mathbf{k}}^{-1}$ as operated in equation (4.31) as well as in every multiplication

involving A' is then conducted in reciprocal space. Therefore first a FFT is applied straightforwardly to Y, $Y \longmapsto Y_{\mathbf{k}}$, then the matrix multiplication $(P_2)_{\mathbf{k}}^{-1} Y_{\mathbf{k}}$ is done and as last step the back transformation $(P_2)_{\mathbf{k}}^{-1} Y_{\mathbf{k}} \longmapsto P_2^{-1} Y$ gives the desired preconditioned vector $P_2^{-1} Y$.
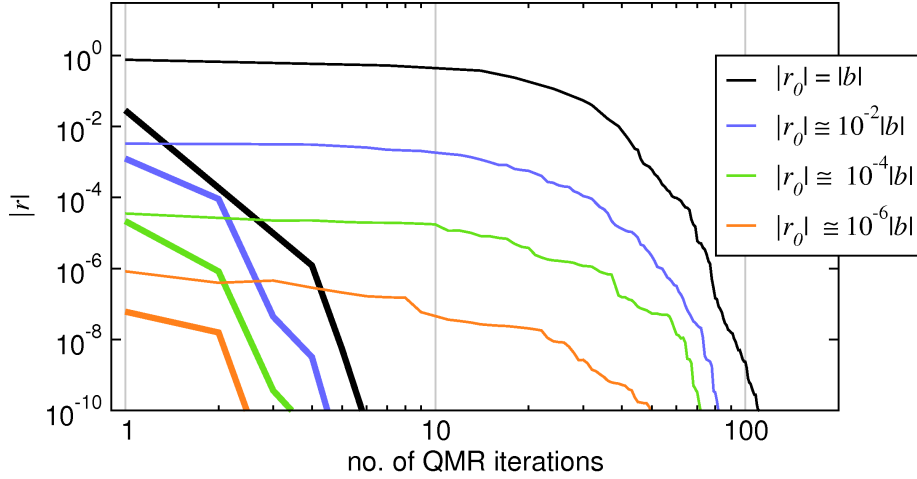

Figure 4.8.: *Convergency of the residual vector $|r|$ as a function of the number of TFQMR iterations combining the initial guess and block-circulant preconditioning for one exemplifying column of a matrix corresponding to Si_{432}. Four different qualities of starting vectors, rated by the initial residual $|r_0| = Ax_0 - b$ are shown for comparison with (thick lines) and without applied preconditioning (thin lines). Note that a log-log scale is used.*

The performance enhancement by the reduction of the required number of TFQMR iterations is shown in Fig. 4.8 for a sample system at the energy point closest to the real axis. This point is of particular importance as convergency is the most demanding at this energy point. For all scenarios with different quality of the initial guess convergency of the residual to zero is reached strikingly faster for block-circulant preconditioning and at least a factor twenty less TFQMR iterations have to be performed.

In order to obtain a more general picture Fig. 4.9 shows the required number of TFQMR iterations at different energy points. Here, the large variability over a range of 100 to 10000 iterations is clearly visible for the unpreconditioned approach. Preconditioning leads to a significant reduction by a factor of 2 to 50 and to a considerable smaller spreading from 60 iterations for the first to 200 iterations for the last energy point.

This reduction of the number of iterations raises the question whether it can be translated into an overall speedup of the algorithm, since additional computational work has to be performed: On the one hand a preconditioning matrix P_2^{-1} has to be created before starting the iterative procedure. On the other hand P_2^{-1} has to be applied at each iteration. Fig. 4.10(b, c) displays the timings for those three steps. While the setup of the initial guess requires only modestly more computational time, about 25% of the total time of iterative solution with BCP is consumed to generate P_2^{-1}. However, comparing the timing with and without BCP in Fig. 4.8(b, c) reveals that the extra amount of computational work is well invested. The strongly reduced number of iterations − even though each of them takes longer due to the multiplication with P_2^{-1} according to (4.30) − translates for this particular energy point into a dramatic speedup of the algorithm by a factor 5 to 9 as shown in Fig. 4.8(b, c).
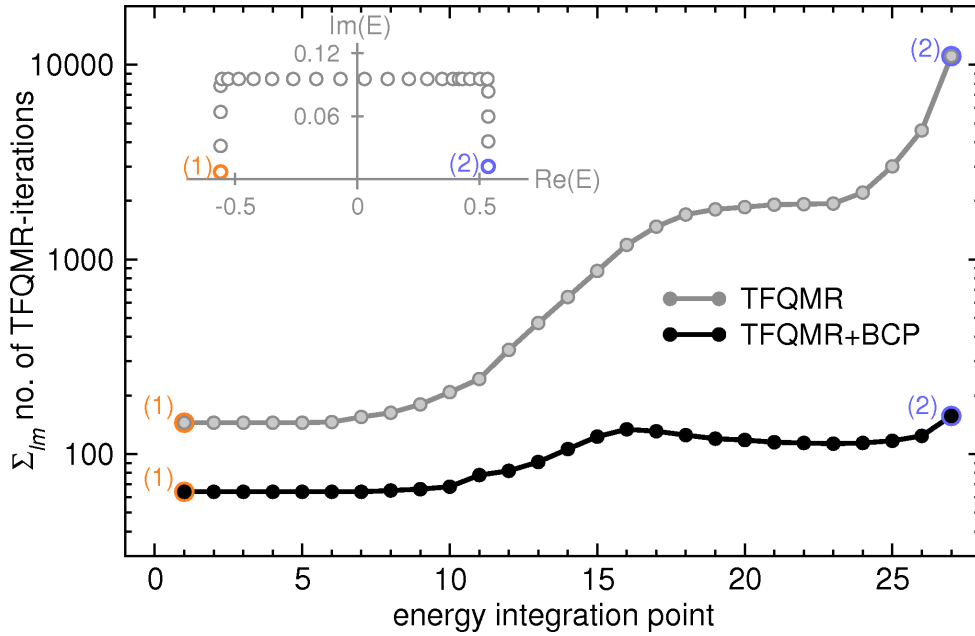
Figure 4.9.: *Sum over TFQMR and TFQMR+BCP iterations for all lm components for one arbitrary atom in the test system Ni_5Pd_{251} for all 27 energy integration points. A random displacement from the ideal atomic sites on the order of 1% has been introduced. The inset shows the distribution of the energy integration points in the complex plane, the color coded points illustrate the corresponding positions of energy points in the main graph.*

		t_f	ΣN_{it}
$Si_{504}P_8$	$BCP^{2\times2\times2}_{16sites}$	12min36.81s	2.49×10^6
	$BCP^{8\times8\times8}_{2sites}$	32min24.85s	13.11×10^6
	no BCP	40min56.27s	18.69×10^6
$Ge_{125}Sb_{250}Te_{500}$	$BCP^{5\times5\times5}_{8sites}$	15min57.45s	8.91×10^6
	$BCP^{10\times10\times10}_{1sites}$	32min41.46s	26.68×10^6
	no BCP	40min57.86s	38.63×10^6
$Ge_{125}Sb_{250}Te_{500}$:DOS	$BCP^{5\times5\times5}_{8sites}$	8min43.00s	4.99×10^6
	no BCP	1h4min48.35s	38.90×10^6

Table 4.2.: *Acceleration of the block-circulant preconditioning (BCP) scheme for different choices of subblock-representations. For the calculation of the density of states a temperature broadening of T=200 K has been used while all other calculations have been performed with T=600 K and an angular momentum cut-off of $l_{max} = 2$. t_f labels the overall execution time of KKRnano and ΣN_{it} is the total number of iterations performed for all energy points, l components and sites.*
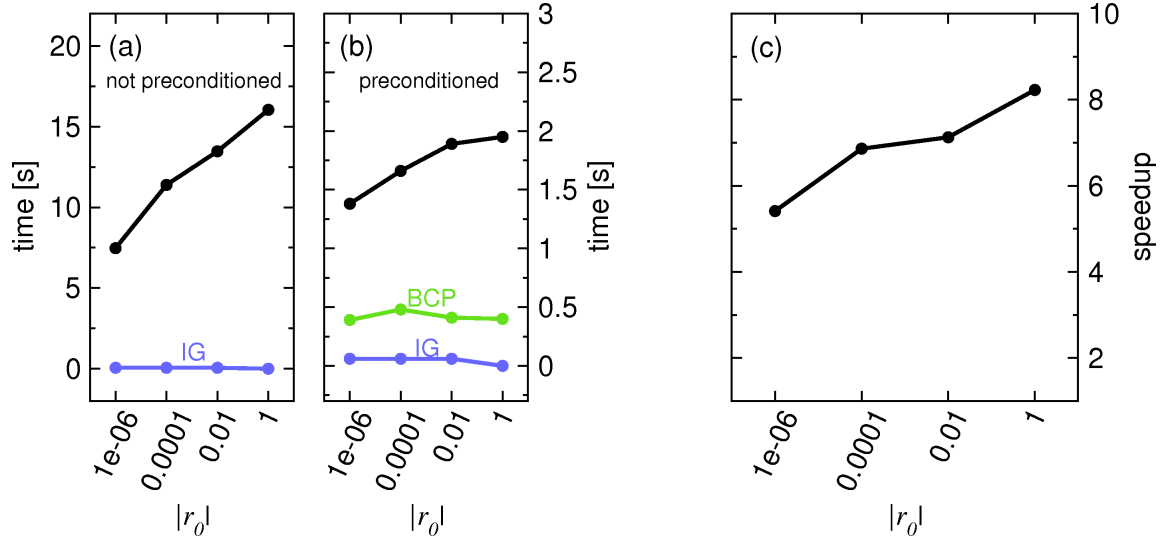
Figure 4.10.: *The timing on a single CPU to obtain the iterative solution (black) of one exemplifying column of a matrix corresponding to Si_{432} (same system as used for the analysis of convergency in Fig. 4.8). In (a) without and in (b) with applied preconditioning shown as a function of the quality of the initial guess $|r_0|$. For comparison the timings for setting up the initial guess (b, c) and the block-circulant preconditioning (BCP) matrix (see (c)) are given. Note that the scale is different by a factor of seven in Figure (a) and (b). In (c) the speedup by application of the BCP scheme is illustrated.*

After demonstrating the striking acceleration by preconditioning a remark which is important for the practical application of this scheme should be made. The construction of the preconditioning matrix as described above allows for different choices of subblocks. Depending on the size of the supercell and complexity of the lattice several representations are possible - partly accounting for the predominant chemical occupation of the lattice, partly not. The latter then resembles purely the structure of the reference system G^r and disregards variations in Δt. In Tab. 4.2 the performance of both types is displayed e.g. for the diamond-lattice Si with substitutional P-impurities. For the subblocks consisting of 16 sites with eight atoms and eight interstitials the difference of interstitial sites as well as Si and P atoms is regarded. Hence, the preconditioning matrices are found to be of high quality leading to a reduction of the total number of TFQMR iterations by more than a factor of seven with respect to the unpreconditioned solver. Taking the overall execution time of KKRnano as a measure the reduction results in a speedup by more than a factor of three. Opposed to that smaller subblocks lead to significantly lower acceleration. The same trend holds for the phase change material $Ge_1Sb_2Te_4$ (GST). GST is treated in rocksalt structure and A sites are exclusively occupied by Te atoms, while Vacancies, Ge and Sb are randomly distributed on B sites. While subblocks of eight sites which represent A and B sites individually lead to a speedup by a factor of three, subblocks of single sites perform significantly worse.

The strong dependency on the choice of the subblock representation is caused by two effects: Smaller subblocks with e.g. only two sites can lead to an averaging over interstitial

and occupied sites and consequently preconditioning matrices which are clearly less accurate approximations to the actual system. The quality of the approximating matrix P is further reduced as integral parts can be disregarded by including too few neighboring subblocks. The latter discrepancy is of course not a shortcoming of the scheme itself but of the practical implementation in KKRnano. On the contrary, block-sizes cannot be chosen arbitrarily large: The preprocessing for preconditioning of the iterative solution scales cubically with the number of atoms in the subblocks.

Therefore, in the practical application block-sizes should be ideally on the order of ten sites. Further, as apparent from this sample calculations chemical ordering should be − if present − exploited for the setup of the preconditioning matrices and blocks should chosen to be sufficiently large to represent the reference Green function properly. For cubic cells and moderately large reference clusters of less than 100 sites the choice of e.g. $N_{\mathrm{bl}}^{\mathrm{fcc}} = 4$, $N_{\mathrm{bl}}^{\mathrm{NaCl}} = 8$, or $N_{\mathrm{bl}}^{\mathrm{zincblende}} = 16$ leads to an excellent acceleration. It is important to note that the block-circulant preconditioning (BCP) has also clear limits in its application. As soon as large relaxation or amorphous systems are considered not only the chemical order but also the geometrical order, resembled by different reference systems, will come into play and prohibit a proper description by the block-circulant ansatz. However, in many systems in three-dimensional periodic environment no larger relaxation occur, which makes the BCP scheme universally applicable for a broad class of materials. In the framework of this thesis the BCP has been tested for many different metallic and semiconducting systems including small relaxations and resulted in all cases to a very successful acceleration.

A further strength of the block-circulant preconditioning is the performance enhancement for density of states (DOS) calculations which is shown in Tab. 4.2. In order to obtain meaningful results the DOS calculations have to be executed on energy integrations paths close to the real axis. As previously shown, the iterative solutions for those energy points are in general notoriously difficult to converge. BCP drastically changes this picture: In the exemplifying calculation a speedup of a factor of eight in overall computational time is reached. Thereby, the application of BCP significantly extends the range of achievable calculations.
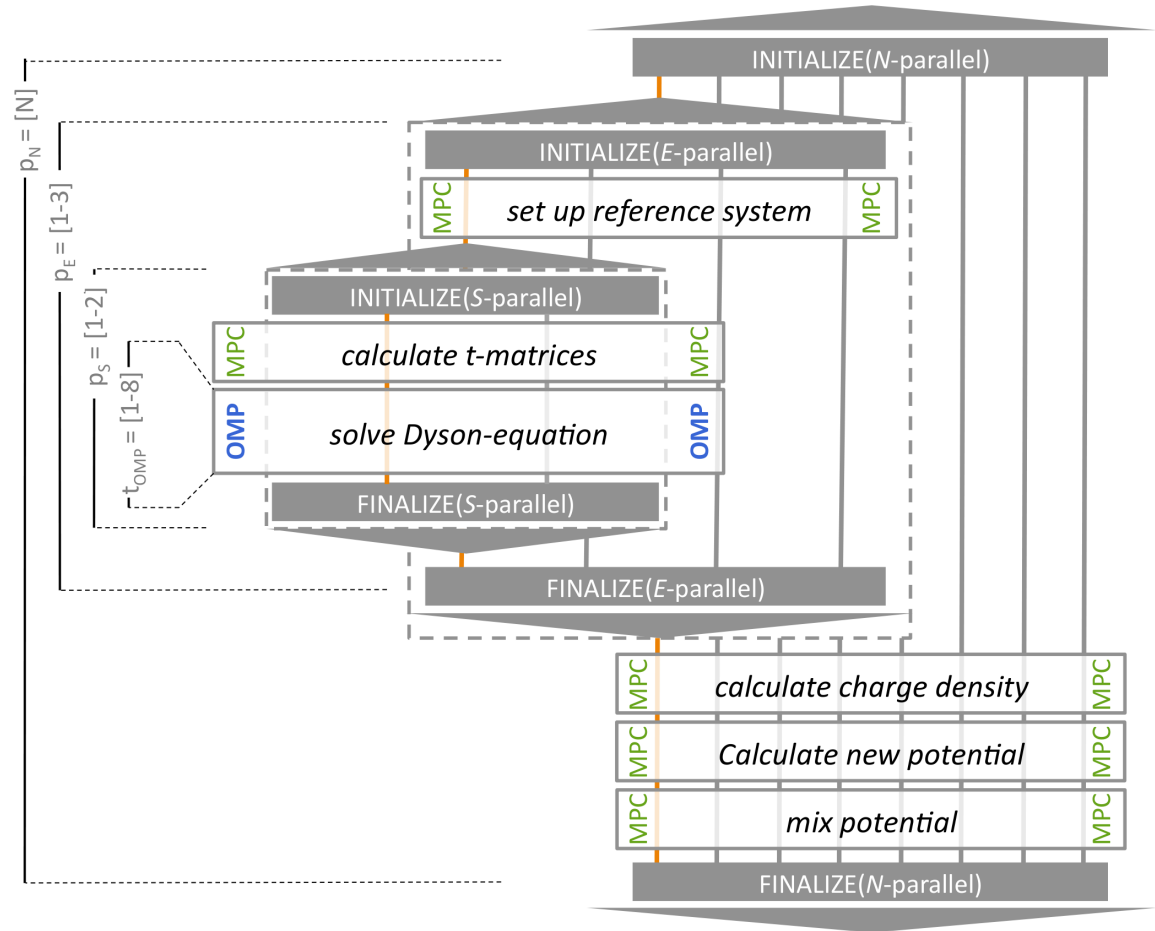
Figure 4.11.: *Schematic representation of the parallelization levels over atoms (N), over spins (S), over energy points: (E) as implemented in KKR**nano**. Important MPI communication is marked by MPC and OMP labels the procedures being OpenMP parallelized. p_N, p_E, p_S and t_{OMP} specify the range of processors/threads which can be used for the atom, energy, spin and OpenMP parallelization, respectively.*

4.3 Massive parallelization

The repetition rate of state-of-the-art supercomputers is about three to four per decade (e.g. at the Research Center Jülich: 2004: JUMP, 2006: BlueGene/L, 2008: BlueGene/P, 2009: BlueGene/P upgrade, 2009: JUROPA). Therefore it is of utmost importance to maintain a high portability of KKR**nano**. In addition the hardware architectures become increasingly parallel and hybrids of shared and distributed memory access. In order to guarantee the portability to existing platforms as JUGENE and JUROPA as well as to future computational architectures, we defined three cornerstones of programming:

- Parallelization up to at least ten thousand processors

- Memory requirements per processor below 512 MB

- Optional OpenMP parallelization

While a massive parallelization and low memory demands are obvious prerequisites to perform calculations on JUGENE, an additional level of OpenMP parallelization gives us a high flexibility to use in a shared memory approach e.g. on JUGENE 2048 MByte of memory or on JUROPA up to 24 GByte of memory. Furthermore OpenMP loops require no communication between OpenMP processes and open the possibility to parallelize procedures with extended communication. To achieve these goals KKR**nano** has overall four levels of parallelization. While the base frame of our method − the atom parallelization − is always active, all other levels of parallelization can be used optionally on top.

4.3.1 Atom-Parallelization

The entire program has been parallelized in real space over atoms, or to be more precise over sites, which potentially includes vacancies or interstitials. KKR**nano** is constructed such that the loop over atoms is the one of highest hierarchy, which considerably simplifies the distribution of work to one process per site. A simplified work-flow diagram of KKR**nano** is drawn in Fig. 4.11.

The atom or N parallelization requires communication between the split processes, which is marked in Fig. 4.11 and the most important ones will be briefly discussed in the following. First, in complex geometrical structures, where atoms have a non-uniform reference cluster, the computation of all reference structure constants on all processors becomes cumbersome. The information is however needed to build up the matrix $A = \Delta t G^{\mathrm{r}} - I$. To circumvent a computational bottleneck in this initial step of obtaining the reference structure constants, each of the N processors is used to calculate not more than one reference Green function. Those results are then broadcasted with low load imbalance to all N-processes. Additional input is required to generate A: the locally calculated t-matrices which are communicated in all-to-all communications. As mentioned earlier the iterative solution requires no communication during the iterative procedure, which is one of the key advantages of our method. However, the charge densities, which are calculated from the diagonal elements of the Green functions, have to be communicated to enable the adjustment of the Fermi level and the calculation of the Madelung potentials. Finally, in order to mix the potentials of two self-consistency steps by methods beyond simple mixing the convergency history on all sites has to be taken into account. For this purpose several reduction operations are performed with MPI.

Knowing about the central communication task in KKR**nano**, we can analyze the parallel performance on a solid basis. In fact, extensive MPI communication is often the main limiting factor for efficient scaling in many massively parallel applications. It is substantial for the following discussion to define two quantities which are used to rate the performance. First, the execution time or short time represents the real time which is used from starting the program to finalizing it. In contrast to this the CPU-time is defined as the execution time multiplied with the number of CPU's (processors) exploited by the parallel application. In KKR**nano** the weak scaling − i.e. increasing the system size and the number of processors on the same footing − reveals that the communication pattern used are not critically slowing down the algorithm (see Fig. 4.12): using the parallel performance analysis tool scalasca [94] the actual time spent for MPI communication can be precisely measured. Having this
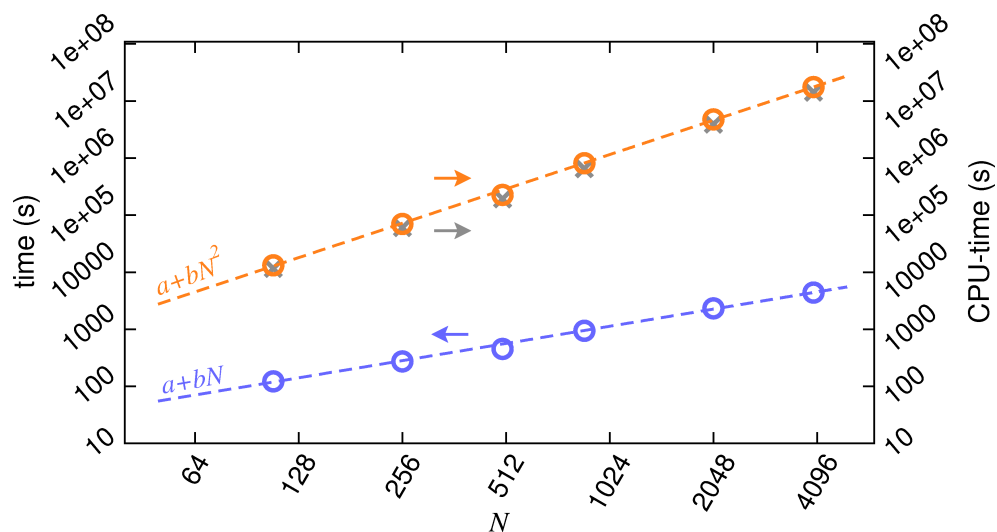
Figure 4.12.: *Weak scaling, which is the scaling observed by equally increasing the system size and the number of used processors of KKRnano on JUGENE. Timings are obtained for one self-consistency step using block-circulant preconditioning of Ni_xPd_{1-x} with $x \approx 3\%$, one $\mathbf{k}$-point and $l_{\max} = 3$ for different system sizes. The required CPU time (orange), CPU time without communication (gray) and execution time (blue) are shown in comparison. Dashed lines are fits to the underlying data points.*

information at hand, it turns out that the CPU time with and without communication do not show decisive differences. Both are scaling proportional to N^2 which fits the theoretically expected scaling law $\propto N_{\text{it}} N_{\text{cl}} N^2$. This scaling behavior underlines that the assumption that the total number of iterations is not varying crucially by changing the system size N is indeed valid for the actual application. The efficiency of the N parallelization can be best seen from the execution time of KKR**nano**, as shown in blue in Fig. 4.12. Here, the scaling is well described by a linear fit in N. Since a N^2 scaling of the CPU time is observed and N processes are involved in the parallelization, observing this linear dependency proves that the atom parallelization is highly efficient, even close to the ideally possible scaling.

4.3.2 Optional levels of parallelization

The excellent parallel performance observed for the atom parallelization motivated to even increase the level of parallelization in KKR**nano**. With the background of existing electronic structure methods the $\mathbf{k}$-space parallelization would be the first natural choice. In case of KKR**nano**, which is especially designed to handle mainly systems of more than one thousand atoms, the reciprocal space is in most cases sufficiently small to be described by a single $\mathbf{k}$-point which rules out any k-point parallelization. However, several other parallelization strategies can be exploited: In the following we will discuss the spin, energy, angular momentum as well as low-level OpenMP parallelization in KKR**nano** and show its performance.
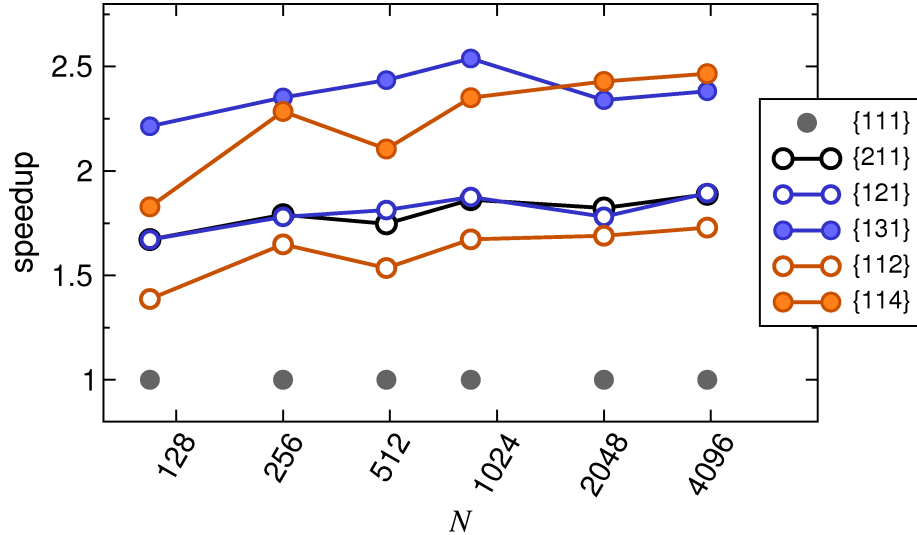
Spin-Parallelization

Figure 4.13.: *Speedup of KKRnano applying all levels of parallelization separately on a BlueGene/P [82] architecture for Ni_xPd_{1-x}. The details are the same as described in caption of Fig. 4.12. Labels $\{p_S\ p_E\ t_{OMP}\}$ with p_S, p_E, and t_{OMP} specify the number of processes/ threads used per spin, energy, and OpenMP parallelization, respectively. For the energy parallelization dynamic load balancing as explained in the text has been adopted.*

The first optional level of parallelization divides the work performed in the spin-up and the spin-down loop equally to two MPI processes, which includes the setup of the Δt-matrices, the solution of the Dyson equation and computation of the electron density. For the sake of completeness it should be pointed out, that optional extensions of the KKR algorithm to non-collinear spin-systems or relativistic calculations both couple the spin channels and would prevent a straightforward spin parallelization. Nevertheless, all systems calculated in this thesis can be well described in the collinear and non-relativistic approach, making the spin parallelization applicable. The speedup shown in Fig. 4.13 displays the high efficiency of this level of parallelization. For all examined system sizes we find an acceleration in execution time of 1.7 to 1.9 compared to an ideal value of 2. Here, as for all optional levels of parallelization, the relative increase in efficiency for larger systems can be attributed to the fact that the ratio of parallelized parts and non-parallelized overhead grows with N. Further, only little additional MPI communication has to be performed to synchronize the results of both spin channels and does not present a limitation for the performance.

Energy parallelization

As second optional parallelization KKRnano possesses the option to run parallel over the energy integration points. Previously discussed in chapter 3 the energy integration is performed on a complex contour avoiding large fluctuations of the integrand on the real axis. Thereby the number of energy points can be reduced to typically 20 to 40. However, since in KKRnano the Dyson equation is solved iteratively, the number of iterations required

for energy points close or far away from the real axis can vary significantly. Fig. 4.14 shows a typical example for the work load at different energy points. Depending on the electronic temperature, the material properties and whether preconditioning is applied, the computational time required to solve the Dyson equation for the last energy point closest to the real axis is in the range between 10 and 40% of the total execution time. Therefore, in **KKRnano** for dividing the work of all energy points usually not more than three MPI processes are utilized.
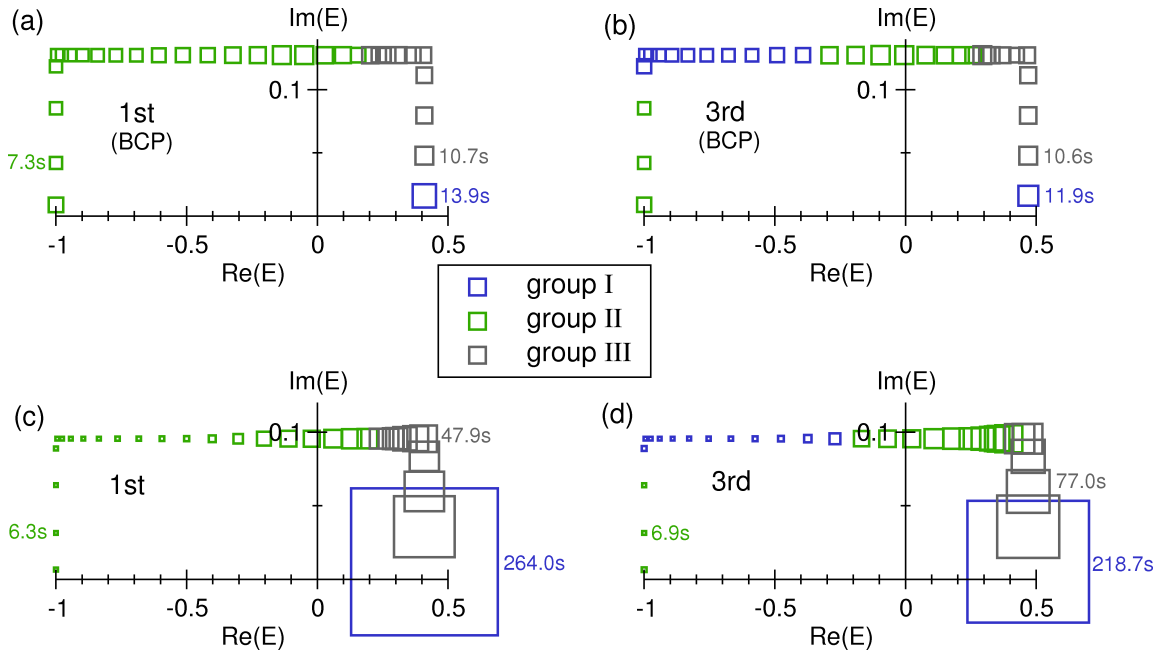
Figure 4.14.: *Distribution of energy integration points as used in the energy parallelization. In (a) and (b) block-circulant preconditioning (BCP) is applied for $Si_{212}P_4$ with an electronic temperature of 800 K, in (c) and (d) no preconditioning is used for the same system with 600 K. The work load for each point is represented by the size of the squares. (a) and (c) show the distribution of the work load by first guess for the first self-consistency step, (b) and (d) the dynamically load balanced ones at the third self-consistency step. For selected energy points the execution time is shown.*

Apparent from Fig. 4.14 the work load for different energy integration points shows huge variations. Furthermore the work load is potentially changing during the self-consistency process which is resembled by the difference of 1st and 3rd step in Fig. 4.14. This effect can be introduced by several reasons: First, shifts of the Fermi level and correspondingly in the energy contour might influence the convergency crucially. Secondly the quality of the preconditioning matrix and in particular of the initial guess can strongly vary for different steps. In any case it would be a highly non-trivial and likely inefficient task to define universal rules for the work load for a given self-consistency step.

In order to acquire a well balanced load distribution the work load per energy point is measured for each self-consistency step and the energy points are redistributed to the E parallel processors to optimize the balance. This dynamical load balancing can result in

		1ˢᵗ	2ⁿᵈ		3ʳᵈ	
		load	pred.	load	pred.	load
BCP	group I	4.9% (1)	34.5%	34.0%	34.0%	33.6% (12)
	group II	61.9% (21)	35.8%	35.9%	35.9%	35.9% (11)
	group III	33.1% (10)	29.7%	30.1%	30.1%	30.5% (9)
	group I	24.4% (1)	33.8%	33.7%	33.7%	29.3% (13)
	group II	28.2% (21)	36.1%	36.3%	36.3%	39.2% (14)
	group III	47.5% (10)	30.1%	30.0%	30.0%	31.6% (5)

Table 4.3.: *Efficiency of the dynamical load balancing as implemented in KKRnano. For three self-consistency steps the actual work load and the prediction of the load balance for the following step is shown with and without block-circulant preconditioning (BCP). The test system is $Si_{212}P_4$ with the parameters given in caption of Fig. 4.14.*

a complex reordering of the distribution as exemplified in Fig. 4.14. The efficiency of this optimization step is of course crucially determined by the capability to predict the work load for the subsequent step. Tab. 4.3 shows that both the prediction of work load and the redistribution are performed efficiently. Both with and without preconditioning the distribution of the work load comes close to the ideal of one third for each of the three MPI processes.

Exploiting this scheme the E parallel loop performs equally well for 2 processes as the spin parallelization, reaching a speedup of 1.9 for 4000 atoms (see Fig. 4.13). Including 3 processes a speedup between 2.4 and 2.6 can be expected. These small deficiencies are mainly introduced by the non-energy parallel overhead specified in Fig. 4.11. The additional communication pattern required to perform the actual energy integration do not crucially limit the speedup. While for the practical application in self-consistency cycles KKRnano is usually executed with $p_E \leq 3$ processes for the E parallelization, there are scenarios where higher p_E is desirable. E.g. for the calculation of density of states with high energy-resolution, the work load is almost uniformly distributed and considerable higher $p_E \leq 16$ can be chosen.

Parallelization over L expansion

Analyzing the timings of the calculations including atom, spin, and energy parallelization shows that the iterative solution of the Dyson equation still remains the major computational effort. One strategy to enhance the level of parallelization is to split the work performed in this iterative procedure. Since all (l, m) columns of the matrix X can be iterated independently, an angular momentum parallelization over (l, m) would be straightforward and could assemble another $(l_{\max} + 1)^2$ processes. For this parallelization collective matrix-matrix multiplications cannot anymore be applied, but matrix-vector multiplications

have to be used, which would slow-down the algorithm by a factor of two[3]. Although our test calculations revealed that this loss can be outbalanced by the usage of more than four processes, the significant decrease of the overall efficiency can be hardly motivated.

Open-MP Parallelization

On the same level but concordant with the usage of matrix-matrix multiplications, our method KKRnano can be executed defining several OpenMP threads. Such an OpenMP approach is of particular advantage since state-of-the-art supercomputers are predominantly build in hybrid-architecture combining 2-32 CPU's in a shared memory environment on one node and up to ten-thousands of those nodes communicating as distributed memory units amongst each other. The shared memory approach gives us a handle to circumvent the notoriously low memory resources on present-day supercomputers.

The OpenMP threads share the work performed in the sparse matrix-matrix multiplications during the iterative procedure, which are computationally still the most demanding parts. More precisely the loop over atoms in Alg. 4.2 (line 1 to 8) can optionally be run with multiple private threads. Additionally, if applied, the construction and application of the preconditioning matrix is distributed to the same number OpenMP threads. As illustrated in Fig. 4.13 this approach leads to a speedup of 1.4 to 1.8 and 1.7 to 2.5 using two and four OpenMP threads, respectively. The best performances of this parallelization are again reached for large systems. Since no additional network communication is required for the OpenMP loop exclusively the ratio of OpenMP parallelized and non-OpenMP parallelized code determines the speedup. Besides the speedup, it is noteworthy that the OpenMP parallelization pushes the memory bound limits of KKRnano up to maximal system size of 8000-16000 atoms in the supercell in full-potential treatment.

The combination of all four levels of parallelization and its speedup is shown in Fig. 4.15 for one exemplifying self-consistency step. Here parallelization of up to 16 processes per atom are probed including spin, energy and OpenMP parallelization. For up to 32768 processors for the system of 4096 atoms a speedup which is close to the ideal situation can be observed. The next step to calculations of 65536 processors comes along with a significant decrease in efficiency, but nevertheless leads to a speedup of eight out of ideally 16. Although the precise performance depends on the computational parameters and is therefore different for every system treated, we find that the usage of $8 \times N$ processors in most of the cases guarantees high parallel performance. Keeping in mind, that KKRnano can be used for the calculation of up to 8000 to 16000 atoms, our method is ideally suited to run on at least one hundred thousand of processors. Here, the limitation in system size N is appearing due to the memory restrictions on present architectures, which leads to the conjecture that on future supercomputing architectures even larger systems on a larger number of processors can be calculated.

---

[3]See section 4.2.1 for further explanation on the preference of matrix-matrix operations.
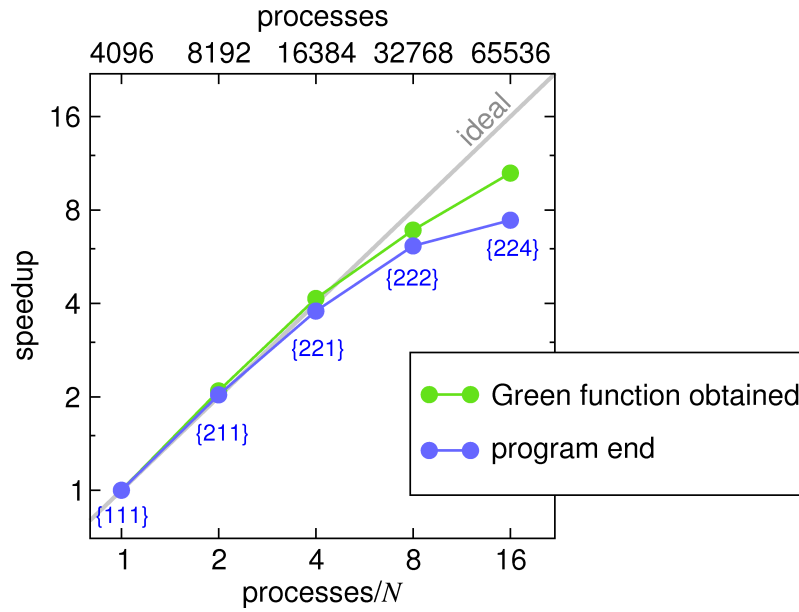
Figure 4.15.: *Speedup of KKRnano (blue) combining subsequently all levels of paralleliza-tion a BlueGene/P[82] architecture for Ni_xPd_{1-x} of 4096 atoms versus ideal speedup (gray line). The details to this calculations are the same as described in caption of Fig. 4.12. Labels $\{p_S\ p_E\ t_{OMP}\}$ specify the number of processes/ threads used per spin (p_S), energy (p_E), and OpenMP parallelization (t_{OMP}). For the energy parallelization the dynamic load-balancing as explained in the text has been adopted. Apparent from the legend, the scaling is obtained at two different positions in KKRnano. First, directly after the solution of the Dyson equation (Green function obtained) and secondly after completing one full self-consistency step (program end).*

4.4 Truncation of interaction

Having the possibility to treat system sizes of many thousands of atoms directly raises the question whether it is worthwhile to exploit the nearsightedness of the density matrix as proposed by W. Kohn [21]. While the diagonal part of the density matrix $\rho(\mathbf{r}, \mathbf{r}')$ is equivalent to the electron density, the full expression for ρ reads

$$\rho(\mathbf{r}, \mathbf{r}') = -\frac{1}{\pi} \operatorname{Im} \int_{-\infty}^{\infty} dE \, f_T(E - E_F) \, G(\mathbf{r}, \mathbf{r}', E). \qquad (4.39)$$

Hence, the Green function gives the density matrix which facilitates a transfer of the nearsighted principle to the KKR Green function approach. Depending on material properties, e.g. the localization of electronic states, the multiple-scattering interaction can be restricted indeed to a local interaction zone (LIZ) of a local cluster with a few hundred up to a few thousand atoms with reasonably small loss of accuracy [32]. This approach has been first introduced to the KKR Green function theory by Wang *et al.* [25] under the name "locally self-consistent multiple scattering" (LSMS). Instead of constructing the method on the basis of LIZ's, Zeller [32] proposed an alternative realization of the nearsightedness principle which allows for an optional use of truncation. In KKRnano we will follow the latter concept.

If such LIZ's or truncations are used, the linear matrix equation becomes even sparser. In fact the number of non-zero entries of the matrix $A = \Delta t G^{\mathrm{r}} - 1$ becomes independent of system size N and proportional to $N_{\mathrm{cl}} \cdot N_{\mathrm{tr}}$. Accordingly, the Dyson equation must be solved only in the individual LIZ's. Restricting ourselves for the sake of simplicity to equations without applied initial guess, the right hand side $b = -\Delta t$ (4.16) has non-zero entries only for the on-site term and remains the same as without truncation. Limiting the solving vector X and the system matrix A to the LIZ, the Dyson equation reads

$$\sum_{\mu} \sum_{L'} A_{LL'}^{\nu\mu} X_{L'L''}^{\mu} = b_{L'L''}, \qquad (4.40)$$

where ν runs over the atoms within the truncation cluster and μ is the index for the screened reference Green function.

In order to perform this operation in practice, N_{tr} instead of N multiplications of each submatrix ν of $A_{LL'}^{\nu\mu}$ have to be performed[4]. As those operations need about 90% of the computational work of one TFQMR iteration, ideally the truncation leads to a reduction of the computational time by N_{tr}/N. Hence, with truncation the algorithm used in KKRnano scales in the ideal case $\propto N_{\mathrm{it}} N_{\mathrm{cl}} N_{\mathrm{tr}} N$ with number of atoms N. Assuming that losses due to MPI communication and overhead are neglected for the moment, the atom parallelization in KKRnano with its distribution of work to N processors potentially leads to an execution time independent of N. Fig. 4.16 reveals that KKRnano operates close to this ideal scenario and except for minor losses due to load imbalance, MPI communication and overhead the execution time indeed remains almost constant for $N > N_{\mathrm{tr}}$. In other words, KKRnano is scaling linearly for $N > N_{\mathrm{tr}}$.

By truncation of long-range interactions obviously errors are introduced in relation to the treatment with full interaction. A first measure for this error is the difference in total energy,

---

[4]This operation corresponds to step 4 in Alg. 4.2.

ΔE: With the applied cut-off of 959 atoms in the interaction zone in the calculations shown in Fig. 4.16, ΔE is below 2 meV per atom. We will comment on the handling of such errors in greater detail in the next subsection and continue with the discussion of the truncation scheme.
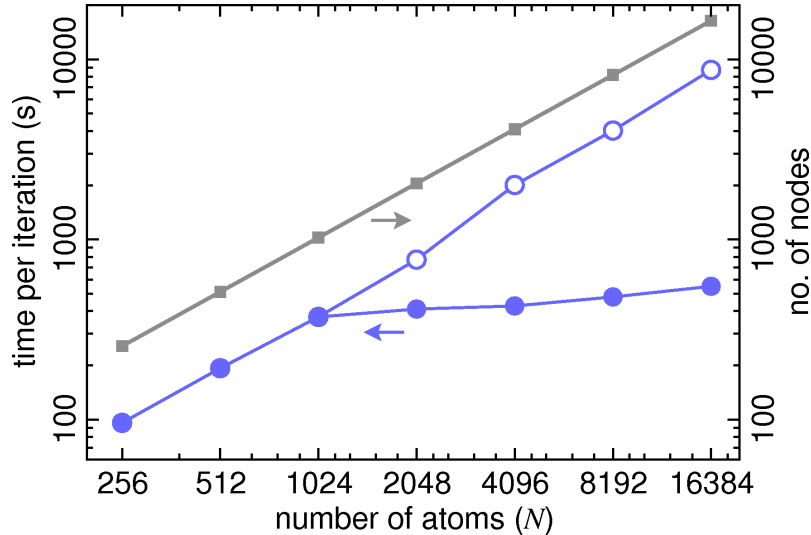

Figure 4.16.: *Scaling of KKR*nano *in double logarithmic representation without (open blue circles) and with applied truncation (filled blue circles). The test system is $Ni_{1-x}Pd_x$ with $x = 5\%$ and one k-point on JUGENE. Here, parallelization over atom, spin and two groups of energy points is used. The number of processing nodes (4 CPU's per node) is shown in dark grey.*

As indicated before, an important advantage of our approach compared to other methods [25] is, that linear scaling can be exploited optionally. For systems where long-range interactions are present and accordingly interaction zones with tens of thousands of atoms would be required to properly describe the electronic structure, in KKR**nano** the truncation can be switched off and the full periodicity of the supercell method can be used instead. Besides the advantageous linear scaling when applying truncation, the high degree of sparsity arising from it opens the possibility to extend the range of system sizes − which is mainly memory bound − to more than 16000 atoms as shown in Fig. 4.16. Since the truncation does not lower the efficiency of the optional levels of parallelization, calculations of 16000 atoms on hundreds of thousands processors become feasible.

4.4.1 Error correction

It should be noted, that the truncation of the multiple-scattering interaction can be further enhanced by refining the spatial cut-off individually for each scattering channel. Following the route of the LSMS method it turns out that scattering above the leading contribution $l = 0$ can be neglected in particular cases after a few shells without crucial loss of accuracy. This refinement leads to matrices of even higher sparsity which could be straightforwardly exploited in the iterative solution in KKR**nano** and potentially leads to additional speedup. However, this additional cut-off is not easy to control and therefore requires precise con-
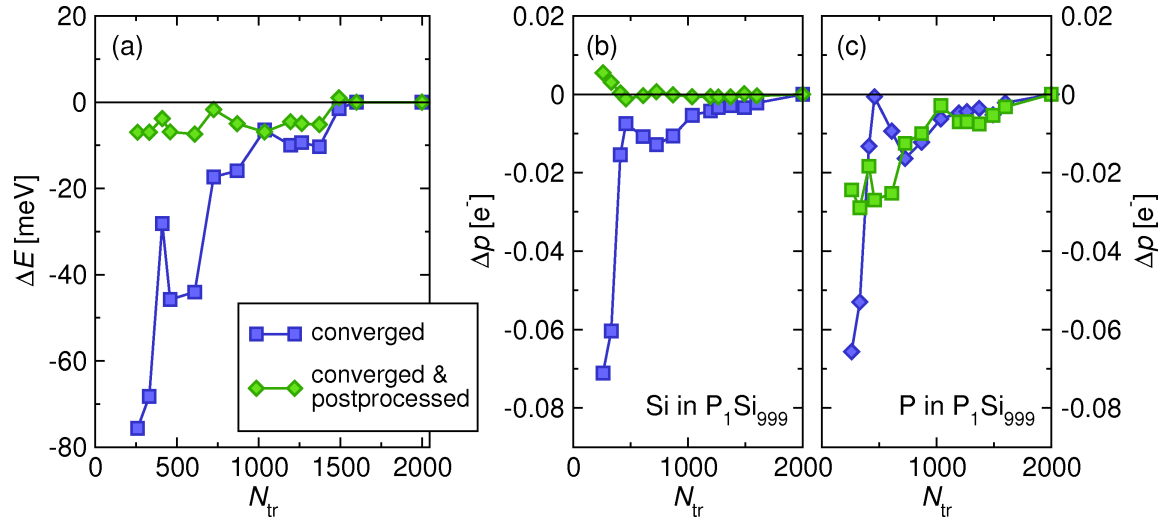
Figure 4.17.: *Error in total energy ΔE per atom (a) and error in the p contribution to the charge (Δp) per Si and P atoms (b) and (c), respectively. For all different sizes of truncation zones N_{tr} converged values are shown versus the postprocessed ones (see text for explanation). All errors are measured relative to the results of a calculation of the same system without truncation.*

vergency checks for every new system to generate accurate results. Due to this downside we do not follow this path in KKRnano and rather investigate how induced errors can be minimized in the truncation scheme.

For a sample system of phosphorus substitutional doping in silicon the error in total energy and charge per site is shown in Fig. 4.17 as function of the size of the local interaction zones N_{tr}. Focusing first on the results obtained by straightforwardly converging the electronic structure with applied truncation, relatively large deviations of up to 80meV per atom in total energy and almost 0.1 electrons in the occupation of p orbitals per atom can be observed for small N_{tr}. Increasing N_{tr} leads nearly monotonously to more accurate results. However, a more advanced procedure can be used to cure for part of this errors. Even for smallest N_{tr} the induced error can be corrected to a great extent by the following scheme: First, perform the self-consistency cycle with applied truncation down to convergency. As a second step this converged potential is used as input for one self-consistency-step which includes the full interaction, i.e. $N_{tr} = N$. The results in Fig. 4.17 clearly show that this procedure leads in particular for $N_{tr} < 1000$ to a reduction of the errors. This pronounced gain in accuracy can be mainly attributed to the improved calculation of the kinetic energy as a functional of the non-truncated charge density obtained in the postprocessing step [95]. The error in total energy is limited to 10% of the original truncation error which also holds for the p charge on Si atoms. The precise calculation of the charge on P impurities appears to be more difficult and the maximal error is only reduced by a factor of two in this case. Overall, error correction renders possible to maintain the same accuracy with smaller truncation zones, which leads to significantly faster calculations.

4.5 Summary

With KKR**nano** we have developed an efficient massively parallel density-functional method
capable to treat supercells with on the order of ten thousand atoms. The combination of
using the screened reference representation and the iterative solution of the Dyson equation
in KKR**nano** facilitated to rule out the usual difficulties and deficiencies in parallelization.
Thereby, highest parallel performance up to one hundred thousand processors is achieved,
which is not suffering limitations by communication or load imbalance. We have successfully
constructed novel schemes such as block-circulant preconditioning to accelerate the iterative
process in KKR**nano** and saved by this optimization overall one order of magnitude in
computational time. This speedup additionally pushes the limits of the maximal system
size even for practical applications up to 8000 atoms. In particular in the limit of that
large systems the optional truncation of interaction as implemented in KKR**nano** can be
exploited and leads to a linear scaling in system size. By taking into account the massive
parallelization and in particular the parallelization over sites or atoms, the execution time
for calculations of systems with increasing size remains constant, as long the number of
processors is increased proportional to the number of atoms. The large supercells which can
be treated with KKR**nano** allow for a statistically sound *ab initio* description of material
properties even of strongly disordered systems. In chapter 6 and 7 we will make use of
this strength in order to analyze long-range, defect-mediated magnetic coupling in dilute
magnetic semiconductors as well as electron localization in phase-change materials.

CHAPTER 5

Important extensions to KKRnano

Resuming at this stage KKRnano is properly designed for calculations of large metallic systems which show no strong correlations. In fact even with this restriction plenty of problems in material science could already be addressed. However, in order to achieve our task of the development of a multiple purpose method which allows for applications to a broader spectrum of materials, additional effort has to be made: Of paramount importance is to enable an accurate treatment of semiconducting or insulating materials. Here, the problems of the standard scheme arise from inaccuracies in the calculation of the charge which can lead to unphysical results if no or only few electronic states are located at the Fermi level. We will prevent this failure of the algorithm by incorporating Lloyd's formula [34] in KKRnano. Another important extension is made by adding Coulomb repulsion by means of the LDA+U algorithm [37] to cure the shortcoming of LDA or GGA-functionals for strongly correlated electrons. Although not required for the correct description of the electronic structure, but an essential advantage of the KKR method above many of the electronic structure methods is the possibility to obtain the individual magnetic interactions with the Lichtenstein formula [38].

All of those extensions are well-established in the field of multiple-scattering KKR theory. However, the aim to use those for electronic structure calculations of many thousand atoms on massively parallel architectures requires elaborate steps of modification and optimization. In this chapter, we will give insight into the efficient massively parallel realization of all above mentioned extensions in KKRnano.

5.1 Lloyd's formula

In order to ensure the correct treatment of band-gap materials the KKR methodology has to be extended by Lloyd's formula. This formula was first derived by Lloyd [34] and is frequently used to accurately describe such materials within the KKR framework, e.g. in [96–98]. In this section we will explain the shortcoming of the standard KKR scheme and introduce the charge normalization by using Lloyd's formula. Thereby we will follow conceptually the references [35, 36]. In large scale calculations the evaluation of Lloyd's formula as it is implemented in existing schemes becomes destructively slow − not even considering the parallel performance of those algorithms. Here, we will analyze the challenges arising from large scale massively parallel calculations of Lloyd's formula and come up with a new algorithm which is ideally suited for such demanding calculations.

As introduced in section 3.1 the charge density in the KKR Green function method can in general be expressed as

$$\rho(\mathbf{r} + \mathbf{R}^n) = -\frac{2}{\pi} \text{Im} \int_{E_B}^{\infty} dE \; f_T(E - E_\text{F}) \, G(\mathbf{r} + \mathbf{R}^n, \mathbf{r} + \mathbf{R}^n, E), \tag{5.1}$$

where the complex contour integration technique is exploited. This means the Fermi-Dirac distribution $f_T(E - E_\text{F})$, which is determined by the Fermi level E_F and the temperature broadening T, is introduced. Due to the rapid decay of $f_T(E - E_\text{F})$ above E_F the integration in (5.1) is then extended from the bottom of the valence states E_B formally up to $E = \infty$ but only a couple of tens integration points are required. (5.1) is then in practice computed by the sum over the energy integration points E_i using

$$\rho(\mathbf{r} + \mathbf{R}^n) = -\frac{2}{\pi} \sum_i \text{Im} \left[w_i \, G(\mathbf{r} + \mathbf{R}^n, \mathbf{r} + \mathbf{R}^n, E_i) \right], \tag{5.2}$$

where in addition the weights w_i enter which are defined by $f_T(E - E_F)$ for each integration point i. The inaccuracy which occurs without using Lloyd's formula is appearing in the Green function G. The general definition is given by

$$\begin{aligned} G(\mathbf{r} + \mathbf{R}^n, \mathbf{r} + \mathbf{R}^n, E) = &-i\sqrt{E} \sum_L^{\infty} R_L^n(\mathbf{r}, E) H_L^n(\mathbf{r}, E) \\ &+ \sum_{LL'}^{\infty} R_L^n(\mathbf{r}, E) G_{LL'}^{nn}(E) R_{L'}^n(\mathbf{r}, E), \end{aligned} \tag{5.3}$$

where the summation is − in the practical evaluation − limited to the lowest $(l_\text{max} + 1)^2$ components:

$$\begin{aligned} \tilde{G}(\mathbf{r} + \mathbf{R}^n, \mathbf{r} + \mathbf{R}^n, E) = &-i\sqrt{E} \sum_L^{l_\text{max}} R_L^n(\mathbf{r}, E) H_L^n(\mathbf{r}, E) \\ &+ \sum_{LL'}^{l_\text{max}} R_L^n(\mathbf{r}, E) G_{LL'}^{nn}(E) R_{L'}^n(\mathbf{r}, E), \end{aligned} \tag{5.4}$$

In both equations $G_{LL'}^{nn}(E)$ are the structural scattering matrices as obtained from the Dyson equation (3.61). $R_L^n(\mathbf{r}, E)$ and $H_L^n(\mathbf{r}, E)$ are the single site regular and irregular solutions[1]. As done in (5.4) the truncated and actual representation of the Green function will be distinguished by denoting it $\tilde{G}$ and G from here on, which also holds for all derived quantities. The inaccuracy in $\tilde{G}$ then directly translates in an approximate charge density

$$\tilde{\rho}(\mathbf{r} + \mathbf{R}^n) = -\frac{2}{\pi} \sum_i \text{Im}\left[\tilde{w}_i\, \tilde{G}(\mathbf{r} + \mathbf{R}^n, \mathbf{r} + \mathbf{R}^n, \tilde{E}_i)\right], \tag{5.5}$$

where also the weight $\tilde{w}_i$ and the energy point $\tilde{E}_i$ are subject to the finite angular momentum cut-off $l_{\max}$. Such shifts in the complex energy plane arise from the requirement of charge neutrality. Considering the total charge

$$\tilde{Q} = \sum_n \int_n d\mathbf{r}\, \tilde{\rho}(\mathbf{r} + \mathbf{R}^n) \tag{5.6}$$

inaccuracies in $\tilde{\rho}$ have to be outbalanced by a shift of the Fermi level to equalize the nuclear charge and $\tilde{Q}$. Metals exhibit a sufficiently large density of states at the Fermi edge to keep this shortcoming of the algorithm negligible. On the contrary, in semiconductors the Fermi level underlies more crucial changes and can be pushed spuriously from the gap into the edge of the valence or conduction band.

In order to cure this potentially severe shortcoming and maintain a multiple-purpose code, we will follow a scheme proposed by Zeller [36] with Lloyd's formula as central ingredient. This formula will be derived first for a general case and then we will continue with the details of the algorithm.

5.1.1 General derivation of Lloyd's formula

A derivation of Lloyd's formula has been given in [35] which will be briefly redrawn here. As introduced in section 3.1 the electron density of states can be expressed by a general energy-dependent Green function

$$G(E) = (E - H)^{-1} \tag{5.7}$$

as

$$n(E) = -\frac{2}{\pi} \text{Im}\left[\text{Tr}\, G(E)\right]. \tag{5.8}$$

If using a transformation e.g. from free-space to a screened reference system the change of the density of states of both systems becomes meaningful. This quantity $\Delta n(E)$ is directly linked to the difference of the two Green functions g and G of both systems and reads:

$$\Delta n(E) = -\frac{2}{\pi} \text{Im}\left[\text{Tr}\, \left[G(E) - g(E)\right]\right]. \tag{5.9}$$

---

[1]For more information how those functions are obtained and defined the reader is referred to section 3.2

Assuming that g and G are connected via a Dyson equation $G = g + g\Delta VG$ (5.9) can be rewritten as

$$\Delta n(E) = -\frac{2}{\pi} \operatorname{Im} \left[\operatorname{Tr} \left[g(E) \, \Delta V \, G(E) \right] \right]. \tag{5.10}$$

From the Dyson equation the identity

$$G = g + g\Delta VG = \frac{1}{1 - g\Delta V} g \tag{5.11}$$

can be easily derived and exploited to reformulate (5.9):

$$\Delta n(E) = -\frac{2}{\pi} \operatorname{Im} \left[\operatorname{Tr} \left[g(E)\Delta V \frac{1}{1 - g(E)\Delta V} g(E) \right] \right]. \tag{5.12}$$

With the formal energy derivative of $g(E)$:

$$\frac{d}{dE} g(E) = -(E - H)^{-2} = -g(E) \cdot g(E), \tag{5.13}$$

$\Delta n(E)$ can be expressed as

$$\Delta n(E) = -\frac{2}{\pi} \operatorname{Im} \left[\operatorname{Tr} \frac{d}{dE} \ln \left[1 - g(E)\Delta V \right] \right]. \tag{5.14}$$

With the matrix identity $\ln \det(A) = \operatorname{Tr} \ln(A)$ (5.14) can be written as

$$\Delta n(E) = -\frac{2}{\pi} \frac{d}{dE} \operatorname{Im} \left[\ln \det \left[1 - g(E)\Delta V \right] \right], \tag{5.15}$$

which leads to an expression for the integrated density of states $N(E)$:

$$\Delta N(E) = -\frac{2}{\pi} \operatorname{Im} \left[\ln \det \left[1 - g(E)\Delta V \right] \right]. \tag{5.16}$$

For further considerations it is convenient to split $g(E)$ into a single site $g_{\mathrm{s}}(E)$ and a multiple scattering contribution $g_{\mathrm{ms}}(E)$ with $g(E) = g_{\mathrm{s}}(E) + g_{\mathrm{ms}}(E)$. According to that Eq. (5.16) can be formally separated in:

$$
\begin{aligned}
\Delta N(E) &= -\frac{2}{\pi} \operatorname{Im} \left[\ln \det \left[1 - g_{\mathrm{s}}(E)\Delta V - g_{\mathrm{ms}}(E)\Delta V \right] \right] \\
&= -\frac{2}{\pi} \operatorname{Im} \left[\ln \det \left[\left(1 - g_{\mathrm{s}}(E)\Delta V \right) \left(1 - g_{\mathrm{ms}}(E) \, \mathcal{V} \right) \right] \right] \\
&= -\frac{2}{\pi} \operatorname{Im} \left[\ln \det \left[1 - g_{\mathrm{s}}(E)\Delta V \right] \right] - \frac{2}{\pi} \operatorname{Im} \left[\ln \det \left[1 - g_{\mathrm{ms}}(E) \, \mathcal{V} \right] \right],
\end{aligned}
\tag{5.17}
$$

where $\mathcal{V}$ is representing

$$\mathcal{V} = \frac{\Delta V}{1 - g_{\mathrm{s}}(E)\Delta V}. \tag{5.18}$$

The clear separation of the single and multiple scattering parts allows to discuss and compute both terms in the last line of (5.17) individually. Note, that for the calculation of the

integrated density of states the Green function G of the actual system is not required. This independency will be of crucial importance later on. Applying these general considerations to the screened KKR Green function method overall four contributions enter the integrated density of states:

$$N(E) = N_\mathrm{s}^\mathrm{r}(E) + N_\mathrm{ms}^\mathrm{r}(E) + \Delta N_\mathrm{s}(E) + \Delta N_\mathrm{ms}(E), \tag{5.19}$$

where $N_\mathrm{s}^\mathrm{r}(E)$ and $N_\mathrm{ms}^\mathrm{r}(E)$ denote the single-site (s) and multiple-scattering (ms) terms of the reference system, which will be represented by $N^\mathrm{r}(E) = N_\mathrm{s}^\mathrm{r}(E) + N_\mathrm{ms}^\mathrm{r}(E)$ in the following.

5.1.2 Renormalization of the charge

Having obtained insight into Lloyd's formula an alternative approach to calculate the total charge Q can be chosen. Instead of using the integration in (5.6) Q can also be obtained by

$$Q = \mathrm{Im} \int_{-\infty}^{\infty} dE \; f_T(E - E_F) \, n(E) = \sum_i \mathrm{Im} \left[w_i \, n(E_i) \right], \tag{5.20}$$

where $n(E)$ is representing the generalized density of states, which is not only defined on the real axis but in the entire complex plane. Since the integrated density of states $N(E)$ is given by Lloyd's formula its energy derivative can provide $n(E)$.

Here, we follow the notation of Drittler *et al.* [99] and continuing studies [36, 100] expressing Lloyd's formula as:

$$
\begin{aligned}
N(E) = N^\mathrm{r}(E) + \frac{2}{\pi} \sum_n &\ln \det \left| \Delta\alpha_{LL'}^n(E) \right| \\
&- \frac{2}{\pi} \ln \det \left| \delta_{LL'}\delta_{nn'} - G_{LL'}^{\mathrm{r},nn'}(E)\Delta t_{LL'}^{n'}(E) \right|,
\end{aligned} \tag{5.21}
$$

with $N^\mathrm{r}(E)$ representing the integrated density of states of the reference system. The matrix $\Delta\alpha_{LL'}^n$ contains the single-site contribution to Lloyd's formula and can be computed by

$$\Delta\alpha_{LL'}^n = \delta_{LL'} + \int_n d\mathbf{r} \; H_{L'}^{\mathrm{r},n}(\mathbf{r}, E) \left[V^n(\mathbf{r}) - V^{r,n}(\mathbf{r}) \right] R_L^n(\mathbf{r}, E), \tag{5.22}$$

where $H_{L'}^{\mathrm{r},n}(\mathbf{r}, E)$ is the irregular solution of the reference system with potential $V^{\mathrm{r},n}(\mathbf{r})$ and $R_L^n(\mathbf{r}, E)$ represents the regular solution of the real system defined by $V^n(\mathbf{r})$. From the definition of $\Delta\alpha_{LL'}^n$ it becomes clear that the calculation can be performed locally for each site and therefore can be straightforwardly conducted in KKR**nano**. It should be noted that for periodic systems, the last contribution in (5.21) has to be extended to an integration over the Brillouin Zone. Although in KKR**nano** this **k**-space integration is performed, its application is straightforward and the **k**-dependency will be disregarded here and in the following for the sake of readability[2].

It remains to address the question why the accuracy of this alternative approach is not or considerably less influenced by the truncation in l. Apparent from (3.60) and (5.22) the

---

[2]The **k**-dependent expression of (5.21) can be found e.g. in [36]

single-site quantities $\Delta t_{LL'}^n$ and $\Delta \alpha_{LL'}^n$ are restricted to a finite $l_{\max}$ as well. However, there is a crucial difference of this truncation to the one in (5.4). Restricting, $\Delta t_{LL'}^n$ and $\Delta \alpha_{LL'}^n$ leads to a mapping of the problem on a subspace of leading contributions with $l \leq l_{\max}$. In this subspace the solution can be obtained exactly. Although inaccuracies in the spatially resolved charge $\rho(\mathbf{r})$ occur due to the mapping onto this subspace, the total charge adds up to the correct value. According to this, Lloyd's formula does not provide any information on $\rho(\mathbf{r})$ but on the energy dependency i.e. the density of states $n(E)$.

The remaining task is to combine the information of the spatial distribution of charge as obtained from the standard procedure using the Dyson equation and the correct density of states from Lloyd's formula. Both parts can be directly related to each other by

$$-\frac{2}{\pi}\lambda_i \sum_n \int_n \mathrm{Im}\left[w_i\, \tilde{G}(\mathbf{r}+\mathbf{R}^n, \mathbf{r}+\mathbf{R}^n, E_i) \right] = \mathrm{Im}\left[w_i\, n(E_i) \right]. \qquad (5.23)$$

Here, the energy points E_i and weights w_i are consistent with the Fermi level obtained from this algorithm of the previous self-consistency step. The left hand side, the spatial integration over the truncated Green function $\tilde{G}(\mathbf{r}+\mathbf{R}^n, \mathbf{r}+\mathbf{R}^n, E_i)$, is decorated by normalization factors λ_i. These factors facilitate the calculation of a normalized charge density by

$$\rho_\lambda(\mathbf{r}+\mathbf{R}^n) = -\frac{2}{\pi}\sum_i \lambda_i\, \mathrm{Im}\left[w_i\, \tilde{G}(\mathbf{r}+\mathbf{R}^n, \mathbf{r}+\mathbf{R}^n, E_i) \right]. \qquad (5.24)$$

This procedure guarantees to obtain the correct total charge by

$$\sum_n \int_n d\mathbf{r}\, \rho_\lambda(\mathbf{r}+\mathbf{R}^n) = Q, \qquad (5.25)$$

which leads to charge neutrality without varying the Fermi level from the correct position. A further minor modification of the scheme which prevents numerical instabilities by avoiding divergencies is presented in Zeller [36] and used in KKRnano. However, the central concept remains the same, the density of states $n(E)$ has to be obtained from Lloyd's formula.

Looking back on the derived algorithm allows to identify the critical parts for an efficient massively parallel computation of large-scale systems. All calculations except for the energy derivative of Lloyd's formula (5.21) can be performed locally for each site and require only little communication. While the single-site contribution in (5.21) represented by $\Delta \alpha_{LL'}^n$ can be obtained locally as done in existing KKR programs [36], the back-scattering parts require a reformulation. In the following the computational treatment of two important terms will be discussed: The back-scattering term of the reference system and of the actual system. In particular the latter has the potential to slow down KKRnano considerably: If a numerical energy derivative is calculated directly from (5.21) using two additional supporting energy points in the complex plane the computational time would increase by at least a factor of three.

5.1.3 The reference system

Apparent from (5.21), the integrated charge density of the reference system $N^{\mathrm{r}}(E)$ has to be obtained, which can be calculated by Lloyd's formula in relation to the free space Green

function g^0, which leads to an additional contribution[3] of

$$\Delta N^{\mathrm{r},0}(E) = \ln \det \left| I - g^0(E)\, t^{\mathrm{r}}(E) \right|. \tag{5.26}$$

In the existing KKR codes this contribution is calculated in **k**-space. In order to circumvent a second inversion of a matrix of size proportional $N \times N$ with the system size N which would have to be solved iteratively, in KKR**nano** (5.26) is solved in real space in the reference cluster. Applying the energy derivative, (5.26) reads:

$$
\begin{aligned}
\frac{d}{dE}\ln \det \left| I - g^0(E)\, t^{\mathrm{r}}(E) \right| &= \frac{d}{dE}\operatorname{Tr} \ln \left| I - g^0(E)\, t^{\mathrm{r}}(E) \right| \\
&= \operatorname{Tr}\frac{d}{dE}\ln \left| I - g^0(E)\, t^{\mathrm{r}}(E) \right| \\
&= \operatorname{Tr}\left(\left| I - g^0(E)\, t^{\mathrm{r}}(E) \right|^{-1} \frac{d}{dE} \left| I - g^0(E)\, t^{\mathrm{r}}(E) \right| \right) \\
&= \operatorname{Tr}\left(X \right).
\end{aligned}
\tag{5.27}
$$

According to the last step in (5.27) the matrix X follow the relation

$$\left(1 - g^0\, t^{\mathrm{r}} \right) X = \frac{d}{dE}\left(1 - g^0\, t^{\mathrm{r}} \right) \tag{5.28}$$

which is a linear matrix equation $AX = b$ likewise to the Dyson equation

$$\left(1 - g^0\, t^{\mathrm{r}} \right) G^{\mathrm{r}} = g^0, \tag{5.29}$$

solved anyhow in KKR**nano** to obtain the reference structure constants G^{r}. Due to the fact that the matrix A is identical for (5.29) and (5.28) the same LU factorization can be exploited by just providing an alternativ right-hand side. This is of particular importance as LU factorizations scale $\mathrm{O}(N_{cl}^3)$ with the number of reference cluster sites N_{cl} and are usually the most cumbersome operation in the computation of the reference t-matrix and Green-function. Still, we have to come up with a scheme to compute the new right hand side

$$\frac{d}{dE}\left(1 - g^0(E)\, t^{\mathrm{r}}(E) \right) = -\frac{dg^0(E)}{dE}\, t^{\mathrm{r}}(E) - g^0(E)\,\frac{dt^{\mathrm{r}}(E)}{dE}. \tag{5.30}$$

The starting point is the definition of the scattering matrix elements $g^0(E)$ in the same notation as described in section 3.3

$$g_{LL'}^{0nn'}(E) = -(1 - \delta_{nn'})\, 4\pi i \sqrt{E} \sum_{L''} i^{l-l'+l''} C_{LL'L''}\, h_{l''}(\sqrt{E}r^{nn'})\, Y_{L''}(\mathbf{r}^{nn'}), \tag{5.31}$$

where $r^{nn'}$ and $\mathbf{r}^{nn'}$ denote the distance and the distance vector between two sites n and n', $C_{LL'L''}$ are Gaunt-coefficients and h_l Hankel functions. Together with the analytical

---

[3]For the sake of simplicity all angular momentum indices will be not shown from here on and only the energy dependency will be labeled.

dependency of the derivative

$$\frac{d}{dE}\left(\sqrt{E}\, h_l(\sqrt{E}r)\right) = \frac{r}{2}\, h_{l-1}(\sqrt{E}r) - \frac{l}{2\sqrt{E}}\, h_l(\sqrt{E}r) \tag{5.32}$$

the energy derivative $\frac{dg^0}{dE}$ can be straightforwardly calculated.

The derivative of the reference t-matrix, $\frac{dt^{\mathrm{r}}}{dE}$ can as well be obtained analytically. The spherically symmetric Schrödinger equation (3.16) which is rewritten here

$$\left[-\frac{1}{r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{r^2} + V(r) - E\right] R_l(r, E) = 0 \tag{5.33}$$

with the repulsive potential of the reference system

$$V(r) = \begin{cases} V^{ref} & r \leq S \\ 0 & r > S \end{cases} \tag{5.34}$$

has the formal solution

$$R_l(r, E) = A_l\, j_l(\sqrt{E}r)\,\Theta(S - r) + B_l\, h_l(\sqrt{E-V}r)\,\Theta(r - S). \tag{5.35}$$

Here, the Bessel-functions j_l and Hankel-functions h_l are the solution for $r \leq S$ and $r > S$, respectively. The coefficients A and B can be obtained straightforwardly by the matching condition of $R_l(r, E)$ and $\frac{d}{dr}R_l(r, E)$ at $r = S$ as well as the overall normalization. Conducting this step and integrating

$$t_l^{\mathrm{r}}(E) = \int_0^S dr\, j(\sqrt{E}r)\, V(r)\, R_l(r, E), \tag{5.36}$$

where $V(r)$ is again representing the reference potential (5.34), leads to the analytical expression

$$t_l^{\mathrm{r}}(E) = \frac{1}{\sqrt{E}}\frac{\sqrt{E}\, j_{l+1}(\sqrt{E}S)\, j_l(\sqrt{E}S) - \sqrt{E-V}\, j_l(\sqrt{E}S)\, j_{l+1}(\sqrt{E-V}S)}{\sqrt{E}\, h_{l+1}(\sqrt{E}S)\, j_l(\sqrt{E}S) - \sqrt{E-V}\, h_l(\sqrt{E}S)\, j_{l+1}(\sqrt{E}S)}. \tag{5.37}$$

Using (5.32), which is also valid for the Bessel-functions j_l, $\frac{d}{dE}t_l^{\mathrm{r}}(E)$ can be straightforwardly obtained.

Summarizing, by developing this new scheme we managed to avoid any additional demanding matrix operations and only included operations which can be computed locally on each site. This part of the algorithm can thereby easily be parallelized over atoms and preserve the parallel performance of KKR`nano`.

5.1.4 The actual system

The bottleneck for large systems of thousands of inequivalent atoms is usually the direct calculation of the determinant $\det\left|\delta_{\mathrm{LL'}}\delta_{\mathrm{nn'}} - \mathrm{G}_{\mathrm{LL'}}^{\mathrm{r,nn'}}(\mathrm{E})\, \Delta\mathrm{t}_{\mathrm{LL'}}^{\mathrm{n'}}(\mathrm{E})\right|$ via an LU factorization, which is both time-consuming and not efficiently to parallelize. Within the next two pages we will derive an algorithm which requires no further expensive computation.

The multi-scattering part of the energy derivative of the Lloyd's formula (5.21), reads after similar transformations as in (5.27) in algebraic representation:

$$
\begin{aligned}
\frac{d}{dE} \ln \det M &= \frac{d}{dE} \mathrm{Tr} \, \ln M \\
&= \mathrm{Tr} \left(\frac{d}{dE} \ln M \right) \\
&= \mathrm{Tr} \left(M^{-1} \frac{dM}{dE} \right),
\end{aligned}
\tag{5.38}
$$

with

$$
M = (I - G^{\mathrm{r}} \Delta t)
\tag{5.39}
$$

and the identity matrix I. The crucial advantage of the above expression is that the derivation with respect to energy and the computational most demanding part − the inversion of the matrix M − can be independently obtained. In the following we will first address the evaluation of $\frac{dM}{dE}$ and then continue by discussing how M^{-1} is obtained in KKRnano.

Calculating the energy derivative

Using the definition in (5.39) we can rewrite:

$$
\frac{dM}{dE} = -\frac{dG^{\mathrm{r}}(E)}{dE} \Delta t(E) - G^{\mathrm{r}}(E) \frac{d\Delta t(E)}{dE}.
\tag{5.40}
$$

While $G^{\mathrm{r}}(E)$ and $\Delta t(E)$ are calculated for the algorithm without Lloyd's formula anyway, additional effort has to be made to obtain $\frac{dG^{\mathrm{r}}(E)}{dE}$ and $\frac{d\Delta t(E)}{dE}$. $\frac{dG^{\mathrm{r}}(E)}{dE}$ can be derived in close relation to the procedure used to solve the Dyson equation which connects the potential-free and screened reference system (4.2) and (5.29). To obtain not G^{r} but its derivative we start from the ansatz

$$
\frac{d}{dE} \left(\left(I - g^0 t^{\mathrm{r}} \right) G^{\mathrm{r}} \right) = \frac{dg^0}{dE}
\tag{5.41}
$$

which can be straightforwardly rewritten as

$$
\left(I - g^0 \, t^{\mathrm{r}} \right) \frac{dG^{\mathrm{r}}}{dE} = \frac{dg^0}{dE} - \frac{d}{dE} \left(I - g^0 \, t^{\mathrm{r}} \right) \, G^{\mathrm{r}},
\tag{5.42}
$$

where the energy derivative in the last term is only performed on the factor $\left(I - g^0 \, t^{\mathrm{r}} \right)$. By that in KKRnano the same LU factorization as needed to solve (5.29) and (5.28) can be used with again a different right hand side. The remaining task to evaluate $\frac{d}{dE} \left(I - g^0 t^{\mathrm{r}} \right)$ has already been computed in (5.27) and following equations to obtain the contribution of the reference system to the integrated electron density.

For the calculation of the energy derivative of $\Delta t = t - t^{\mathrm{r}}$, half of the work is already done in (5.37) and subsequent steps by obtaining $\frac{dt^{\mathrm{r}}}{dE}$. The remaining contribution $\frac{d}{dE} t(E)$ is numerically obtained by

$$
\frac{d}{dE} t(E) = \frac{t(E+\epsilon) - t(E-\epsilon)}{2\epsilon}.
\tag{5.43}
$$

which requires additional computation of $t(E)$ at two energies $E + \epsilon$ and $E + \epsilon$. Nevertheless also this computation is locally well-defined for each site and its execution time carries no weight.

Calculating the inversion

The obstacle is the calculation of M^{-1} which if calculated by direct inversion would scale with N^3 with the system size N. However, similar to the strategy exploited to solve the Dyson equation in KKRnano also this inversion can be performed iteratively. Then, the original expression

$$Y \cdot M = I, \tag{5.44}$$

where $Y = M^{-1}$ and I displays the unity matrix can be rewritten by replacing M by the actual definition of it (5.39) which yields:

$$Y = I + G\Delta t Y. \tag{5.45}$$

An iterative solution exploiting the same optimized scheme as used for the iterative solution of the Dyson equation (4.15)

$$X = \Delta t + G\Delta t X \tag{5.46}$$

would anyhow lead to a doubling of the computational effort and the memory requirements as compared to the using KKRnano without Lloyd's formula.

Comparing (5.45) and (5.46) reveals a short-cut for this computation: Both matrices X and Y can be directly linked by:

$$Y = X \, \Delta t^{-1}. \tag{5.47}$$

This means, the desired solution $Y = M^{-1}$ needed as input for the application of Lloyd's formula can be calculated by a simple matrix-matrix multiplication of the known Δt^{-1} and the solving vector of the Dyson equation which has to be obtained even without application of the Lloyd's formula.

Recalling the developed scheme all additional computations for Lloyd's formula can thereby be conducted performing exclusively inexpensive operations. To prove this point, we have conducted test calculations showing the influence of the usage of Lloyd's formula on the performance of KKRnano which is summarized in Tab. 5.1. It turns out that applying Lloyd's formula leads to additional computational time of clearly less than 10% of the standard procedure. Neither the computation of the contribution of the reference or the actual system to Lloyd's formula nor the small amount of additional MPI-communication present a crucial limitation for the performance. Thereby, also if Lloyd's formula is utilized in large-scale calculations, the high parallel efficiency of KKRnano can be maintained.

		t_{ref}	t_{Dys}	t_f	MPI Comm.
$Si_{212}P_4$	normal	0min13.67s	3min36.93s	4min2.13s	14.62 GByte
	Lloyd	0min20.04s	3min56.01s	4min14.79s	15.74 GByte
$Si_{504}P_8$	normal	0min17.43s	7min54.40s	8min58.20s	87.26 GByte
	Lloyd	0min25.98s	8min25.27s	9min31.18s	93.13 GByte

Table 5.1.: *Timings at different processing steps of KKR**nano** with and without Lloyd's formula, implemented as described above. t_{ref}, t_{Dys} and t_f denote the execution time after having obtained all information on the reference system such as the reference Green-function and t-matrices, the iterative solution of the Dyson equation and finalized the program, respectively. In addition the overall transfered data by MPI communication is analyzed based on performance measurements with scalasca [94]. The test system is in all cases a Si crystal with substitutional P-impurities in the specified concentration and an angular momentum cut-off of $l_{\max} = 2$. The timings are obtained on JUGENE [82] using the atom parallelization as well as two OpenMP threads.*

5.2 LDA+U: treating strong correlations

The shortcoming of the local density approximation (LDA) and the generalized gradient approximation (GGA) to the exchange-correlation functional in *ab initio* calculations of strongly correlated electrons is a long and well known fact. The development of approaches which cure this problem is still a major task in the material science community. Most of such approaches which aim at high accuracy such as the GW approximation (GWA) are intrinsically nonlocal and frequency dependent, which makes the computation a notoriously demanding task [58, 101]. This holds in particular if the system size and complexity of the supercell is increased.

Anisimov *et al.* [37] showed that the GWA can be approximated by the LDA+U scheme for localized states. In the LDA+U method which was developed by Anisimov *et al.* [102] the electrons are conceptually separated in localized and delocalized ones such as in the Anderson model [103]. While the delocalized orbitals are treated using the LDA functional, for the localized ones an additional orbital dependent screened Coulomb repulsion can be controlled by a set of parameters U for each orbital independently. Thereby although it is in principle possible to obtain those parameters from *ab initio* calculations [104], it should be noted that they are usually semi-empirically chosen to best fit experimental results. Once this is the case the clear-cut parameter-free characteristic of density functional calculations is of course partly lost. However, for many well-localized orbitals and many different systems the LDA+U approach has been extremely successfully applied amongst many others in e.g. [105, 106] and the preselection of the U parameter is often not crucially limiting the accuracy of the calculations.

From the computational point of view LDA+U combines two important advantages: First, no frequency-dependency has to be considered. Secondly, all necessary calculations can be performed locally on each site. Especially the latter is essential for a straightforward

parallelization over atoms as utilized in KKRnano. In the implementation of LDA+U we are following the main concepts of Anisimov *et al.* [37] and its realization in the KKR München-Jülich-program [107]. However, to guarantee precise and reliable results for gadolinium doping in gallium nitride, the system which we will focus on in chapter 6, two important extensions and refinements have to be made. Firstly, previous studies of GaN:Gd showed that for both the 4f- and 5d-states a Coulomb repulsion has to be included in order to match the experimental electronic structure. Therefore, we generalized the approach of the München-Jülich-program and allow the application of LDA+U to multiple orbitals per site − in practice one s, p, d, and f orbital each. Secondly, the convergency scheme of the self-consistency steps is revised to suspend numerical instabilities which can arise from the additional LDA+U contribution to the potential.

In the LDA+U framework the exchange-correlation energy is given by the sum of the LDA functional depending on the spin-resolved electron density $\rho^\sigma(r)$ and the specifically added Coulomb repulsion of energy E^U as a functional of the density matrix $\{n^\sigma\}$ plus the double-counting correction E_{dc}.

$$E^{LDA+U}[\rho^\sigma(r), \{n^\sigma\}] = E^{LDA}[\rho^\sigma(r)] + E^U[\{n^\sigma\}] - E_{dc}[\{n^\sigma\}] \tag{5.48}$$

Both additional contributions E^U and E_{dc} will be discussed in the following.

Regarding the density matrix $\{n^\sigma\}$ within the LDA+U scheme, exclusively the block-diagonal submatrices $\{n^\sigma\}_l, l = 0(s), 1(p), 2(d), 3(f)$ have to be considered since only the intra-orbital interactions are relevant. Then $\{n^\sigma\}$ reads:

$$\{n^\sigma\} = \begin{pmatrix} \{n^\sigma\}_s & 0 & 0 & 0 \\ 0 & \{n^\sigma\}_p & 0 & 0 \\ 0 & 0 & \{n^\sigma\}_d & 0 \\ 0 & 0 & 0 & \{n^\sigma\}_f. \end{pmatrix} \tag{5.49}$$

Here, the size of the submatrices depend on the orbital character l and can be universally expressed as

$$\{n^\sigma\}_l = \begin{pmatrix} n^\sigma_{1,1} & \cdots & n^\sigma_{1,2l+1} \\ \vdots & \ddots & \vdots \\ n^\sigma_{2l+1,1} & \cdots & n^\sigma_{2l+1,2l+1} \end{pmatrix} \tag{5.50}$$

and all coefficients $n^\sigma_{mm'}$ can be formally calculated from the spatially integrated and orbitally decomposed Green function:

$$n^\sigma_{m_l m'_l} = -\frac{1}{\pi} \int_{E_B}^\infty dE \, \mathrm{Im} \left[G^\sigma_{nm_l, nm'_l}(E) \right], \tag{5.51}$$

where the indices of $G^\sigma_{nm_l, nm'_l}$ resemble that the density matrix is a pure on-site and intra-orbital quantity.

Using this representation the energy contribution of the Coulomb repulsion reads

$$
\begin{aligned}
E^U[\{n^\sigma\}] = \tfrac{1}{2} \sum_l \sum_{\{m_l\},\sigma} \Big[&\langle m_l, m_l'' | V_{ee} | m_l', m_l''' \rangle \, n^\sigma_{m_l, m_l'} \, n^{-\sigma}_{m_l'', m_l'''} \\
&- \big(\langle m_l, m_l'' | V_{ee} | m_l', m_l''' \rangle - \langle m_l, m_l'' | V_{ee} | m_l''', m_l' \rangle \big) \, n^\sigma_{m_l, m_l'} \, n^\sigma_{m_l'', m_l'''} \Big],
\end{aligned}
\tag{5.52}
$$

where the sum over $\{m_l\}$ extends from 1 to $2l + 1$ and implies the summation over all four dimensions m_l, m_L', m_L'' and m_l'''. The matrix elements $\langle m, m'' | V_{ee} | m', m''' \rangle$ can be calculated on the basis of the effective Slater integrals F^k (see references in [37])

$$
\langle m, m'' | V_{ee} | m', m''' \rangle = \sum_k a_k(m_l, m_l', m_l'', m_l''') \, F^k,
\tag{5.53}
$$

where the summation index k is restricted to $0 \le k \le 2l$ and the coefficients a_k are defined by

$$
a_k(m_l, m_l', m_l'', m_l''') = \frac{4\pi}{2k + 1} \sum_{q=-k}^{k} \langle lm | Y_{kq} | lm' \rangle \, \langle lm'' | Y_{kq}^\star | lm''' \rangle.
\tag{5.54}
$$

This means all matrix elements can be derived analytically using the Gaunt coefficients which incorporate the complex spherical harmonics Y_{kq}.

For the evaluation of the double-counting correction to the energy functional in (5.48) we restrict the algorithm to the atomic limit scheme proposed by Czyżyk and Sawatzky [108]. Then, the occupation of the density matrix $n^\sigma_{mm'}$ for a given orbital l is assumed to be diagonal ($m = m'$) and equal to the average occupation

$$
\overline{n^\sigma} = \frac{1}{2l + 1} \sum_m^{2l+1} n^\sigma_{mm}.
\tag{5.55}
$$

The above step leads to the simplification that the full density matrix of an orbital l can be described by two scalars: the screened Coulomb parameter U_l and the exchange parameter J_l. Using the definition e.g. in [109] the double-counting correction to E_U simplifies to

$$
E_{dc}[\{n^\sigma\}] = \frac{1}{2} \sum_l U_l N_l (N_l - 1) - \frac{1}{2} \sum_l J_l \sum_\sigma N_l^\sigma (N_l^\sigma - 1),
\tag{5.56}
$$

where orbital-dependent densities are defined as $N_l^\sigma = \mathrm{Tr}(\{n^\sigma\}_l)$ and $N_l = N_l^\uparrow + N_l^\downarrow$. The previously introduced parameter U_l and J_l are constructed to be adjustable and the strength of Coulomb repulsion can be directly regulated by varying them.

Besides the correction to the energy functional the changes to the potential have to be addressed. Using the LDA+U scheme an additional part has to be considered in the Hamiltonian:

$$
\hat{H} = \hat{H}^{LDA} + \sum_l \sum_{m,m'} |\mu, k, l, m_l, \sigma\rangle V^\sigma_{m_l, m_l'} \langle \mu, k, l, m_l', \sigma|,
\tag{5.57}
$$

where $\hat{H}^{LDA}$ is representing the overall potential obtained in the usual scheme which includes

the LDA exchange correlation potential. The LDA+U potential is defined by

$$
\begin{aligned}
V_{m_l,m_l'}^{\sigma} = \sum_{\{m_l\},\sigma} & \Big[\langle m_l, m_l'' | V_{ee} | m_l', m_l''' \rangle \, n_{m_l'',m_l'''}^{-\sigma} \\
& - \big(\langle m_l, m_l'' | V_{ee} | m_l', m_l''' \rangle - \langle m_l, m_l'' | V_{ee} | m_l''', m_l' \rangle \big) \, n_{m_l'',m_l'''}^{\sigma} \Big] \\
& - U_l \big(N_l - \tfrac{1}{2} \big) + J_l \big(N_l^{\sigma} - \tfrac{1}{2} \big)
\end{aligned}
\tag{5.58}
$$

with the identical notation as declared above. In KKR`nano` this scheme is embedded in the self-consistency procedure in the following manner.

1. The externally defined parameter U_l and J_l for $l = 0 \ldots l_{max}$ and the LDA+U-contribution to the potential V_s^U serve as input, where s labels the self-consistency step. It is important to note, that U_l and J_l can not only be adjusted for every atom type in the supercell but also for each site individually.

2. A radial dependent reference basis Φ_l is constructed for the existing spherical part of the potential. If LDA+U has not been added in the previous step $s-1$, the potential without LDA+U contribution is considered as starting point.

3. The coefficients $\langle m, m'' | V_{ee} | m', m''' \rangle$ are obtained according to (5.53) and subsequent equations and normalized by $\langle \Phi_l | \Phi_l \rangle$.

4. The regular and irregular solution to the Schrödinger equation are obtained including V_l^U which is defined in (5.58) in the following manner:

   a) V_l^U is cut smoothly close to the core r_C and at the muffin-tin radius R_{MT} to guarantee a continuos potential by the function

   $$
   \Gamma(r) = \Big(1 + \exp\big(\gamma(r - r_{\mathrm{MT}})\big) \Big)^{-1} \cdot \Big(1 + \exp\big(\gamma(r_C - r)\big) \Big)^{-1}
   \tag{5.59}
   $$

   b) As the diagonal elements of the LDA+U interaction potential V_{m_l,M_l}^U constitute the largest perturbation to V^{LDA} and differ only by less than five percent, the average diagonal element

   $$
   \bar{V}_l^{\sigma} = \frac{1}{2l+1} \sum_{m_l} V_{m_l,m_l}^{\sigma}
   \tag{5.60}
   $$

   is added to the spherical part of the potential

   $$
   V_{sph}^{\sigma'}(r) = V_{sph}^{\sigma}(r) + \bar{V}_l^{\sigma} \cdot \Gamma(r).
   \tag{5.61}
   $$

   This modified potential enters the Schrdinger equation for the radial wave-function $R_l(r,E)$:

   $$
   \left[-\frac{1}{r} \frac{\partial^2}{\partial r^2} r + \frac{l(l+1)}{r^2} + V_{sph}^{\sigma'}(r) - E \right] R_l(r,E) = 0
   \tag{5.62}
   $$

c) The LDA+U contribution beyond $\bar{V}_l^\sigma$ is added to the spherical solution similar as the non-spherical potential perturbatively. Both left-over terms are considered in the Lippmann-Schwinger equation

$$
\begin{aligned}
R_{LL'}(r, E) &= R_l(r, E) \\
&+ \int^S G_l^{sph}(r, r', E) \sum_{L''} \Delta V_{LL''} R_{L''L'}(r', E) r'^2 dr',
\end{aligned}
\tag{5.63}
$$

which is being iterated in a Born-series of usually two to four steps. Here, the single-site spherical Green-function is defined by the regular and irregular solution of the spherical Schrödinger equation. The non-spherical contribution to the potential $V_{LL'}^{\sigma,ns}$ and the remaining part of V_l^σ result in the perturbative potential difference, which then reads for $L \geq l^2 + 1$ and $L \leq (l+1)^2$ and subsequently $m_l = L - l^2 + 1$ as:

$$
\Delta V_{LL'} = V_{LL'}^{ns}(r) + \left[V_{m_l,m_l'}^{U,\sigma} - \delta_{m_l,m_l'} \bar{V}_l^\sigma \right] \cdot \Gamma(r).
\tag{5.64}
$$

5. The density matrix $\{n^\sigma\}$ is calculated from the Green function using (5.51). To obtain the integrand in (5.51) the radial integration has been performed over the Green function and the single-site solution which are here projected on the basis Φ_l.

6. The new potential V_{i+1}^U is obtained by equation (5.58). In order to ensure convergency of this algorithm simple mixing is applied to obtain:

$$
V_{i+1}^U = \alpha V_s^U + (1 - \alpha) V_{s+1}^U.
\tag{5.65}
$$

Here α_s is adjustable and depends in the practical application on the self-consistency step s.

7. The additional contribution to the energy functional (5.48) is calculated.

Elaborate tests revealed that if using a continuous simple mixing of the LDA+U contribution to the potential it is likely that numerical instabilities occur which can lead to divergency of the self-consistency cycle. Therefore we modified the mixing scheme: In the first part the electronic structure is converged down to a RMS-error[4] well below 10^{-2} using simple mixing for the usual LDA potential. Here, the LDA+U contribution is fully updated ($\alpha = 1$) only every 10th step, while in intermediate steps no mixing is applied ($\alpha = 0$). In the second part the more insensitive LDA+U contribution can be accounted as reasonable well converged and is frozen ($\alpha = 0$), while the LDA potential is mixed by faster schemes such as Anderson or Broyden mixing. The convergency of this procedure is shown in Fig. 5.1. On the downside this scheme requires usually up to a factor of two more self-consistency steps than the procedure without LDA+U where faster mixing schemes can be applied earlier. However, it enables reliable convergency of LDA+U calculations.

---

[4]See Fig. 5.1 for the definition of the RMS-error.
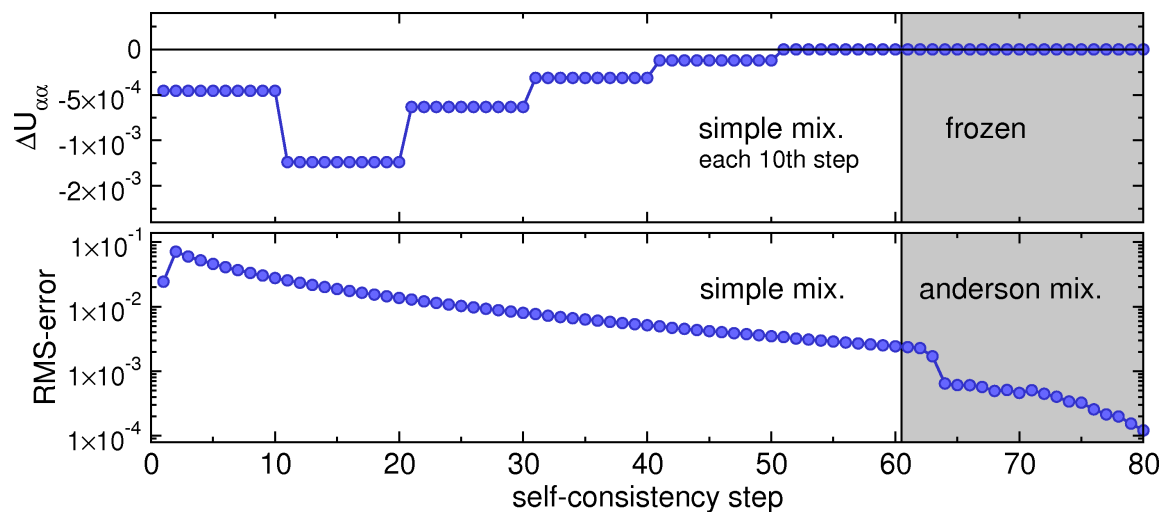
Figure 5.1.: *Adaptive mixing scheme in KKR**nano** for LDA+U calculations. In the upper panel the deviation Δn of one exemplifying diagonal density matrix component corresponding to Gd f states in GaN from its converged value is shown. For the same calculation the lower panel depicts the RMS-error which is the variance of previous and actual potential $\propto \|V^{i-1}(r) - V^i(r)\|$. Detailed explanation of the used mixing scheme is given in the text.*

5.3 Lichtenstein formula

Aiming at the description of magnetic order of disordered system or defects it is inevitable to have a scheme at hand which allows to extract physically transparent parameters from the *ab initio* calculations. With respect to magnetic ordering the exchange coupling parameter J_{ij} describing the interaction between two spins at sites i and j are those desired parameter. Given the fact that spin-orbit effects are small in the systems we are going to analyze in this thesis the J_{ij} alone provide a model description in terms of the Heisenberg Hamiltonian

$$H = -\sum_{ij} J_{ij} \, \mathbf{e}_i \cdot \mathbf{e}_j, \tag{5.66}$$

where $\mathbf{e}_i$ and $\mathbf{e}_j$ are the unit vectors of the spins on sites i and j.

In DFT algorithms which are relying on wave-function representation, the J_{ij}'s can only be extracted by relating the total energies of different magnetic states to each other, which comes along with significant drawbacks: First, the computational effort is doubled if e.g an antiferromagnetic and ferromagnetic configuration has to be compared. Secondly, already if only a few magnetic sites are involved it is difficult to extract the pair-wise interactions J_{ij}. For hundreds or thousands of magnetic atoms in a supercell it becomes impossible on this basis.

In Green function based methods J_{ij} can be obtained more elegantly 'on the fly'. Generally speaking, the magnetic force theorem can be exploited to probe the influence of small angle rotations on the energy. Without derivation, which can be found in the original report by

Lichtenstein *et al.* [38], the Lichtenstein formula reads

$$J_{ij} = \int_{E_B}^{E_F} dE \; j_{ij}(E) \tag{5.67}$$

with

$$j_{ij}(E) = \frac{1}{4\pi} \text{Im Tr} \left[\Delta^i(E) \, G_\uparrow^{ij}(E) \, \Delta^j(E) \, G_\downarrow^{ji}(E) \right]. \tag{5.68}$$

Here, the differences of spin up and down channel of the t matrices on site i is denoted by $\Delta^i = t_\uparrow^i - t_\downarrow^i$. The off-diagonal elements of the Green function are used as obtained in (4.10). It is important to note that although in the standard procedure in KKRnano only the diagonal parts of the Green function are required, the off-diagonal contributions are known since they are by-products of the solution of the Dyson equation. Therefore no significant numerical effort has to be made to evaluate (5.67). However, additional MPI communication is required by distributing the t-matrices and off-diagonal elements of the Green function, where the overall amount of transfered data depends on the range of probed exchange interactions. This introduces a considerable but not decisive slow-down of about 10-30% of the algorithm: Regarding that the calculation of the J_{ij}'s is only executed on the last self-consistency step for fully converged potentials, it becomes clear that the minor slow-down is not a limiting factor.

For the analysis of magnetic materials the computation of the Lichtenstein formula reveals an interesting side effect: As implied in (5.67) not only the final integrated value J_{ij} can be evaluated but also its energy dependency $j_{ij}(E)$. It will turn out in our analysis of magnetic defects in semiconductors in chapter 6 that this additional information can provide valuable insight about the character of the exchange interaction.

5.4 Summary

After accomplishing the three previously described fundamental extensions to KKRnano −
Lloyd's formula, LDA+U and Lichtenstein formula − we can conclude that our objective
to develop a broadly applicable electronic structure code is achieved. We engineered and
implemented all of those schemes under the premise of high efficiency for both large scale
applications and massively parallel computation. By that we were able to minimize the
additional computational effort and maintain the high parallel performance of KKRnano.

It is noteworthy to highlight the importance of those extensions for this thesis: Both of
the following applications, GaN:Gd as well as the phase change material GeSbTe, crucially
depend on the accurate description of the charge with Lloyd's formula. The LDA+U correc-
tion will be of utmost significance to describe the electronic structure of gadolinium atoms
correctly in chapter 6. Finally, the Lichtenstein formula facilitates to disclose the underly-
ing magnetic coupling mechanism of defects in GaN:Gd, thereby lifting the discussion to a
higher level.

Magnetic defects in GaN:Gd

Over the last two decades great effort has been made both experimentally and theoretically to search for dilute magnetic semiconductors (DMS) with a Curie temperature above room temperature. This search is triggered by the enormous technological advantages which can potentially be made connecting the properties of semiconductors with spintronics in DMS [39, 40]. The vast majority of materials which have been considered as possible candidates are III-V as well as II-VI semiconductors with magnetic transition metal dopants such as manganese [41, 42, 44, 46]. Although several groups reported on ferromagnetism in DMS above room temperature (see e.g. references [43, 45]), it turned out that the reproducibility of those samples remains a challenge. Even smallest variations in the preparation technique induced dramatic differences in the magnetic properties, often resulting in considerably lower Curie temperatures [46, 110, 111] which impeded any technological use.

Instead of doping with $3d$ transition metals a different strategy of lanthanide doping and in particular the gadolinium doping of gallium-nitride attracted attention. In pioneering work by Teraguchi *et al.* [112] in 2002 and Dhar *et al.* [1] in 2005 ferromagnetic ordering above room temperature in GaN:Gd was detected (see Fig. 6.1(a)). While Teraguchi *et al.* [112] restricted the analysis to relatively high Gd concentrations of 4%, Dhar *et al.* [1] observed ferromagnetic signatures in samples with doping ratios of one Gd on more than 10.000 Ga atoms. Interestingly the saturation magnetization is only weakly decreasing upon reducing the Gd concentration (Fig. 6.1(b)). The direct consequence is that colossal magnetic moments of up to 4000 μ_B per Gd atom have been measured in the dilute limit of Gd doping.

Starting with this discovery the tracing of the origin of those colossal magnetic moments as well as of the ferromagnetic coupling is challenging for the scientific community. The first proposed model was considering an induced spin-polarized cloud on the GaN matrix around each of the incorporated Gd atoms [1]. This concept was quickly disproved by the finding that the magnetic properties are rather of extrinsic type than triggered by Gd itself: Khaderbad *et al.* [47] observed two important dependencies on the preparation technique. First, samples prepared by Gd implantation reveal stronger magnetic signatures than the ones grown by molecular beam epitaxy. Secondly, annealing shows a clear reduction of the
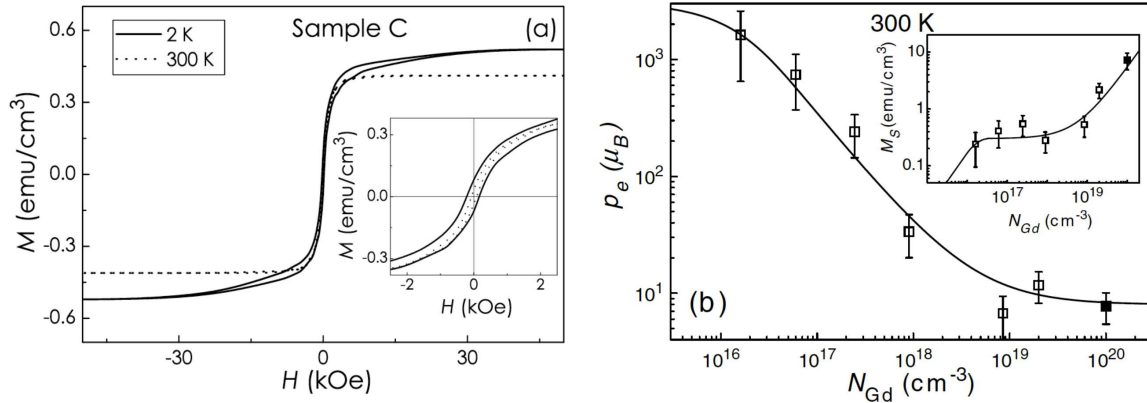
Figure 6.1.: *Fig.(a) and (b) from Dhar* et al. *[1]. In (a) the magnetization loops as a function of magnetic field (B) measured at 2 K and 300 K are shown. The inset is a magnification of the regime of small B to highlight the hysteresis for both temperatures. In (b) the magnetic moment per Gd (p_e) is displayed as a function of Gd concentration for 300 K. The inset specifies the saturation magnetization M_S for the same set of samples. Details are provided in the original publication [1].*

saturation magnetization. Both results lead to the conclusion that extrinsic defects such as interstitials or vacancies likely play an important role. In accordance with this, x-ray magnetic circular dichroism measurements of the Gd L_3 edge underlined that only a small fraction of the overall magnetization is originated by the Gd sites [48, 113].

Dalpian and Wei [49] suggested a different mechanism based on a strong coupling of gadolinium f states and unoccupied gallium s states which then lifts the spin-degeneracy in the conduction band. By adding donors to GaN:Gd the majority states of the spin-split band become occupied and lead to an additional magnetic moment. Due to the fact that the splitting in the conduction band decreases linearly by reducing the Gd concentration [114], this model case cannot provide an explanation of the behavior in the dilute limit.

While the models above have been disproved, two interesting concepts, which are based on extrinsic defects and which have the potential to explain the magnetic properties of GaN:Gd, are discussed controversially. On the one hand, Mitra and Lambrecht [50] suggested the electron donation by including oxygen or nitrogen on interstitial lattice sites. Indeed, their calculations revealed a significant spin-splitting of the defect states in the band-gap and moreover a short-range ferromagnetic coupling to Gd. On the other hand, *ab initio* calculations showed that hole doping by Ga-vacancies (V_{Ga}), which support a magnetic moment of 3 μ_B each, leads to ferromagnetic ordering [51, 52].

We performed first-principles calculations considering both kinds of defects in order to contribute to the solution of the GaN:Gd puzzle. On this path KKR**nano** unifies important advantages: The treatment of Gd and extrinsic defects in large supercells allows for a proper sampling of the configurational space spanned by those defects. Additionally, the magnetic exchange interactions can be identified for all pairs of atoms individually by using Lichtenstein's formula, which allows for an identification of the crucial mechanisms.

In this section, first the computational treatment of the host material GaN:Gd will be described and its electronic structure discussed. The following analysis of the magnetic

coupling will then be individually conducted for oxygen- and nitrogen-interstitials as well as for gallium-vacancies.

6.1 Computational treatment of gadolinium doped gallium nitride

For the computational study of the incorporation of gadolinium defects in a gallium-nitride (GaN) lattice first structural issues have to be considered. In the key experiments on Gd doping of GaN by Dhar *et al.* [1] and Teraguchi *et al.* [112] layers of a few hundred nanometer thickness are grown on silicon carbide substrates[1] by molecular beam epitaxy. As shown in a follow-up work of Hite *et al.* [115] the magnetic properties of GaN:Gd do not depend on the film thickness and are therefore not an interfacial or surface effect but driven by the volume. Therefore, we performed in line with existing theoretical work on GaN:Gd [50, 51, 116] calculations in three-dimensional periodically repeated supercells. Experimentally, thick layers of GaN can be grown both in wurzite and zincblende structure revealing in both cases matchable magnetic properties [117]. This finding can be directly related to the close similarities between the two lattice types: Both of them can be constructed out of tetrahedrally bound building blocks and differ for each atom type only on the third-nearest neighbor sites. In addition also the properties of native defects − such as formation energies − have been reported by Neugebauer and Van de Walle [118] and [119] to depend only insignificantly on the realization in a zincblende or wurzite structure.

Considering these arguments we will treat GaN:Gd in the cubic zincblende lattice with a lattice constant of 4.46 Å. In order to create potentials and shape-functions in the full-potential treatment within the KKR methodology which are well converged with respect to the angular-momentum cut-off, it is important to limit the deformation of the voronoi cells. Therefore, for the calculations with KKR**nano** we include additional empty sites on the high symmetry tetrahedral positions of the GaN zincblende lattice − a procedure previously used e.g. in [120]. This setup results in a space-filling of voronoi sites which is equivalent to a bcc lattice and hence guarantees a proper description within multiple-scattering theory.

Up to this point all structural issues addressed a regular GaN lattice. Now, the additional complexity and difficulties introduced by Gd doping will be discussed. Experimental studies revealed that doping of GaN with rare-earth elements in general [121] and specifically Gd [122] takes place preferentially substitutional on the Ga site with a formation energy of roughly 5eV [123]. For Gd on Ga sites only small relaxations have been obtained from theoretical investigations [52, 124] which can be explained by the fact that Ga and Gd are isovalent with one $4p$ and one $5d$ electron, respectively and therefore bind with almost comparable strength. Since all of our calculations aim to mimic the dilute limit of Gd doping, geometries which have the potential to create larger relaxations, arising e.g. from anyhow energetically costly Gd clusters [124], are unlikely to occur. Both facts permit to place the substitutional Gd dopants on the ideal Ga sites, which we do from here on.

Different exchange-correlation functionals have been exploited in great detail to properly describe optical transitions in GaN by means of density-functional theory, e.g. by Rubio *et al.* [125]. The authors contrasted the experimentally obtained band-gap in GaN of about

---

[1]Precisely on the (0001) Si surface of 6HSiC substrates.
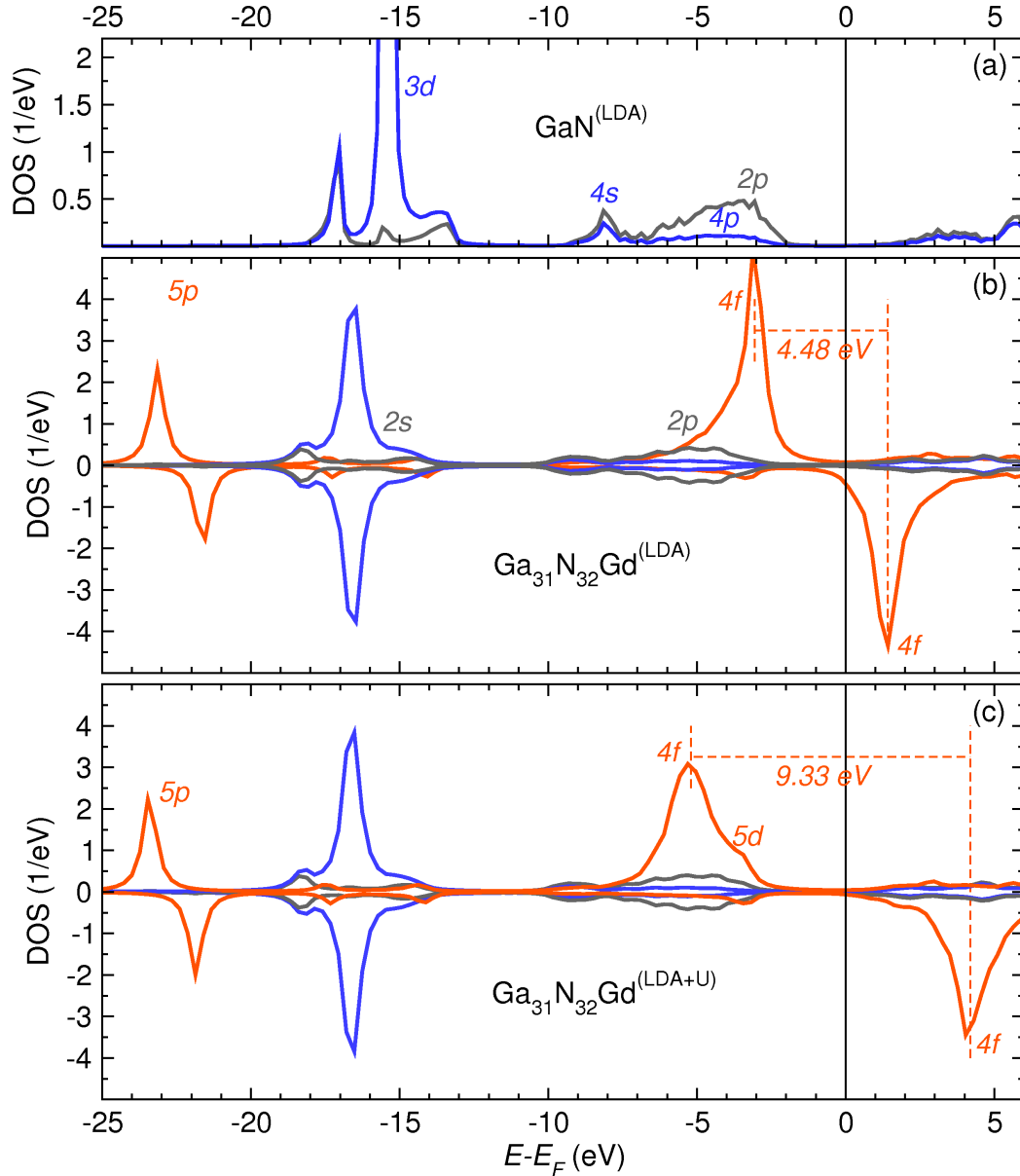
Figure 6.2.: *Atom resolved density of states (DOS) for GaN (a), $Ga_{31}N_{32}Gd_1$ using the LDA exchange-correlation functional (b) and additional Coulomb-repulsion on the Gd f and d states (for details see explanation in text). The minority and majority spin channels are represented by negative and positive DOS values, respectively. For the sake of clarity the local DOS on interstitial sites is not shown and the average DOS is illustrated. In addition the position of characteristic orbitals is labeled for Ga, N and Gd in (a), (b) and (c) respectively. The characteristic splitting of the Ga 3d states is originated by the hybridization to N 2s states. Note, that the actual contour extends down to even lower energies than shown, incorporating the Gd 5s states.*

3 eV with calculations using the local density approximation (LDA) which resulted in a band-gap of 2 eV (see also Fig. 6.2(a) where the atom resolved density of states as obtained for GaN by KKRnano is shown.). This shortcoming of LDA can be partly improved by the application of GW schemes [125], which are however still limited to structures of the order of ten inequivalent atoms due to its inherent non-locality and frequency dependency and the resulting extensive computational demands (see e.g. [101] and references therein). Since we are not primarily interested in optical properties of GaN, but on the influence of defect levels either in the gap or at the valence band maximum, the usage of LDA presents a reasonable approximation.

However, for an accurate treatment of the strongly localized Gd f states we apply an additional orbital dependent Coulomb repulsion U using the LDA+U scheme as introduced in chapter 5.2. Following previous electronic structure calculations of GaN:Gd [126–128] we use $U_f = 8.0$ eV and $J_f = 1.2$ eV for the f states and an additional repulsion of $U_d = 3.4$ eV with $J_d = 0.0$ eV on the Gd $5d$ states. The impact of the Coulomb repulsion on the Gd $4f$ states is clearly visible in Fig. 6.2, increasing the splitting of majority and minority f states from 4.5 eV to 9.3 eV. Thereby, the position of the f-minority states relative to the valence band maximum is shifted to approximately 6 eV − in good agreement with existing GW studies [123]. The application of U_d can be regarded as a second order effect, which leads to a slightly increased splitting of the Gd $5d$ states and thereby prohibits a further reduction of the band-gap. In particular, the significant increase in splitting of the f states prevents unphysical hybridization and magnetic coupling at the band edges which is crucial for the following analysis.

For all three types of defect calculations investigated in this chapter the following sets of computational parameter and convergency procedures have been applied: All supercells consist of 1024 voronoi sites − this means $4 \times 4 \times 4$ zincblende subblocks of 16 sites. The calculations are performed using a single k point at the Γ point, with reference clusters containing 27 sites and in full-potential representation with an angular momentum cut-off of $l_{\max} = 3$, which is essential to incorporate the f states of Gd. For the energy contour integration we used overall 44 energy points, 7 Matsubara poles and an electronic temperature of 700 K. Within this contour the Ga $4s$, $4p$ and $3d$ states, the N $2s$ and $2p$ states, as well as the Gd $5s$, $6s$, $5p$, $5d$ and $4f$ states are included as valence states. A first overview on part of those valence states is given in Fig. 6.2 while more detail on the bands of high importance for the material properties will be given in the following sections. All remaining lower lying states are treated as core states − namely [Ar], [He] and [Kr]$4d$ for Ga, N and Gd, respectively. Calculations involving oxygen interstitials in section 6.3 are performed with the same classification of oxygen states as used for N. It is important to note that the self-consistency calculations within the LDA+U scheme require a careful setting of the mixing procedure. Here, we exploit first the two-level mixing scheme as introduced in chapter 5.2 down to a relative potential changes of less than 1%. As a second step the additional Coulomb repulsion contribution to the potential is fixed and the self-consistency iterations are continued with Broyden or Anderson mixing schemes [129, 130]. For the acceleration of the self-consistency cycle we utilized initial guess from the previous self-consistency step as introduced in section 4.2.2. The block-circulant preconditioning schemes (see section 4.2.3) have not been applied since the development of this approach and the project on GaN:Gd overlapped in time.

6.2 Nitrogen-Interstitials in gadolinium doped gallium nitride

The fundamental quantity which has to be considered in order to rate the occurrence of a given defect is its formation energy. Even in highly non-equilibrium conditions such as the experimental growth of GaN:Gd layers by molecular beam epitaxy (MBE) or ion bombardement the formation energy remains an important indicator whether defects are likely to form or not. Detailed reports on formation energies of native defects in GaN by Neugebauer and Van de Walle [118] and Limpijumnong and Van de Walle [131] reveal that, amongst all other native defects in their neutral charge state, nitrogen interstitials (N_I) require relatively small formation energies of $F_{N_I} \approx 3.0$ eV.

Driven by this motivation Mitra and Lambrecht [126] performed calculations on the magnetic properties of N_I in GaN:Gd with two important results: First, N_I's display a sizable spin-polarization and second, a sequence of Gd and N_I spin configurations shows a tendency towards ferromagnetic coupling. However, the description of magnetic coupling in those calculations is intrinsically limited: On the one hand, the calculations have been performed in unit cells consisting of 64 atoms, which practically results in a limitation to only one or two inequivalent defects in the sample. In particular when codoping of Gd and N_I is considered the phase space of defect configurations can hardly be modeled in those cells. Additionally finite-size effects such as spurious interactions to periodic images are difficult to control. On the other hand, the ferromagnetic tendency was found by means of total energy arguments, which makes the search for the origin of this coupling notoriously difficult: In the presence of multiple magnetic defects in the unit cell it often becomes impossible to trace back the obtained total energy differences of two spin states into element- or site-specific contributions to the magnetic coupling.
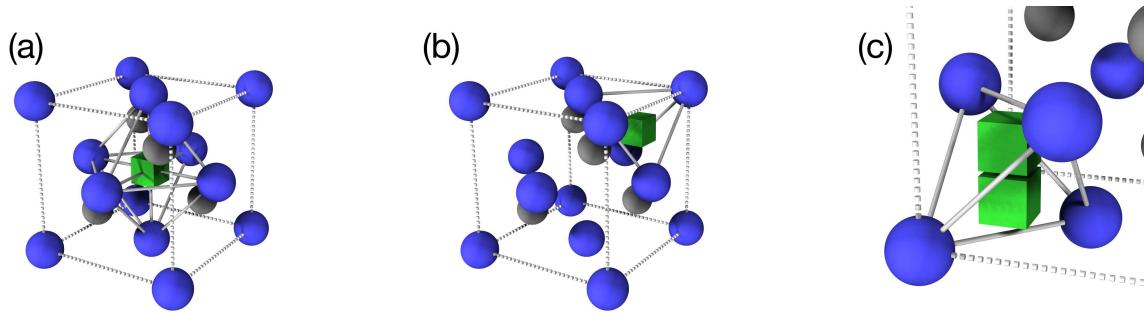
Figure 6.3.: *Schematic view on a nitrogen interstitial (green) on the octahedral site (a), the tetrahedral site (b) of a Gallium (blue) Nitride (gray) zincblende lattice. In (c) the local structure of a nitrogen split interstitial (green) in the same lattice is shown.*

To overcome these limitations and to render possible a microscopic understanding of the magnetic coupling in those samples we performed calculations of a $Ga_{248}N_{256}Gd_8N_{I,32}$ supercell (see Fig. 6.4) using KKR**nano**. Computational as well as structural details of GaN and Gd defects are given in section 6.1. It remains to find the local structural configuration of a N_I. The local structure of the energetically most favorable configuration was found to be not the tetrahedral or octahedral but the so-called split-interstitial [131, 132]. All different

interstitial configurations are schematically drawn in Fig. 6.3. While the octahedral and tetrahedral sites are placed on the high symmetry points in the lattice the structure of the split-interstitial configuration requires more explanation. Here, the additional N atom is closely bound to a regular N atom of the GaN matrix. Both atoms remain enclosed by the tetrahedrally ordered Ga atoms and lead to considerable outward directed relaxations of those Ga atoms [2].
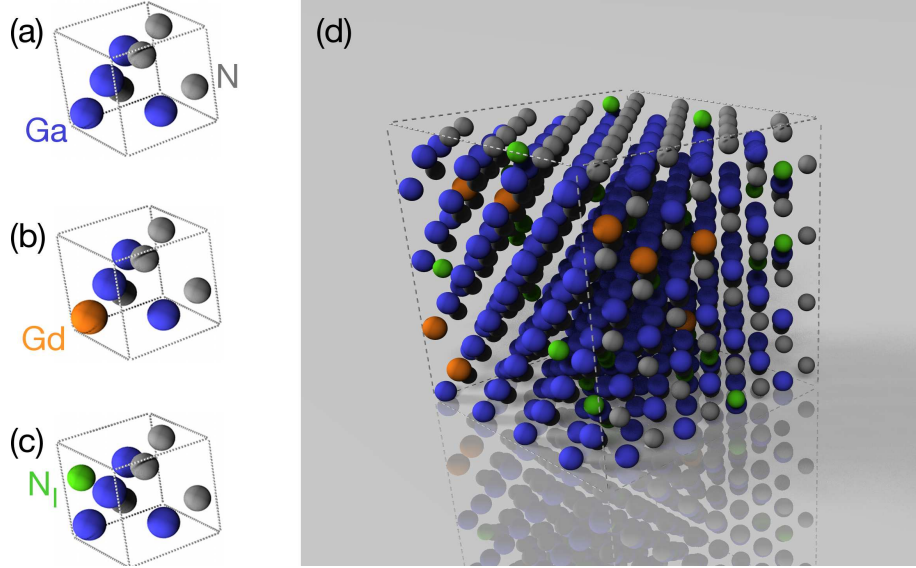
Figure 6.4.: *Illustration of the fundamental building blocks of the investigated structure: GaN in zincblende structure without defects (a), with Gd on the Ga site (b) and a nitrogen interstitial (N_I) on a randomly selected interstitial site. In (d) the actual calculated structure $Ga_{248}N_{256}Gd_8N_{I,32}$ is shown. For all panels Ga, N, empty interstitial, Gd and N_I are labeled in blue, gray, transparent, orange and green, respectively.*

While the ground-state split-interstitial configuration is known, as soon as those defects are not any more well isolated but neighboring each other, a cumbersome calculation of relaxations is required. This scenario is likely appearing in cells of high defect concentrations which are crucial to explain colossal magnetic moments. In order to avoid computationally extremely demanding relaxations we restrict our calculations to interstitials on the tetrahedral and octahedral sites. Making this choice also we maintain the possibility for a one-to-one comparison to the previous work by Mitra and Lambrecht [126] on N_I.

At this point it is important to check wether this simplification affects the properties of the system crucially. The quantity which most sensitively depends on different local geometries of this defect is the magnetic moment. Following the arguments of the Stoner criterion larger hybridization to neighboring atoms can lead to a significantly reduced magnetization. In order to verify that the more densely packed split-interstitial configuration with more hybridized N states still exhibits a magnetic moment of comparable size Cheiwchanchamnangij [133] performed first-principle calculation of a single split-interstitial using the FP-LMTO method [134]. Those calculations underlined that the size of the magnetic moment on the

---

[2]For details on this local geometry the reader is referred to Limpijumnong and Van de Walle [131]

split-interstitial is in good agreement with the magnetic moments of N_I on tetrahedral or octahedral sites [133].

In order to probe the outstanding importance of the extrinsic defects we added four N_I in the previously discussed structure on tetrahedral and octahedral positions for each Gd atom in the sample at a Gd concentration of about 3%.
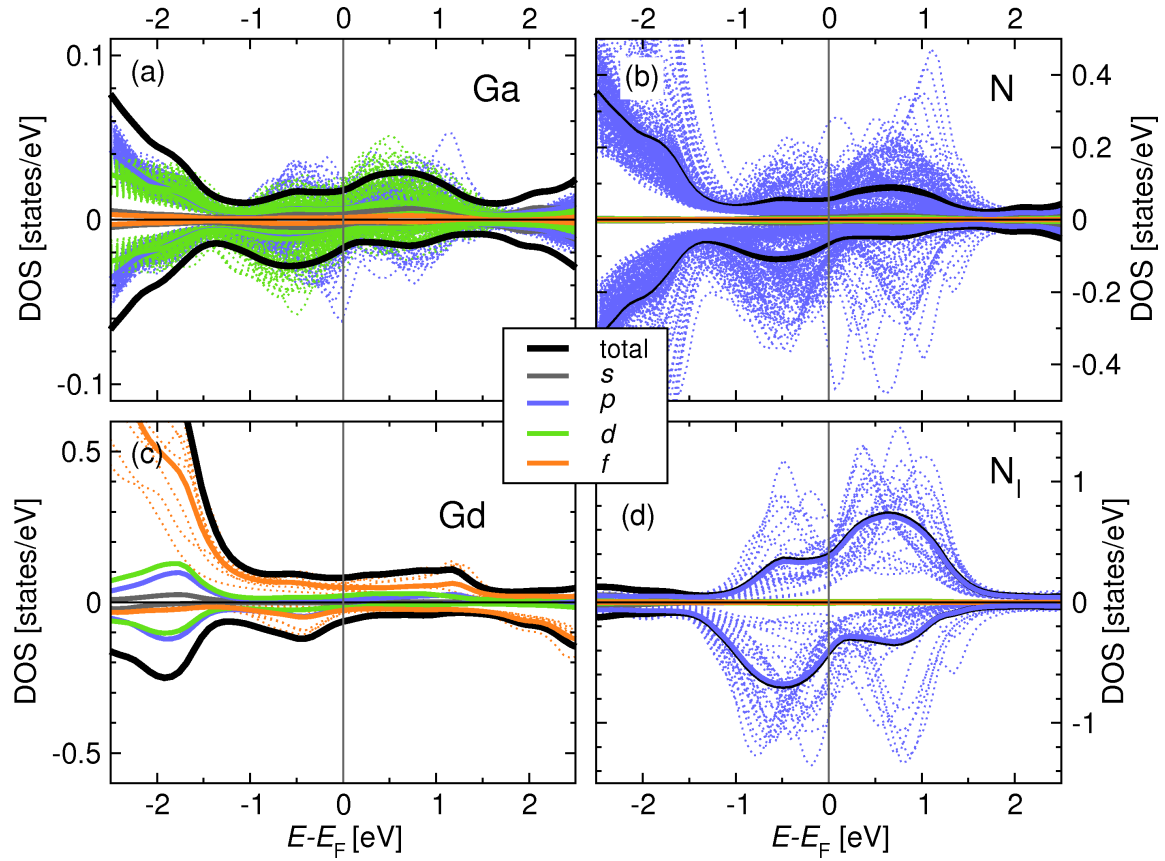
6.2.1 Electronic structure
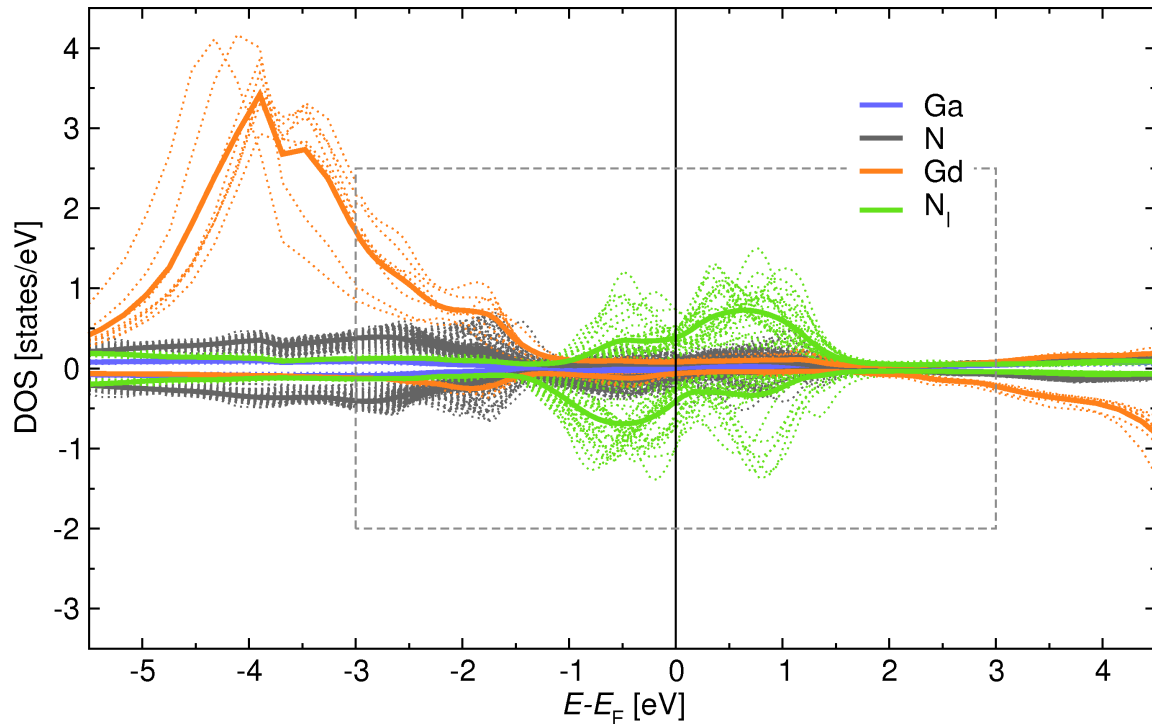
Figure 6.5.: *Local density of states (DOS) of $Ga_{248}N_{256}Gd_8N_{I,32}$ calculated with a temperature broadening of 600 K. The minority and majority spin channels are represented by negative and positive DOS values, respectively. Both mean (solid lines) and values on all sites individually (dotted lines) are displayed in the color coding as specified in the inset. In the region marked with dashed lines the calculation has been performed with a higher energy resolution. Contributions of states localized on interstitial sites are small and not shown for the sake of clarity.*

Before analyzing the magnetic properties specifically we will focus on the electronic structure we obtain from a supercell calculation of $Ga_{248}N_{256}Gd_8N_{I,32}$. As noted before both types of defects, Gd and N_I are randomly placed on Ga and tetrahedral or octahedral interstitial sites, respectively. In Fig. 6.5 a first overview over the local density of states (LDOS) on all individual sites is given. The first fundamental observation is that the high temperature broadening (in this case 600 K) which we introduce to accelerate the iterative scheme[3]

---

[3]Since the block-circulant preconditioning has not been applied for this calculations the application of a relatively high temperature broadening is crucial to accelerate the computation.

leads to a significant smoothening of all involved bands. This broadening is in particular striking for the more localized states such as the Gd $4f$ states and N_I defect states, which without smearing would exhibit a spiky LDOS [126].

In order to discuss the spin-polarization element-specifically it is important to note an additional computational detail. As starting potential of the self-consistency cycle for Ga, N and N_I degenerate potentials for spin-up and spin-down channels have been chosen, while spin-split Gd potentials with a predefined spin-direction are used. On all Gd atoms this initial spin direction is conserved upon self-consistency resulting in a collective spin polarization as shown in Fig. 6.5. Besides the Gd sites a strong spin-splitting can be observed on the N_I sites on which the states are positioned in the bandgap. In addition large variations of the LDOS occur on the N_I sites. However, this variation is not equally distributed amongst the two configurations − parallel or antiparallel to Gd − but the majority of N_I's align antiferromagnetically to Gd.

To enable a more detailed discussion of the electronic structure the LDOS is redrawn in Fig. 6.6 resolving the partial LDOS in a smaller energy window. As expected the three additional electrons of the unfilled $2p$-shell of the N_I are the determining states in the gap. An important detail is that there is no significant net charge transfer to or from the N_I sites. From Fig. 6.5(d) and Fig. 6.6(d) we can observe a distinct hybridization of those gap defect states with predominantly N sites which induces spin-polarized p states on those N sites. Taking into account the different scale of Fig. 6.6(a) and Fig. 6.6(b) shows that this effect is still visible but clearly less pronounced on Ga sites, where a considerable part of those states is of p and d type. However, on both Ga and N sites the distribution of LDOS over all lattice sites is highly inhomogeneous, leaving part of the sites with no states in the gap. Keeping in mind that the concentration of N_I's in the supercell is relatively high this finding is a first indication that the influence of N_I's on their surrounding is of short-range nature. In order to complete this picture the effect of the presence of N_I's on the Gd sites has to be discussed. On the basis of Fig. 6.6(c) we can deduce that the hybridization of Gd and N_I's remains relatively small indicating a weak interaction. Which, however, leads to a preferential antiferromagnetic coupling of the N_I moments to the Gd spins − as can be seen by the LDOS of N_I and Gd.

As pointed out in section 6.2 both N_I's on octahedral and tetrahedral sites are taken into account in our supercell calculations, which facilitates to analyze their difference. Both types of defects feel four nearest neighboring atoms in tetrahedral geometry, which are N atoms for the octahedral and Ga atoms for the tetrahedral site (both shown in Fig. 6.7). In order to underline this difference we will abbreviate the tetrahedral and the octahedral interstitial configuration with I_N and I_{Ga}, respectively.

Before discussing the electronic structure of I_N and I_{Ga} it is important to note that the LDOS as shown in Fig. 6.7 is projected on a uniform spin direction. This means on the sites with spins pointing in which point not in direction of the total spin of all N_I sites, the majority and minority LDOS is interchanged for this analysis. This procedure allows to draw a meaningful average LDOS which reveals a clear shift in energy between I_N and I_{Ga} sites. The total charge being accumulated on the I_{Ga} site is by about 0.2 electrons larger than on I_N's which explains part of the shift. Further the shift in energy results in slightly larger magnetic moments on I_{Ga} with $1.04\mu_B$ comparing to I_N with 0.85 μ_B. These findings can be explained considering the different strength of hybridization to the

Figure 6.6.: *Local density of states (DOS) of $Ga_{248}N_{256}Gd_8N_{I,32}$ calculated with a temperature broadening of 600 K. The minority and majority spin channels are represented by negative and positive DOS values, respectively. For all Ga (a), N (b), Gd (c) and N_I sites the mean value of the orbital decomposition into s, p, d and f states is shown in relation to the mean total DOS (all in solid thick lines according to the color coding defined in the inset). In addition the individual partial DOS of p and d states for Ga (a), p states for N (b), f states for Gd (c) and p states for N_I (d) are displayed in dashed lines in the corresponding color coding.*
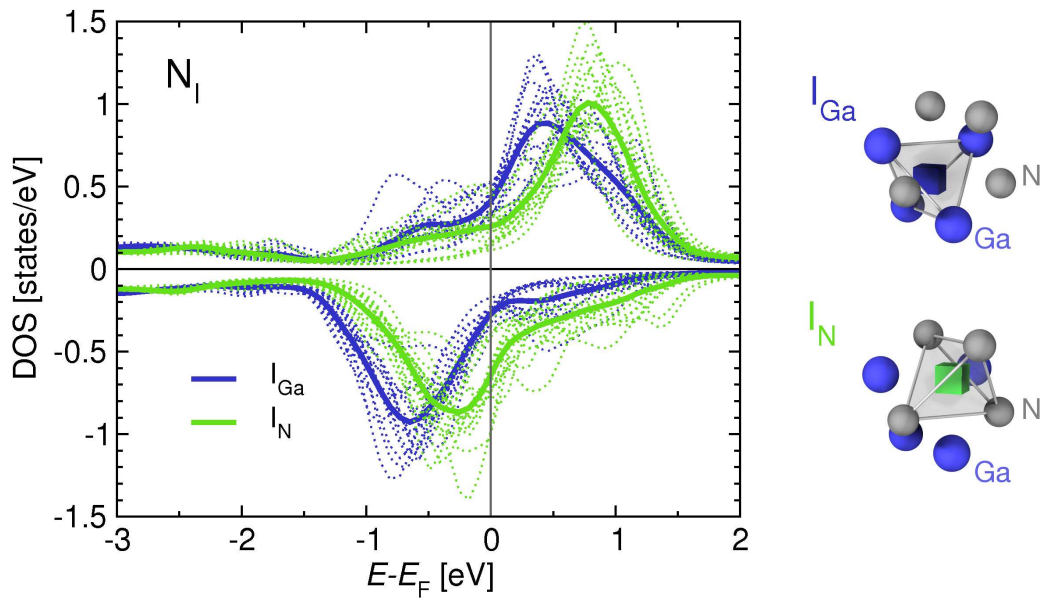
Figure 6.7.: *Local density of states (LDOS) of nitrogen (N_I) on the two inequivalent interstitial sites in the zinc-blende lattice. Here, individual and mean LDOS are shown in dashed and solid lines, respectively. The different sites are labeled according to its nearest neighboring atoms as I_N (tetrahedral) and I_{Ga} (octahedral), both shown in the right panel. Note, that the minority and majority spin channels are represented by negative and positive DOS values, respectively.*

Ga and N sites which has been discussed previously and illustrated in Fig. 6.6(a, b). I_N's experience a larger hybridization and by that lose in comparison to I_{Ga}'s a larger amount of charge preferably to the N sites. However, the hybridization of interstitials on I_{Ga} sites to the surrounding bulk Ga states introduces a more pronounced downward shift than the hybridization of I_N's to the N bulk states, which should also have an effect on the energetics of the defects. Before briefly discussing the extracted total energies it is important to note, that those values can only provide a first estimate since no relaxations have been taken into account. The local energy[4] on I_N sites turns out to be slightly more favorable − by 0.38 eV − as compared to I_{Ga}. This picture changes, once considering the adjacent sites being involved in hybridization: Ga sites around I_{Ga} gain on the order of 0.5 eV due the presence of the interstitial while for N sites around I_N no significant change in energy can be observed. Since all more distant shells do not play an important role, at least if relaxations are neglected, overall the I_{Ga} or octahedral position is the more favorable one.

6.2.2 Magnetic moments

The electronic structure on the different sites of the GaN:Gd lattice with N_I already provided first insight on the general trends of the magnetization of the sample. The spin polarization of N_I is found to be intrinsic and thereby also present in samples without Gd. If Gd is present as in our calculations, judging from the LDOS shown in Fig. 6.5 and Fig. 6.6 leads to the first insight that the predominant part of interstitials seem to couple antiferromagnetically to Gd. In this section we will focus on a more quantitative analysis of the magnetization which is shown for all elements in Fig. 6.8 and table 6.1.

Inherent from the pronounced splitting of the f states all Gd atoms carry a magnetic moment close to $7\mu_B$ (see Fig. 6.8(a)). Driven by the significant hybridization of N sites with N_I the induced magnetic moment on N atoms can reach up to $0.1\mu_B$. This effect is by one order of magnitude weaker for Ga atoms. Here, already the hybridization is significantly smaller since considerable less p charge of the valence band is located on Ga sites as compared to N sites (see density of states of GaN in Fig. 6.2(a)). In any case, the most volatile quantity is the magnetic moment on N_I sites reaching from almost -1.5μ_B to 1.0μ_B. These large variations predominantly occur in regions in the supercell with a local N_I concentration above average. As shown in Fig. 6.8(b) two groups can be defined: N_I with less than two other N_I's in its three neighboring shells of interstitial sites are all aligned ferromagnetically to the Gd spins. For N_I's with a higher local defect concentration the distribution of direction and absolute value of magnetic moment is broad. This clear trend indicates a strong coupling amongst N_I's which is at least not uniformly ferromagnetic.

Our view on the spin polarization of N_I and Gd in GaN allows to comment on previous explanations of the reported colossal magnetic moments. Dhar *et al.* [1] proposed a model, where Gd atoms induce a large spin-polarized cloud on the GaN matrix. Since the spin-polarization on Ga and N sites is in all cases small and not caused by the interaction to Gd but to N_I's, the explanation by Dhar *et al.* [1] can be excluded and it is likely that the effect is of extrinsic nature.

The results we have obtained up to now provide sufficient insight to deduce a rough

---

[4]In the KKR framework the local energy for each of the Voronoi sites is obtained which then by summing over all sites leads to the total energy of the system. For further reading on this issue we refer to [72].
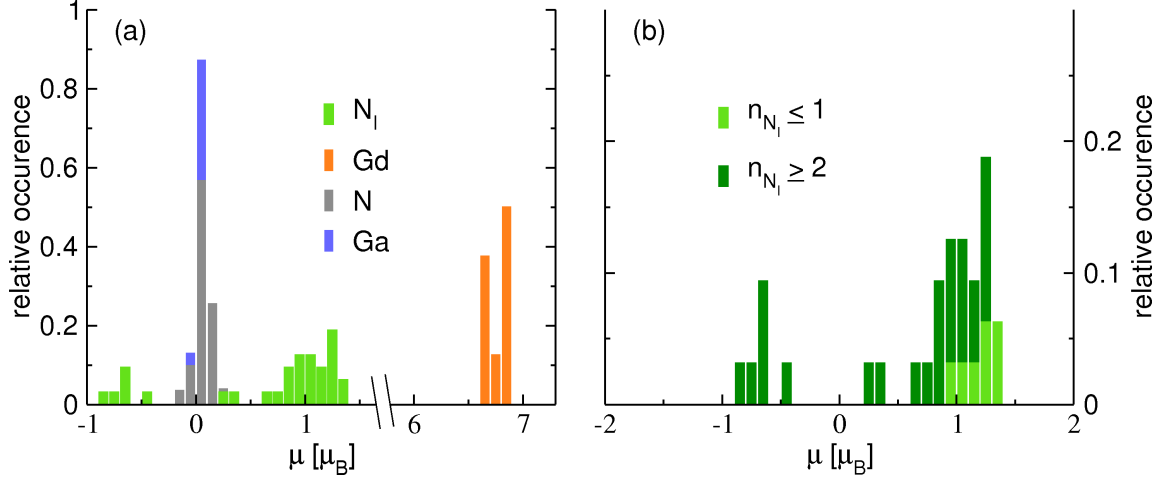
Figure 6.8.: *(a) Site-specific relative probability of size and sign of the magnetic moments in $Ga_{248}N_{256}Gd_8N_{I,32}$. To obtain the relative probability the occurrence in all bins has been normalized by the total number of sites of the same type. In (b) the magnetic moments on N_I sites are split into two groups with respect to the occupation of its innermost three shells of interstitials. The group $n_{N_I} \leq 1$ experiences at most one N_I on those near interstitial sites, accordingly $n_{N_I} \geq 2$ at least two.*

	$\overline{\mu}$	σ_μ	$\overline{\|\mu\|}$	$\sigma_{\|\mu\|}$	$\Sigma\mu$	$\Sigma\|\mu\|$
Gd	6.76	0.08	6.76	0.08	54.09	54.09
N_I	0.57	0.74	0.86	0.34	18.26	27.59
N	0.05	0.08	0.07	0.07	13.95	16.96
Ga	0.01	0.01	0.01	0.01	2.28	2.80
I	0.01	0.01	0.01	0.01	3.50	4.09
Σ					92.08	105.53

Table 6.1.: *Element specific average and standard deviation of magnetic moments ($\overline{\mu}$ and σ_μ) and absolute magnetic moments ($\overline{\|\mu\|}$ and $\sigma_{\|\mu\|}$) for Gd, N_I, N, Ga, and Interstitials (I). In the rightmost columns the sum of magnetic moments and absolute magnetic moments over all sites of equivalent atom type are given ($\Sigma\mu$ and $\Sigma\|\mu\|$, respectively). Total sum of μ and $\|\mu\|$ are shown in the lowermost line. All values are in units of Bohr magneton (μ_B).*

estimate on the required concentration of defects. The sum of absolute magnetic moments in the supercell in our calculated configuration is 105 μ_B, where roughly 1.5 μ_B can be accounted to each of the N_I's (see Tab. 6.1). Hence, in order to generate colossal magnetic moments of up to 4000 μ_B per Gd atom as observed in experiments [1] thousands of N_I's would be required. In this limit an occupation of the Ga site is according to Fig.1 and Fig.3 of [1] one Gd in 10^5 Ga atoms. For each Ga site two interstitial positions are availible in the lattice and about 1000 N_I's would have to be placed on 0.5% of these sites to create colossal moments. From this estimate it becomes clear that a coupling amongst Gd atoms alone cannot stabilize the ferromagnetic state. Rather the Gd-N_I and in particular N_I-N_I can be expected to play a decisive role.

6.2.3 Magnetic exchange coupling

One of the advantages of the KKR formalism above most other electronic structure methods is the possibility to elegantly extract the magnetic exchange coupling constants (J_{ij}) between individual atoms i and j by means of the Lichtenstein formula [38]. Details on this formalism and how it is implemented in KKR`nano` are given in chapter 5.

The Lichtenstein formula can be applied straightforwardly in a standard self-consistency iteration using the same energy contour integration in the complex plane. Using such contours final integrated values J_{ij} can be conveniently obtained. However, such contours include Matsubara frequencies and energy points far from the real axis which lead to the fact, that the energy dependent contributions to J_{ij} have no clear physical meaning. By integration along the real axis including a temperature broadening but no Matsubara-poles we can obtain the J_{ij} and the energy dependency of the magnetic exchange coupling. This additional information will crucially contribute to explain our results. Here, it is important to note that the integration over the complex contour with poles turned out to be of higher accuracy than the integrational scheme along the real axis − even if a factor of ten more integration points are used. Accordingly, we use the same complex contour as in the self-consistency steps which includes the Matsubara frequencies to calculate the J_{ij}'s. On the contrary once the energy dependency of the magnetic exchange interactions becomes of interest, we have used the integration along the real axis with 250 energy integration points, a temperature broadening of 1600 K and the Γ point as representation for the Brillouin-zone[5]. It is important to consider that not all spins in the supercell are necessarily aligned in parallel. For the sake of a universal definition the sign of the coupling constants J_{ij} is inverted once spins on i and j are antiparallel[6]. This means that, whenever $J_{ij} > 0$ we have a ferromagnetic interaction, while $J_{ij} < 0$ signals an antiferromagnetic interaction.

Out of ten possible combinations of interacting sites Ga, N, Gd and N_I we first considered the magnetic exchange interaction amongst strongly spin-polarized Gd and N_I sites. In fact the strength of spin-polarization is an important measure for the strength of the magnetic coupling since the difference of the t-matrices of spin-up and down channel directly enter the Lichtenstein formula (5.67) in chapter 5.3. It turns out that the Gd-Gd

---

[5]Here, the high temperature broadening is important to increase the k-points convergency. The J_{ij} have been tested for smaller temperature and larger sets of k-points showing only small variations.

[6]In order to validate this procedure an additional test calculation has been performed: The resulting coupling constants as obtained by inverting the spins are in good agreement with the outcome of calculations of the same cell where all spins of significant size are forced to be aligned in the same direction.
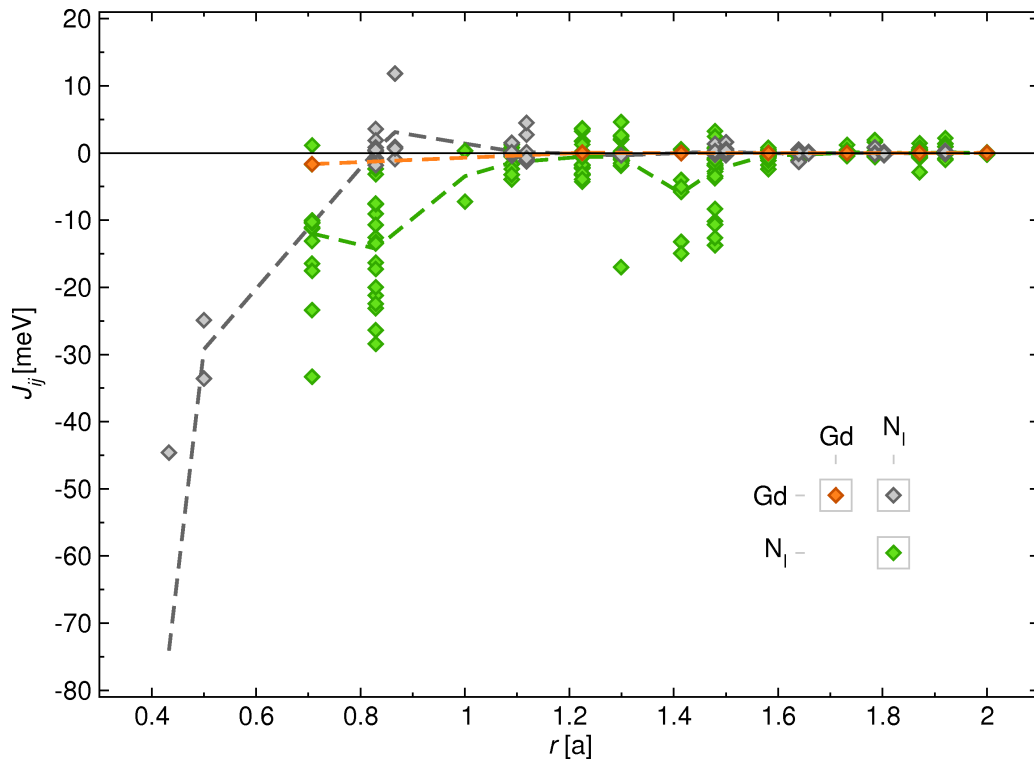
Figure 6.9.: *Magnetic exchange interactions J_{ij} as obtained from the Lichtenstein-formula between Gd and Gd (orange), N_I and Gd (gray) and N_I and N_I sites (green) for $Ga_{248}N_{256}Gd_8N_{I,32}$. Average exchange interactions (dashed lines) are shown for each inter-atomic distance and for all three different cases. The inset summarizes the color-coding for all three types of interactions.*

interaction is remarkably small over the entire range of interatomic distances (Fig. 6.9). Only nearest-neighboring Gd atoms with a distance of $r = \frac{a}{\sqrt{2}}$ possess a considerable but still small antiferromagnetic coupling of about -2 meV. The coupling of Gd to N_I shows a more complex behavior. At small interatomic distances − i.e. if N_I is placed on the adjacent tetrahedral and octahedral interstitial sites $r = 0.43a$ and $r = 0.5a$ − the coupling appears to be strongly antiferromagnetic. For slightly larger distances the coupling is weakly ferromagnetic and the coupling strength decays quickly to zero for interatomic distance larger than a. For the interaction amongst N_I's we find a pronounced antiferromagnetic preference. Within the group of all interactions N_I-N_I is of longest range providing J_{ij}'s of -10 meV for interatomic distances up to $r = 1.5a$. Striking are the large variations for different participating defects. We will discuss this finding in the context of the local atomic configuration of the participating defect after completing the general picture on the exchange interactions with the Gd-N and N_I-N coupling.
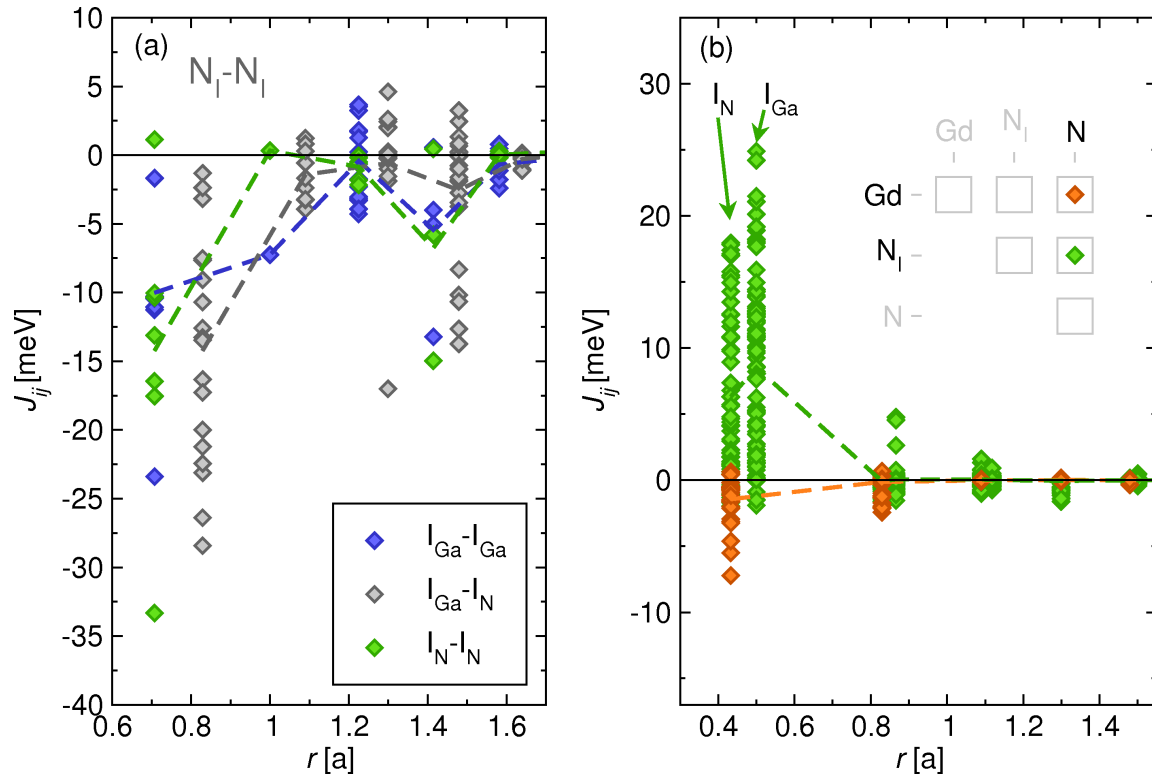
Figure 6.10.: (a) Magnetic exchange interactions J_{ij} between interstitial nitrogen sites from Fig. 6.9 resolved with respect to the two different interstitial sites I_N and I_{Ga} as defined in Fig. 6.7. For all three interaction types the mean exchange interaction is displayed in dashed lines with the corresponding color coding. In (b) magnetic exchange interactions of Gd and N as well as N_I and N sites for a unit cell of $Ga_{248}N_{256}Gd_8N_{I,32}$ are drawn. For both groups of interactions the mean exchange interaction is given in dashed lines.

As previously discussed a small but significant spin-polarization of up to 0.1 μ_B can be observed on N atoms adjacent to N_I. Apparent from Fig. 6.11(b) both Gd and N_I support a strong but very short-range ($r \leq 0.5a$) coupling to N atoms. While the Gd-N interaction

favors in all cases antiferromagnetic alignment, N_I-N couples ferromagnetically with on average $J_{ij} \approx 8$ meV. Additionally no dependency of the N_I-N interaction on the tetrahedral ($r_{ij} = \frac{\sqrt{3}}{4}a$) and octahedral ($r_{ij} = \frac{1}{2}a$) position appears. However, the determining factor is the short-rangeness of the Gd- and N_I-N interaction which can only support a magnetic network in the limit of very high concentrations but not in the dilute limit.

Summarizing this first view on the important exchange interactions we find that including N_I in a GaN:Gd sample cannot support ferromagnetism. This holds in particular for the limit of dilute Gd concentrations, where the strong antiferromagnetic N_I-N_I is dominant and which is of high importance to explain the experimental results. However, on a small scale and for large defect concentrations for part of the configurations ferromagnetic coupling could result from the Gd-N_I and N_I-N exchange. This scenario is one possible explanation for the results by Mitra and Lambrecht [126]. Here, the spin-alignment in supercells of 64 atoms has been studied by means of total energy arguments and revealed a tendency towards ferromagnetic alignment of Gd and N_I defects.

It became clear from our analysis that the N_I-N_I coupling has decisive character, which motivates to delve into the reasons for the large variations (see Fig. 6.9) at fixed interatomic distances. For that purpose we probe the dependency of the exchange coupling on two degrees of freedom: on the site dependency and on the local defect concentration.

The first one is charted in Fig. 6.10(a). This viewgraph clarifies that no significant site dependency occurs except for the short-ranged interaction of ($r = \frac{a}{\sqrt{2}}$). Here, N_I's positioned on I_N sites tend to couple twice as strong as the group of N_I's on I_{Ga}.
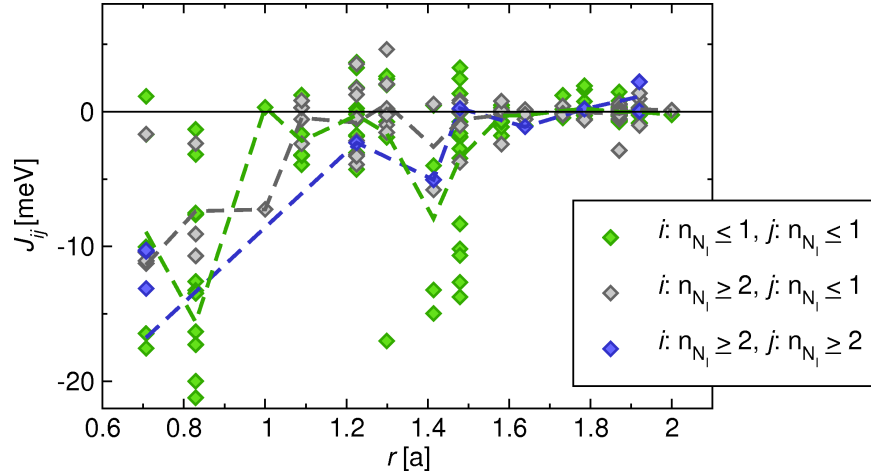
Figure 6.11.: *Dependency of the magnetic exchange interactions J_{ij} amongst nitrogen interstitials (N_I) on the local density of N_I defects. For each pair i and j the occupation of closest three shells n_{N_I} is taken as a measure for the local concentration of N_I. According to the occupation of those sites all N_I's are classified in two groups with either $n_{N_I} \leq 1$ or $n_{N_I} \geq 2$. Individual and mean values of magnetic exchange interactions within those groups and the intermediate interaction are shown in the color coding as defined in inset.*

A more important effect is found relating the J_{ij}'s to the local concentration of N_I's as shown in Fig. 6.11: We find the clear tendency that more isolated ($n_{N_I} \leq 1$) N_I's create stronger magnetic exchange coupling. For an understanding of this concentration depen-

dency and the search for possibilities to engineer a ferromagnetic network it is crucial to reveal the exchange mechanisms. Therefore, the different types of exchange mechanisms will be briefly reviewed in the following.

Exchange mechanisms in dilute magnetic semiconductors

All magnetic exchange phenomena which occur in dilute magnetic semiconductors can be conceptually divided into three fundamental exchange mechanisms. On the one hand Zener's double and Zener's *p-d*-exchange both supporting ferromagnetic ordering. On the other hand the superexchange which predominantly leads to antiferromagnetism. In practice those mechanisms are of course coexisting and competing. However, we will begin the discussion by providing a physical picture for all those phenomena individually where we follow conceptually the work of Sato *et al.* [46] and Belhadji [135].

For further discussion it is intuitive to introduce a simple molecular model with two states ϵ_1 and ϵ_2. Allowing for a finite hopping t between those states a coupling occurs which can be described by the off-diagonal elements of the Hamiltonian of this toy model:

$$\begin{pmatrix} \epsilon_1 & t \\ t & \epsilon_2 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}. \tag{6.1}$$

The eigenvalues of this matrix − which are the eigenenergies of the system − then read:

$$\lambda = \frac{\epsilon_1 + \epsilon_2}{2} \pm \sqrt{\left(\frac{\epsilon_1 - \epsilon_2}{2}\right)^2 + t^2}. \tag{6.2}$$

Up to this point no assumptions have been made, which provides the flexibility to derive both the double exchange and the superexchange mechanisms from (6.2).

Double exchange In order to create non-zero double exchange a finite density of states at
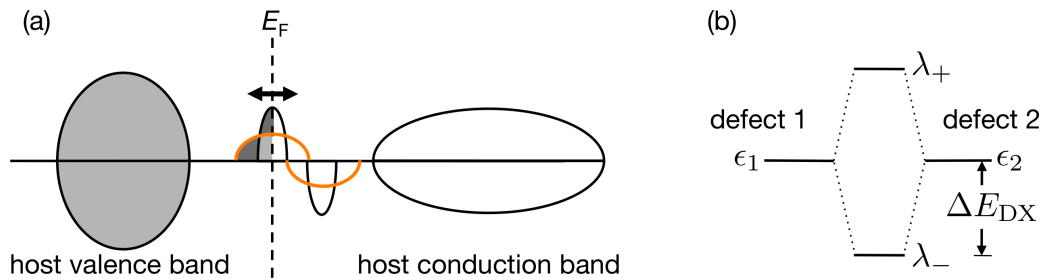
Figure 6.12.: *Schematic view on the double exchange mechanism illustrated by the density of states of a bandgap material with half-filled majority defect band (a). In (b) the magnetic coupling and its influence on the defect states is shown. For details see text.*

the Fermi level (E_F) has to be present. E.g. if majority or minority defect levels in the gap are only partly occupied such as illustrated in Fig. 6.12(a). Given this fact hybridization of the partly filled states and accordingly a broadening of the affected defect band directly translates in a gain of kinetic energy. Since a hybridization appears only between bands resembling the same spin direction this effect favors ferromagnetism. Expressed in terms of

the molecular model both involved majority states ϵ_1 and ϵ_2 are degenerate ($\epsilon_1=\epsilon_2$) without applied coupling. Then (6.2) results in the generation of an anti-bonding and a bonding state both shifted by $\pm t$ (see also Fig. 6.12(b)), which is the energy gain due to double exchange $\Delta E_{DX} = t$.

Zener's *p-d* exchange Another scenario likely appearing in narrow band-gap semiconduc-
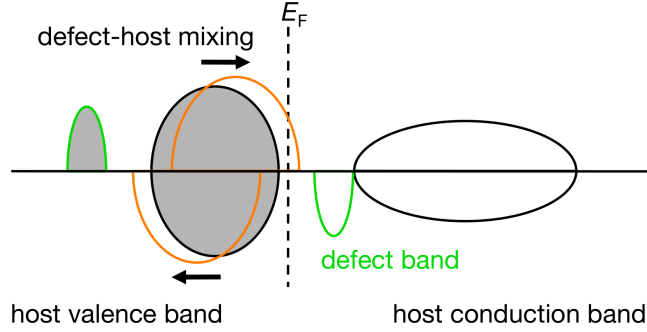
Figure 6.13.: *Schematic view on the Zener's p-d exchange mechanism illustrated by the density of states of a bandgap material with half-filled majority defect band (green) with large spin-splitting. Defect and host band mix resulting in a spin-splitting of the valence band (orange).*

tors leads to Zener's *p-d* type of exchange. It is important to note that the nomenclature originally was used for transition metal defects (*d*) in semiconductors, but the same arguments hold for defects of *p*- or *f*-character. We will anyhow maintain using the original name for the sake of simplicity. In Zener's *p-d* exchange the majority defect state is fully occupied and positioned lower in energy than the host valence band while the minority states remain unoccupied (Fig. 6.13). Without interaction of the defects with the host material, E_F lies in the gap. Including those interactions can crucially change the picture: Both valence bands and defect states of the same spin channel mix resulting in additional states in the energy range of the host valence bands. If E_F remains in the gap charge neutrality would be violated − accordingly E_F is pushed down in energy into the valence band. On the same footing majority as well as minority defect and host states are shifted apart, leading to an antiferromagnetic alignment of host and defect. In sum this antiferromagnetic coupling can mediate an effective ferromagnetic interaction amongst the defects. Since toy models for Zener's *p-d* exchange are less intuitive than for double- and super-exchange and the mechanism is of minor importance in GaN:Gd the reader is referred to Sato *et al.* [46] for further interpretation.

Superexchange While double- and Zener's *p-d* exchange presume a finite density of states at E_F superexchange has no such requirement. The superexchange mechanism was originally introduced to explain the coupling of two, not directly neighbored, cations via a non-magnetic anion [136]. In DMS this complication does not enter and the defect states alone are sufficient to capture the main phenomenon of superexchange. Assume two defects in antiferromagnetic spin alignment as visualized in Fig. 6.14(a). If further the hopping t between those two defects is reasonable small ($|\epsilon_1 - \epsilon_2| \gg t$). the eigenvalues defined by
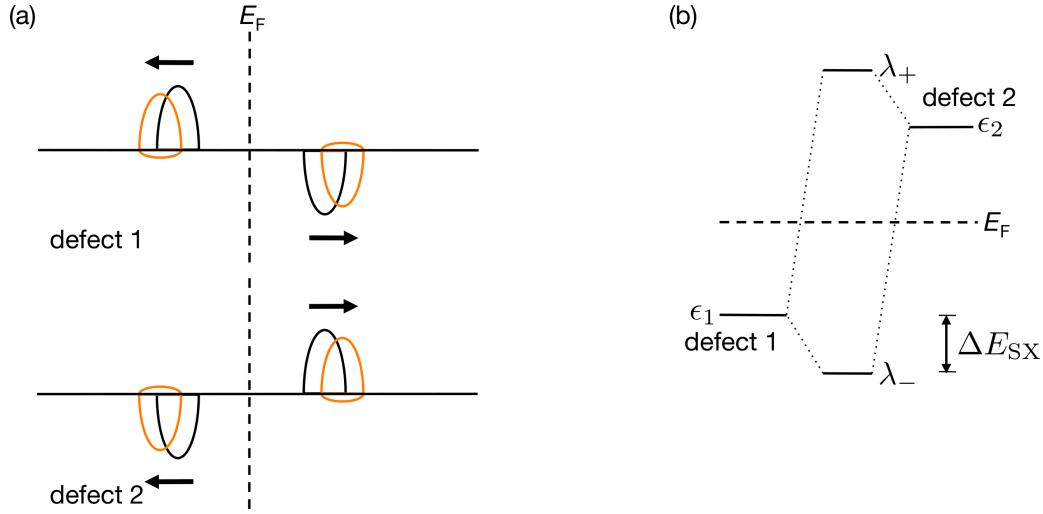
Figure 6.14.: *Schematic representation of the superexchange mechanism. In (a) the density of states of a bandgap material with half-filled defect band, where majority and minority states are well separated is shown. (b) depicts the magnetic coupling and its influence on the defect states. For details see text.*

(6.2) can be expanded in a Taylor-series up to first order and simplify to:

$$\lambda_+ = \epsilon_2 + \frac{t^2}{\epsilon_1 - \epsilon_2} \tag{6.3}$$

and

$$\lambda_- = \epsilon_1 - \frac{t^2}{\epsilon_1 - \epsilon_2}. \tag{6.4}$$

Hence, λ_- represents the bonding state of two antiferromagnetically coupled defect spin (see Fig. 6.14(b) for a graphical representation). Accordingly the energy which can be attained by superexchange reads for $(|\epsilon_1 - \epsilon_2| \gg t)$ as[7]:

$$\Delta E_{\mathrm{SX}} = \frac{t^2}{\epsilon_1 - \epsilon_2}. \tag{6.5}$$

After having considered the idealized scenarios for all three types of exchange we can accomplish an identification of the crucial exchange mechanisms in GaN:Gd with N_I defects. The five different groups of interactions shown in Fig. 6.9 and Fig. 6.10 provide a rich set of information for many different local configurations and several interatomic distances. To condense this vast amount of information to the essential, exclusively mean interactions at selected interatomic distances will be discussed. Before that two quantities have to be introduced which will provide important leads on the type of interaction. As noted previously the energy integration to obtain the magnetic exchange coupling constants is performed along the real axis using a temperature broadening but no additional Matsubara poles in the contour. By that the energy-resolved magnetic exchange-coupling $j_{ij}(E)$ is

---

[7]See Fig. 6.14 for a graphical representation of ΔE_{SX}.

calculated at each energy integration point E. Based on this input the energy-resolved integrated exchange interaction $J_{ij}(E')$ can be calculated by:

$$J_{ij}(E) = \int_{-\infty}^{E} j_{ij}(E')dE', \tag{6.6}$$

where in practice the integration is not started at $E = -\infty$ but at the bottom of the valence energy window. Accordingly the magnetic exchange coupling constants J_{ij} can be defined as

$$J_{ij} = J_{ij}(E_{\mathrm{F}}). \tag{6.7}$$

In Fig. 6.15 $j_{ij}(E)$ and $J_{ij}(E)$ are plotted versus the density of states for the four most important groups of interactions.

For the N_I-N_I interaction the distance of $r_{ij} = 0.829a$ is exemplary for the strong anti-ferromagnetic coupling amongst those defects. As can be seen in Fig. 6.15(a) the majority and minority defect levels of N_I split by $\simeq$1.5 eV. Keeping in mind the large temperature broadening of the DOS, the splitting results in almost fully occupied majority states and almost depleted minority states. Thereby the electronic structure complies with the requirements for dominant superexchange. Indeed, a strong antiferromagnetic contribution ($j_{ij}(E)$) is created from the majority p states, changing sign at E_{F}. Besides this most important signature a small asymmetry which superimposes the super-exchange contribution can be observed. First, between -1.5 and -1.0 eV a ferromagnetic contribution builds up. Secondly, after extending the integration to higher energy, e.g. $J_{ij}(E = 1 \text{ eV})$ has a positive value of more than 10 meV. Both findings indicate that the splitting of the defect levels is not fully completed and a small ferromagnetic double-exchange contribution competes with the antiferromagnetic superexchange. We will come back to this double-exchange peak at about 1 eV in the next section, when O interstitials are considered. Overall, we can nevertheless conclude that the N_I-N_I interaction is mainly driven by the superexchange mechanism leading to antiferromagnetic ordering.

The Gd-Gd interaction (see Fig. 6.9) appears to be non-zero only for the nearest-neighbor distance $r_{ij} = 0.707a$. The energy-resolved contribution to the coupling constant J_{ij} are shown for this short-range distance in Fig. 6.15(b). Here, $j_{ij}(E)$ oscillates at the position of the Gd d and f states but vanishes for energies larger than $E = -1.0$ eV. By that the superexchange amongst the Gd states can be identified to be responsible for the weak antiferromagnetic coupling.

In case of Gd-N_I interactions we focus on the explanation of the most striking characteristic − the strong antiferromagnetic coupling at short distances shown in Fig. 6.9. Evaluating the results of Fig. 6.15(c) the energy dependency bears a striking resemblance to the N_I-N_I coupling. By the same arguments the super-exchange mechanism is clearly deciding for the short-range Gd-N_I interactions. However, this contribution likely decays more quickly than the initially smaller ferromagnetic contributions leaving a small ferromagnetic coupling at distances $r_{ij} > 0.8$.

The N_I-N exchange demonstrates a completely different pattern. As shown in Fig. 6.10 a significant coupling only occurs to nearest shell of N atoms. Comparing the mean LDOS and $j_{ij}(E)$ clarifies that the crucial contribution to exchange is the double-exchange of the defect states and the defect induced spin-polarized states on the adjacent N atoms.
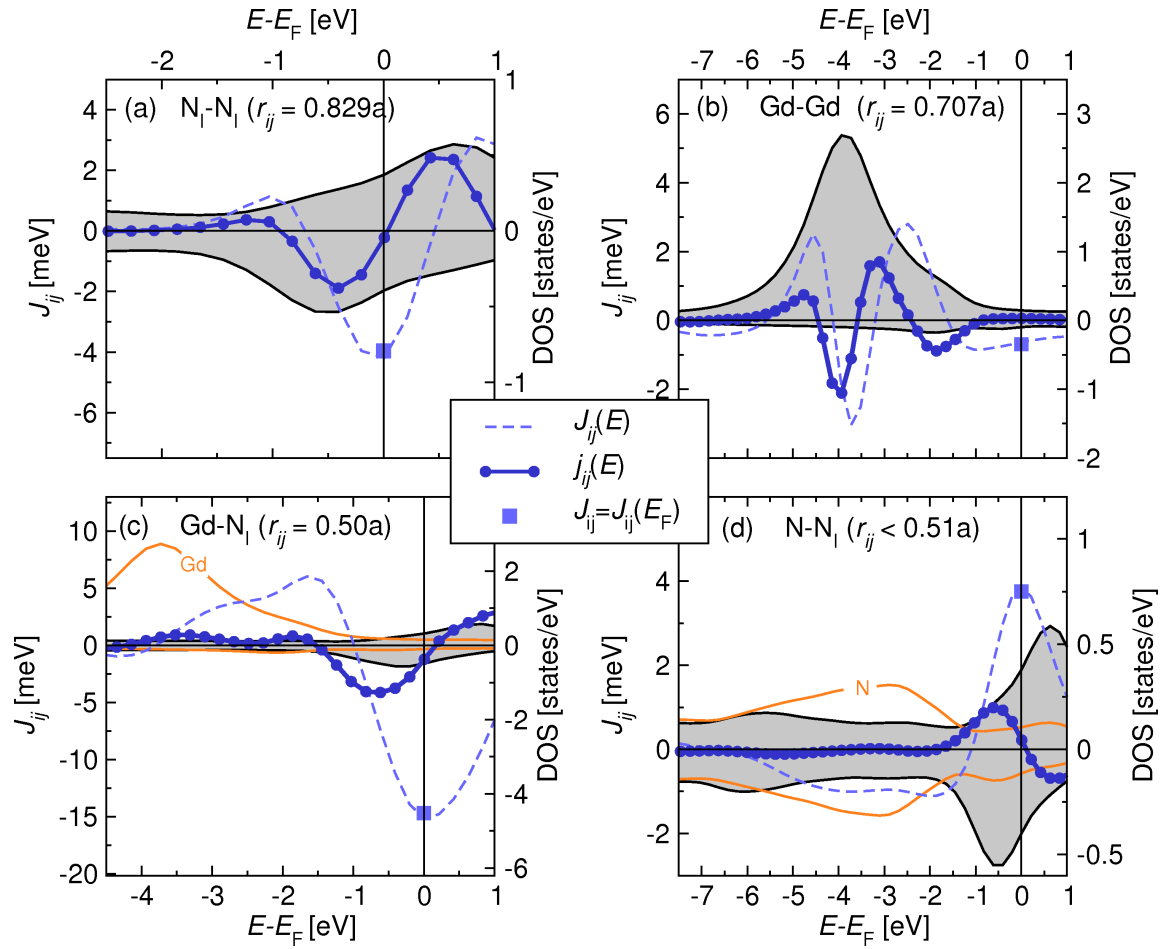
Figure 6.15.: *Mean energy-resolved exchange interaction ($j_{ij}(E)$) and mean energy-resolved integrated exchange interaction ($J_{ij}(E)$) at selected interatomic distances r_{ij} for N_I-N_I (a), Gd-Gd (b), Gd-N_I (c), and N_I-N (d). In all panels the mean local density of states (DOS) of the interacting sites is shown as gray-shaded area, where the minority and majority spin channels are represented by negative and positive DOS values, respectively. In (c) and (d) two different sites are interacting and the DOS of the second component Gd in (c) and N_I in (d) is displayed in orange lines.*

In summary, our calculations showed that N_I's in GaN:Gd reveal that those defects are strongly spin-polarized. The polarization of the neighboring sites is however a very localized phenomenon. We found that Gd-N_I and in particular N_I-N_I are the decisive interactions for the magnetic ordering in the sample. The clear antiferromagnetic interaction amongst tetrahedral or octahedral N_I's leaves thereby no room for an explanation of the experimentally observed ferromagnetic signatures by means of N_I's.

6.3 Oxygen-Interstitials in GaN:Gd

The experimental observation of residual oxygen in the GaN:Gd [113, 137] brings up the question on the influence of oxygen defects on the magnetic properties of the material. Recently, systematic studies on oxygen codoping have been published by Roever *et al.* [138] and revealed that the presence of oxygen enhances the tendency towards ferromagnetic coupling in GaN:Gd. From theoretical point of view, the structural properties of oxygen defects in GaN have been investigated by Mattila and Nieminen [139]. Their first-principles calculations revealed that under equilibrium conditions and high temperatures oxygen tends to locate substitutionally on the N sites. However, for non-equilibrium growth e.g. by oxygen-implantation oxygen-interstitials (O_I's) can − although having a considerable higher formation energy − occur in a significant concentration in the sample [139, 140].

Mitra and Lambrecht [50] performed calculations considering the magnetic coupling of such O_I's to Gd in GaN. In this section we will extend this analysis to a large supercell of statistically distributed Gd- and O_I defects. The chemical composition of the supercell is $Ga_{253}N_{256}Gd_3O_{I,20}$ guaranteeing sufficient sampling of the configurations to extract statistically sound properties. By this choice of a slightly smaller concentrations of Gd and O_I as compared to N_I in section 6.2 two characteristics are modeled. First, the formation energy of O_I is larger than for N_I. Secondly, the number of Gd atoms in the cell is reduced since one of the findings in our analysis of N_I defects is that the dilute limit of Gd concentration is the fundamental regime to explain the material properties. Except for this point, the treatment, including the computational parametrization and the assumption made considering ideal lattice positions, is similar to the calculation of N_I's. Due to these similarities we will discuss the electronic structure, magnetic properties and exchange coupling mechanism of O_I's in close connection to the results obtained for N_I's in section 6.2.

6.3.1 Electronic structure

The electronic structure of $Ga_{253}N_{256}Gd_3O_{I,20}$ is shown in Fig. 6.16 resolving the DOS locally and with respect to its orbital character. The first observation is that the p states of the group of O_I become spin-polarized with a non uniform spin direction with respect to the Gd spins. By introducing oxygen on interstitial sites one additional p-electron is donated as compared to N_I, which results in a mean total charge on O_I of 6.4 electrons as compared to 5.3 electrons for N_I. More effects appearing for O instead of N doping become visible on the atoms surrounding the defects. As apparent from Fig. 6.16(a) and (b) the value of the induced spin-polarization on N and Ga sites is larger than for N_I in Fig. 6.6 (a) and (b). Here, it is important to keep in mind that the plotted average values cannot be used for a one-to-one comparison of O_I and N_I, because of the different defect concentrations (20 O_I versus 32 N_I).

Another outcome is that the position of the defect states with respect to the valence bands is shifted crucially. On the one hand for N_I's those states are located in the middle of the bandgap − well separated from the valence and conduction band edge. On the other hand O_I defect states are lying closer to the valence band maximum which results in a smooth transition of defect-induced state to valence band states on N- and Ga atoms. This shift directly translates in a lower position of E_F − best observable relating the Gd d states to E_F for O_I and N_I in Fig. 6.16(c) and Fig. 6.6(c), respectively.
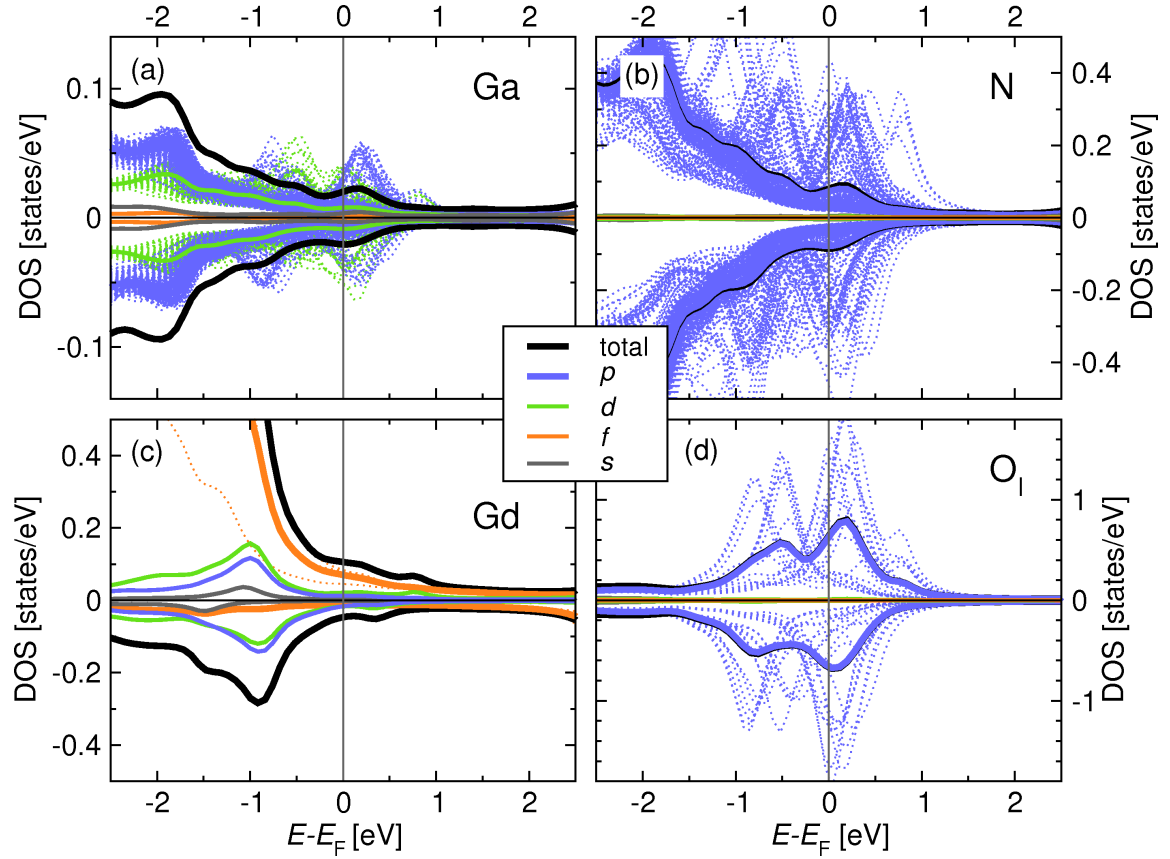
Figure 6.16.: *Local density of states (DOS) of $Ga_{253}N_{256}Gd_3O_{I,20}$ calculated with a temperature broadening of 600 K. Here, the minority and majority spin channels are represented by negative and positive DOS values, respectively. For all Ga (a), N (b), Gd (c) and O_I sites the mean value of the orbital decomposition into s, p, d and f states is shown in relation to the mean total DOS (all in solid thick lines according to the color coding defined in legend). In addition the individual partial DOS of p and d states for Ga (a), p states for N (b), f states for Gd and p states for O_I are displayed in dotted lines.*
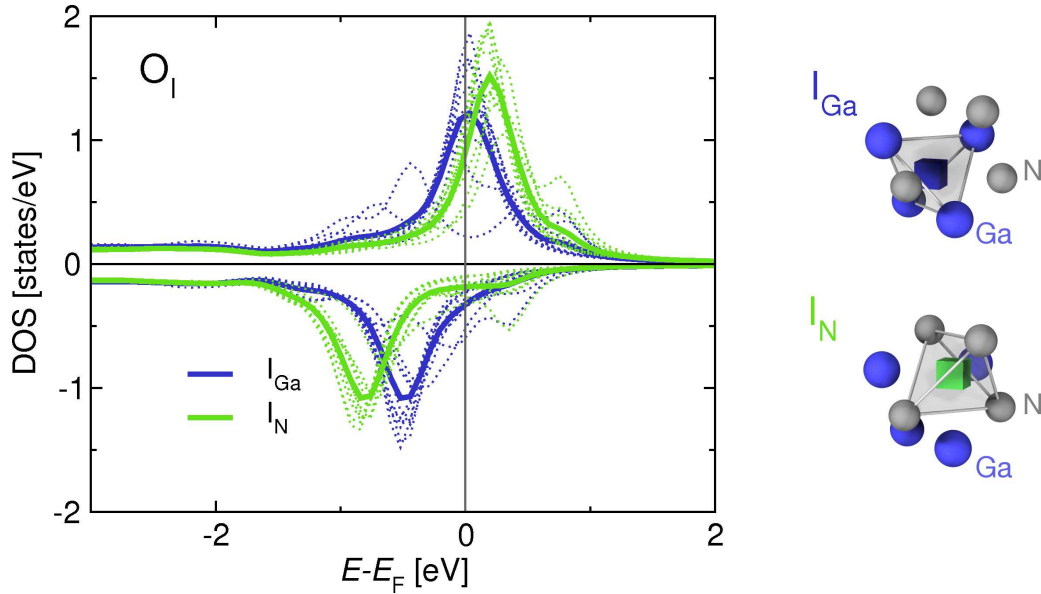
Figure 6.17.: *Local density of states (LDOS) of oxygen (O_I) on the two inequivalent inter-stitial sites in the zinc-blende lattice. Here, individual and mean LDOS are shown in dotted and solid lines, respectively and the minority and majority spin channels are represented by negative and positive DOS values. The different sites are labeled according to its nearest neighboring atoms as I_N (tetrahedral) and I_{Ga} (octahedral), both shown in the right panel.*

Since all oxygen defects are randomly placed on the high-symmetry tetrahedral and octa-hedral sites our calculations provide sufficient input to evaluate the different binding mech-anisms on those sites. In order to disclose the general trends and not discuss the variations due to differing local defect configurations the mean DOS of tetrahedral (I_N) and octahe-dral sites (I_{Ga}) are plotted in Fig. 6.17 [8]. Comparing both O_I configurations, on I_N sites a considerable enhanced spin-splitting occurs. In particular, the majority states are shifted to lower energies resulting in larger magnetic moments: For oxygen on I_{Ga} average magnetic moments of 0.52 μ_B appear while on I_N on average 0.99 μ_B is reached. Considering the local energies of O_I sites on I_N and I_{Ga}, we find that the tetrahedral position (I_N) is the favorable one by 0.85 eV. I.e. O_I significantly stronger prefers the I_N site than N_I's if only the local energy of the interstitial site is taken into account. This difference can be explained by the strong O-N bonding which is unambiguously more pronounced if O_I experiences N on the nearest neighboring shell. Similar to our finding for N_I's the surrounding matrix of Ga and N atoms would favor rather the octahedral (I_{Ga}) interstitial by 0.5 eV per atom. Disre-garding relaxation, the competing energy contributions of the interstitial site with 0.85 eV and the four neighboring shell atoms with 0.5 eV each still leads to a preferential binding of oxygen on the I_{Ga} (octahedral) site.

---

[8]To obtain a meaningful average value all local spins are aligned to a uniform spin direction. I.e. on the sites with spins pointing in which point not in direction of the total spin of all N_I sites, the majority and minority LDOS is interchanged for this analysis.

6.3.2 Magnetic exchange mechanisms
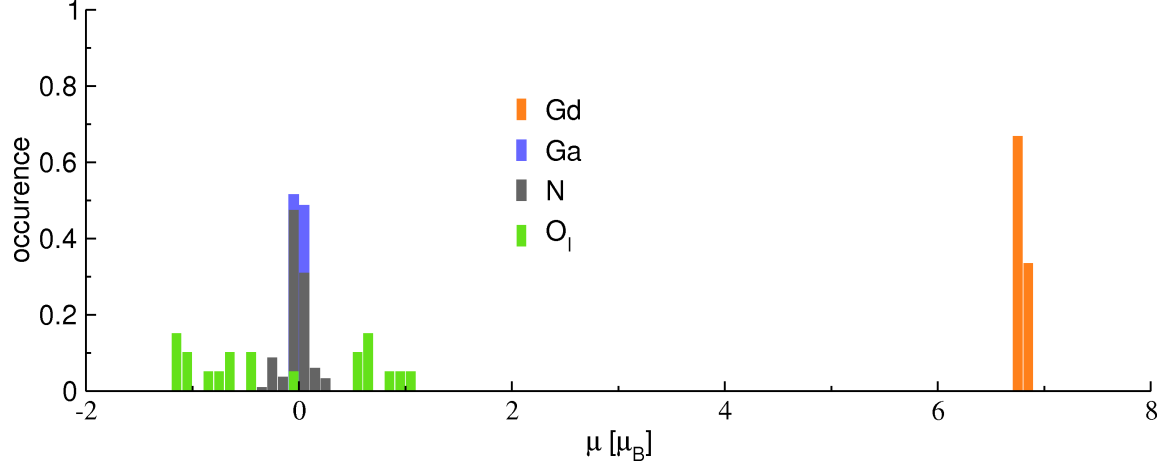
Figure 6.18.: *Site-specific relative probability of size and sign of magnetic moments in* $Ga_{253}N_{256}Gd_3O_{I,20}$. *To obtain the relative probability the occurrence in all bins has been normalized by the total number of sites of the same type.*

Before evaluating the magnetic exchange interactions in O_I doped GaN:Gd, first the strength of spin-polarization will be quantified. In Fig. 6.18 the distributions of the magnetic moments are displayed for all elements in the unit cell. As expected from the previous analysis of the electronic structure the Gd magnetic moment remains close to 7 μ_B. The difference in strength of hybridization from O_I to Ga and N sites is clearly resembled by the larger magnetic moments on N sites. Most noticeable is the broad distribution amongst O_I between -1.2 μ_B and 1.2 μ_B, which is in close agreement to the characteristics of N_I sites.

Out of the full set of all possible exchange interactions[9] part of them can be safely excluded on this basis: First Ga sites experience only a marginal spin-polarization which results in a weak magnetic interaction to Gd, N, Ga and O_I. The strongest of those interactions, Ga-O_I, remains even for the nearest-neighbor distance on the order of 0.1 meV. Secondly interaction amongst the weakly spin-polarized N atoms turn out to be always well below 1 meV and of short-range nature. Based on the data presented in Fig. 6.19 the remaining groups of magnetic exchange interactions J_{ij} will be analyzed in greater detail.

We will begin the analysis considering the three types of interactions involving O_I which are displayed in Fig. 6.19(a). First, for the Gd-O_I exchange only a few configurations exist in the unit-cell which surely limits its statistical evaluation. Nevertheless, interactions of up to a few meV can be observed, which show a weak ferromagnetic tendency at smaller distances. Thus, in the examined range of distances no striking difference of O_I-Gd and N_I-Gd coupling can be observed (see Fig. 6.9).

On the contrary, the characteristics of the interactions amongst the interstitial defects are clearly changing. While for N_I-N_I a strong antiferromagnetic interaction is found over a wide range of distances, from the O_I-O_I exchange a no clear preference between ferro- or

---

[9]All magnetic exchange interaction J_{ij} have been evaluated in similar fashion as for the samples with N_I by using the self-consistent complex contour integration including Matsubara poles with a temperature broadening of 700 K in order to achieve high accuracy.
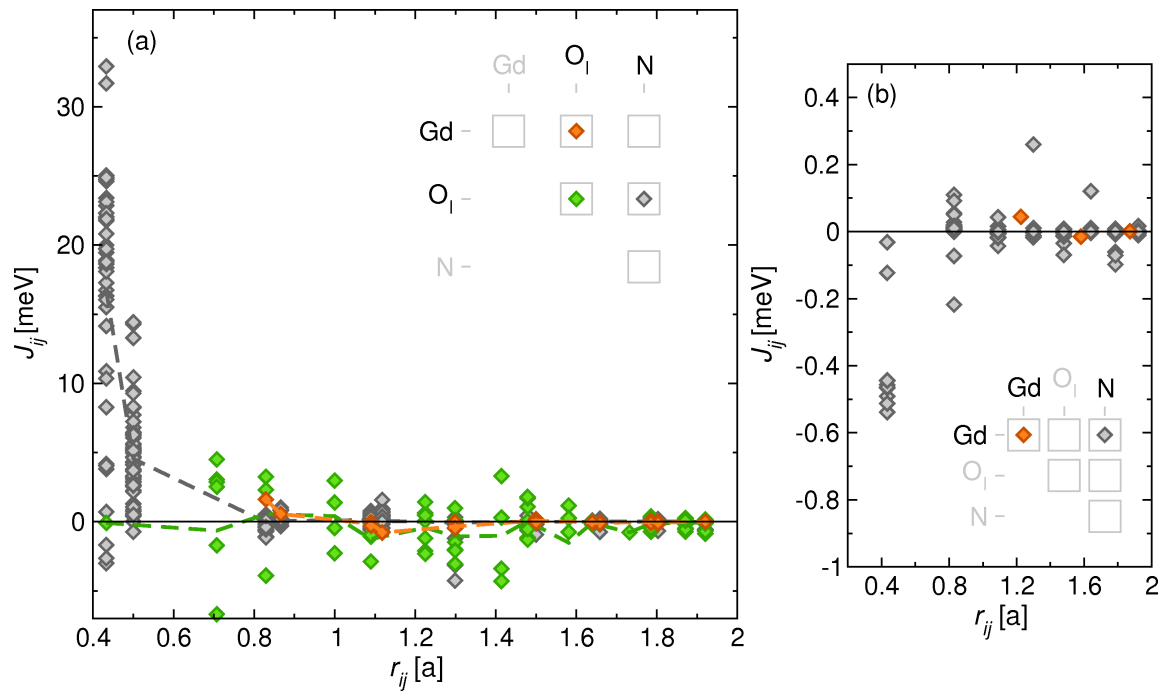
Figure 6.19.: (a) Magnetic exchange interactions J_{ij} as obtained from the Lichtenstein formula between O_I and Gd (orange), O_I and N (gray) as well as O_I and O_I sites (green) for $Ga_{253}N_{256}Gd_3O_{I,20}$. The average exchange interactions for each distance r_{ij} are given in dashed lines for each of the interaction types in the corresponding color. In (b) the J_{ij}'s of interacting Gd and Gd (orange) and Gd and N (gray) sites is displayed. Both in (a) and (b) the inset summarizes the color coding for the different interaction types. Note the different scale of (a) and (b).

antiferromagnetic spin-alignment can be deduced. However, for smaller distances configuration occur which show a signifcant ferromagnetic coupling. This coupling appears e.g. at the interatomic distance of $r_{ij} = 0.829a$ where all probed configurations exhibit a ferromagnetic coupling of 3-5 meV can be reached. In order to track down the reasons for the significantly different pattern of exchange for N_I-N_I and O_I-O_I both will be discussed on the basis of energy-resolved exchange later on.

To complete the picture how O_I is magnetically incorporated in GaN:Gd the interaction of O_I with N sites is analyzed. O_I-N possesses a ferromagnetic coupling limited to the innermost shell of N atoms. This includes two different interatomic distances due to the presence of both tetrahedral and octahedral O_I. Here, the underlying mechanism is similar to N_I-N and arises from double-exchange.

It remains the task to discuss the coupling between Gd atoms as well as Gd with its surrounding GaN matrix. Apparent from Fig. 6.19(b) the longer-ranged interactions amongst Gd which are present in this unitcell do not exceed values of 0.1 meV. In analogy to our results obtained from the N_I doped supercell Gd-Gd exchange cannot support magnetic ordering of any kind. The left-over interaction from Gd to N is very short-ranged and weakly antiferromagnetic.

Summarizing, out of all groups of interactions no crucial differences to the supercell containing N_I appear except for the most important one, the O_I-O_I exchange. Instead of the antiferromagnetic coupling of N_I-N_I which prevents the existence of any long-range ferromagnetic regime in GaN:Gd, the O_I-O_I coupling can favor $-$ at least for part of the configurations at short-range distances $-$ a ferromagnetic alignment. In order to explain this effect the energy-resolved contribution to the exchange coupling $j_{ij}(E)$ as well as the integrated value $J_{ij}(E)$ [10] are illustrated in direct comparison O_I-O_I versus N_I-N_I in Fig. 6.20.

$j_{ij}(E)$ shows a dip-peak-dip structure where the peak represents ferromagnetic contributions and is located at E_F. The antiferromagnetic (dip) contributions beginning at the lower edge of the temperature broadened defect majority states can be attributed to superexchange. For an interpretation of the peak it is important to note that the minority states on O_I are approximately half filled. Accordingly, the maximal LDOS-values is positioned at E_F. Under this circumstances double-exchange originates a strong ferromagnetic contribution. The importance of the filling of the minority states is underlined by contrasting the LDOS of O_I and N_I. From Fig. 6.20(b) it becomes clear that the shift of the minority band and the different occupation directly translates into a shift and quenching of $j_{ij}(E)$.

Considering the limit of dilute Gd concentrations the O_I-O_I interactions are clearly determining the magnetic order. Studying the magnetic properties of O_I's we have found that the the O_I-O_I interaction might support $-$ at least for part of the defect configurations $-$ a weak ferromagnetic coupling. However, it is unlikely that unintentional doping with residual oxygen can result in sufficiently large defect concentrations required to stabilize a magnetic state, which excludes that solely O_I's can create a stable ferromagnetic network. In any case, our results show that the presence of O_I's can lead to a strengthening of an existing ferromagnetic network due to the coupling to neighboring N atoms and to other O_I's. This finding can foster the interpretation of the recent experimental work on the enhancement of ferromagnetism in GaN:Gd due to oxygen codoping by Roever *et al.* [138].

---

[10]For the definition of $j_{ij}(E)$ and $J_{ij}(E)$ see (6.6) $-$ here, they are obtained by integrating along the real axis without Matsubara-poles with 700 K.
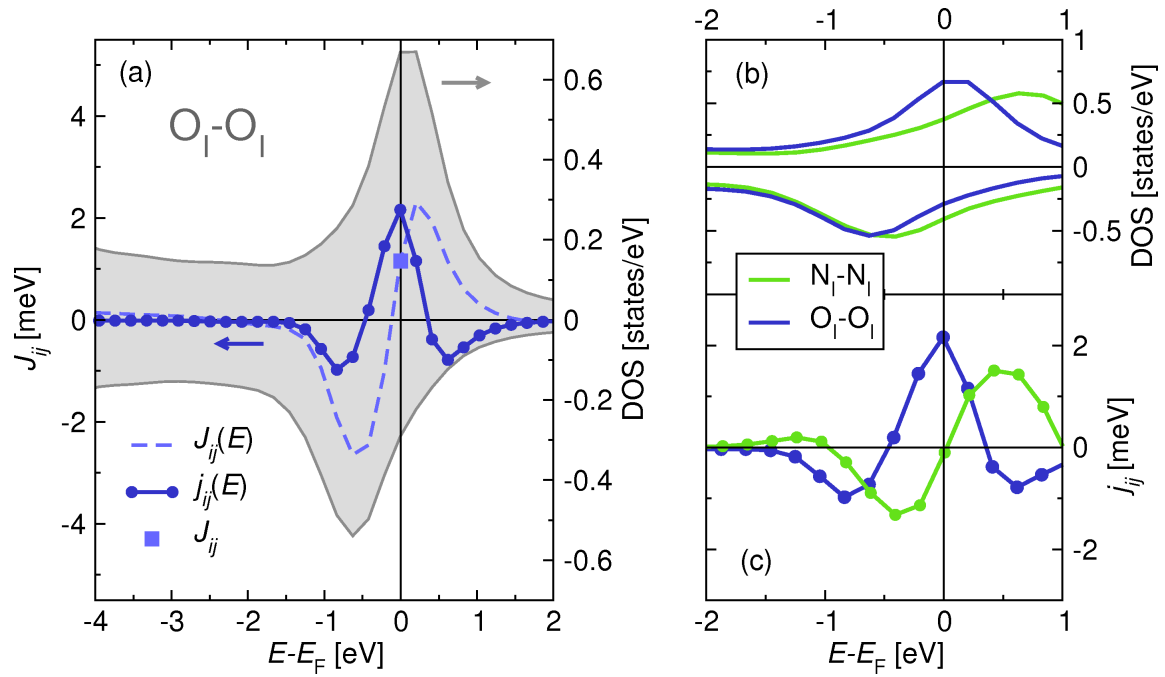
Figure 6.20.: *(a) Mean local density of states on O_I sites versus the mean energy-resolved exchange interaction amongst O_I (O_I-O_I) $j_{ij}(E)$, and its mean energy resolved integrated exchange interaction $J_{ij}(E)$. The square shows the mean exchange coupling constant $J_{ij} = J_{ij}(E_F)$. All values represent the coupling at a distance $r_{ij} = 0.829a$ in $Ga_{256}N_{256}Gd_3O_{I,20}$. In (b) the mean LDOS and $j_{ij}(E)$ of O_I and O_I-O_I is shown in comparison to the mean LDOS and $j_{ij}(E)$ of N_I and N_I-N_I for the same interatomic distance.*

Further, a view on the energy resolved contribution $j_{ij}(E)$ to J_{ij} (Fig. 6.20) allows to draw conclusions on the impact of electron (n) or hole (p) doping. In particular n-doping by e.g. O_N has the interesting effect, that the upward shift of the Fermi level would directly lead to a strengthening of the ferromagnetic O_I-O_I exchange interactions.

6.4 Gallium vacancies in GaN:Gd

In 2008 Liu *et al.* [51] and Gohda and Oshiyama [52] reported on first-principles calculations revealing ferromagnetic coupling in GaN:Gd when including gallium vacancies (V_{Ga}). Despite this consensus it is important to realize that V_{Ga}'s in GaN possess a large formation energy of 9 eV [132] in the neutral and spin-carrying charge state. In order to rate whether V_{Ga}'s remain a realistic scenario to explain ferromagnetism as well as colossal magnetic moments [1] we will go one step beyond the existing calculations [51, 52]: Based on the magnetic exchange interactions between individual sites we will identify crucial exchange mechanisms, concentration thresholds as well as Curie temperatures of the ferromagnetic network.
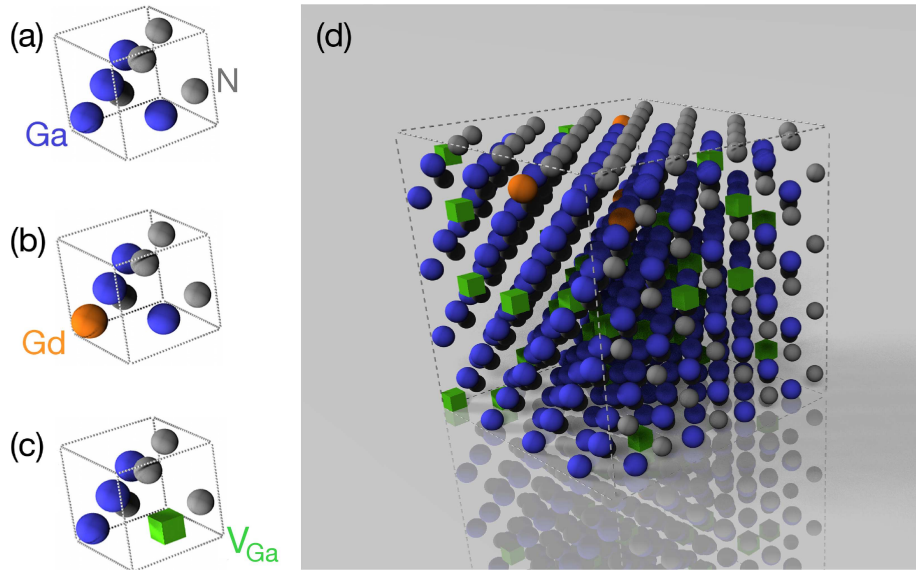
Figure 6.21.: *Illustration of the fundamental building blocks of the investigated structure: GaN in zincblende structure without defects (a), with Gd on the Ga site (b) and a Gallium vacancy (V_{Ga}) on a randomly selected Gallium site. In (d) the actually calculated structure $Ga_{220}N_{256}Gd_4$ is shown. For all panels Ga, N, interstitial, Gd and V_{Ga} are labeled in blue, gray, transparent, orange and green, respectively.*

The geometric structure which we have considered is schematically drawn in Fig. 6.21. 32 V_{Ga}'s and four Gd atoms are randomly placed on the ideal Ga positions of a GaN matrix containing 512 atoms on 1024 sites. We restrict ourselves to the zincblende lattice, which is well-justified by the finding that the magnetic properties of V_{Ga}'s in wurzite or zincblende show no significant differences [51]. For the sake of simplicity we will not consider the outward relaxation of the N atoms surrounding V_{Ga} of about 10% of the bonding-distance [116, 141]. For more detailed information on the computational representation of the supercell the reader is referred to section 6.1.
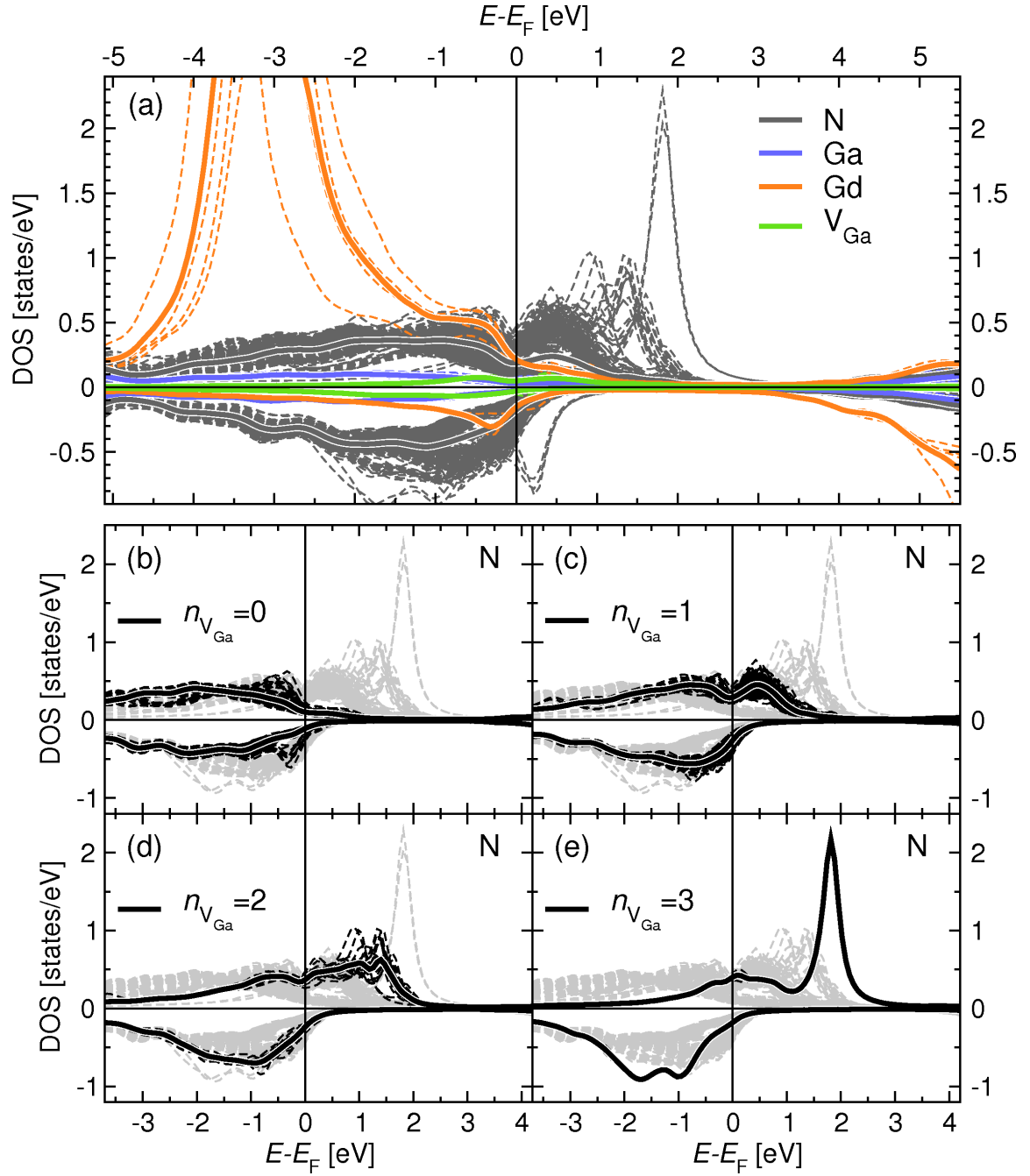
Figure 6.22.: (a) Local density of states (DOS) of Ga, N, V_{Ga} and Gd sites in $Ga_{220}N_{256}Gd_4$. DOS of all sites is shown in dashed lines, the average DOS for all groups of elements in solid thick lines. (b-e) shows the DOS on N sites for different number of V_{Ga} on the nearest-neighboring Ga sites ($n_{V_{Ga}}$). Dashed black lines represent DOS on all N sites with $n_{V_{Ga}}$=0, 1 , 2, and 3 in (b), (c), (d), and (e), respectively and solid black lines the average amongst those groups. Gray lines show the DOS on the remaining N sites. In all panels negative and positive DOS values represent the minority and majority spin channel.

6.4.1 Electronic structure

The local density of states in Fig. 6.22(a) shows the consequence of hole doping with V_{Ga}'s. The Fermi level E_F is pushed into the valence band leaving a considerable part of the nitrogen p states unoccupied in the gap. Those unoccupied states are located on the nearest neighboring positions around the V_{Ga}'s which generate a deficiency of three electrons in the valence band each. In addition, a strong spin-polarization can be observed which collectively resembles an antiferromagnetic ordering with respect to the initially aligned Gd spins. This most striking effect on the N sites is considered in greater detail in Fig. 6.22(b)-(e). It turns out that hole doping affects mainly the surrounding matrix of atoms and can therefore be accounted as a well localized effect. In fact, mainly the occupation of the nearest neighboring Ga sites determines the electronic structure on N sites while the influence of the presence of vacancies on the next nearest neighboring Ga shell can be regarded as small (see Fig. 6.23 for an illustration of both shells). Due to that N atoms can be conceptually grouped according to the number of nearest neighboring V_{Ga}'s. Evident from Fig. 6.22(b) N atoms which do not experience V_{Ga}'s on their nearest neighboring sites ($n_{V_{Ga}} = 0$) have an electronic structure close to the one in an ideal GaN crystal (see Fig. 6.2).
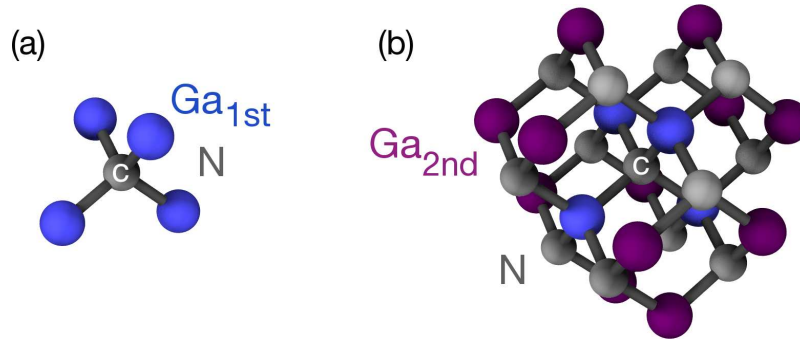
Figure 6.23.: *(a) Local configuration around a N atoms (gray marked with c) including nearest-neighboring Ga atoms (blue). In (b) the same illustration is shown as in (a) but the surrounding atoms around the central atom N are drawn up to the second nearest-neigbor Ga sites (violet).*

This picture distinctively changes for $n_{V_{Ga}} \neq 0$: By introduction of one V_{Ga} on one of the four nearest-neighboring sites the tetrahedral environment is destroyed. In terms of molecular orbitals the binding mechanism cannot anymore be described by sp^3 but rather by sp^2 orbitals. Alongside the three sp^2 orbitals one nonbonding p_z-orbitals, usually called a 'dangling' bond state, completes the new basis set. The fact that sp^2 orbitals connect rather in a plane than tetrahedrally causes the large outward relaxations of N atoms observed by Neugebauer and van de Walle [141]. Although these relaxations are disregarded in our calculations this simple molecular model allows to interpret how the DOS of our calculations is affected by V_{Ga}'s. For $n_{V_{Ga}} = 1$ in Fig. 6.22(c) almost all sp^2 states are occupied and degenerate in both spin channels. On the contrary, the non-bonding states split leaving the minority states partly unfilled. Considering less frequent configurations in the supercell of more than one V_{Ga} neighboring an N site, the fraction of non-bonding states increases resulting in an enhanced spin-polarization. This trend culminates for $n_{V_{Ga}} = 3$ as charted

in Fig. 6.22(e). Here, the fraction of non-bonding states forms a larger part of the total DOS than the actual bonding states and both types of nitrogen states are clearly separated in energy − even in the majority band.
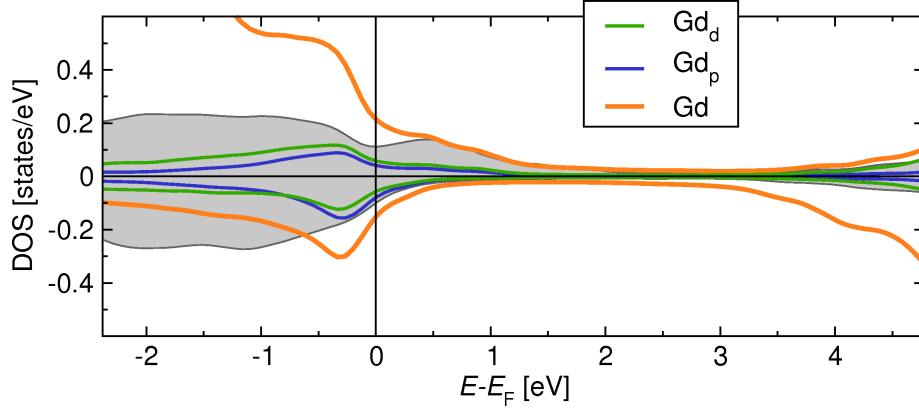

Figure 6.24.: *Orbital resolved local density of states (DOS) on Gd sites in $Ga_{220}N_{256}Gd_4$. Gray filled areas stand for the total DOS of the system per atom and negative and positive DOS values represent the minority and majority spin channel.*

Via the previous discussion we have obtained an intuitive picture on the electronic structure on N sites being in the vicinity of Gallium vacancies. Another important question is how Gd is embedded in the magnetic coupling mechanisms of the supercell. Here, it is worthwhile noting that V_{Ga}'s do induce spin-polarization of similar strength even without the presence of Gd atoms in the crystal as our calculations of smaller unit cells revealed in accordance to results by Dev *et al.* [116]. The question on the role of Gd is anyhow addressed in an elaborate discussion by Liu *et al.* [51]. In agreement with this article we obtain that the main peak of gadolinium d states is located close to the Fermi energy (Fig. 6.24). In addition we find a considerable contribution of p states in the same energy range, which are induced by hybridization to the surrounding N sites. Before the impact of those states on the magnetic coupling will be discussed we will first look at the distribution of magnetic moments throughout the supercell.

6.4.2 Magnetic moments

The statistical distribution of the magnetic moments of Gd, Ga, V_{Ga} and N sites is summarized in Fig. 6.25(a). Due to the clear separation of the gadolinium f states and the almost complete occupation of the hybridization induced p- and d states of the valence band a uniform magnetic moment close to 7 μ_B is found on Gd sites. Although the Fermi energy is pushed slightly into the valence band, both spin-channels at Ga sites remain nearly degenerate with an average magnetic moment of 0.01 μ_B. On the contrary on all V_{Ga} sites a magnetic moment of about 0.07 μ_B is formed which cannot be neglected a priori. The most important deviations are visible on the N sites: The magnetic moment exhibits no smooth distribution but rather discreet values. With increasing number of vacancies on the surrounding tetrahedral sites, $n_{V_{Ga}}$, the absolute values of the magnetic moments stepwise rise from 0 μ_B via 0.5 μ_B and 1.2 μ_B up to 1.8 μ_B (best visible from Fig. 6.25(b)). An-
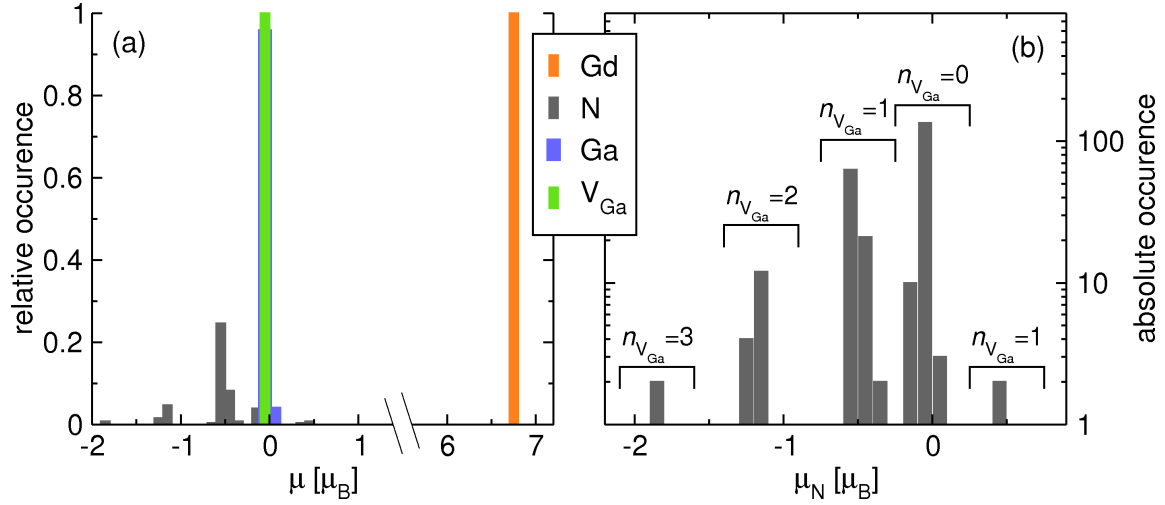
Figure 6.25.: (a) Site-specific relative probability of size and sign of magnetic moments in $Ga_{220}N_{256}Gd_4$ with 36 V_{Ga} in the unit cell. To obtain the relative probability the occurrence in all bins have been normalized by the total number of sites of the same type. (b) Absolute occurrence of magnetic moments on nitrogen sites, μ_N, on a logarithmic scale. Based on the size of μ_N N sites can be conceptually divided into four groups according to the presence of $n_{V_{Ga}} = 0, 1, 2, 3$ Ga vacancies on the four nearest-neighbor Ga sites.

other important feature appears considering the concentration of V_{Ga}'s in the sample. 32 V_{Ga}'s representing a vacancy concentration of 12.5% introduce a non-zero sizable magnetic moment of roughly $120-$ or in other words almost half of all $-$ N sites. This means, relatively small defect concentrations can originiate an almost collective spin-polarization in the sample.

In order to highlight this effect, the distribution of magnetic moments is shown in real space representation in Fig. 6.26. This viewgraph indeed underlines that in part of the supercell a collective spin-polarization can be observed while other parts containing no vacancies remain non-polarized. This phenomenon poses the question on the magnetic coupling in those spatially confined spin-polarized regions, which will be addressed in the next subsection.

6.4.3 Magnetic exchange interactions

From the analysis of the spin-polarization throughout the supercell it became clear that the N sites in the vicinity of V_{Ga}'s play a key role for magnetic coupling. Therefore we will focus on the magnetic exchange coupling involving N sites, while we first briefly comment on the remaining types of interactions[11].

All interactions involving Ga sites remain insignificant due to their nearly non-polarized electronic structure which leaves us with the exchange between Gd-Gd and Gd-V_{Ga}. With 1.5% Gd doping in the supercell the Gd-Gd coupling is clearly of secondary importance. At the few interatomic distances around one lattice constant which are realized in the

---

[11]For the qualitative discussion all exchange coupling constants are generated by adding all contributions $j_{ij}(E)$ below E_F.
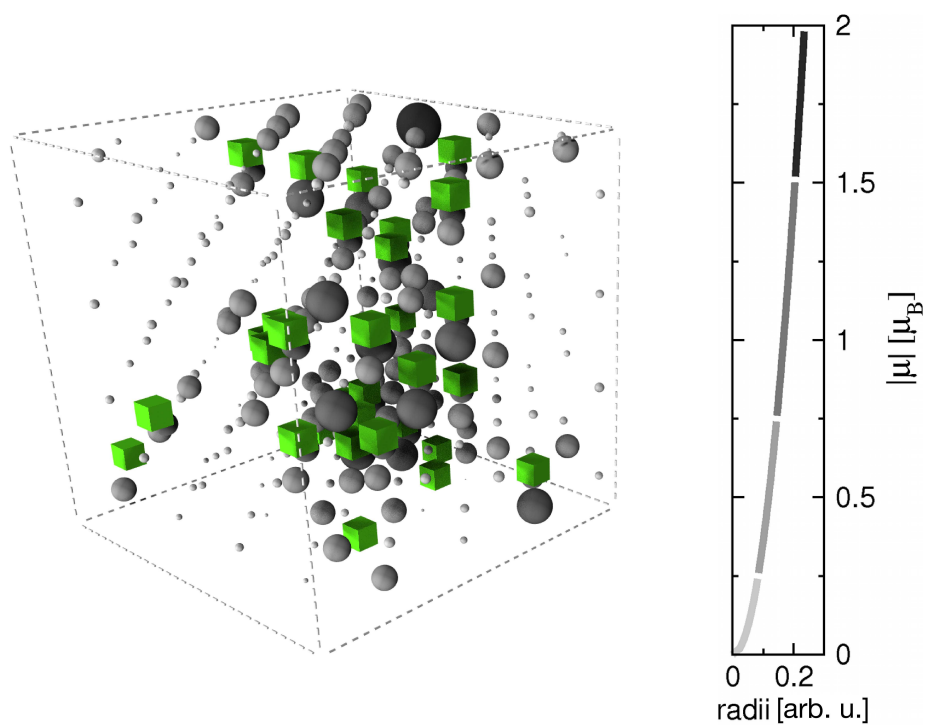
Figure 6.26.: *Real space visualization of absolute values of local magnetic moments ($\|\mu\|$) on N sites (gray), where the color coding and radius of the spheres corresponds to the size of magnetic moment (see legend). Green boxes illustrate V_{Ga}.*
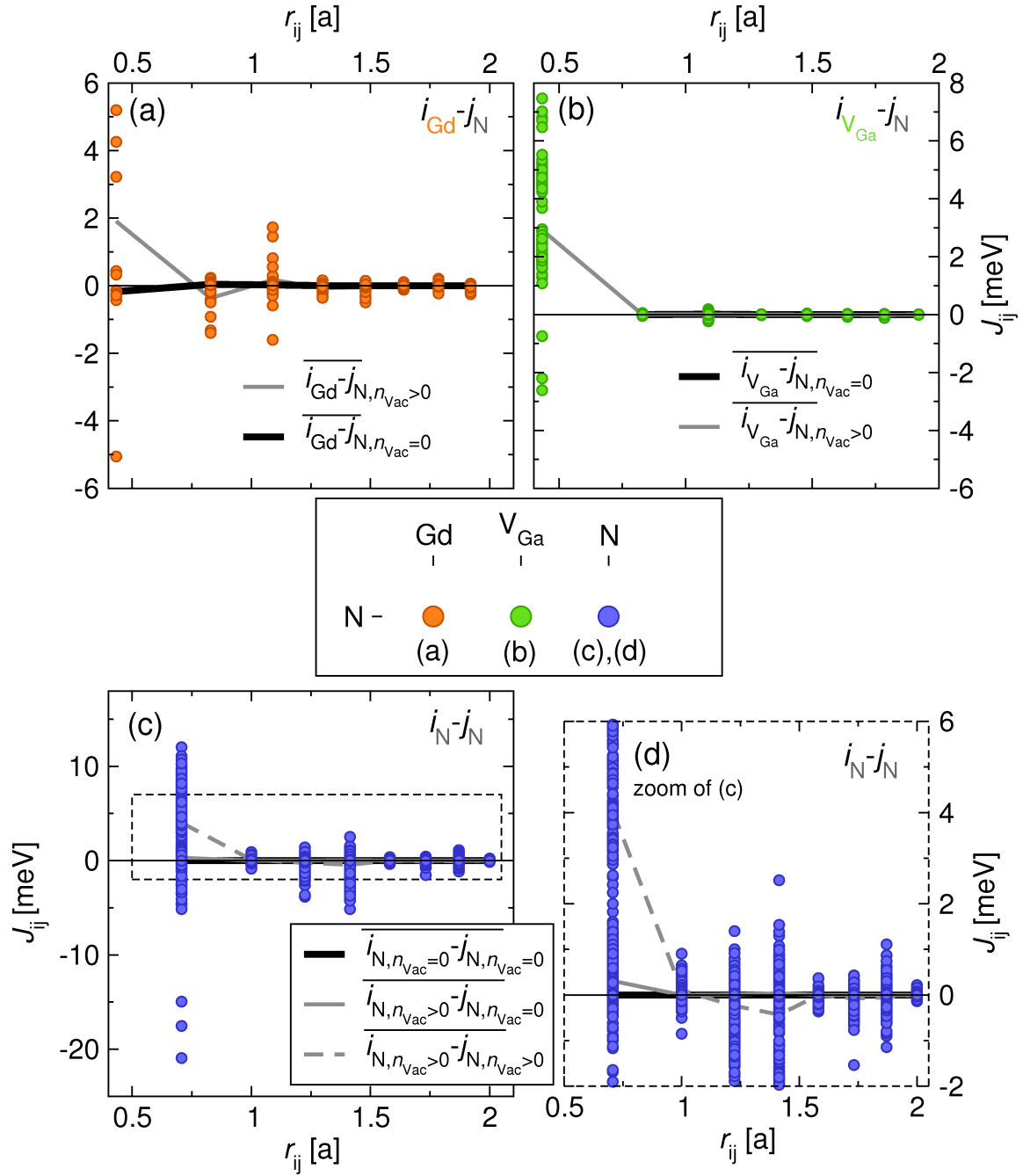
Figure 6.27.: *Magnetic exchange interactions J_{ij} between (a) N and Gd, (b) N and V_{Ga} and (c),(d) N and N sites for $Ga_{220}N_{256}Gd_4$. In addition to the full set of data of all interacting atoms of given kind, the mean exchange interaction to and amongst two groups of N atoms are shown (thick lines). In (a) and (b) the mean exchange constants to atoms without ($n_{Vac}=0$) and with at least one nearest-neighbor Ga vacancy ($n_{Vac}>0$) are shown in black and gray, respectively. Accordingly, in (c) and (d) the mean interaction between N atoms with $n_{Vac}=0$ and $n_{Vac}=0$, $n_{Vac}=0$ and $n_{Vac}>0$ as well as $n_{Vac}>0$ and $n_{Vac}>0$ are displayed in black, gray, and gray dashed lines, respectively.*

supercell, a weak antiferromagnetic coupling of less than 0.2 meV appears. Better statistics is available for the Gd-V_{Ga} exchange, which turns out to be on the order of 0.2 meV on the nearest-neighboring shell and well below 0.02 meV for all longer-range distances.

The exchange coupling constants affecting N sites, i.e. Gd-N, N-N and N-V_{Ga}, are shown in Fig. 6.27. Conceptually we will distinguish two types of N sites − the ones being almost non-magnetic having no V_{Ga} on the neighboring shell (n_{Vac}=0) as well as the clearly spin-polarized ones with at least one V_{Ga} on the adjacent sites (n_{Vac} >0). Apparent from the mean values presented in Fig. 6.27(a)-(c) is that all interactions with N atoms having n_{Vac}=0 are almost identical to zero. The presence of at least one V_{Ga} clearly changes this picture: The Gd-N interaction appears to be short-range ferromagnetic and on the order of 4 meV for the nearest neighbor distance $r_{ij} = 0.433a$ regarding N-sites with n_{Vac} =1. An exception is the coupling of Gd to a N site surrounded by three V_{Ga}'s showing a clear preference for antiferromagnetic ordering. However, this single configuration can be regarded as unphysical and by that as an exotic artefact of the randomly occupied supercell. Besides this interaction N sites and vacancies possess as displayed in Fig. 6.27(b) on a short-scale a considerable ferromagnetic coupling of on average 3 meV while for all further distances the interaction is identical to zero.

The largest deviations and absolute values of coupling constants are appearing for the N-N interaction (see Fig. 6.27(c) and magnified in Fig. 6.27(d)). It becomes clear that the presence of adjacent vacancies is crucial to create significant exchange interactions: Even when one N with n_{Vac} >0 and one N with n_{Vac} =0 are involved the J_{ij}'s remain well below 1 meV. On the contrary a large mean ferromagnetic interaction of 4-5 meV can be observed amongst the group of n_{Vac} >0. It is important to clarify that most of the deviations from the mean value ranging from -20 meV to 12 meV are originated by configurations of n_{Vac} ≥1. The most relevant class of interacting N sites with n_{Vac} =1 shows always a ferromagnetic coupling of on average 4.25 meV with a standard deviation of about 2.6 meV. This relatively large standard deviations are caused by the fact that this group of interactions can actually be subdivided in two geometrically different groups. The difference between those groups is whether the site which is placed on the tetrahedral lattice almost between the interacting N atoms is occupied by a Ga or a vacancy. Accordingly, the former and latter group is named 'off V_{Ga}' and 'via V_{Ga}', respectively. Considering the average exchange interactions shown in Tab. 6.2, it turns out that interactions across vacancies (via V_{Ga}) are by a factor of three larger than the ones across Ga sites.

Besides this important effect, the data which are presented in Tab. 6.2 enable us to address some issues regarding the computational treatment. First, although a ferromagnetic alignment of Gd and N spins is preferential according to the evaluated exchange interactions, in the self-consistency procedure the system converged to the metastable antiparallel configuration. For the sake of consistency we have converged the same system again after aligning all spins. As can be seen from Tab. 6.2 the results on J_{ij}'s of all kind of interactions for both calculations are in good agreement. As a general rule we observe that for the parallel as compared to the antiparallel spin alignment about 5% larger values emerge. Secondly, the results evaluated by two different energy integration contours are contrasted. These test calculations reveal that the integration along the real axis without the inclusion of Matsubara poles lead J_{ij}'s which show the same trends but are reduced by roughly a factor of two. Thirdly, we contrast the mean exchange interactions obtained from larger

	align.	N-N			Gd-N
		all	via V_{Ga}	off V_{Ga}	
$\mathrm{Ga}_{220}\mathrm{N}_{256}\mathrm{Gd}_4$	$\mathrm{Gd}_\uparrow\,\mathrm{N}_\downarrow$	4.29 meV	7.11 meV	2.12 meV	4.23 meV
		2.80 meV^R	-	-	1.58 meV^R
	$\mathrm{Gd}_\uparrow\,\mathrm{N}_\uparrow$	4.56 meV	7.45 meV	2.34 meV	4.49 meV
$\mathrm{Ga}_{24}\mathrm{N}_{32}\mathrm{Gd}_4$	$\mathrm{Gd}_\uparrow\,\mathrm{N}_\downarrow$	3.26 meV	6.58 meV	1.27 meV	-
	$\mathrm{Gd}_\uparrow\,\mathrm{N}_\uparrow$	3.35 meV	6.72 meV	1.33 meV	-
		2.20 meV^R	-	-	-
$\mathrm{Ga}_{28}\mathrm{N}_{32}$	$\mathrm{N}_\uparrow$	2.63 meV	6.71 meV	0.18 meV	-

Table 6.2.: *Mean exchange interactions between N and N as well as Gd and N as obtained by the Lichtenstein-formula for different system sizes and spin alignments. $Gd_\uparrow\,N_\uparrow$ and $Gd_\uparrow\,N_\downarrow$ label a parallel and antiparallel orientation of Gd and N spins. The exchange interactions amongst N atoms are restricted to N sites with precisely one V_{Ga} on the next neighboring site. This group of interactions can be further split into two subgroups: on the one hand interacting sites having a vacancy in between (via V_{Ga}) and on the other hand where a Ga atom is present on the intersite (off V_{Ga}). A schematic view on both groups is provided in Fig. 6.29(a). For part of the configurations the integration is performed with two different contours both starting at $E-E_{\mathrm{F}} = -49\ eV$: While for both a temperature broadening of 700 K has been applied, either 4 Matsubara poles with overall 37 energy points or 250 energy points along the real axis without additional Matsubara poles have been selected. Values obtained by the latter option are marked with ^{R}.*

(1024 sites) and from smaller (128 sites) supercells. From both calculations the same favorable spin configurations can be deduced, but in small cells the strength of the interactions is underestimated due to finite-size effects or insufficient statistical sampling.
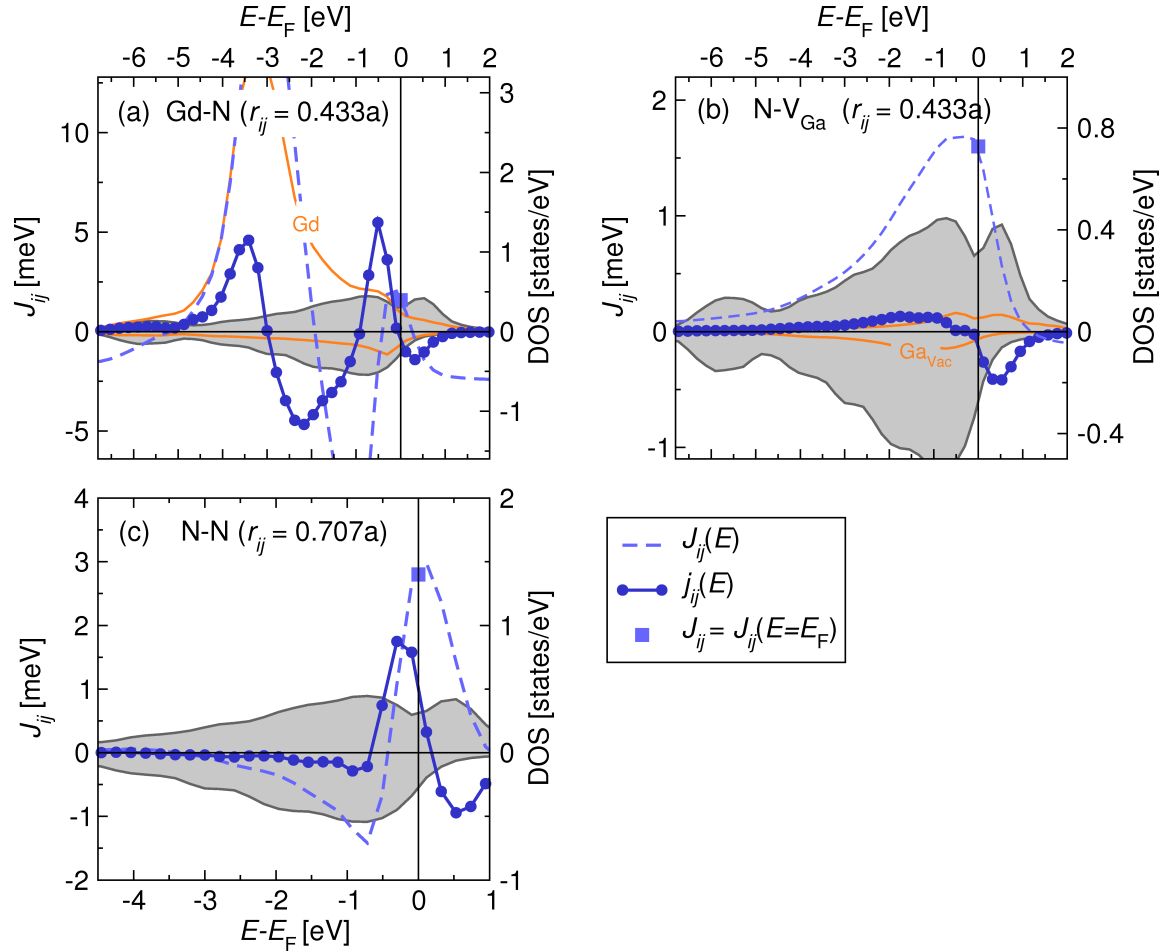
Figure 6.28.: *Energy resolved mean exchange interactions $j_{ij}(E)$, its integration up to E, $J_{ij}(E)$ and its final value J_{ij} for interactions to or from N sites with $n_{Vac} = 1$ in $Ga_{220}N_{256}Gd_4$. For the sake of comparison the mean DOS of N with $n_{Vac} = 1$ is displayed as gray shaded area in (a)-(c). Additionally, in (a) and (b) the mean DOS of Gd and V_{Ga} is shown in orange lines. All quantities have been obtained with a temperature-broadening of 700 K and 250 energy integration points.*

The previously presented tests validate to continue the qualitative discussion of the energy resolved exchange interactions on the basis of the electronic structure of antiparallel aligned Gd and N spins and by using the integration along the real axis. In Fig. 6.28 the energy-resolved exchange interactions of Gd-N, N-N and N-V_{Ga} are shown for the deciding interatomic distances. The following discussion is based on the fundamental concepts of exchange mechanism as derived in section 6.2.3. The ferromagnetic N-N exchange (Fig. 6.28(c)) builds up in the region of E_F-1 eV to E_F due to band broadening and can therefore be attributed to the double-exchange mechanism. In contrast to this the energy dependency of the N-V_{Ga}

exchange is crucially different since the main contribution cumulates well below E_F. In fact, the DOS of V_Ga's follows one to one the DOS of the surrounding N atoms. Hence, states being located on V_Ga's can rather be classified as extensions to the states on N atoms. Obviously, the N-V_Ga interaction is effectively an additional contribution to the N-N interaction mediated by its extended states. Inspection of Fig. 6.28(a) reveals the exchange mechanisms of the Gd-N interactions for $r_{ij} = 0.433a$. Beginning with E_F-5 eV a characteristic oscillation of $j_{ij}(E)$ can be observed switching from positive to negative and closer to E_F back to positive contributions. Considering the orbital decomposed DOS on Gd sites in Fig. 6.24 the interplay of gadolinium d states and non-bonding states on the N sites is decisive for the Gd-N interaction. Here, double-exchange exceeds the superexchange between those states which causes the short range ferromagnetic tendency.

It remains the task to compare our findings with existing theoretical work [51, 52]. Both of them are based on total energy arguments and report in agreement with our calculations on a preferential ferromagnetic alignment of V_{Ga}'s. In addition they reveal a ferromagnetic (FM) coupling between Gd and the vacancy complexes. On the one hand Liu *et al.* [51] obtained a strong preference of the ferromagnetic alignment by on the order of $\Delta_\mathrm{FM-AFM}$ =500 meV. On the other hand Gohda and Oshiyama [52] found $\Delta_\mathrm{FM-AFM} \approx 40$meV. Although the absolute values of energy differences obtained in [51] are clearly larger, the overall trend of the coupling of Gd and vacancies is in both studies conform to our results. However, keeping in mind the short-rangeness of the Gd-N and Gd-V_Ga interactions, except in the limit of large Gd concentrations no crucial impact of the Gd-V_Ga and Gd-N interaction can be expected. This holds in particular regarding that the N-N interaction is considerably larger than the ones incorporating Gd, which we find in agreement with Gohda and Oshiyama [52]. Therefore we will restrict the following discussion to the experimentally interesting dilute limit by explicitly not considering the presence of Gd but rather focusing on the ordering of Ga vacancies amongst each other.

6.4.4 Ferromagnetic network

The pronounced ferromagnetic coupling of spin-polarized nearest-neighbor N atoms leads us directly to two questions of statistical nature: What is the minimal concentration of V_Ga's which enforces not a superparamagnetic but a ferromagnetic network? And, up to which temperature can such a network resist thermal fluctuations?

To address the first question the consideration of purely geometrical arguments is sufficient. The nearest neighboring ferromagnetic coupling of N atoms which have at least one V_Ga on the adjacent sites can be translated into an effective coupling of two V_Ga: From Fig. 6.29(a) it becomes clear that this effective coupling ranges up to the 4th nearest neighboring Ga site at a distance of $\sqrt{2}$a. Based on this input we have calculated the percolation threshold for ferromagnetic coupling mediated by N atoms using the following scheme: For a each of concentration of dopants a series of supercells is randomly set up. In a second step it is probed how many of the dopands are magnetically coupled to each other and form a cluster. The magnetic moment of the largest of those coupled clusters is taken as measure for the magnetization of the sample. Since many of those supercells are numerically probed at fixed concentration, statistical fluctuations of the cluster magnetic moment can be evaluated. The largest statistical fluctuations of this cluster moments then determine
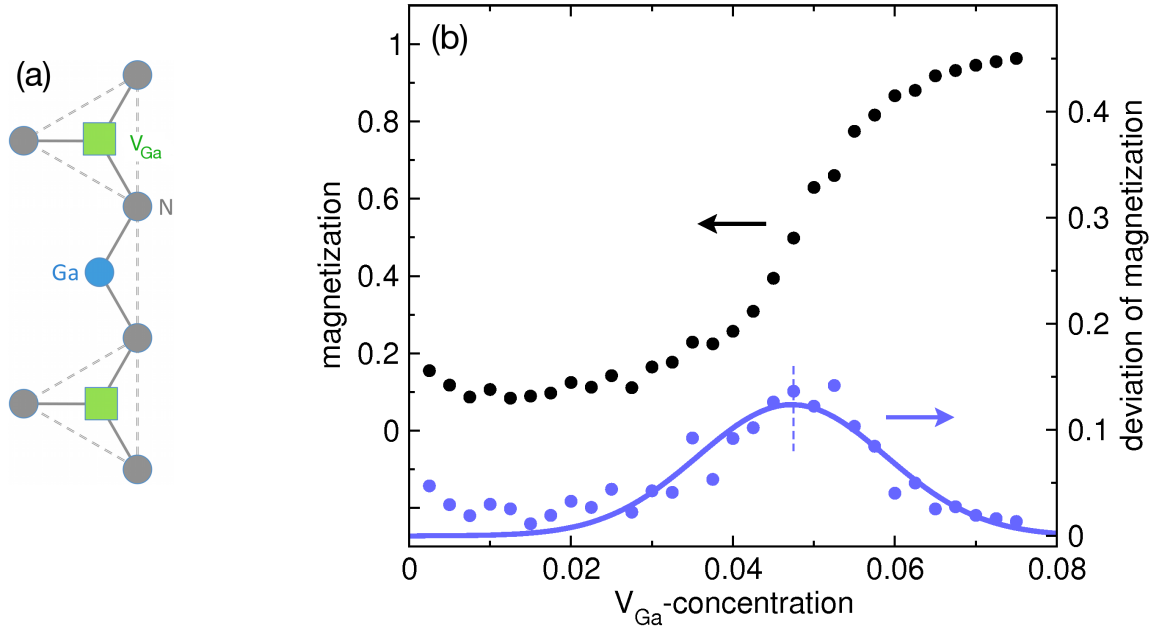
Figure 6.29.: (a) Two-dimensional projection of tetrahedral environment of N atoms in presence of gallium vacancies. Here, all N atoms are coupled assuming a ferromagnetic alignment if at least one V_{Ga} is on the neighboring site. This coupling results in an effective interaction distance of two V_{Ga} of $\sqrt{2}a$ i.e. the distance of the 4th nearest neighbor in the fcc lattice. (b) Normalized magnetic moment (magnetization) of the largest magnetically coupled cluster (black) and deviation of this value (blue) as obtained within a statistical approach to determine the percolation threshold (see appendix A for details). Here, exclusively the coupling between V_{Ga}'s which have a distance of less or equal $\sqrt{2}a$ has been taken into account in a periodically repeated supercell of $Ga_{4000}N_{4000}$. All statistical analysis is performed as a function of the concentration of V_{Ga}. The deviation of magnetization has been fitted to a Gaussian-distribution (solid line) and its maximum (dashed line) corresponds to the percolation threshold.

the percolation threshold[12]. As shown in Fig. 6.29(b) the percolation threshold is reached for roughly 4.8% V_{Ga} doping on Ga sites. This value is consistent to calculations of the percolation threshold in the fcc lattice by Sato *et al.* [46], which for interactions up to fourth neighbors give a percolation threshold of about 5%. For comparison with these data we interpret the N-N interactions as explained above in Fig. 6.29(a) as an effective V_{Ga}-V_{Ga} interaction, which indeed reaches up to four neighbors the Ga fcc lattice. Interestingly, our choice of composition in the *ab initio* supercell, which is 32 V_{Ga} on 256 Ga sites, is already well above the percolation threshold, which explains why almost all spins induced by V_{Ga}'s are aligned in parallel.
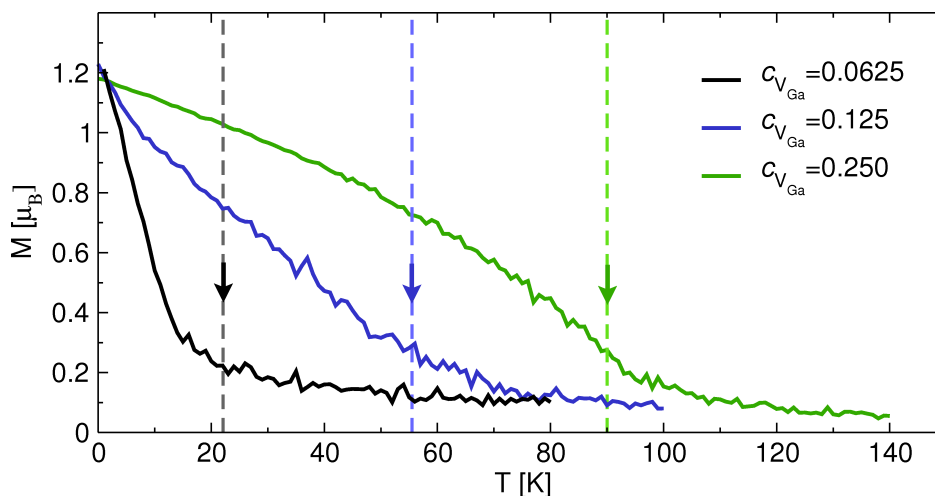
Figure 6.30.: *Temperature dependency of the magnetic moment per V_{Ga} in periodically repeated zincblende supercells with 16000 sites as obtained by Monte-Carlo simulation of coupled V_{Ga} for different vacancy concentrations $c_{V_{Ga}}$. For three concentrations $c_{V_{Ga}}$ 15 different Monte-Carlo simulations with 500 Monte Carlo sampling steps have been conducted. Shown is the mean magnetization of all 15 simulations. Arrows and vertical lines indicate the Curie-temperatures (T_C) for the all three $c_{V_{Ga}}$, which are defined as the temperature of largest relative fluctuation of the magnetization amongst all 15 simulations. For further details on see explanation in text.*

For a statistical analysis of the temperature dependency of the ferromagnetic state we performed Monte-Carlo simulations exploiting the Metropolis algorithm [142]. Focusing on the dilute limit of Gd doping we considered exclusively the mean N-N and N-V_{Ga} exchange interactions of shortest range. Additionally, we restrict ourselves in this model calculation to the most important group of N sites with one V_{Ga} on the adjacent sites ($n_{Vac} = 1$). This means if N sites with $n_{Vac} > 1$ occur they are treated anyhow as sites with $n_{Vac} = 1$. Under these assumptions the average strength of interaction is obtained from the results for the parallel aligned Gd and N spins in Tab. 6.2 which are 7.45 meV for the N-N coupling via a vacancy ($r_{ij} = \frac{\sqrt{2}}{2}a$), 2.34 meV for a coupling via a gallium site ($r_{ij} = \frac{\sqrt{2}}{2}a$) and 2.09 meV for the N-V_{Ga} interaction ($r_{ij} = \frac{\sqrt{3}}{4}a$) (see Fig. 6.27 for comparison). For the Monte-Carlo simulations zincblende supercells with 16000 sites and accordingly 4000 Ga sites containing

---

[12]For a detailed description of this scheme the reader is referred to appendix A.

randomly placed V_{Ga} with a concentration of $c_{V_{Ga}} = 0.0625$, $c_{V_{Ga}} = 0.125$ and $c_{V_{Ga}} = 0.25$ are constructed. For the sake simplicity we assume that the magnetization of N sites is 0.5 μ_B if one or more neighboring V_{Ga} are present and 0.0 μ_B otherwise. Additionally, a universal magnetization of all V_{Ga} sites of 0.1 μ_B is assumed. Then the spins of N's and V_{Ga}'s are coupled if geometrically allowed by the two average exchange interactions as mentioned above.

The results of the Monte-Carlo simulations are summarized in Fig. 6.30. In this viewgraph the magnetization per V_{Ga} defect (M_{Vac}) is presented which is composed of V_{Ga} and N magnetic moments. All calculated configurations are above the percolation threshold of $c_{V_{Ga}} \approx 4.8\%$ and accordingly show a finite magnetization at $T = 0$ K. It turns out that at zero Kelvin M_{Vac} has the same size for all concentrations. This is the result of two competing effects. The simplification that each N atom adjacent to V_{Ga}'s has a magnetic moment of 0.5 μ_B leads to a reduction of M_{Vac} due to N sites which experience more than one V_{Ga} on the nearest-neighboring shell. Instead of four N site which can be fully accounted to each V_{Ga}, only 3.6, 3.3, and 2.7 spin-polarized N sites exist in the generated supercells for $c_{V_{Ga}} = 0.0625$, $c_{V_{Ga}} = 0.125$ and $c_{V_{Ga}} = 0.25$, respectively. On the contrary the smaller the concentration of V_{Ga} is, the more V_{Ga}'s are not coupled to the main cluster of V_{Ga}'s, which of course reduces the statistically analyzed magnetic moment.

Increasing the temperature the ferromagnetic ordering in samples of higher concentration has a significantly higher stability with respect to thermal fluctuations. This is caused by the fact that in those supercells of higher $c_{V_{Ga}}$ different subclusters of V_{Ga}'s are coupled not by one but by multiple exchange interactions. Consequently, this redundancy leads to a significantly higher Curie temperature of $T_C = 90.0$ K for $c_{V_{Ga}} = 0.25$ while the ferromagnetic network collapses already at $T_C = 22.1$ K for $c_{V_{Ga}} = 0.0625$.

Considering the assumptions used in the Monte-Carlo approach and the further approximations in the first-principles calculations − such as neglecting relaxations − it cannot be the aim to predict accurate T_C's for this system. However, this does not affect our qualitative results which led us to an important new insight: The short-range coupling between spin-polarized N sites can lead − already for relatively small vacancy concentrations − to a distinct and thermally stable ferromagnetic ordering. By that finding we revealed that V_{Ga}'s − especially if they are clustered − are a potential candidate to explain both striking experimentally observed features: room-temperature ferromagnetism and colossal magnetic moments [1, 112].

Beyond this key result of this chapter, it is interesting to discuss how further experimentally observed attributes of GaN:Gd could be integrated in the V_{Ga} model. Reconsidering the distribution of magnetic moments in the unit cell as shown in Fig. 6.26 leads to an interesting conclusion: the distribution of V_{Ga}'s can actually be highly inhomogeneous and still originate a ferromagnetic order although part of the GaN matrix is left magnetically inactive. This opens room to form conjectures: The magnetic order in GaN:Gd can e.g. potentially be supported by high vacancy concentrations at extended defects such as grain boundaries. Another possibility is that clustered vacancies are induced along the pathways of Gd ions during the ion bombardement process which is used in part of the experimental studies. This is an ansatz which could explain the comparative studies reported by Dhar *et al.* [143] revealing significantly larger magnetic moments in Gd implanted samples than for expitaxially grown layers. This hypothesis is compatible with further experimentally

observed attributes: Roever *et al.* [138] reported on a large fluctuation of the magnetization amongst a series of samples grown with the same preparation scheme. Assuming that the magnetic coupling is created by the network of one or two dimensional extended defects, statistical fluctuations in the local geometry likely result in strong deviations of the macroscopic properties. Such a network is also highly sensitive to thermally induced healing processes of defects, which might explain the drastic decrease of the magnetization on the time-scale of months [138]. Besides these effects it is important to regard for the experimentally measured strikingly high resistance in GaN:Gd [144]. In contradiction, from the relatively high values of density of states at the Fermi level obtained from our calculations with V_{Ga}'s, rather low resistances would be expected. Again, the proposed scenario of vacancy complexes in spatially selected areas is conform to this finding, since the remaining large undoped parts might lead to transport properties which are close to the ones for pure GaN.

6.5 Summary

In this chapter we addressed the experimentally raised and still unresolved question on the origin of high-temperature ferromagnetic coupling in gadolinium doped gallium nitride [1]. By utilizing the capability to describe thousands of atoms per unit cell with KKR**nano** we performed extensive studies on the magnetic coupling of nitrogen and oxygen interstitials as well as gallium vacancies. Here, the large scale of the treated samples has allowed for an explicit and elaborate analysis of the electronic structure and revealed a significant magnetization of the defects. We gained important new insight in the magnetic coupling amongst defects and between defects and gadolinium by utilizing the Lichtenstein formula. Further we did not only identify ferro- and antiferromagnetic coupling trends but we were able to go conceptually one step beyond: we discovered the underlying coupling mechanisms by arguing on the basis of the energy resolved contributions to the exchange coupling. Our studies clearly revealed that nitrogen interstitials show a distinct antiferromagnetic coupling and therefore can be ruled out as origin for the experimental observations. Also oxygen interstitials at most only provide a weak ferromagnetic coupling which cannot be the determining factor for the magnetic order. For gallium vacancies we identified a different mechanism of coupling. The vacancies induce large magnetic moments on all surrounding nitrogen sites, which then couple strongly ferromagnetically both amongst themselves and to the gadolinium dopants. Based on the information on extracted exchange coupling constants we evaluated by statistical methods that already small concentrations of vacancies can lead to distinct ferromagnetic ordering. Hence, our calculation revealed strong indications that gallium vacancies are causing the ferromagnetic coupling of colossal magnetic moments in GaN:Gd.

Disorder and localization in $Ge_1Sb_2Te_4$

A very successful realization of the high demands for fast rewritable storage media with persistent bits have been achieved in phase-change-materials (PCMs). Ternary alloys in different composition of germanium, antimony and tellurium (GST) take a key position in technical application and are routinely used in rewritable digital-versatile-discs (DVD) and Blu-ray^{TM}-discs. Here, the macroscopic functional procedure is similar for all used compounds: Due to different optical properties of the amorphous (A) and crystalline phase (C) locally confined bits in state A or C can be identified optically. Then laser pulses of different length and intensity can be functionalized to switch those bits from A to C and vice versa. On the one hand rapid heating above melting temperature by a short laser pulse of high intensity followed by rapid cooling forces the system locally into A. On the other hand the bit can be switched back to C by applying a long laser pulse of low intensity [53, 54, 145]. Remarkable is the speed of the switching process between the phases below one nanosecond [54]. This allows for todays standard writing and reading speed of tens of MB per second.
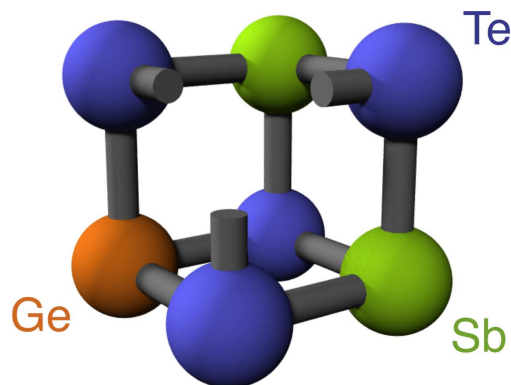
Figure 7.1.: *Schematical visualization of $Ge_1Sb_2Te_4$ in the rocksalt phase. The broken bonds represent the direction to the vacancy.*

While this technology has left the prototype status a decade ago, major steps in the microscopical understanding of e.g. the switching process and electronic structure have been taken only throughout the last years [146–151]. Recent experimental work by Siegrist *et al.* [55] focused on one of the most frequently used compounds $Ge_1Sb_2Te_4$ and revealed striking new transport properties of GST by a disorder-induced localization in the crystalline phase. Dependent on the annealing temperature applied to the sample the crystalline phase was found to show a drop in resistivity by several orders of magnitude. Thereby this compound is a promising candidate to combine optically and electronically probed and switched data storage. However, this recent finding raises the question on the origin of the effect, which we will address with large-scale density-functional calculations.

The ground state of $Ge_1Sb_2Te_4$ has been identified by Wełnic *et al.* [147] as a distorted rocksalt-structure, which conceptually can be understood as a superposition of two fcc-sublattices translated in [100]-direction by half of the lattice constant. The most common structural model for the rocksalt-structure of $Ge_1Sb_2Te_4$ is to assume that the first sublattice is completely occupied by Te, while on the second one Ge, vacancies and Sb atoms are randomly distributed (Fig. 7.1). In this crystalline phase distortions from the ideal structure mainly affect the Ge-Te bonds, which relax by 4 to 5% [147]. Those relaxations are remarkable small considering that 25% of the sites of the second sublattice are randomly distributed vacancies. Lencer *et al.* [150] observed that the presence of relaxations in fact impedes resonant p-bonding and thereby reduces the ability for phase-changing. Due to this intrinsic tendency for small relaxations, we will restrict ourselves to undistorted geometries in the following analysis.

7.1 Electronic structure
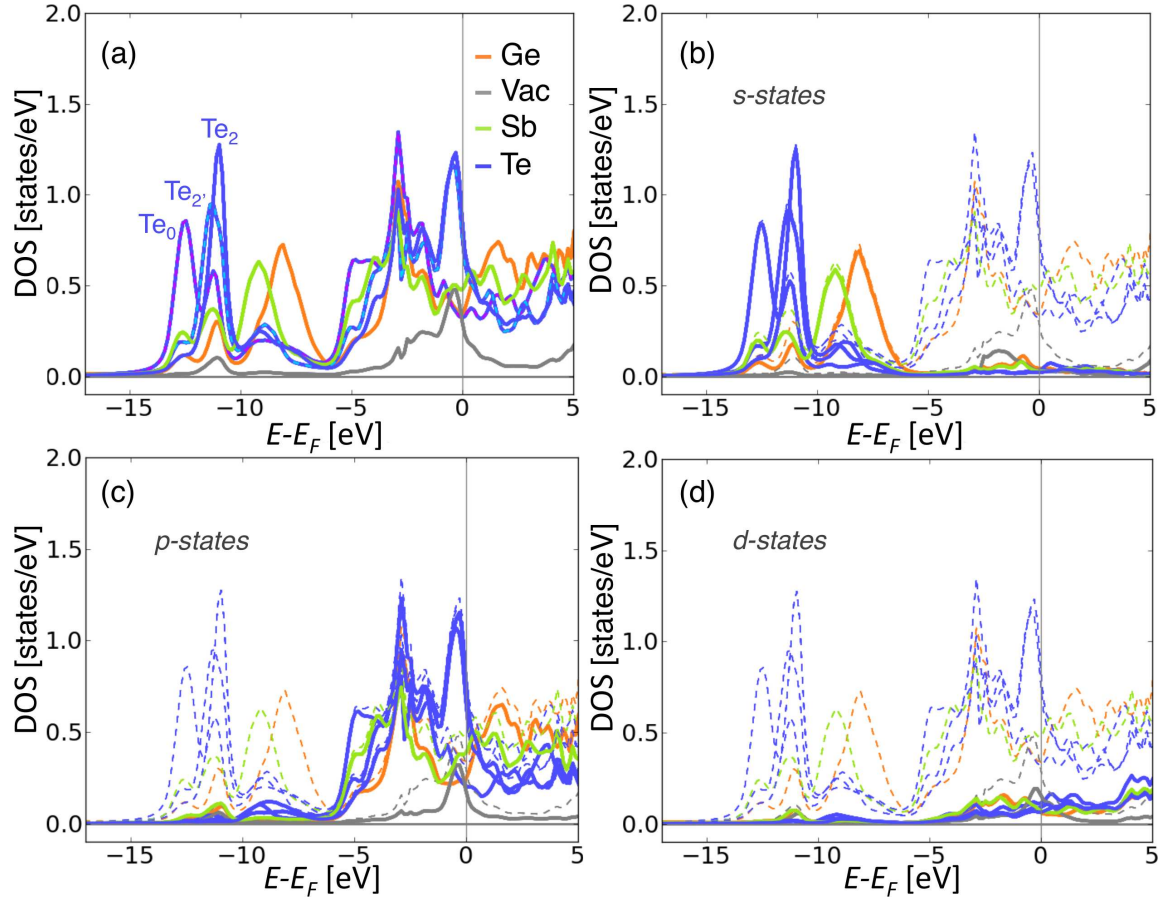
Figure 7.2.: *(a) Local density of states (LDOS) of the regular ordered $Ge_8Vac_8Sb_{16}Te_{32}$ supercell calculated with 64 k-points in Brillouin-zone and a temperature broadening of T=800 K. Te atoms are present in three neighbor configurations: zero vacancies, Te_0 (dashed-magenta), as well as two inequivalent with two vacancies, Te_2 and $Te_{2'}$ (dashed light blue) on next-nearest neighbor sites. In (b)-(d) the partial LDOS for s, p and d states (solid lines) are shown in relation to the total LDOS (dashed lines) for the same system as in (a). In all panels Ge, vacancy, Sb, and Te-contributions are shown in orange, gray, green, and blue, respectively.*

The electronic structure for a perfectly ordered crystal of $Ge_1Sb_2Te_4$ as calculated with KKRnano in LDA is shown in Fig. 7.2. This and all following calculations in this chapter have been performed with a lattice constant of $a = 6.04$Å, with an angular momentum cut-off of $l_{max}=2$ and 33 atoms in the reference cluster. The $4s$ and $4p$ states of Ge are included as valence bands in the energy contour while the lower lying states [Ar]$3d$ are treated as core states. Accordingly for Sb and Te the $5s$ and $5p$ states are accounted for as valence states with [Kr]$4d$ as core. For convergence, the energy integration has been performed with a temperature broadening of 1000 K over four Matsubara-poles and overall 24 energy points.

This set of parameters is used for all calculations throughout this chapter.

For the first calculation of a smaller unit cell which contains 64 atoms 64 k-points in the Brillouin-zone have been included. In this ordered 3D-periodic geometry all Ge, vacancy and Sb sites are equivalent, but due to the different composition of nearest neighbors the local density of states (LDOS) on Te atoms can be separated into three types: (i) Te_0, binding to no vacancies, two Ge and four Sb atoms ($n_{Vac}=0$, $n_{Ge}=2$, $n_{Sb}=4$); (ii) Te_2, with ($n_{Vac}=2$, $n_{Ge}=2$, $n_{Sb}=2$); and (iii) $Te_{2'}$, with ($n_{Vac}=2$, $n_{Ge}=0$, $n_{Sb}=4$). Those three groups of configurations show a significantly different LDOS − especially the localized s states and the states around the Fermi energy (Fig. 7.2(a)).

Focussing first on the s-contribution to the LDOS (Fig. 7.2(b)) a couple of features can be directly related to the compositional properties. Firstly, a clear chemical shift of the peak of the s LDOS is observed in relation to the composition of the neighboring atoms. The Te_0 s states occur at lower energy than the corresponding states on Te_2 and $Te_{2'}$. The lack of atoms on the neighboring sites acts as a repulsive potential on the Te sites and leads to a corresponding shift in energy. A second clear observation can be made about the $5s$ states of $Te_{2'}$ and Te_0 atoms, which have in difference to the Te_2 configuration two neighboring Ge atoms and show a clear hybridization to the slightly higher lying $3s$ states of Ge[1].
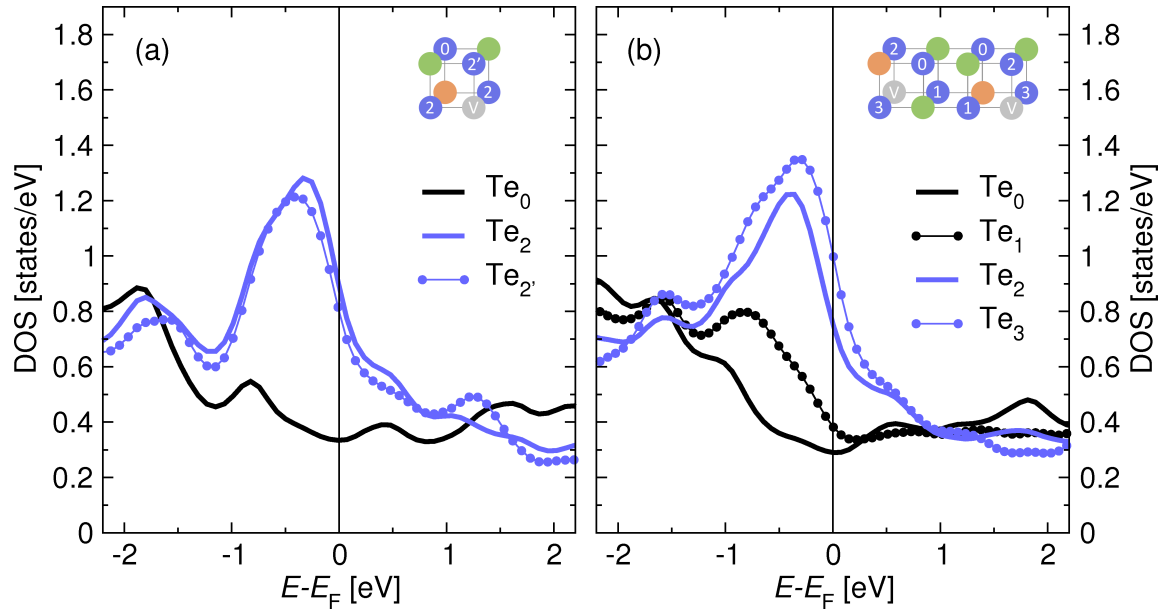
Figure 7.3.: *(a) Zoom in at the energy interval around E_F of the DOS on Te sites as shown in Fig. 7.2. The 64 atoms unit cell is constructed out of eight identical subblocks of eight atoms − one of those subblocks is shown schematically in the inset. Here, the color-coding is Ge (orange), Sb (green), vacancies (gray and labeled with 'V') and Te (blue). Te sites are labeled according to their number of nearest neighbor vacancies. (b) DOS of a 64 atoms unit cell, which consists of 4 subblocks of 16 atoms each. In the inset the local structure of one of those subblocks is shown in the same color-coding and labeling as described above for (a). All computational parameters are equivalent to the ones specified in Fig. 7.2.*

---

[1]Due to stronger relativistic mass enhancement the $5s$ states of Sb and Te experience a larger shift in energy relative to the $5p$ states as compared to the $4s$ states relative to $4p$ states.

As expected and shown in Fig. 7.2(d) the occupation of higher lying $4d$ and $5d$ states is small and does not play an important role. The electronic transport properties of the system are crucially determined by the states around the Fermi energy E_F, which are for $Ge_1Sb_2Te_4$ almost exclusively p states. For further evaluation the energy interval close to E_F is magnified and redrawn in Fig. 7.3(a). Three important features can be observed around E_F: First, we find in agreement with [147] that $Ge_1Sb_2Te_4$ exhibits no gap at the Fermi energy. Second, Te atoms with two vacancies in their vicinity (Te_2, $Te_{2'}$) reveal a peak in the LDOS at the Fermi energy, while Te_0 atoms have significantly lower values of LDOS. Third, on the vacancy sites a considerable p-type contribution to the total LDOS appears. Each of those trends holds also for one and three vacancies on the nearest neighbor sites to Te as proven by the DOS of a second regular structure shown in Fig.7.3(b).
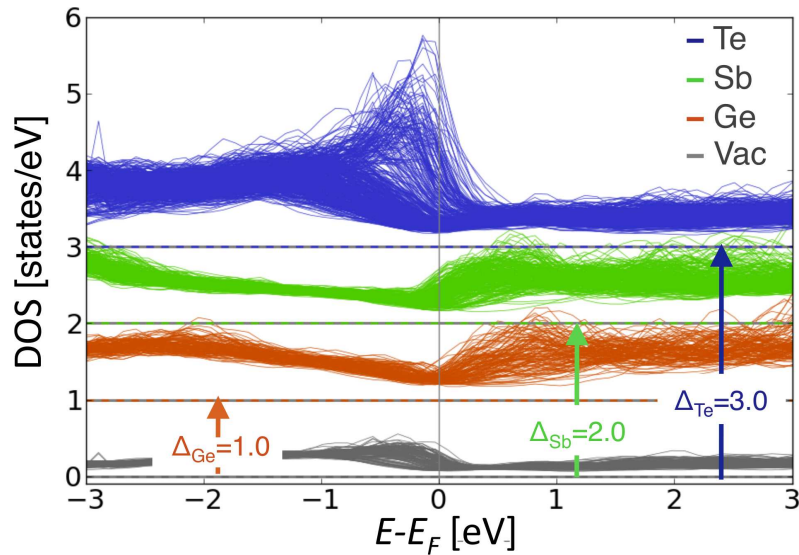
Figure 7.4.: *Local density of states (LDOS) of Ge, Va, Sb, and Te sites in a $Ge_{125}Vac_{125}Sb_{250}Te_{500}$ supercell with random distribution of Ge, Vac, and Sb on the sublattice obtained with a temperature broadening of $T = 400$ K. For the sake of visibility LDOS-values are shifted by 1.0, 2.0, and 3.0 eV for Ge, Sb, and Te, respectively. The new individual zero axes including the offset are shown as gray-colored dashed lines, where the color corresponds to the ones given in the legends.*

In order to obtain a qualitative picture for the variation of the LDOS around the Fermi energy a $Ge_{125}Vac_{125}Sb_{250}Te_{500}$ supercell with random atom distribution on the GeVacSb-sublattice has been calculated using KKR**nano** with the Γ-point k-space sampling. As displayed in Fig. 7.4 the LDOS of Ge, vacancy and Sb at the Fermi level (E_F) fluctuates in a range of 0.5 states/eV, while it can deviate on different Te atoms by up to 2.5 states/eV. A finding, which can be related to the fact that the 1st sublattice consists entirely of Te atoms while the 2nd sublattice has a random distribution of atoms: Hence, Te atoms experience different configurations of atoms on their nearest-neighbor site, while Ge, Vacancies and Sb are always surrounded by six Te atoms and see the randomness of the sample only by their next-nearest neighbor atoms. Another striking characteristic is that the occurrence of peaks around the Fermi energy varies significantly comparing Te and Vac with Ge and Sb sites.
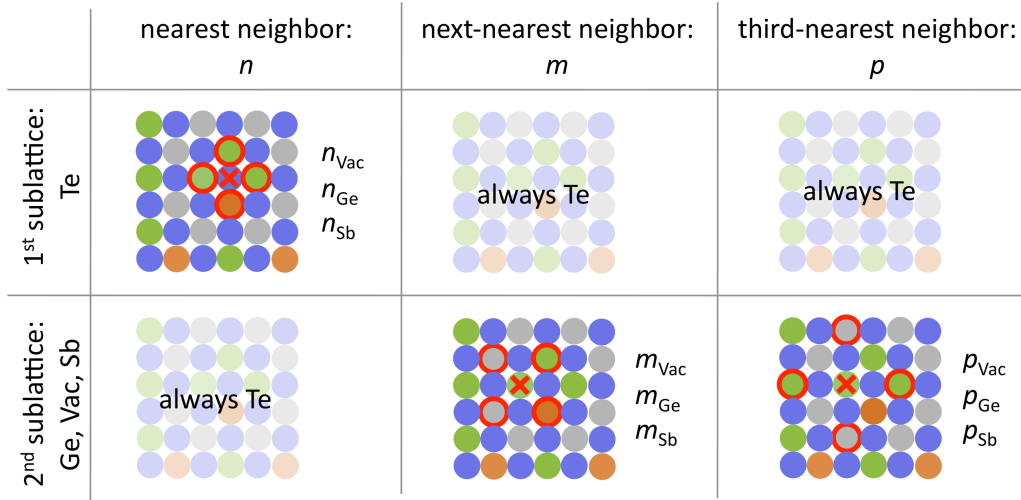
Figure 7.5.: *Two-dimensional projection of the rocksalt-lattice and its disorder on neighboring sites of both the 1st-sublattice and the 2nd-sublattice occupied with Te and with Ge, Vac, Sb, respectively. The occupation with Ge, Vac and Sb on the nearest neighboring, next-nearest neigboring and third-nearest neighboring sites is described by the set of parameters n_{Ge}, n_{Vac}, n_{Sb} as well as m_{Ge}, m_{Vac}, m_{Sb} and p_{Ge}, p_{Vac}, p_{Sb}, respectively.*

On the one hand Te sites and less pronounced also Vac sites possess a peak below E_F, on the other hand Ge and Sb sites exhibit a peak above E_F. Motivated by these findings we will address the issue of variation of the electronic states first on the Te sites and then on the 2nd sublattice in detail.

In order to reveal the origin of the observed variations in the electronic structure the results will be analyzed site-dependent with respect to the local chemical composition. For that purpose we define the parameter n_{Ge}, n_{Vac}, n_{Sb} and m_{Ge}, m_{Vac}, m_{Sb} and p_{Ge}, p_{Vac}, p_{Sb} to describe the occupation of the nearest, next-nearest and third-nearest neighbor sites, respectively − as visualized in Fig. 7.5.

On the basis of the number of neighboring vacancies (n_{Vac}) the LDOS of all Te atoms can be divided into several groups. This criterion discloses a clear trend for all partial LDOS values: Although relatively small in absolute value the s and d LDOS at E_F considerably decreases by increasing number of neighboring vacancies n_{Vac}. For the main and decisive part of the total LDOS, the p LDOS, the effect is reversed: The larger n_{Vac} the higher values for p LDOS at E_F appear. The mean values of the p LDOS nicely follow this trend and clearly reveal that the peak below E_F becomes sharper and shifts towards E_F with increasing n_{Vac}. Both effects are directly linked to the missing hybridization with Ge- or Sb- atoms in the presence of vacancies. Less hybridization with neighboring atoms directly corresponds to the sharpening of the peak. Further, due to the lack of bonding this electronic state resembled by the observed peak experiences a shift to higher energy. Beyond this nicely visible trends a considerable variation within a group of constant n_{Vac} remains, which will be discussed in the following.

To detect further "second order correlations" we take a closer look at the exemplary group defined by $n_{Vac} = 4$. Besides a large variety of minor effects induced by the individual

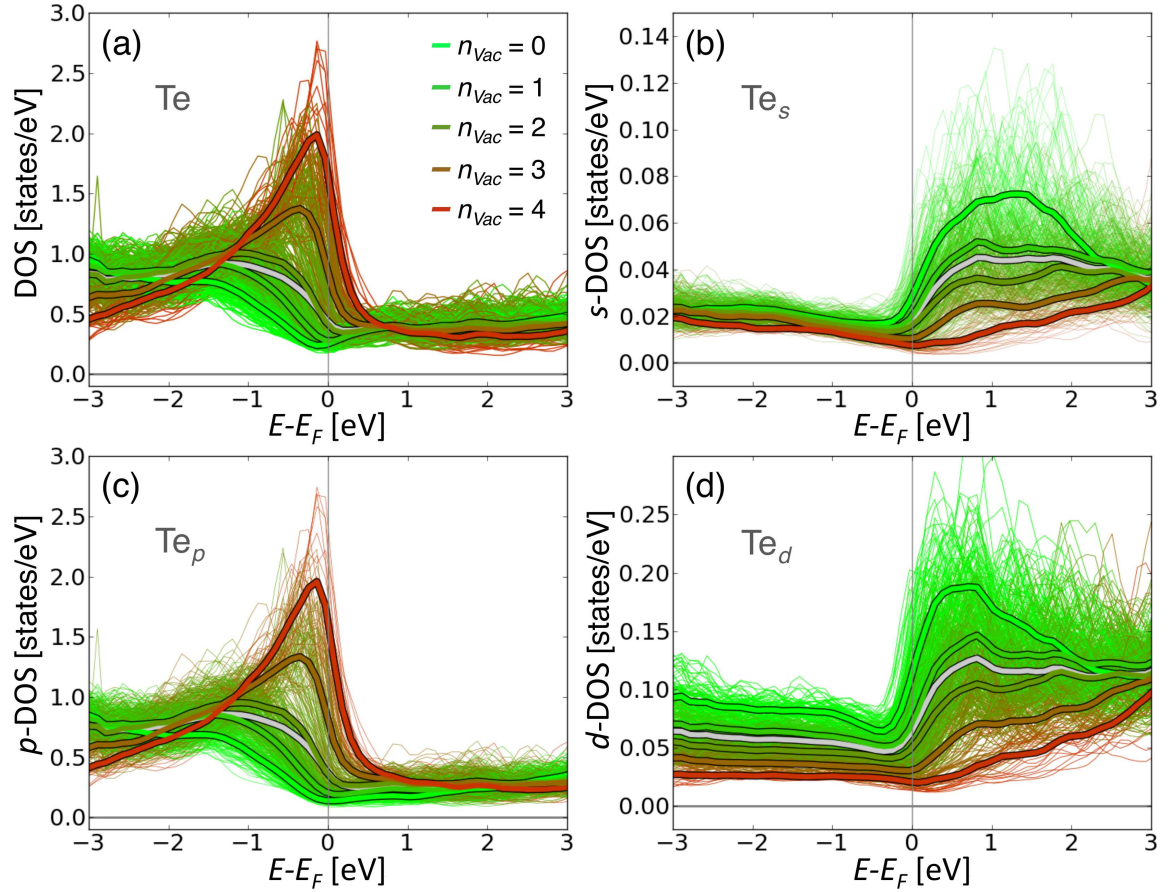**Figure 7.6.:** *Local density of states (LDOS) of Te atoms in a $Ge_{125}Vac_{125}Sb_{250}Te_{500}$ supercell. In all subplots five different types of Te atoms are grouped by their number of next-nearest neighbor sites occupied by vacancies and color-coded as shown in (a). For each of those groups the average LDOS is shown as thick line in the corresponding color. The total LDOS and partial s, p, and d LDOS are shown in (a), (b), (c) and (d), respectively. Note the different scale of the LDOS in different panels.*
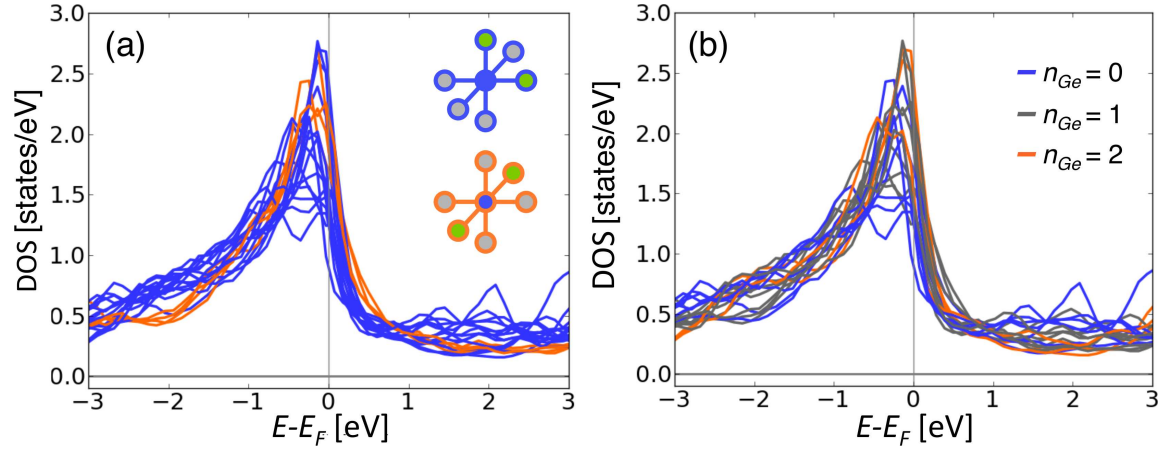
Figure 7.7.: *Local density of states (LDOS) of Te atoms with four vacancies on the nearest neighbor sites ($n_{\mathrm{Vac}} = 4$) with T =400 K. In (a) the different impact of the two possible local vacancy configurations (shown in inset, where gray filled circles stand for vacancies and green filled circle for Ge or Sb atoms) on the LDOS of Te is displayed in the corresponding color. In (b) the color coding resembles the number of Ge atoms on the next-nearest neighbor sites.*

occupation of more distant shells two main influences can be identified: The p-orbitals are known to be strongly directional, which motivates to analyze the relative geometric position of the broken bonds. For the group of Te atoms with four vacancies on the nearest-neighbor sites ($n_{\mathrm{Vac}} = 4$) two configurations occur and its corresponding DOS is shown in Fig. 7.7(a) (for a schematic visualization see inset of Fig. 7.7(a)). Those two possible configurations reveal indeed deviating characteristics: All configurations with opposing vacancies (orange) in the supercell display an almost similar behavior over the entire range of energy. For this configuration high LDOS-values, which are all in the upper third of the complete group of $n_{\mathrm{Vac}} = 4$, are found. In addition to this, the impact of the occupation on the remaining − until now not specified − nearest neighbor sites of Te atoms is taken under scrutiny. Here, in particular at the Fermi energy, the absence of Ge and consequently the presence of Sb atoms is leading to smaller values of the LDOS at the Fermi energy. This trend can be related to the fact that Ge ($4s^2p^2$) and Sb ($5s^3p^2$) are not isovalent and Sb can likely compensate for a larger fraction of vacancy-induced unbound orbitals.

The effect of a strongly directional bonding of p states can not only be studied by considering the LDOS on Te sites but also the LDOS on Ge, Vac, and Sb sites. Those can be partitioned into groups defined by two different criteria: the occupation of vacancies on the next-nearest-neighbor sites m_{Vac} and on the third-nearest neighbor sites p_{Vac}. Arguing on the basis of locality of the electronic states, a stronger impact of m_{Vac} should be visible. As shown in Fig. 7.8 the opposite tendency can be seen. Due to directional bonding the presence of a vacancy on the site behind the nearest-neighbor Te site has a significantly higher impact on the LDOS at the Fermi energy. In fact, from the average values for all local configuration the main trends are determined by p_{Vac}, which is a direct consequence and measure for directional bonding. On the one hand on Ge and Sb sites large peak above E_{F} is present for $p_{\mathrm{Vac}} = 0$ which is successively suppressed with higher $p_{\mathrm{Vac}} = 0$ and shifted

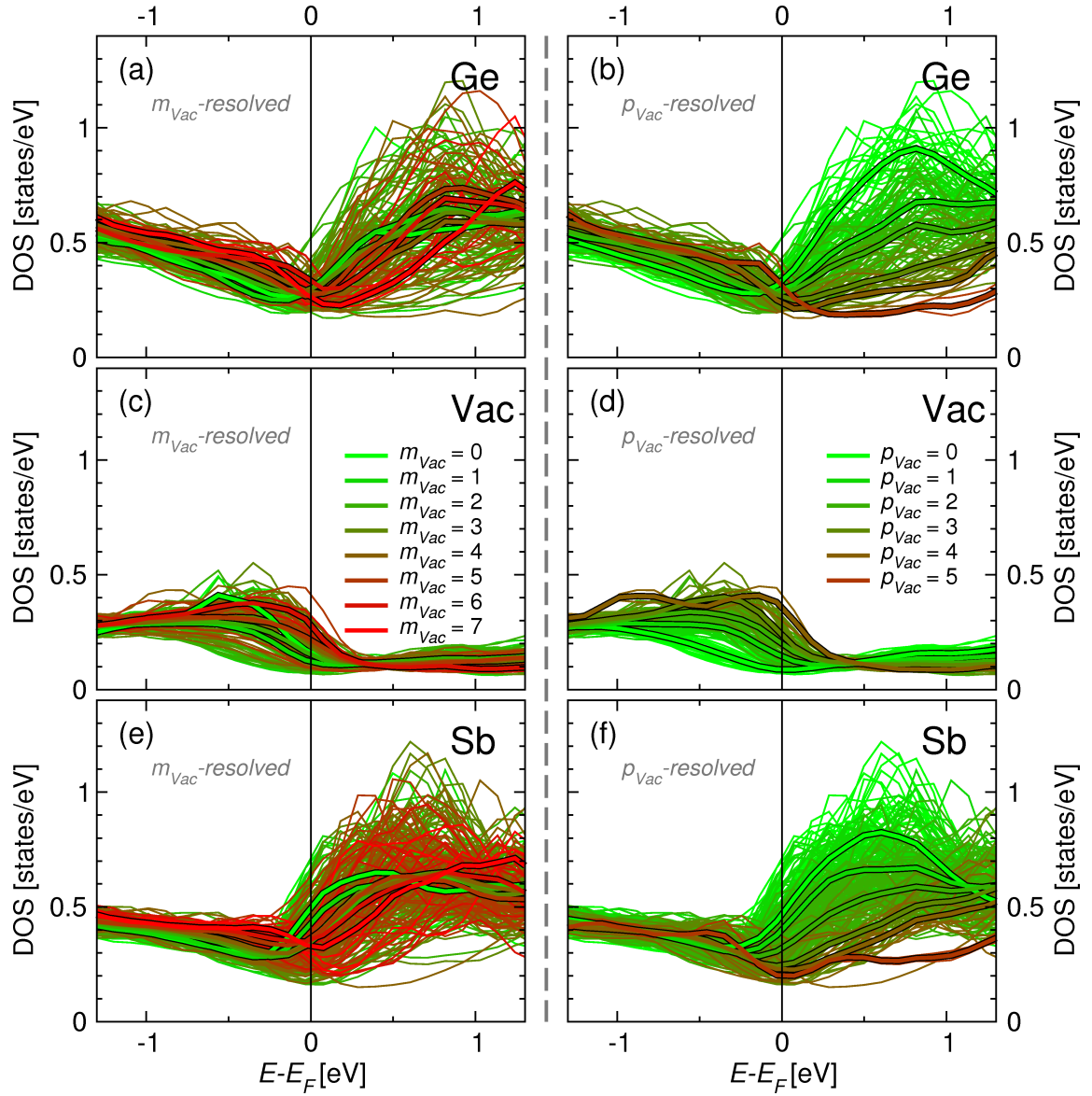**Figure 7.8.:** *Local density of states (LDOS) for Ge, Vac and Sb sites in a $Ge_{125}Vac_{125}Sb_{250}Te_{500}$ supercell, color-coded according to the vacancy occupation of the next-nearest- $(m_{Vac})(a,c,e)$ and third-nearest neighbor $(p_{Vac})(b,d,f)$. The LDOS of all sites is shown individually in thin lines, the average throughout a group of fixed m_{Vac} or p_{Vac} is represented by thick lines in the corresponding color.*

to higher energies. On the other hand on Vac sites a peak sharpens and shifts towards E_F with increasing p_{Vac} − in similar way as shown for Te sites in Fig. 7.6. In all cases the presence of vacancies in the vicinity leads to a shift of those peaks to higher energies − an effect which can be attributed to reduced hybridization. This shift affects the occupied peak on Te and Vac sites which correspond to bonding states, while on the Ge and Sb sites the unoccupied anti-bonding states are influenced. The tails of those bonding and anti-bonding states of the 1st and 2nd sublattice meet at E_F and prohibit the formation of a gap (see Fig. 7.4, 7.6 and 7.8). However, those states can be clearly separated, which allows to introduce the concept of a pseudo-gap in GeSb$_2$Te$_4$ − in similar fashion as usually done in the analysis of quasi-crystaline materials (e.g. by Fujiwara and Yokokawa [152]). Following this argument, for high n_{Vac} the peak on Te sites is pushed towards or even into the energy range of the pseudo-gap. These electronic states which are located in the pseudo-gap consequently experience a exponential spatial decay and can therefore be accounted as highly localized in real space. This observation thereby is a fingerprint for localization in GeSb$_2$Te$_4$.
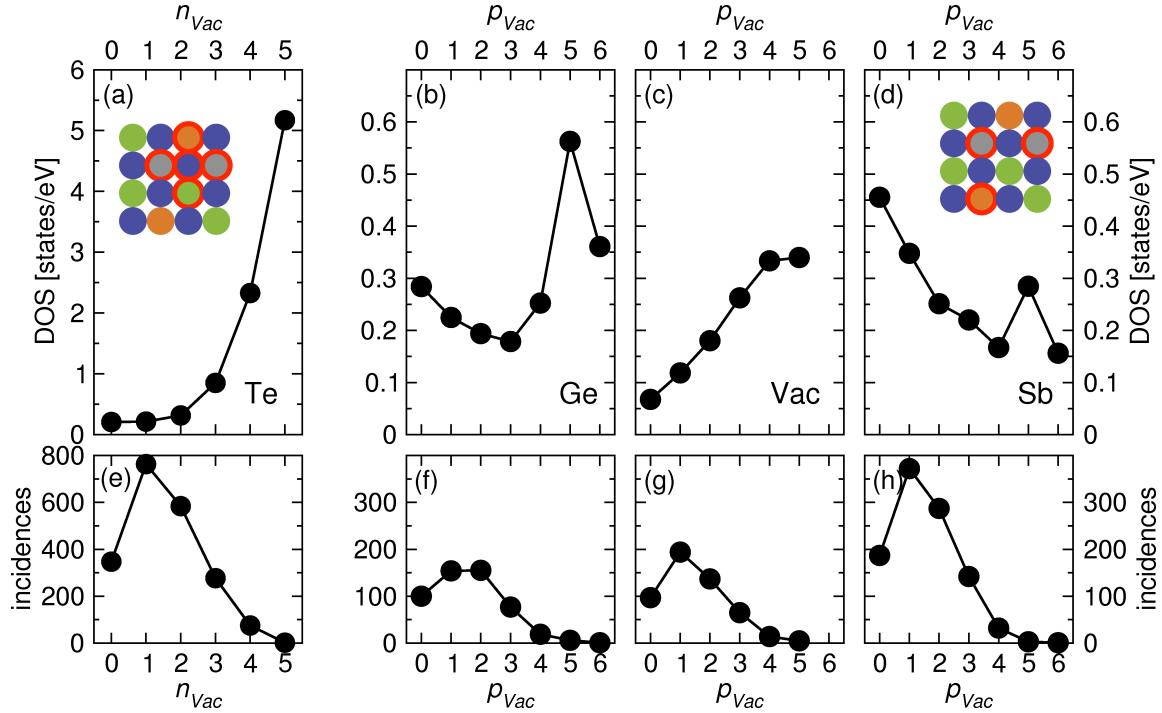
Figure 7.9.: *(a)-(d) Distribution of average local density of states for a Ge$_{512}$Vac$_{512}$Sb$_{1024}$Te$_{2048}$ supercell as a function of the concentration of vacancies in the surrounding material. In case of Te (a) the number of vacancies on the six next-nearest neighbor sites n_{Vac} is used as a reference. For Ge, Vac, and Sb the next-nearest neighbor sites are all occupied by Te atoms and consequently the number of vacancies on the twelve third-nearest neighbor sites p_{Vac} is plotted. Both cases are visualized in the inset of (a) and (c), where sample configurations of Te (blue), Ge (orange), Vac (gray), and Sb (green) are depicted. Note that the scale of (a) and (b)-(d) differs by an order of magnitude. In (e)-(h) the number of incidences of all different vacancy configurations in the supercell is shown for Te (e), Ge (f), Vac (g) and Te (h) separately.*

In order to conduct not only a qualitative but also a quantitive analysis, better statistics are required. Although not excluded per se in smaller cells, e.g. local geometries with five vacancies on next-nearest neighbor sites of Te atoms more likely appear for structures larger than 1000 atoms. Therefore we increase the system size to 4096 sites, which results in a supercell configuration $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$ possessing an edge length of 4.8 nm in each direction. At such cell sizes the question rises whether truncation as implemented in KKR**nano** can be exploited. As discussed in section 4.4 a very accurate description of the total energy and the charge distribution can be achieved with local-interaction zones of one to two-thousand sites. However, for the application to GST, where the localization and delocalization of electronic states is of central interest, the cut-off of long-range interaction is difficult to control and therefore has not been applied for the following calculations.

In the previous section a strong dependency of the local vacancy configuration on the LDOS values at the Fermi energy ($LDOS(E_F)$) has been discovered and discussed for unit cells consisting of one thousand sites. For the larger unit cell of 4096 sites this trend is confirmed by the average values for the $LDOS(E_F)$, displayed in Fig. 7.9. On Te sites a clear non-linear increase of $\rho(E_F)$ is observed, resulting in a by a factor of 20 to 30 enhanced $LDOS(E_F)$ with $n_{Vac} = 5$, as compared to $n_{Vac} = 1$. As shown in Fig. 7.9(e) the configuration of $n_{Vac} = 5$ is a rare event of just three incidences, since the mean vacancy occupation of the neighboring site is only $n_{Vac} = 1.5$ out of six nearest-neighbor sites. However, already for $n_{Vac} = 4$ eighty different configurations occur and lead to very reasonable statistics. Since a significantly higher impact of the third-nearest (p_{Vac}) than for the next-nearest neighbor occupation m_{Vac} has been revealed, we restrict the following analysis to p_{Vac}. In agreement with the previously discussed qualitative findings, Ge, vacancies and Sb reveal a significantly weaker dependency on the number of vacancies on their 6 third-nearest neighbor sites p_{Vac}. Accordingly, the mean value of occupation is $p_{Vac}=1.5$. While only a weak correlation of $LDOS(E_F)$ on Ge atoms with p_{Vac} is present, a distinct trend can be observed for vacancies and Sb: $LDOS(E_F)$ of vacancies tend to increase linearly with p_{Vac} up to $p_{Vac}=5$. In case of Sb increasing p_{Vac} translates into a clear reduction of $LDOS(E_F)$. Both effects are mediated by the nearest neighbor Te atoms.

To gain more insight on the spatial distribution of the LDOS, high values are highlighted on the corresponding sites in the supercell in Fig. 7.10 by choosing the radii of the atomic spheres to be proportional to $LDOS(E_F)$. By that, in consistency with previous findings, large spheres with high values of the $LDOS(E_F)$ exclusively occur on Te sites. Additionally, a small-scaled grouping of the otherwise randomly placed spots with high valued $LDOS(E_F)$ is present. This effect can be attributed to simple geometrical arguments. Suppose two Te atoms A and B in the usual next-nearest neighbor relation. Then, both atoms share three out of their six nearest neighbor sites. Hence, a large number of vacancies on A directly leads to an increased probability for a large number of vacancies on B.

For further interpretation it is important to analyze the distribution of charge on the different sites. Fig. 7.11 shows the charge in each of the Wigner-Seitz cells and its net charge transfer. Focussing first on the average values of charge, a clear transfer of almost one electron towards the vacancy sites can be observed. This gain of charge is compensated by the loss of 0.2 electrons on each of the Sb and Te sites. Additionally, Sb (Te) have one (two) electrons more in their free atomic configuration than Ge, which results in a charge flow of about 0.2 electrons towards Ge. The second order effect is seen by analyzing
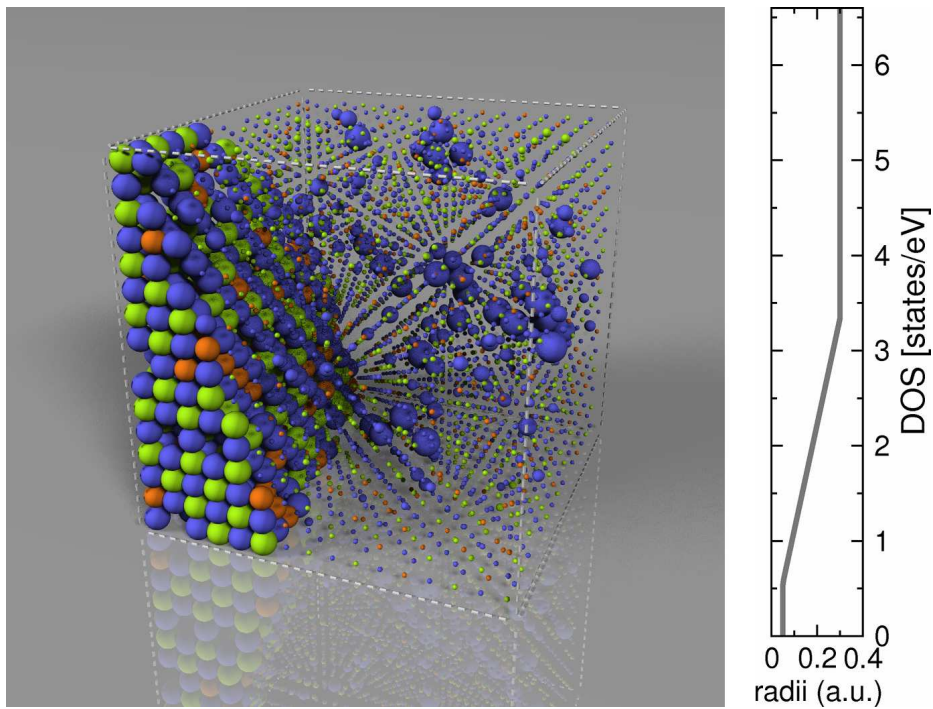
Figure 7.10.: *Spatial distribution of local density of states for a $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$ supercell. In the left lower part the chemical information, in the upper right part the value of the LDOS is displayed. Here, large (small) radii of the spheres correspond to high (low) DOS values as specified in the right panel. For both parts of the plot Ge, Vac, Sb, and Te are shown in orange, transparent, green, and blue, respectively.*
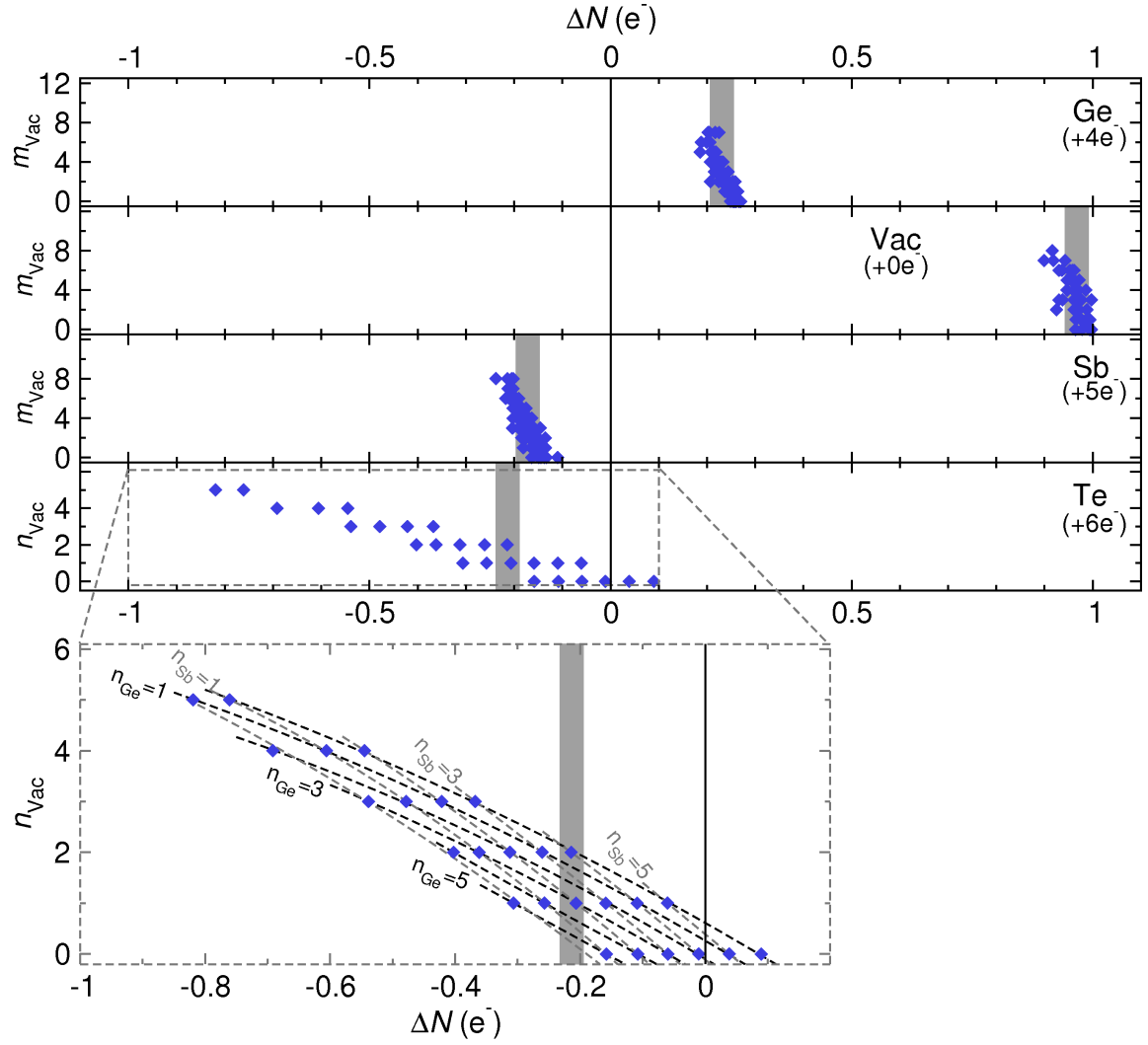
Figure 7.11.: *Net charge transfer ΔN relative to the free atomic configuration for Ge ($3s^2p^2$), vacancies, Sb ($4s^2p^3$), and Te ($4s^2p^4$) in a $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$ supercell. For each composition of nearest neighbors (n_{Vac}, n_{Ge}, n_{Sb}), respectively next nearest neighbors (m_{Vac}, m_{Ge}, m_{Sb}) one mean value is obtained. This value is plotted as a function of the vacancy occupation of the nearest disordered shell n_{Vac} or m_{Vac}. The total average charge transfer of all configurations is displayed as thick gray line. For Te the region of interest is emphasized and lines of constant n_{Ge} and n_{Sb} are shown from n=0 to 5 in black and gray dashed lines, respectively.*

the average values for each group of nearest (n_{Vac}, n_{Ge}, n_{Sb}) or next-nearest neighbor configurations (m_{Vac}, m_{Ge}, m_{Sb}). Here, Ge, Vac and Sb show a weak but distinct correlation with m_{Vac}: The higher m_{Vac} the more charge is donated by Sb and the less additional charge is accumulated on Vac and Ge. The charge on Te atoms turns out to be the most volatile quantity − as affected by disorder on the neighboring sites. This amount of lost charge is determined in first order by the number of vacancies n_{Vac}, which are the largest acceptors of charge. Accordingly, high n_{Vac} leads to large charge transfer away from the Te-atom. Furthermore, for a fixed n_{Vac}, the occupation of the remaining sites shows a clear trend: The more Sb atoms are present, the more charge is donated to the vacancies by those n_{Sb} atoms and the fewer is lost on the Te site. As indicated in Fig. 7.11 those identified trends manifest in a clear and almost undistorted and regular grid of isolines for n_{Vac}, n_{Ge}, and n_{Sb}. In fact, replacing single bonds results in an almost constant contribution, which leads to the conclusion that each of the p-bonds can be − to a certain extent − considered as isolated.

This attribute completes the picture and is in line with our previous findings of the existence of states in the pseudo-gap, and that most of the properties are determined by the atomic configuration within a shell of third-nearest neighbor sites − a strong indication for a considerable localization of the electronic states in GST. In the next section we will focus on the analysis of the influence of disorder on the observed localization.

7.2 Finite size scaling

A recent experimental breakthrough was the discovery of a disorder-induced localization in $Ge_1Sb_2Te_4$ by Siegrist *et al.* [55] which was interpreted as Anderson localization. A theoretical proof for such a localization of states is challenging. We will apply two methods, the analysis of the inverse participation ratio [153] and the probability distribution of the local density of states as introduced by Schubert *et al.* [56]. Common for both approaches is a statistical evaluation of the electronic states at the Fermi energy for different system sizes. For further discussion it is important to note, that for the analysis of localization via the inverse participation ratio usually eigenfunctions, wave-function coefficients or fluctuations of energy levels are taken into account [154, 155]. From KKR multiple scattering theory only the information on the LDOS can be extracted, which is in that sense an integration over all wave-functions. By performing this integration information on the localization of the individual states is lost. Therefore, for this application based on LDOS values the inverse participation ratio is expected to have limited informative value on the actual localization of electronic states. However, from the finite-size scaling of the inverse participation ratio we can still deduce important properties of the electronic states. We will apply this technique before continuing with the analysis using the approach of Schubert *et al.* [56] which is developed to assess localization from the distribution of the DOS on the lattice sites.

In order to study finite-size effects in our system we selected a series of supercells composed of $Ge_{64}Vac_{64}Sb_{128}Te_{256}$, $Ge_{125}Vac_{125}Sb_{250}Te_{500}$, $Ge_{256}Vac_{256}Sb_{512}Te_{1024}$ as well as $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$, which are visualized in Fig. 7.12(a)-(d). To ensure a proper statistical relevance from small to large cells 5, 4, 3, and 1 randomly generated different configurations have been calculated. As illustrated in Fig. 7.12(g),(h) the probability distribution for the occupation of neighboring Te sites meets for all system sizes the ideal theoretical distribution.

In further post-processing steps we will refer from now on to the $LDOS(E_F)$ by $\rho(E_F)$. For the calculation of all $\rho(E_F)$ with KKR**nano** the LDOS precisely at $E = E_F$ has been determined with a temperature broadening of 50 K. Such small broadening has been chosen to minimize the influence of temperature broadened states lying close to E_F and has become only applicable to this system sizes by the block-circulant preconditioning of the iterative solution as discussed in chapter 4.2.3. Although irrelevant for the calculation of the inverse participation ratio, it is important for the analysis of the probability distribution to normalize $\rho(E_F)$ by its mean value $\rho_0(E_F)$ to

$$\tilde{\rho}_{i \in N}(E_F) = \frac{\rho_{i \in N}(E_F)}{\rho_0(E_F)} = \frac{N \cdot \rho_{i \in N}(E_F)}{\sum_i^N \rho_i(E_F)}, \tag{7.1}$$

where i runs over a group of lattice sites N. Here, we calculate this quantity both for the entire system with N sites and atom-specific for Ge, Sb, Va and Te with accordingly smaller N. It is important to note, that in all treated cells and subgroups of elements only small deviations of $\rho_0(E_F)$ are present.
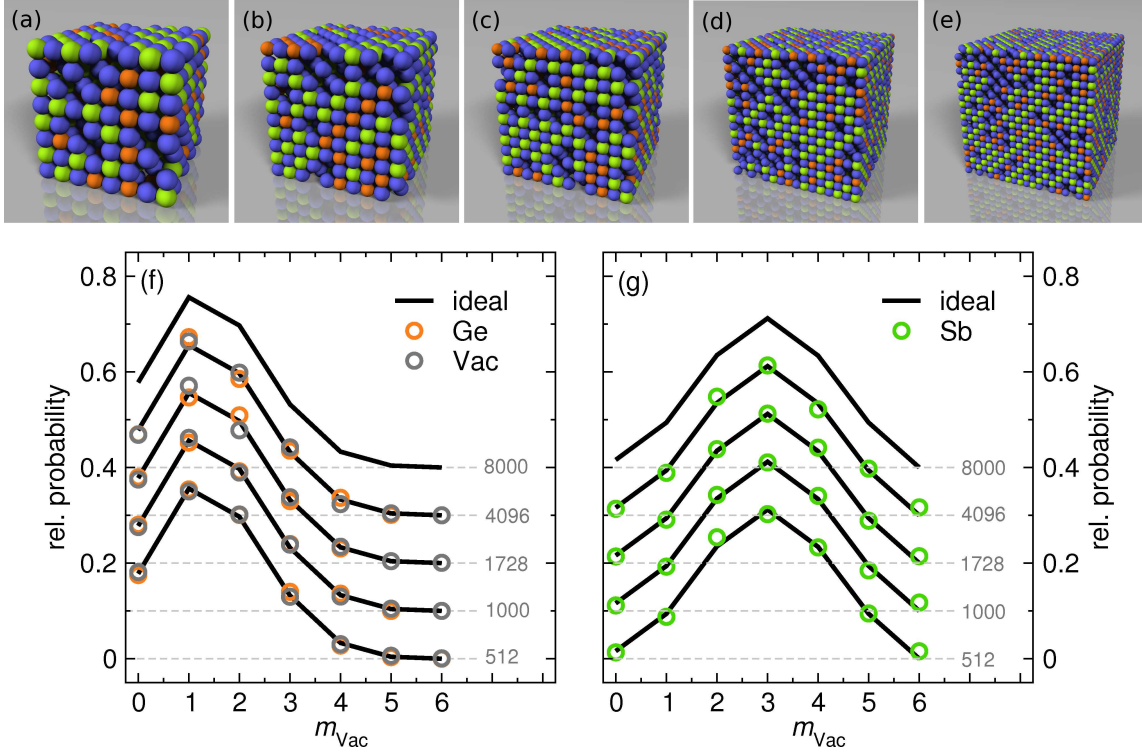
Figure 7.12.: *(a)-(e) Illustration of differently sized unit cells of the phase change material* $GeSb_2Te_4$ *with one sublattice being fully occupied with Te (blue) and the second sublattice with 25% Ge (orange), 25% vacancy sites (transparent) and 50% Sb (green). The following unit cells are shown:* $Ge_{64}Vac_{64}Sb_{128}Te_{256}$ *(a),* $Ge_{125}Vac_{125}Sb_{250}Te_{500}$ *(b)* $Ge_{216}Vac_{216}Sb_{432}Te_{864}$ *(c),* $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$ *(d) and* $Ge_{1000}Vac_{1000}Sb_{2000}Te_{4000}$ *(e). The distributions of next-nearest neighbor configurations of Ge and vacancy (f) as well as Sb (g) as seen by the Te atoms are shown for all sizes of unit cells. The probability distribution is successively shifted by 0.1 starting from the unit cell size of 512 atoms. In all cases the ideal random values are plotted versus the average values over 5, 4, 3, 1, and 1 inequivalent finite-sized realizations for unit cells of 512, 1000, 1728, 4096 and 8000 atoms, respectively.*
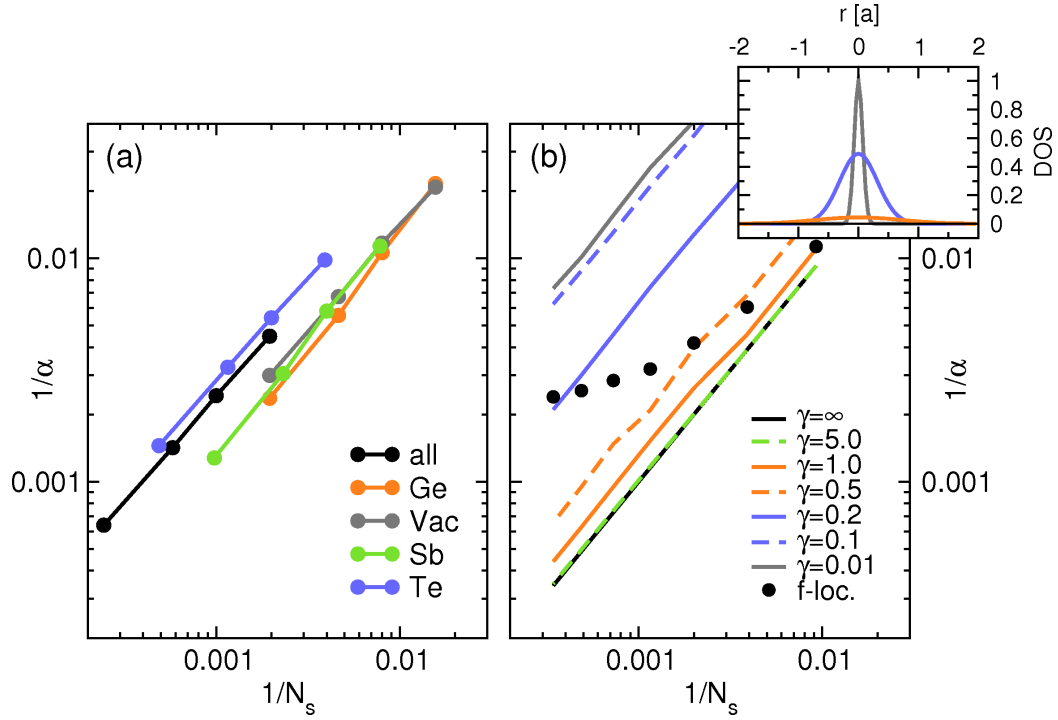
Figure 7.13.: (a) Inverse participation ratio α at E_{F} for $Ge_1Sb_2Te_4$ supercells of in total 512, 1000, 1728 and 4096 sites. For the sake of comparison in (b) Gaussian distributions of states $\rho(E_{\mathrm{F}}) = \rho_0 \exp(r^2/\gamma)$ are placed randomly on 5% of the sites of a fcc-lattice with lattice constant $a = 1$. Inset of (b) shows the distribution for $\gamma = 0.01, 0.2$ and 1.0 with the same color-coding as in (b). In the special case labeled with 'f-loc.' the number of Gaussian states (γ=0.2) distributed on the lattice is set to 125 for all system sizes.

With the obtained information the inverse participation ratio α^{-1} can be calculated as:

$$\alpha^{-1} = \frac{\sum_{i=1}^N |\tilde{\rho}_i|^2}{\left(\sum_{i=1}^N |\tilde{\rho}_i|\right)^2}, \tag{7.2}$$

where $\tilde{\rho}_i$ is the short notation for $\tilde{\rho}_{i \in N}(E_{\mathrm{F}})$. For an understanding of the behavior of α^{-1}, it is worth to briefly discuss the following idealized scenarios: Suppose a completely delocalized electron in a single ideal Bloch-state spanning over all sites $i \in N$. Then $\tilde{\rho}_i$ is equal ρ_0 on all sites and α^{-1} decays as N^{-1} with increasing system size N. In the other extreme a single electronic state is assumed to be completely localized on one site j out of the N sites, which results in $\tilde{\rho}_i = \delta_{ij} \cdot \tilde{\rho}_c$. In this particular case α^{-1} is equal to one and therefore independent of system size.

Having those two extrema in mind, we examine the inverse participation ratio α^{-1} of the calculated systems. In Fig. 7.13(a) α^{-1} is plotted on a log-log-scale over N^{-1} for each element as well as for all lattice sites. In all cases our calculated values can be well described by a straight line in the (α^{-1}, N^{-1}) space. As a consequence α^{-1} converges with increasing N to zero. However, comparing α_{Te}^{-1} with all remaining elements a significant

offset can be identified, which roughly is described by $\alpha_{Te}^{-1} = 2 \cdot \alpha_{Ge,Vac,Sb}^{-1}$. To be able to rate this behavior we modeled a series of gaussian states with different level of localization on an fcc-lattice. The number of considered gaussians is straightforwardly determined by the fixed concentration of gaussians and the system size. The evaluation of α^{-1} is based on the resulting LDOS on each of the fcc sites in the exact same fashion as described above and plotted in Fig. 7.13(b). Starting with $\gamma = 0.01$ which corresponds to completely localized states and ending with $\gamma = \infty$ which resembles fully delocalized states, a linear dependency appears on the log-log-scale. From that observation we can deduce that for states being well described by gaussians higher levels of localization lead to higher values of α^{-1} considering a constant system size. Under this assumption the electronic states on Te sites turn out to be considerably more localized that the electronic states on the second sublattice. Although the numerical tests based on gaussian states clearly resemble a scenario of distinct localization, the fact that the number of those states increases proportionally to N leads to the behavior described above. Only if the number of states is assumed to be constant for all N, α^{-1} converges to non-zero values in the limit of large N as exemplified in Fig. 7.13(b). Such a setting occurs in several studies based on Anderson-model Hamiltonians and can be interpreted as a proof for localization. We do not find such a behavior for the finite-size scaling in GST. This is − in particular because we base our analysis on the DOS and not on single wave-function coefficients − no proof for the absence of localization. Nevertheless, the striking difference in α^{-1} for Te and the group of Ge, Vac, and Sb is still pointing at an enhanced localization of the states on Te sites.

An alternative approach to derive a proof for Anderson-localization from the density of states has been given in [56]. Here, the shape of the distribution of the LDOS is the key attribute. The shape is still closely connected to the inverse participation ratio but it allows to extract considerably more information on the electronic states. Since $\tilde{\rho}_i$ varies by more than one order of magnitude (see Fig. 7.9) the histogram bins, which are utilized in the following analysis, are defined in logarithmic partitioning. Here, twenty bins are used to represent a range of $\tilde{\rho}_i$ from 0 to 7 states/eV. The relative occurrence at each of those bins defines the resulting histograms $\Lambda_X(\tilde{\rho}_b)$, where $\tilde{\rho}_b$ labels the central value of the interval which is represented by the bin b. The histograms $\Lambda_X(\tilde{\rho}_b)$ are shown in Fig. 7.14 for each element separately for several unit cell sizes X. According to Schubert *et al.* [56] a significant shift of this distribution upon finite-size scaling acts as a proof for Anderson localization. In greater detail, in case of Anderson localization the most probable values of the distribution are shifted to smaller values with increasing system size N. It becomes immediately clear from the distribution plotted in Fig. 7.14 that no shift with the system size can be deduced by eye. Before we will highlight possible differences in the distributions for differently sized supercells, a few important remarks can be made about the universal and size-independent shape.

Comparing the distributions of the different elements, a significantly deviating behavior of the LDOS on Te sites can be observed: While for Ge, vacancies and Sb mean and most probable value differ by less than 0.2 states/eV, Te exhibits a large asymmetry of those two values by 0.5 states/eV. This strong asymmetry in the distribution is a direct consequence of the steep non-linear increase of $\tilde{\rho}_i$ with increasing number of neighboring vacancies (see Fig. 7.9(a) for reference). Additionally the large non-uniformity in space is a first indication of localization.
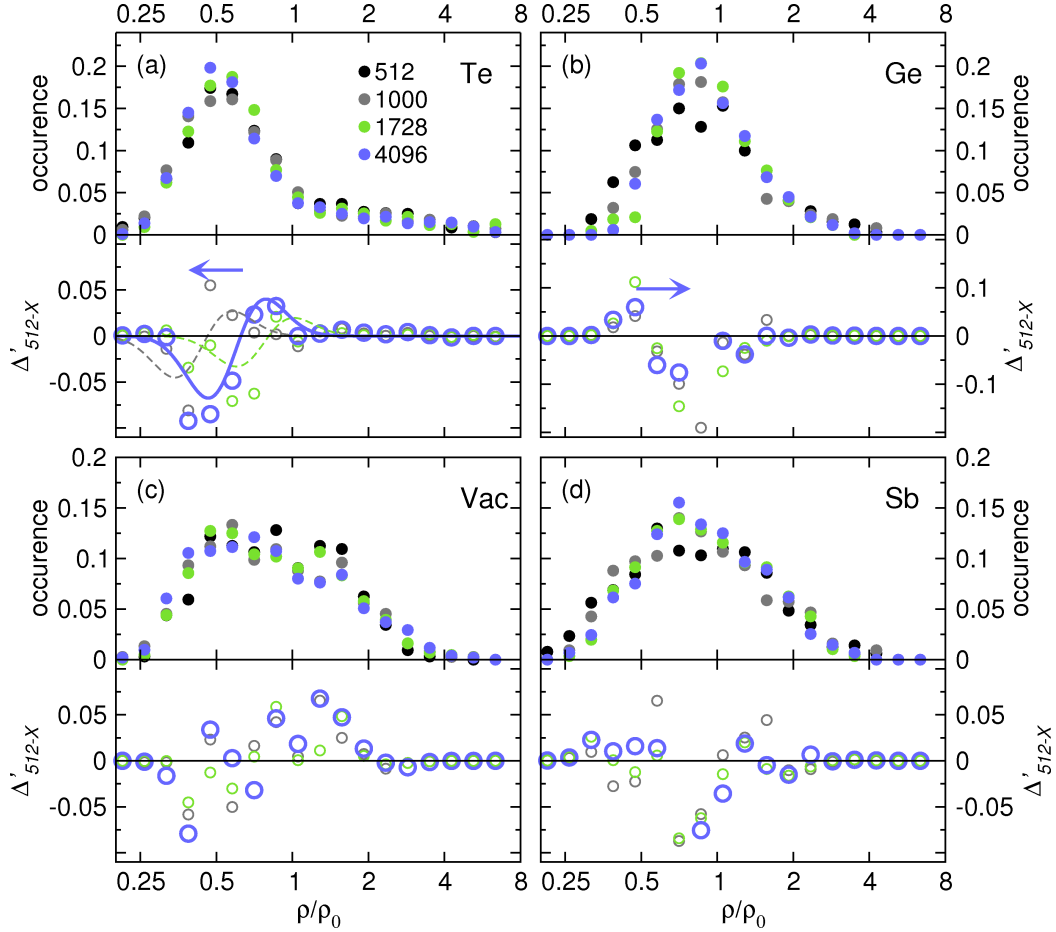
Figure 7.14.: *Distribution of local density of states (LDOS) at E_F for $Ge_{.25}Vac_{.25}Sb_{.5}Te$ supercells of in total 512, 1000, 1728 and 4096 sites in 5, 4, 3, and 1 different randomly generated configurations. All LDOS values are normalized by its mean LDOS ρ_0. Both the distributions (upper panel) as well as its differences (lower panel) are separately displayed for the LDOS on Te (a), Ge (b), vacancy (c) and Sb (d) sites. In all upper panels of the subplots the color coding is used as specified in (a). For all lower panels the difference as defined in the text is given in all cases relative to the smallest supercell size: $\Delta'_{512-1000}$ (gray), $\Delta'_{512-1728}$ (green), and $\Delta'_{512-4096}$ (blue). Lines are least-square fits according to (7.6) to the shown data points.*

In order to increase the resolution with respect to the differences of the distributions statistically relevant parts are accentuated by weighting the distribution Λ_X of supercell size X by its actual size at the bin b. In other words, bins in which only few events occurred will be suppressed:

$$\Lambda'_X(\tilde{\rho}_b) = \Lambda_X(\tilde{\rho}_b) \cdot \frac{\frac{1}{n_X} \sum_X \Lambda_X(\tilde{\rho}_b)}{\frac{1}{n_b} \sum_b \Lambda_X(\tilde{\rho}_b)}, \tag{7.3}$$

where n_X denotes the number of equally sized supercells which have been calculated to improve the statistic[2] and n_b is the number of bins for setting up the histogram. For the following application $n_b = 20$ has been used. The difference of two such post-processed distributions then reads:

$$\Delta'_{X_1 - X_2} = \Lambda'_{X_1}(\tilde{\rho}_b) - \Lambda'_{X_2}(\tilde{\rho}_b). \tag{7.4}$$

This function is drawn in the lower panels of Fig. 7.14, where $X_2 = 1000$, 1728, and 4096 is set into relation to $X_1 = 512$. First, we consider the statistically best sampled data-sets with are the ones for Te, which are placed on 50% of the sites in the supercells. Here, a characteristic behavior appears: Left (right) from the highest probable values a dip (peak) in relation to $X_1 = 512$ emerges, which resembles a shift in the distribution to the left. While for Ge sites the scenario is less clear but still points at a shift to the right upon increasing the system size, for vacancy- and Sb sites no trend can be observed on the basis of the existing statistics.

Focussing on the most systematic differences for the distributions of the LDOS on Te sites, we have to render an analytical expression for $\Delta'_{X_1 - X_2}$. Single distributions $\Lambda_X(\tilde{\rho}_b)$ and in particular $\Lambda'_X(\tilde{\rho}_b)$, since the tail of the distribution is strongly suppressed, can be fitted to a modulated Gaussian:

$$\Lambda'_X(\tilde{\rho}_b) \approx a_1 \frac{1}{\tilde{\rho}_b} \exp\left[-\frac{(\ln(\tilde{\rho}_b) - a_3)^2}{a_2} \right], \tag{7.5}$$

where a_1, a_2, and a_3 are the fitting parameter and a_3 is of particular importance being a measure for a shift in $\tilde{\rho}_b$. Therefore, and because the variations of the shape of the distributions is small we select a universal a_1 and a_2 for a fit to $\Delta'_{X_1 - X_2}$ with individual parameter $a_3^{X_1}$ and $a_3^{X_2}$. The fixed parameter a_1 and a_2 have been obtained by least square fits to the element-specific $\Lambda'_X(\tilde{\rho}_b)$ which were averaged over all system sizes X. With this assumptions $\Delta'_{X_1 - X_2}$ can be expressed as

$$\Delta'_{X_1 - X_2} \approx a_1 \frac{1}{\tilde{\rho}_b} \cdot \left(\exp\left[-\frac{(\ln(\tilde{\rho}_b) - a_3^{X_1})^2}{a_2} \right] - \exp\left[-\frac{(\ln(\tilde{\rho}_b) - a_3^{X_2})^2}{a_2} \right] \right), \tag{7.6}$$

with now only two free parameters $a_3^{X_1}$ and $a_3^{X_2}$. The results of the least square fits using this function are shown in 7.14(a). The distinct shape of this fits agrees for all three sets of data points. Starting from small values of $\tilde{\rho}_b$ the curve drops to negative values, experiences a crossing to positive values and then saturates to zero. This behavior is in fact originated by shifted distributions: Both $\Lambda'_X(\tilde{\rho}_b)$ and by that directly also $\Lambda_X(\tilde{\rho}_b)$ for $X = 1000$, 1728, and 4096 are shifted to the left with respect to $\Lambda'_{512}(\tilde{\rho}_b)$ and $\Lambda_{512}(\tilde{\rho}_b)$, respectively. Although this

---

[2]More details on this supercells are given in Fig. 7.12.

fitting procedure is of limited accuracy and obeys statistical fluctuations, an overall trend of an increasing shift in $\tilde{\rho}_b$ of $a_3^{512} - a_3^{1000} = 0.019$, via $a_3^{512} - a_3^{1728} = 0.024$ to $a_3^{512} - a_3^{4096} = 0.038$ can be observed.

With the criteria of Schubert *et al.* [56] at hand, the most important attribute of the distributions displayed in Fig. 7.14 is that a small shift of the distribution can indeed be observed for the LDOS of Te sites. From our analysis it becomes clear that this shift is well above statistical fluctuations and hence is an important sign for disorder induced localization. Summarizing, indications of disorder-induced Anderson localization can be drawn by utilizing the finite-size scaling approach being strongest on the Te sublattice.

7.3 Simulated annealing

The findings in the previous sections state that a locally enhanced density of vacancies (vacancy-complexes) induces spatially very confined states predominantly on the adjacent Te sites. In this section we will provide evidence that those states disappear, when the vacancy-complexes are diluted.
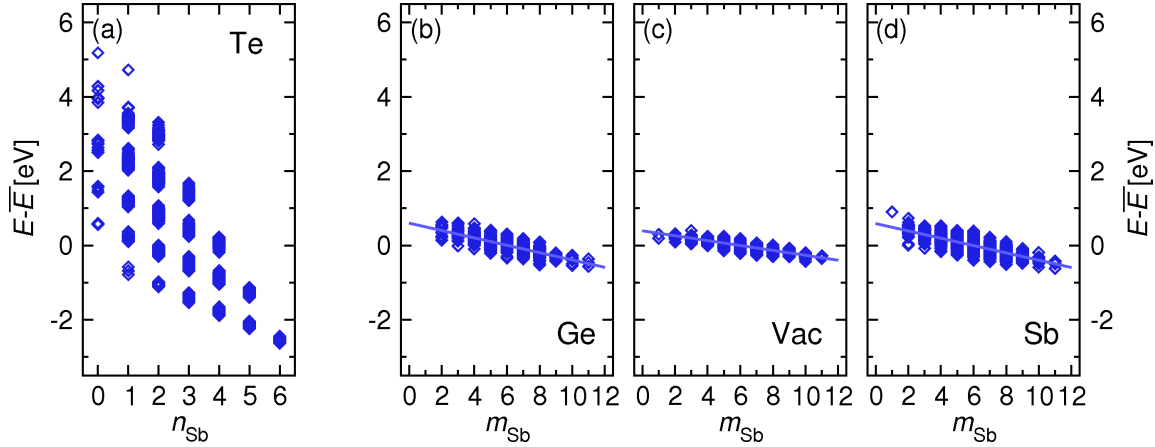
Figure 7.15.: *Distribution of the total energy per site E (blue) relative to the average total energy $\overline{E}$ of the same atomic type and as a function of number of neighboring Sb atoms. The relative total energy is shown for Te (a), Ge (b), the vacancies (c), and Sb (d) and is based on a supercell calculation of 4096 sites. The x-axis value n_{Sb} and m_{Sb} stand for the number of Sb atoms on the nearest neighbor (a) and on the next-nearest neighbor site (b)-(d), respectively. Straight blue lines in (b)-(d) are a linear regression to the according data.*

In order to gain insight which effect can be the driving force for such a process of destruction of vacancy-complexes the total energy of local configurations is playing a fundamental role. Within the KKR-method the total energy as obtained from the self-consistent density-functional cycle can be − contrary to plane-wave methods − accounted to the individual Voronoi sites. Those site- or atom-resolved total energies obtained with **KKRnano** are plotted in Fig. 7.15. As done before the influence of neighboring vacancies appears to be clearly different for both sublattices in GST. On the one hand the total energies of Ge, Vac and Sb are varying by ± 0.7 eV. For all elements large numbers of next-nearest neighboring vacancies lead to higher total energies and are thereby energetically not favored. On the other hand this effect is crucially enhanced for Te. Here the maximal (minimal) total energies deviate from the average total energy by $+5$ eV $(-3$ eV$)$. In addition a conspicuous grouping of total energy values seems to be present. For the sake of clarity Fig. 7.15(a) is redrawn in Fig. 7.16 showing the mean total energies for each group of nearest-neighbor configurations $(n_{Ge}, n_{Vac}, n_{Sb})$. In the same manner as done for the dependency of the charge on $(n_{Ge}, n_{Vac}, n_{Sb})$ in Fig. 7.11 in section 7.1 a regular grid spanned by almost linear curves can be overlaid on the mean data points. Since structural relaxations have been neglected in our approach the impact of relaxations on the total energy has to be classified. Wełnic *et al.* [147] found that the change in bond-length does not exceed five percent and comes along

with a change in total energy just under 0.1 eV per atom. Relative to the strong influence of the local configuration in a range of 8 eV the relaxation energy will not affect the trends we observe. In the physical picture of GST we have up to now, a larger n_{Vac} will likely result in an increased gain of energy due to structural relaxations. Therefore the entire grid of energy configurations will be probably quenched by one or two eV without changing the order of the grid.
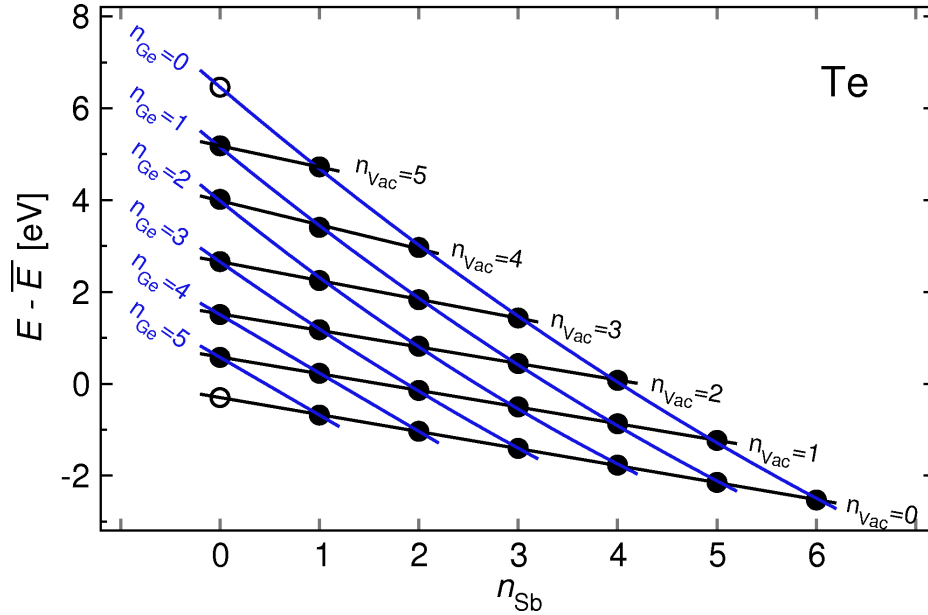

Figure 7.16.: *Average relative total energy $E - \overline{E}$ of Te atoms (full black dots) extracted from a 4096 site supercell calculation with KKR**nano** as a function of number of neighboring Sb atoms. The average values are based on the raw data presented in Fig. 7.15(a). Lines are fits to data points with constant number of next-nearest neighbor Ge atoms n_{Ge} (blue) and vacancies n_{Vac} (gray). For all n_{Vac} and $n_{\mathrm{Ge}} = 4$ and 5 a linear, for the remaining lines a quadratic regression has been performed. The unlikely, but missing point for $n_{\mathrm{Sb}} = 0$ and $n_{\mathrm{Ge}} = 0, 6$ (open black dots) have been obtained by those fits.*

Continuing with the analysis of Fig. 7.16 the most important finding is that vacancy complexes are highly unfavorable with respect to the total energy. This observation is in particular interesting once related to the recent report by Siegrist *et al.* [55] who studied the dependency of the transport properties of $Ge_1Sb_2Te_4$ on the annealing temperature experimentally. Interestingly, in these experiments the resistivity is strongly reduced upon annealing − the higher the annealing temperature the clearer this effect gets. To reveal the origin of this result we performed simulated annealing computational Monte-Carlo studies, which underlie several assumptions to mimic just the main effects of temperature induced kinetics:

- The Te sublattice is assumed to be unchanged during the entire annealing process.

- Migration is assumed to be always mediated by a vacancy and takes place on short scales. This means one out of two exchanged atoms has to be a vacancy and exclusively

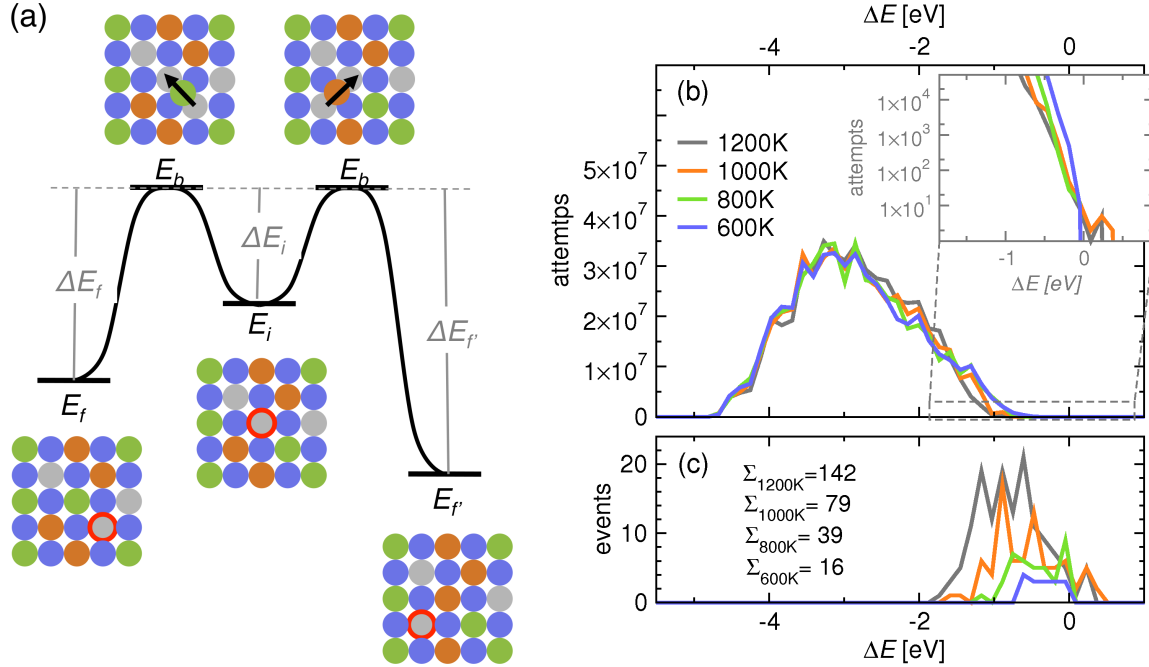nearest neighbor pairs (on the GeVacSb-sublattice) of atoms and vacancies can move.

Figure 7.17.: (a) Illustration of the concept of simulated annealing. E_i displays the energy of the initial configuration, E_f and $E_{f'}$ the possible final states after migration. All three configurations are separated by a migration-barrier E_b, which acts as reference to obtain the relativ barrier ΔE. As inset all configurations of Te (blue), Ge (orange), Vac (gray), and Sb (green) as well as intermediate ones during migration are schematically drawn. Atoms, to which the E_i, E_f and $E_{f'}$ are related to, are marked with a red circle, moving atoms with a black arrow. Total attempts (b) and successful atom hopping (c) generated by the Monte-Carlo scheme applied to $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$ for four different temperatures and as a function of ΔE. In (c) the total number of achieved atom hoppings, Σ, for the different temperatures is indicated.

- The migration paths including relaxation effects are mimicked by a universal energy barrier E_b as shown in Fig. 7.17(a). In order to obtain a measure for the total energy we restrict the model to nearest-neighbor effects which is a reasonable assumption according to all previous results. Hence, the total energies of the nearest neighbor atoms are explicitly added which leads to two parts in the overall energy E of the considered configuration:

$$E = E^X + \frac{1}{6}\sum_j^6 E_j^{Te}, \tag{7.7}$$

where X stands for Ge, Vac, or Sb. Here only contributions induced by e.g. different binding to neighboring atoms are supposed to enter and not e.g. the contribution of the core states to the total energy. Therefore both energy contributions E^X and E^{Te} are the total energies relative to their mean value as shown in Fig. 7.15. Keeping in mind the nearest-neighbor restriction all Ge, Vac, and Sb atoms are equivalent

as uniformly surrounded by 6 Te atoms, which results in $E^X = 0$. Accordingly, the energy of a configuration is solely be defined by its surrounding Te atoms. Because the barrier E_b cannot be extracted from the performed *ab initio* calculations, we choose E_b slightly above the highest possible configuration energy E. Choosing a higher energy barrier will not crucially affect the resulting structural properties but will enhance the required time-steps drastically.

- To rate whether a migration event takes place or not, we exploit the random-walk-algorithm probing one configuration at a time. The probability for migration Γ is temperature-dependent and given by:

$$\Gamma(\Delta E, T) = \exp\left(-\frac{\Delta E}{kT}\right), \tag{7.8}$$

with $\Delta E = E_b - E$.

Fig. 7.17(b) and (c) shows the impact of the simulated annealing on the lattice. Clearly, only less favorable configurations with total energies above average are affected by migration. As expected the number of approved switching events increases exponentially with temperature. This effect results in a less frequent occurrence of high valued total energies for larger annealing temperatures. Those energetically unfavorable configurations are likely to migrate when being close in energy to E_b. Once switched, there is a high probability that after migration configurations are created which are all lower in energy than the initial state. The exponentially reduced probability for switching leads to a trapping in those energetically more favorable configurations.

Plotting the distribution relative to the number of vacancies on nearest neighbor sites n_{Vac} drawn in Fig. 7.18(a) it becomes clear that the number of vacancy complexes with $n_{\text{Vac}} \geq 4$ is considerably reduced when temperature and time-steps are increased. As none of the vacancies leave the sample on the same footing a strong increase of $n_{\text{Vac}} = 2$-configurations can be observed. In Fig. 7.18(b) and (c) the occurrence of those most affected vacancy-configurations is shown with respect to the Monte-Carlo time-steps. For all probed annealing temperatures convergency of $n_{\text{Vac}} = 2$ and $n_{\text{Vac}} = 4$ and the entire distribution (not shown) is reached after about 300 million time-steps (attempts).

Based on the annealed configuration (T=1000 K) new density-functional calculations have been performed with KKRnano. Here, the same computational parameters have been used, as for the initial structure (see section 7.1 and 7.2). From the comparison of the LDOS at the Fermi energy as a function of the local vacancy configuration of the annealed and initial structure it becomes clear that the fundamental trends remain unchanged (see Fig. 7.19). This observation allows for the conclusion that a smooth transition of material properties can be expected and no phase transition has been induced by the structural reordering. Beyond that, additional information on the short-range and mid-range ordering of vacancy complexes by the nearest neighbor (n_{Vac}) and third-nearest neighbor (p_{Vac}) occupation can be deduced from Fig. 7.19(e-h). As previously discussed vacancy complexes resembled by high n_{Vac} are clearly suppressed upon annealing. On the other hand the distribution of p_{Vac} is unaffected by annealing, which leads to the conclusion that annealing does not influence the mid- and long-range ordering of vacancies.

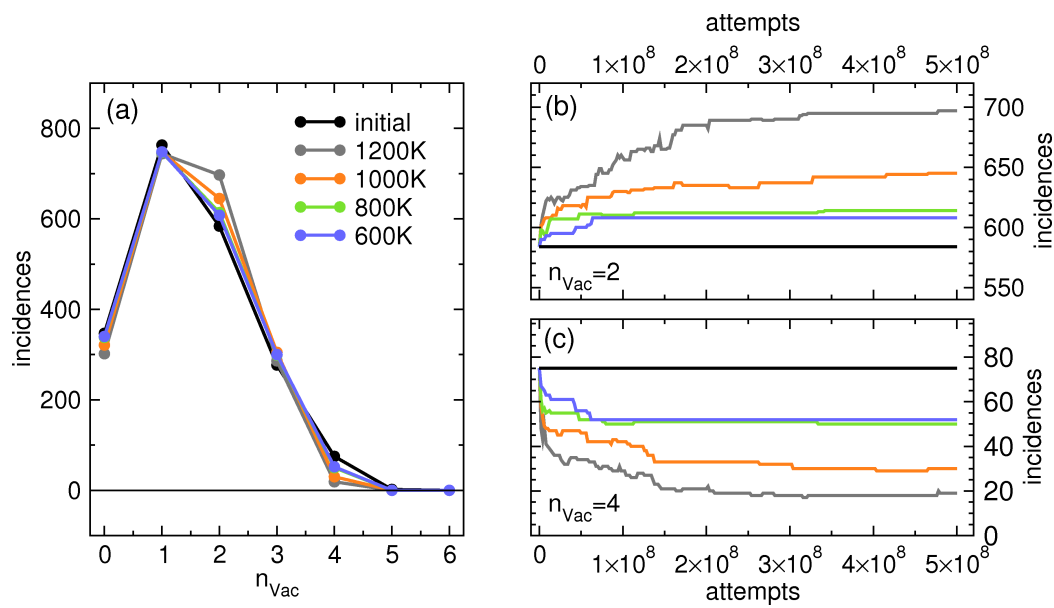For a quantitative analysis of the influence of annealing on the transport properties which

Figure 7.18.: (a) Distribution of number of vacancies n_{Vac} occurring on the six nearest neighbor sites to Te atoms in $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$. The distribution of n_{Vac} is shown for the initial configuration corresponding to Fig. 7.15(a) and after simulated annealing for 600 K, 800 K, 1000 K, and 1200 K for $5 \cdot 10^8$ Monte-Carlo-steps each. In (b) and (c) the evolution of the distribution during the simulated annealing is shown exemplified by $n_{Vac} = 2$ (b) and $n_{Vac} = 4$ (c). The same color-coding as in (a) is used for different annealing temperatures.
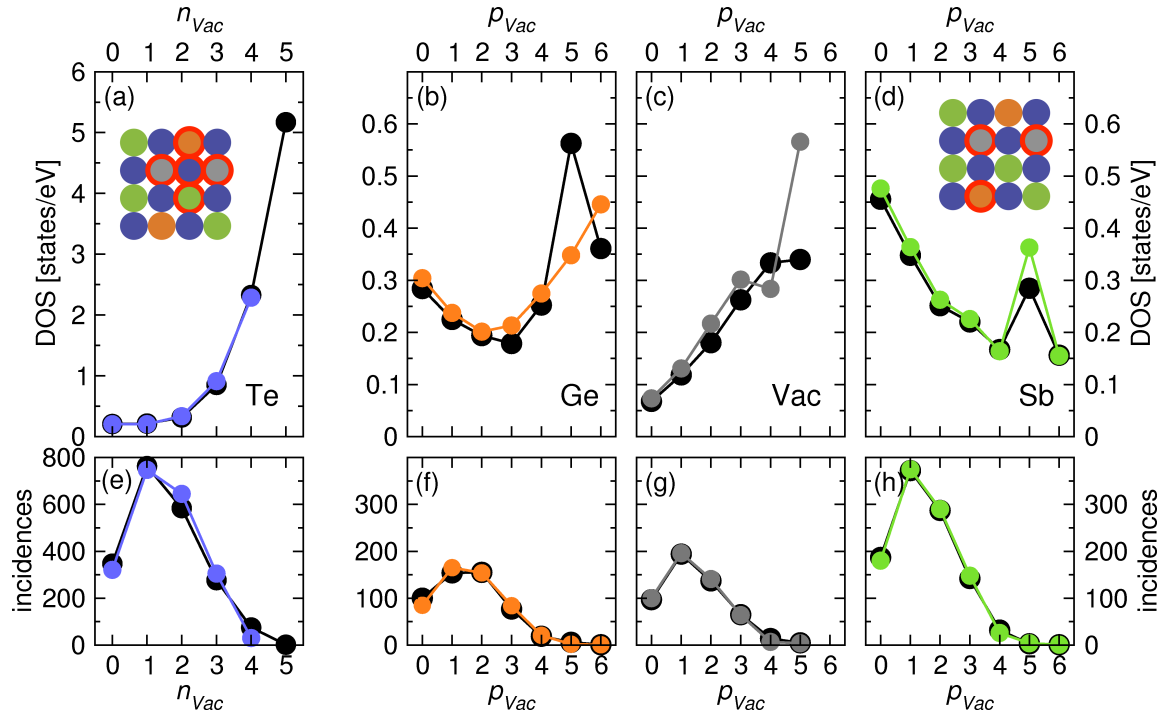
Figure 7.19.: *(a)-(d) Distribution of average local density of states at E_F, LDOS(E_F), for a $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$ supercell as a function of the concentration of vacancies in the surrounding material. In case of Te (a) the number of vacancies on the six next-nearest neighbor sites n_{Vac} is used as a reference. For Ge, Vac, and Sb the next-nearest neighbor sites are all occupied by Te atoms and consequently the number of vacancies on the twelve next next-nearest neighbor sites m_{Vac} is plotted. Both cases are visualized in the inset of (a) and (c), where sample configurations of Te (blue), Ge (orange), Vac (gray), and Sb (green) are depicted. Note that the scale of (a) and (b)-(d) are by one order of magnitude different. In (e)-(h) the number of incidences of vacancy configurations in the supercell is shown, where (e)-(h) are related to the above shown plots. Black dots correspond to the initial random structure, colored (blue, orange, gray, green) dots represent the annealed (1000 K) structure.*

are crucially defined by the DOS at the Fermi energy, we apply the approach used above for the finite-size scaling (section 7.2). The distribution of LDOS of pre- and post-annealed structures is in similar fashion parametrized defining $n_b = 30$ exponential bins (for details see section 7.2). The resulting probability distributions Λ_i (initial) and Λ_a (annealed) are shown element-specifically in Fig. 7.20. Λ_i and Λ_a on Te sites exhibit a distinct shift to the right upon annealing. Opposed to that, from Λ_i and Λ_a for Ge, Vac and Sb sites no clear trend is noticeable. Highlighting the differences and reducing the statistical noise on the distribution according to 7.4 underlines this first evaluation for Te, Vac and Sb sites. In contrast to that this more detailed analysis reveals that for Ge sites a shift to the right can be observed, which is however subject to higher statistical fluctuations. This leads us to the conclusion that Te sites − 50% of the compound − tend to a less asymmetric distribution and thereby to an increased homogeneity of the sample, while Ge sites − 12.5% of the compound − show a weak opposed tendency. The more pronounced and more important process on Te sites can be interpreted as reduced localization and accordingly the reduction of scattering centers. Following this line our finding might explain the smoothly enhanced conductivity upon annealing reported in recent experimental work by Siegrist *et al.* [55]. Here, it is important to note that in the experiment a considerably lower annealing temperature was used than in our simulated annealing. For temperatures higher than approximately 500 K the rocksalt-crystal would in fact experience a phase transition to the hexagonal structure as reported in [55] in Figure 1. The discrepancy to the applied 1000 K in the simulated annealing studies presented here is originated by the underlying assumptions of the kinetic Monte-Carlo analysis. Both neglecting structural relaxations and assuming migration barriers to be constantly high increases the required activation temperature.

Returning to the results of the simulated annealing studies, we find that the change of the distribution of the LDOS at the Fermi energy arises from the fact that vacancy-complexes of $n_{Vac} \geq 4$ are dissolved by annealing. Breaking those complexes leads to less regions of highest LDOS-values, according to a smaller spread between mean and highest probable LDOS which produces the shift to the right. In order to visualize this result the distribution of LDOS in real space is drawn in Fig. 7.21. In the annealed structure considerably less centers of high LDOS are present, which emphasizes the important role of the vacancy distribution in the sample.
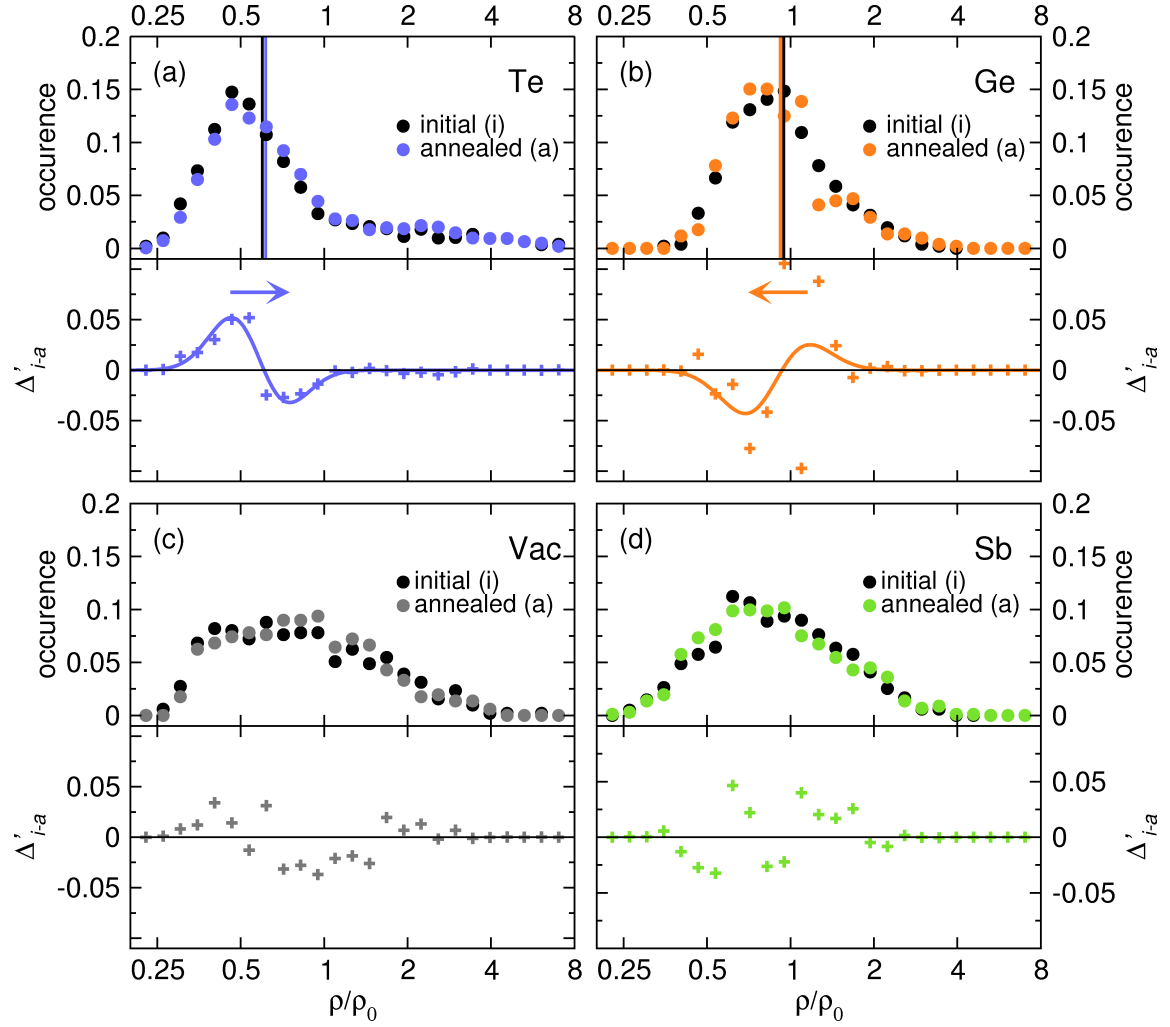
Figure 7.20.: *Distribution of local density of states (LDOS) at E_{F} for $Ge_{.25}Vac_{.25}Sb_{.5}Te$ supercells of 4096 sites in randomly generated configuration (i) and annealed configuration (a). All LDOS values are normalized by its mean LDOS ρ_0. Both distributions (upper panel) and their differences (lower panel) are separately displayed for the LDOS on Te (a), Ge (b), vacancy (c) and Sb (d) sites. For all lower panels the difference as defined in the text is given in all cases relative to the initial configuration (i) supercell Δ'_{i-a}. Lines are least-square fits according to (7.6) to the shown data points.*
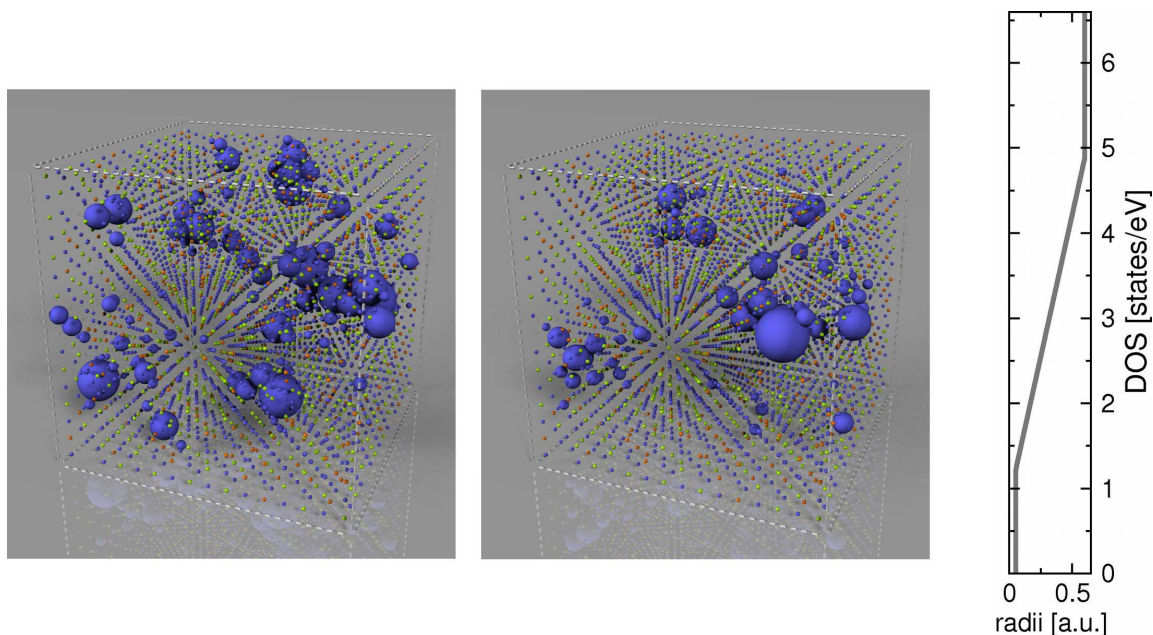
Figure 7.21.: *Spatial distribution of local density of states for a $Ge_{512}Vac_{512}Sb_{1024}Te_{2048}$ supercell for the initial randomly placed configuration (left) and the configuration after simulated annealing with 1000 K and $5 \cdot 10^8$ Monte-Carlo time-steps (right).*

7.4 Summary

In this chapter we have studied the influence of disorder on the electronic states of the phase change material $GeSb_2Te_4$. Our investigations were motivated by recent experimental findings of Siegrist *et al.* [55], who reported a pronounced localization in $GeSb_2Te_4$ and a strong dependency of this localization on annealing. We provide an explanation for both observed effects, which we base on a detailed analysis of the electronic states by utilizing the strength of KKRnano. It turns out that all fundamental local properties of the sites can be well understood by taking into account only the chemical configuration on few neighboring shells. In particular the distribution and ordering of vacancies plays the crucial role for localization. In the vicinity of regions with high vacancy density the Ge-Te and Sb-Te bonding states lie in a pseudo band-gap and are therefore highly localized. In addition those vacancy complexes possess very high density of states at the Fermi energy and act as strong scattering centers in electronic transport. Combining this finding with an analysis of the annealing process with the help of kinetic Monte-Carlo-Simulations we are able to shed light onto the experimentally observed transport properties and its dependency on the annealing temperature: the simulated annealing reveals that the higher the annealing temperatures become the more vacancy complexes are dissolved. As a direct consequence we find that annealing significantly reduces the number and strength of scattering centers which directly leads to higher conductivity.

CHAPTER 8

Conclusion

The primary aim of this thesis was the development of a multiple purpose large scale *ab initio* electronic structure method in order to enable the accurate description of spatially confined, disordered or defectious systems. Here, it is important recall that up to now such large scale methods with high parallel efficiency and linear scaling behavior do not exist for metallic systems. We have achieved our goal by exploiting the computational power of state-of-the-art supercomputers and by developing a new algorithm in the framework of the full potential KKR Green function approach, called KKRnano. Amongst the most important modifications to the standard KKR scheme, the introduction of a screened reference system and the optional truncation of long-range interactions directly led to the desired highly sparse matrix representation. Although our approach to solve the Dyson equation not by direct inversion but rather iteratively enabled an efficient and highly parallel computation, it revealed on the downside unfavorably slow convergency − partly requiring hundreds to thousands of iterations. Besides other optimization steps, two improvements contributed most to the crucial reduction of iterations and an outstanding speedup of KKRnano: First, we have developed a fast scheme to obtain an accurate initial guess and secondly, we have incorporated preconditioning on the basis of a block-circulant representation. By combining all these improvements to the algorithm we have accomplished an $\mathrm{O}(N^2)$ and, if long-range interactions can be disregarded, even an $\mathrm{O}(N)$-scaling with number of atoms (N) in the unit cell. Due to this beneficial scaling we managed to extend the applicability of KKRnano to supercells of more than ten thousand atoms. We have aligned all our efforts to design the above summarized new algorithms for the computation on massively parallel architectures, which results in an high parallel efficiency on the latest generation of supercomputers. By that we were able to introduce multiple levels of parallelization, and we accomplished that KKRnano can be operated with excellent parallel performance on up to hundred thousand processors.

Magnetic defects in Gd doped GaN

As first application of KKRnano we approached the unresolved question on the origin for ferromagnetic high-temperature coupling in gadolinium doped gallium nitride [1]. By utilizing the capability to describe thousands of atoms per unit cell we performed extensive studies on the magnetic coupling of nitrogen and oxygen interstitials as well as gallium vacancies. Here, the large scale of the treated samples allowed for an explicit and elaborate analysis of the electronic structure and the magnetization of the defects. We gained important new insight on the magnetic coupling amongst defects and from defects to gadolinium by utilizing the Lichtenstein formula for the exchange interactions. Further we did not only identify ferro- and antiferromagnetic trends but we were able to go conceptually one step beyond: we could determine the underlying magnetic coupling mechanisms by studying the energy resolved contributions to the exchange coupling. Our results clearly revealed that nitrogen interstitials show a distinct antiferromagnetic coupling and therefore can be ruled out as origin for the experimental observations. Also oxygen interstitials, which provide for part of the defect concentrations a weak ferromagnetic coupling, can only assist magnetic order but cannot be the determining factor for it. Gallium vacancies, however, provide an important coupling mechanism. The vacancies induce due to dangling bond formation large magnetic moments on all surrounding nitrogen sites, which then couple ferromagnetically both amongst themselves and to the gadolinium dopants. Based on the information on extracted exchange coupling constants we evaluated by statistical methods that already small concentrations of vacancies can lead to distinct long-range ferromagnetic ordering. Hence, our calculations revealed strong indications that gallium vacancies are causing the ferromagnetic coupling of colossal magnetic moments in GaN:Gd.

Disorder and localization in $Ge_1Sb_2Te_4$

We have studied with KKRnano the influence of disorder on the electronic states of the technologically important phase change material $Ge_1Sb_2Te_4$. Our work was motivated by two recent experimental findings by Siegrist *et al.* [55], who reported a pronounced localization in $Ge_1Sb_2Te_4$ and a strong dependency of this behavior on annealing. We provided an explanation for both observed effects, which we based on a elaborate calculation and analysis of the electronic structure and which have been − in this form − only possible by the ability to treat supercells of thousands of atoms. Our results led to the general conclusion that all fundamental local properties of the sites can be well understood taking into account only the chemical configuration on few neighboring shells. This observation holds for the localization of states, for which we highlighted the particular importance of the distribution and ordering of vacancies: in the vicinity of regions with high vacancy density the Ge-Te and Sb-Te bonding states lie in a pseudo band-gap and are therefore highly localized. In addition, we performed a series of calculations addressing the finite size scaling of the electronic structure. The statistical analysis of this calculations indeed revealed a dependency on the size of the sample which serves as indication for disorder induced localization. Besides the issue of localization the vacancy complexes likely play an important role for the electronic transport: we found that such vacancy complexes possess very high density of states at the Fermi level and accordingly can act as strong scattering centers. By combining this finding with our analysis of the annealing process by means of kinetic Monte-Carlo-Simulations

we were able to shed light onto the experimentally observed transport properties and their dependency on the annealing temperature. Our model calculations of the annealing process revealed that the more larger vacancy complexes are dissolved, the higher the annealing temperature becomes. As a direct consequence annealing reduces simultaneously both the number and strength of scattering centers with the result of a strongly reduced resistance.
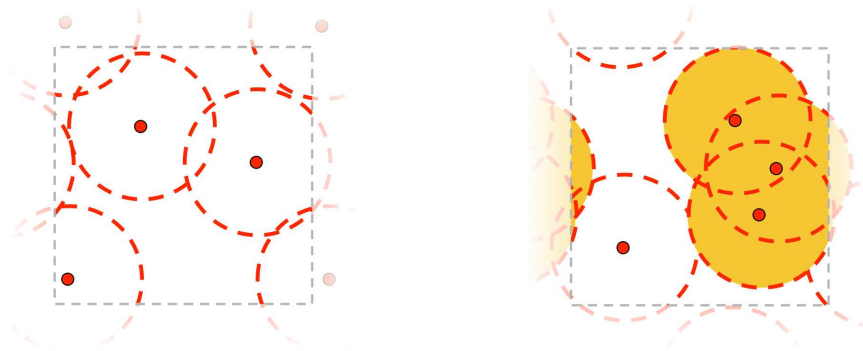
Percolation threshold

Figure A.1.: *Schematic representation of the classification of clusters as defined to obtain the percolation threshold. Two supercells with three (left) and four (right) magnetic defects are shown. On the left none of the defects couple as no second defect is found within the interaction zone. On the right three out of four defects are accounted as coupled and result in a fraction of 75% of saturation magnetization.*

A qualitative measure if a present magnetic coupling can maintain a ferromagnetic structure or remains in a (super)paramagnetic state at a given concentration of magnetic atoms. This geometric problem of magnetic percolation has been previously addressed in the literature [156–158]. We follow the basic concepts of this reports and provide the details on the numerical evaluation we've performed. Assume that around each magnetic atom a critical radius exists, which defines two binary regions: in case other magnetic atoms are located within this radius both atoms are accounted as being coupled. Beyond that critical radius the magnetic interaction is artificially set to zero. Based on this definitions magnetically coupled clusters can be found as exemplified in Fig. A.1.

In the next conceptual step the sizes of those magnetically coupled clusters are determined, typically in periodically repeated unit cells on the order of ten thousand atoms or more. Out of this set of clusters the largest – i.e. the one coupling most atoms – defines an integer value
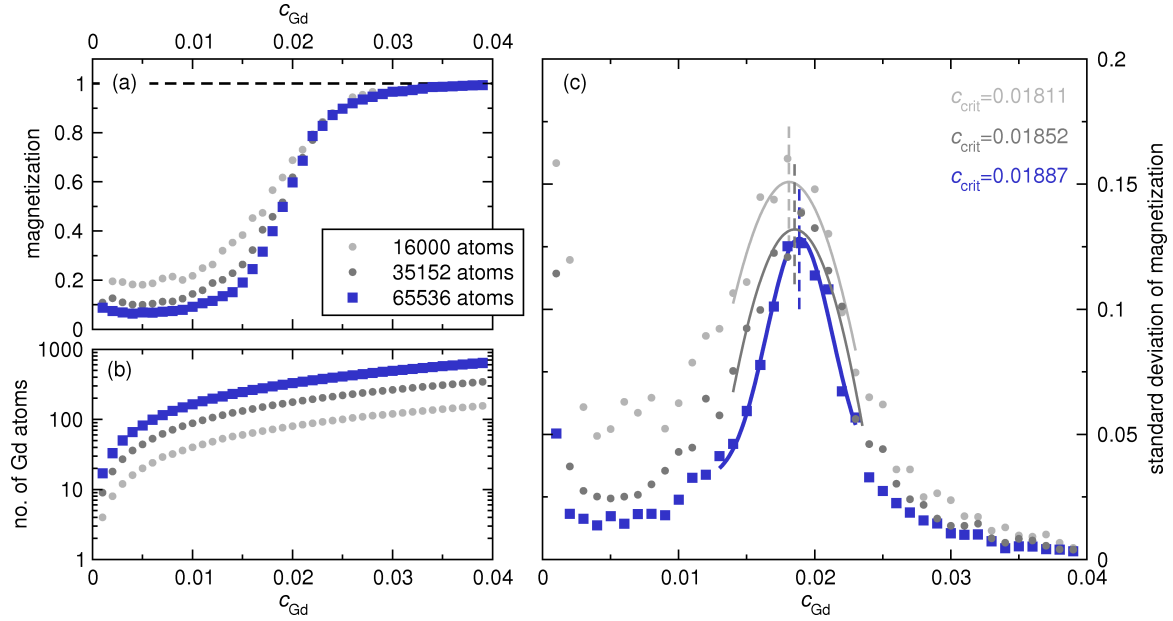
Figure A.2.: *Sample calculation of the percolation threshold exemplified by substitutional Gd in GaN with overall 16000, 35152 and 65536 sites in the supercell. In (a) the average magnetization of 100 different samples at a given concentration is shown for the three system sizes. The saturation magnetization is illustrated in broken black lines. The actual number of Gd-atoms in the sample is given in (b). Figure (c) shows the standard deviation of the magnetization. Solid lines are gaussian-fits to the data in the intervall $[c_{crit} - 0.005, c_{crit} + 0.005]$. The maxima of these obtained fits are the resulting percolation thresholds and are marked with a broken line in the corresponding color.*

of magnetization in the unit cell. Clearly this value underlines strong statistic fluctuations, which in fact carry useful information on the magnetic ordering. In order to determine the size of those fluctuations at each given concentration of magnetic defects several unit cells are randomly generated and its mean value as well as the standard deviation of the magnetization obtained. This set of data which is shown in Fig. A.2 for an exemplatory system serves as basis to evaluate the actual percolation threshold.

It is intuitive to define the transition point - the percolation threshold - as the concentration which shows the largest fluctuations in the magnetization. Apparent from Fig. A.2 this concentration can be obtained accurately from least square fits. Further it becomes clear that the calculation of the percolation threshold in the above presented procedure is relatively insensitive to finite size effects: As long as interactions on the usual length scale of up to $\mathring{A}$ are considered usually ten thousand of atoms are sufficient to guarantee precise values. This finding is validated by evaluation of the percolation threshold from different sample sizes from 16000 to 65536 sites in the supercell in Fig. A.2. Here, even for long-range magnetic interaction of $r_{ij} = 9.92\mathring{A}$ only insignificantly varying percolation thresholds have been obtained.

Bibliography

[1] S. Dhar, O. Brandt, M. Ramsteiner, V. F. Sapega, and K. H. Ploog, Phys. Rev. Lett. **94**, 037205 (2005).

[2] R. Zeller and P. H. Dederichs, Phys. Rev. Lett. **42**, 1713 (1979).

[3] O. Gunnarsson, O. Jepsen, and O. K. Andersen, Phys. Rev. B **27**, 7144 (1983).

[4] J. Kudrnovský, I. Turek, V. Drchal, P. Weinberger, N. E. Christensen, and S. K. Bose, Phys. Rev. B **46**, 4222 (1992).

[5] M. Aldén, I. A. Abrikosov, B. Johansson, N. M. Rosengaard, and H. L. Skriver, Phys. Rev. B **50**, 5131 (1994).

[6] R. J. Elliott, J. A. Krumhansl, and P. L. Leath, Reviews of Modern Physics **46**, 465 (1974).

[7] J. S. Faulkner and G. M. Stocks, Phys. Rev. B **21**, 3222 (1980).

[8] D. A. Rowlands, Reports on Progress in Physics **72**, 086501 (2009).

[9] G. Kresse and D. Joubert, Phys. Rev. B **59**, 1758 (1999).

[10] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, and D. Sánchez-Portal, Journal of Physics Condensed Matter **14**, 2745 (2002).

[11] R. Car and M. Parrinello, Physical Review Letters **55**, 2471 (1985).

[12] J. Vandevondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, Computer Physics Communications **167**, 103 (2005).

[13] M. D. Segall, P. J. D. Lindan, M. J. Probert, C. J. Pickard, P. J. Hasnip, S. J. Clark, and M. C. Payne, Journal of Physics Condensed Matter **14**, 2717 (2002).

[14] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, Journal of Physics Condensed Matter **21**, 5502 (2009).

[15] http://www.top500.org.

[16] J. R. Chelikowsky, N. Troullier, and Y. Saad,
Physical Review Letters **72**, 1240 (1994).

[17] J. R. Chelikowsky, N. Troullier, K. Wu, and Y. Saad,
Phys. Rev. B **50**, 11355 (1994).

[18] K. Hirose, T. Ono, Y. Fujimoto, and S. Tsukamoto, *Unknown*, edited by Imperial
College Press, London (2005).

[19] T. L. Beck, Reviews of Modern Physics **72**, 1041 (2000).

[20] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen,
Phys. Rev. B **71**, 035109 (2005).

[21] W. Kohn, Physical Review Letters **76**, 3168 (1996).

[22] C.-K. Skylaris, P. D. Haynes, A. A. Mostofi, and M. C. Payne,
The Journal of Chemical Physics **122**, 084119 (2005).

[23] D. R. Bowler, R. Choudhury, M. J. Gillan, and T. Miyazaki,
Physica Status Solidi B Basic Research **243**, 989 (2006).

[24] T. Ozaki, Phys. Rev. B **82**, 075131 (2010).

[25] Y. Wang, G. M. Stocks, W. A. Shelton, D. M. C. Nicholson, Z. Szotek, and W. M.
Temmerman, Phys. Rev. Lett. **75**, 2867 (1995).

[26] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[27] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[28] J. Korringa, Physica **13**, 392 (1947).

[29] W. Kohn and N. Rostoker, Phys. Rev. **94**, 1111 (1954).

[30] B. Drittler, *KKR-Greensche Funktionsmethode für das volle Zellpotential*, Berichte
des Forschungszentrums Jülich (Forschungszentrum Jülich, 1991).

[31] R. Zeller, P. H. Dederichs, B. Újfalussy, L. Szunyogh, and P. Weinberger,
Phys. Rev. B **52**, 8807 (1995).

[32] R. Zeller, Journal of Physics Condensed Matter **20**, C4215+ (2008).

[33] A. V. Smirnov and D. D. Johnson, Phys. Rev. B **64**, 235129 (2001).

[34] P. Lloyd, Proceedings of the Physical Society **90**, 207 (1967).

[35] R. Zeller, J. Phys. Condens. Matter **16**, 6453 (2004).

[36] R. Zeller, J. Phys. Condens. Matter **20**, 035220 (2008).

[37] V. I. Anisimov, F. Aryasetiawan, and A. I. Lichtenstein,
J. Phys. Condens. Matter **9**, 767 (1997).

[38] A. I. Lichtenstein, M. I. Katsnelson, V. P. Antropov, and V. A. Gubanov,
Journal of Magnetism and Magnetic Materials **67**, 65 (1987).

[39] G. A. Prinz, Science **282**, 1660 (1998).

[40] A. H. MacDonald, P. Schiffer, and N. Samarth, Nature Materials **4**, 195 (2005).

[41] J. K. Furdyna, Journal of Applied Physics **64**, 29 (1988).

[42] H. Ohno, Science **281**, 951 (1998).

[43] P. Sharma, A. Gupta, K. V. Rao, F. J. Owens, R. Sharma, R. Ahuja, J. M. O.
Guillen, B. Johansson, and G. A. Gehring, Nature Materials **2**, 673 (2003).

[44] T. Jungwirth, J. Sinova, J. Mašek, J. Kučera, and A. H. MacDonald,
Rev. Mod. Phys. **78**, 809 (2006).

[45] A. Bonanni, Semiconductor Science Technology **22**, 41 (2007).

[46] K. Sato, L. Bergqvist, J. Kudrnovský, P. H. Dederichs, O. Eriksson, I. Turek,
B. Sanyal, G. Bouzerar, H. Katayama-Yoshida, V. A. Dinh, T. Fukushima,
H. Kizaki, and R. Zeller, Rev. Mod. Phys. **82**, 1633 (2010).

[47] M. A. Khaderbad, S. Dhar, L. Pérez, K. H. Ploog, A. Melnikov, and A. D. Wieck,
Applied Physics Letters **91**, 072514 (2007).

[48] A. Ney, T. Kammermeier, E. Manuel, V. Ney, S. Dhar, K. H. Ploog, F. Wilhelm,
and A. Rogalev, Applied Physics Letters **90**, 252515 (2007).

[49] G. M. Dalpian and S.-H. Wei, Phys. Rev. B **72**, 115201 (2005).

[50] C. Mitra and W. R. L. Lambrecht, Phys. Rev. B **80**, 081202 (2009).

[51] L. Liu, P. Y. Yu, Z. Ma, and S. S. Mao, Phys. Rev. Lett. **100**, 127203 (2008).

[52] Y. Gohda and A. Oshiyama, Phys. Rev. B **78**, 161201 (2008).

[53] M. Wuttig and N. Yamada, Nat. Mater. **6**, 824 (2007).

[54] T. Matsunaga, J. Akola, S. Kohara, T. Honma, K. Kobayashi, E. Ikenaga, R. O.
Jones, N. Yamada, M. Takata, and R. Kojima, Nat. Mater. **10**, 129 (2011).

[55] T. Siegrist, P. Jost, H. Volker, M. Woda, P. Merkelbach, C. Schlockermann, and
M. Wuttig, Nat. Mater. **10**, 202 (2011).

[56] G. Schubert, J. Schleede, K. Byczuk, H. Fehske, and D. Vollhardt,
Phys. Rev. B **81**, 155106 (2010).

[57] J. Heyd, G. E. Scuseria, and M. Ernzerhof,
The Journal of Chemical Physics **118**, 8207 (2003).

[58] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998).

[59] A. H. MacDonald and S. H. Vosko,
Journal of Physics C: Solid State Physics **12**, 2977 (1979).

[60] U. von Barth and L. Hedin, Journal of Physics C Solid State Physics **5**, 1629 (1972).

[61] M. Levy, Phys. Rev. A **26**, 1200 (1982).

[62] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989).

[63] U. von Barth and L. Hedin, J. Phys. C: Solid State Phys. **5**, 1629 (1972).

[64] M. M. Pant and A. K. Rajagopal, Solid State Communications **10**, 1157 (1972).

[65] K. Capelle, Brazilian Journal of Physics **36**, 1318 (2006).

[66] D. Ceperley and B. J. Alder, *Unknown*, edited by Ceperley, D. & Alder, B. J. (1980).

[67] S. H. Vosko, L. Wilk, and M. Nusair, Canadian Journal of Physics **58**, 1200 (1980).

[68] J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. **77**, 3865 (1996).

[69] T. H. Dupree, Annals of Physics **15**, 63 (1961).

[70] G. J. Morgan, Proceedings of the Physical Society **89**, 365 (1966).

[71] N. Papanikolaou, R. Zeller, and P. H. Dederichs,
Journal of Physics Condensed Matter **14**, 2799 (2002).

[72] P. Mavropoulos and N. Papanikolaou, in *Computational Nanoscience: Do It
Yourself!* (John von Neumann Institute for Computing, Jülich, 2006) pp. 131–158.

[73] S. Lounis, *Theory of Magnetic Transition Metal Nanoclusters on Surfaces*, Ph.D.
thesis, RWTH Aachen (2007).

[74] R. Zeller, Journal of Physics C Solid State Physics **20**, 2347 (1987).

[75] F. Meier, L. Zhou, J. Wiebe, and R. Wiesendanger, Science **320**, 82 (2008).

[76] S. Lounis, P. Mavropoulos, P. H. Dederichs, and S. Blügel,
Phys. Rev. B **72**, 224437 (2005).

[77] H. Shiba, Progress of Theoretical Physics **46**, 77 (1971).

[78] C. W. M. Castleton, A. Höglund, and S. Mirbt,
Modelling Simul. Mater. Sci. Eng. **17**, 084003 (2009).

[79] M. E. Gruner, G. Rollmann, P. Entel, and M. Farle,
Phys. Rev. Lett. **100**, 087203 (2008).

[80] Z. Wu, J. B. Neaton, and J. C. Grossman, Phys. Rev. Lett. **100**, 246804 (2008).

[81] http://www.netlib.org/benchmark/performance.ps.

[82] http://www.fz-juelich.de/jsc/jugene.

[83] http://www.fz-juelich.de/jsc/juropa.

[84] http://www.nccs.gov/computing-resources/jaguar/.

[85] R. W. Freund and N. M. Nachtigal, Numerische Mathematik **60**, 315 (1991), 10.1007/BF01385726.

[86] Y. Saad, *Iterative method for sparse linear systems* (Society for Industrial and Applied Mathematics, Philadelphia, 2003).

[87] Y. Saad and M. H. Schultz, SIAM Journal on Scientific and Statistical Computing **7**, 856 (1986).

[88] C. Lanczos, J. Res. Natl. Bur. Stand **49**, 33 (1952).

[89] P. Sonneveld, SIAM Journal on Scientific and Statistical Computing **10**, 36 (1989).

[90] R. W. Freund, SIAM Journal on Scientific Computing **14**, 470 (1993).

[91] Y. Saad, SIAM J. Sci. Comput. **14**, 461 (1993).

[92] M. Benzi and M. Tuma, SIAM Journal on Scientific Computing **19**, 968 (1998).

[93] M. Bolten, A. Thiess, I. Yavneh, and R. Zeller, Linear Algebra Appl. **436**, 436 (2012).

[94] Z. Szebenyi, B. J. N. Wylie, and F. Wolf, in *Lecture Notes in Computer Science* (2008) pp. 99–123.

[95] P. Mavropoulos, (2011), private communication.

[96] D. A. Rowlands, A. Ernst, B. L. Györffy, and J. B. Staunton, Phys. Rev. B **73**, 165122 (2006).

[97] T. Hoshino, M. Asato, R. Zeller, and P. H. Dederichs, Phys. Rev. B **70**, 094118 (2004).

[98] P. Bose, A. Ernst, I. Mertig, and J. Henk, Phys. Rev. B **78**, 092403 (2008).

[99] B. Drittler, M. Weinert, R. Zeller, and P. H. Dederichs, Phys. Rev. B **39**, 930 (1989).

[100] R. Zeller, Journal of Physics Condensed Matter **17**, 5367 (2005).

[101] C. Friedrich, M. C. Müller, and S. Blügel, Phys. Rev. B **83**, 081101 (2011).

[102] V. I. Anisimov, J. Zaanen, and O. K. Andersen, Phys. Rev. B **44**, 943 (1991).

[103] P. W. Anderson, Phys. Rev. **124**, 41 (1961).

[104] M. Cococcioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005).

[105] C. Ederer and N. A. Spaldin, Phys. Rev. B **71**, 060401 (2005).

[106] J. Seidel, L. W. Martin, Q. He, Q. Zhan, Y.-H. Chu, A. Rother, M. E. Hawkridge, P. Maksymovych, P. Yu, M. Gajek, N. Balke, S. V. Kalinin, S. Gemming, F. Wang, G. Catalan, J. F. Scott, N. A. Spaldin, J. Orenstein, and R. Ramesh, Nature Materials **8**, 229 (2009).

[107] H. Ebert, Solid State Communications **127**, 443 (2003).

[108] M. T. Czyżyk and G. A. Sawatzky, Phys. Rev. B **49**, 14211 (1994).

[109] O. Gunnarsson, O. K. Andersen, O. Jepsen, and J. Zaanen, Phys. Rev. B **39**, 1708 (1989).

[110] Ü. Özgür, Y. I. Alivov, C. Liu, A. Teke, M. A. Reshchikov, S. Doğan, V. Avrutin, S.-J. Cho, and H. Morkoç, Journal of Applied Physics **98**, 041301 (2005).

[111] C. N. R. Rao and F. L. Deepak, J. Mater. Chem. **15**, 573 (2005).

[112] N. Teraguchi, A. Suzuki, Y. Nanishi, Y.-K. Zhou, M. Hashimoto, and H. Asahi, Solid State Communications **122**, 651 (2002).

[113] A. Ney, T. Kammermeier, V. Ney, S. Ye, K. Ollefs, E. Manuel, S. Dhar, K. H. Ploog, E. Arenholz, F. Wilhelm, and A. Rogalev, Phys. Rev. B **77**, 233308 (2008).

[114] C. Mitra, *Ab-initio calculations for dilute magnetic semiconductors*, Ph.D. thesis, Case Western Reserve University (2009).

[115] J. K. Hite, R. M. Frazier, R. Davies, G. T. Thaler, C. R. Abernathy, S. J. Pearton, and J. M. Zavada, Applied Physics Letters **89**, 092119 (2006).

[116] P. Dev, Y. Xue, and P. Zhang, Phys. Rev. Lett. **100**, 117204 (2008).

[117] F. Lo, A. Melnikov, D. Reuter, A. D. Wieck, V. Ney, T. Kammermeier, A. Ney, J. Schörmann, S. Potthast, D. J. As, and K. Lischka, Applied Physics Letters **90**, 262505 (2007).

[118] J. Neugebauer and C. G. Van de Walle, Phys. Rev. B **50**, 8067 (1994).

[119] J. Neugebauer and C. G. Van de Walle, MRS Proceedings **339**, 687 (1994).

[120] L. Bergqvist, K. Sato, H. Katayama-Yoshida, and P. H. Dederichs, Phys. Rev. B **83**, 165201 (2011).

[121] J. Filhol, R. Jones, M. J. Shaw, and P. R. Briddon, Applied Physics Letters **84**, 2841 (2004).

[122] M. Hashimoto, S. Emura, R. Asano, H. Tanaka, N. Teraguchi, A. Suzuki, Y. Nanishi, T. Honma, N. Umesaki, and H. Asahi, Physica status solidi (c) **0**, 26502653 (2003).

[123] A. Svane, N. E. Christensen, L. Petit, Z. Szotek, and W. M. Temmerman, Phys. Rev. B **74**, 165204 (2006).

[124] T. Cheiwchanchamnangij, A. Punya, and W. R. L. Lambrecht,
MRS Online Proceedings Library , 1290 (2011).

[125] A. Rubio, J. L. Corkill, M. L. Cohen, E. L. Shirley, and S. G. Louie,
Phys. Rev. B **48**, 11810 (1993).

[126] C. Mitra and W. R. L. Lambrecht, Phys. Rev. B **78**, 134421 (2008).

[127] H. J. Trodahl, A. R. H. Preston, J. Zhong, B. J. Ruck, N. M. Strickland, C. Mitra,
and W. R. L. Lambrecht, Phys. Rev. B **76**, 085211 (2007).

[128] P. Larson and W. R. L. Lambrecht, Phys. Rev. B **74**, 085108 (2006).

[129] C. G. Broyden, Math. Comput. **19**, 577 (1965).

[130] D. G. Anderson, J. ACM **12**, 547 (1965).

[131] S. Limpijumnong and C. G. Van de Walle, Phys. Rev. B **69**, 035207 (2004).

[132] C. G. van de Walle and J. Neugebauer, Journal of Applied Physics **95**, 3851 (2004).

[133] T. Cheiwchanchamnangij, (2010), private communication.

[134] M. Methfessel, M. van Schilfgaarde, and R. A. Casali, in *Lecture Notes in Physics
Vol. 535*, edited by H. Dreyssé (Springer-Verlag, Berlin, 2000) p. 114.

[135] B. Belhadji, *Ab-initio calculations for dilute magnetic semiconductors*, Ph.D. thesis,
RWTH Aachen (2009).

[136] J. B. Goodenough, Phys. Rev. **100**, 564 (1955).

[137] A. Bedoya-Pinto, J. Malindretos, M. Roever, D. D. Mai, and A. Rizzi,
Phys. Rev. B **80**, 195208 (2009).

[138] M. Roever, J. Malindretos, A. Bedoya-Pinto, A. Rizzi, C. Rauch, and F. Tuomisto,
ArXiv e-prints (2011), arXiv:1103.4256 [cond-mat.mtrl-sci] .

[139] T. Mattila and R. M. Nieminen, Phys. Rev. B **54**, 16676 (1996).

[140] J. C. Zolper, R. G. Wilson, S. J. Pearton, and R. A. Stall,
Applied Physics Letters **68**, 1945 (1996).

[141] J. Neugebauer and C. G. van de Walle, Applied Physics Letters **69**, 503 (1996).

[142] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller,
Journal of Chemical Physics **21**, 1087 (1953).

[143] S. Dhar, T. Kammermeier, A. Ney, L. Pérez, K. H. Ploog, A. Melnikov, and A. D.
Wieck, Applied Physics Letters **89**, 062503 (2006).

[144] S. Dhar, L. Pérez, O. Brandt, A. Trampert, K. H. Ploog, J. Keller, and
B. Beschoten, Phys. Rev. B **72**, 245203 (2005).

[145] N. Yamada, E. Ohno, K. Nishiuchi, N. Akahira, and M. Takao, J. Appl. Phys. **69**, 2849 (1991).

[146] A. Kolobov, P. Fons, A. Frenkel, A. Ankudinov, J. Tominaga, and T. Uruga, Nat. Mater. **3**, 703 (2004).

[147] W. Wełnic, A. Pamungkas, R. Detemple, C. Steimer, S. Blügel, and M. Wuttig, Nat. Mater. **5**, 56 (2006).

[148] Z. Sun, J. Zhou, and R. Ahuja, Phys. Rev. Lett. **96**, 055507 (2006).

[149] C. Lang, S. A. Song, D. N. Manh, and D. J. H. Cockayne, Phys. Rev. B **76**, 054101 (2007).

[150] D. Lencer, M. Salinga, B. Grabowski, T. Hickel, J. Neugebauer, and M. Wuttig, Nat. Mater. **7**, 972 (2008).

[151] K. Shportko, S. Kremers, M. Woda, D. Lencer, J. Robertson, and M. Wuttig, Nat. Mater. **7**, 653 (2008).

[152] T. Fujiwara and T. Yokokawa, Phys. Rev. Lett. **66**, 333 (1991).

[153] N. C. Murphy, R. Wortis, and W. A. Atkinson, Phys. Rev. B **83**, 184206 (2011).

[154] F. Evers and A. D. Mirlin, Phys. Rev. Lett. **84**, 3690 (2000).

[155] F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).

[156] L. Bergqvist, O. Eriksson, J. Kudrnovský, V. Drchal, P. Korzhavyi, and I. Turek, Phys. Rev. Lett. **93**, 137202 (2004).

[157] K. Sato, W. Schweika, P. H. Dederichs, and H. Katayama-Yoshida, Phys. Rev. B **70**, 201202 (2004).

[158] S. Hilbert and W. Nolting, Phys. Rev. B **70**, 165203 (2004).

Acknowledgements

First I like to express my gratitude to Prof. Dr. Stefan Blügel for giving me the privilege to write this thesis at the Peter-Grünberg Insitut and for supervising this thesis. Special thanks for many important discussions and ideas as well as the continuous enthusiasm for this work.

I am thankful to Prof. Dr. Riccardo Mazzarello for his willingness to examine this work as second referee and for the fruitful discussions on the localization effects in phase change materials.

Particular thanks go to Dr. Rudolf Zeller for supervising my work and providing me access to his outstandingly rich fund of experience with the KKR method and for his important ideas especially concerning the development of KKRnano.

I am thankful to Prof. Dr. Peter Dederichs for his exceptional and profound interest for all subtopics of this thesis and many important discussions and suggestions which led to crucial improvements of this work.

It is my pleasure to thank Prof. Dr. Walter Lambrecht for his kind hospitality during my research stay in Ohio and for many discussions and advises regarding GaN:Gd. In this context, I like to gratefully acknowledge the financial support by the DAAD provided for this project.

I like to thank Prof. Dr. Matthias Bolten and Prof. Dr. Irad Yavneh for the fruitful discussions and efficient collaboration on the development of the block-circulant preconditioning technique.

Thanks go to Dr. Dominic Lencer and Dr. Martin Salinga for providing important information on their previous studies and on the experimental insight on phase change materials.

I like to express many thanks to Dr. Phivos Mavropoulos, Prof. Dr. Yura Mokrousov, Prof. Dr. Marjana Lezaic, David Bauer, Dr. Rudolf Zeller and Prof. Peter Dederichs for carefully proofreading this thesis.

I am gratefully acknowledging the opportunity to join the German Research School for simulational sciences and benefit from the interdisciplinar scientific environment − a possibility for which I like to thank in particular Prof. Dr. Heiner Müller-Krumbhaar.

Many thanks go to the large number of members of the Peter-Grünberg Institut and the Insitute for Advanced Simulation at the Forschungszentrum Jülich, the German Research

Curriculum Vitae

Personal data

Name:	Alexander Reinhold Thieß
Date of birth:	26. September 1982 in Neustadt i.H., Germany
Address:	Institute for Advanced Simulation, Forschungszentrum Jülich, D-52425 Jülich

Employment

since 06/2011	Researcher at the Institute for Advanced Simulation at the Forschungszentrum Jülich.
07/2010 - 10/2010	Visiting scientist at the Case Western Reserve University in Cleveland, Ohio
07/2009 - 06/2011	Doctoral Candidate at the German Research School for Simulation Sciences, Aachen/Jülich
01/2008 - 06/2011	Researcher on a PhD position in the Theory Group of Prof. Dr. S. Blügel at the Peter Grünberg Institut and the Institute for Advanced Simulation at the Forschungszentrum Jülich.
10/2007 - 12/2007	Researcher at the Institute for Applied Physics, University of Hamburg

Education

09/2007	Diploma in Physics, mark: 1,0
10/2006 – 09/2007	Diploma Thesis in the Theory Group of Prof. Dr. S. Heinze at the Institute for Applied Physics, University of Hamburg:
01/2005	Prediploma in Physics, mark: 1,1
10/2002 – 09/2007	Studies in Physics at the University of Hamburg
09/2001 – 07/2002	Compulsory community service
08/1992 – 06/2001	Dietrich Bonhoeffer Gymnasium, Quickborn (Holstein), Degree: Abitur (mark: 1,5)
08/1989 – 07/1992	Grundschule Mühlenberg, Quickborn (Holstein)

Awards

06/2010	DAAD scholarship for a period of three month starting from 1st of July 2010 at the Case Western Reserve University, Ohio (USA)
07/2009	Prize for the best students poster at the International Colloquium on Magnetic Films and Surfaces (ICMFS), Berlin 2009
07/2008	Prize for the best students talk at the European Congress of Surface Science (ECOSS), Liverpool 2008

List of Publications

A. Thiess, R. Zeller, M. Bolten, P. H. Dederichs, and S. Blügel,
Massively parallel density functional calculations for thousands of atoms: KKRnano,
submitted to Phys. Rev. B

M. Bolten, A. Thiess, I. Yavneh, and R. Zeller,
Preconditioning systems arising from the KKR Green function method using block-circulant matrices,
Linear Algebra Appl. **436**, 436 (2012).

A. Punya, T. Cheiwchanchamnangij, A. Thiess, and W. R. L. Lambrecht,
First-principles Study of Nitrogen Vacancies in GdN,
MRS Proceedings, 1290 (2011)

A. Thiess, Y. Mokrousov, and S. Heinze,
Competing magnetocrystalline anisotropy energies in suspended transition metal trimers,
Phys. Rev. B **81**, 054433 (2010).

G. Bihlmayer, M. Heide, Y. Mokrousov, A. Thiess, S. Heinze and S. Blügel,
Spin-orbit induced phenomena at surfaces: wires and surfaces.,
John von Neumann Institute for Computing Symposium 2010, Proceedings, IAS Series Vol. 3 (2010).

A. Thiess, Y. Mokrousov, S. Heinze and S. Blügel,
Magnetically hindered chain formation in transition-metal break junctions,
Phys. Rev. Lett. **103**, 217201 (2009).

Y. Mokrousov, A. Thiess, and S. Heinze,
Structurally driven magnetic state transition of biatomic Fe chains on Ir(001),
Phys. Rev. B **80**, 195420 (2009).

A. Thiess, Y. Mokrousov, S. Blügel and S. Heinze
Theory and application of chain formation in break junctions,
Nano Letters **8**, 2144 (2008).