

The LDA+DMFT approach to strongly correlated materials

Eva Pavarini, Erik Koch, Dieter Vollhardt, and Alexander Lichtenstein (Eds.)

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation

Band / Volume 1

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation (IAS)

German Research School for
Simulation Sciences GmbH

Lecture Notes of the Autumn School 2011
Hands-on LDA+DMFT

Eva Pavarini, Erik Koch, Dieter Vollhardt, and Alexander Lichtenstein (Eds.)

The LDA+DMFT approach to strongly correlated materials

Autumn school organized by the
DFG Research Unit 1346
Dynamical Mean-Field Approach
with Predictive Power
for Strongly Correlated Materials

at the Forschungszentrum Jülich
on 4 – 7 October 2011

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation

Band / Volume 1

ISSN 2192-8525

ISBN 978-3-89336-734-4

Bibliographic information published by the Deutsche Nationalbibliothek.
The Deutsche Nationalbibliothek lists this publication in the Deutsche
Nationalbibliografie; detailed bibliographic data are available in the
Internet at <http://dnb.d-nb.de>.

Publisher: Forschungszentrum Jülich GmbH
Institute for Advanced Simulations

Cover Design: Grafische Medien, Forschungszentrum Jülich GmbH

Printer: Druckerei Gehler, Düren-Birkesdorf

Copyright: Forschungszentrum Jülich 2011

Distributor: Forschungszentrum Jülich
Zentralbibliothek, Verlag
D-52425 Jülich
Phone +49 (0)2461 61-5368 · Fax +49 (0)2461 61-6103
e-mail: zb-publikation@fz-juelich.de
Internet: <http://www.fz-juelich.de>

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation Band / Volume 1

ISSN 2192-8525
ISBN 978-3-89336-734-4

The complete volume is freely available on the Internet on the Jülicher Open Access Server (JUWEL) at
<http://www.fz-juelich.de/zb/juwel>

Neither this book nor any part of it may be reproduced or transmitted in any form or by any
means, electronic or mechanical, including photocopying, microfilming, and recording, or by any
information storage and retrieval system, without permission in writing from the publisher.

Contents

Preface

Introduction

- 1 Dynamical Mean-Field Approach for Strongly Correlated Materials
Dieter Vollhardt
- 2 Theory and Practice of Density-Functional Theory
Peter Blöchl
- 3 Model Hamiltonians and Basic Techniques
Frank Lechermann
- 4 Wannier Functions and Construction of Model Hamiltonians
Jan Kuneš
- 5 Introduction to Dynamical Mean-Field Theory
Marcus Kollar
- 6 The LDA+DMFT Approach
Eva Pavarini
- 7 The Constrained RPA Method for Calculating the Hubbard U from First-Principles
F. Aryasetiawan, T. Miyake, and R. Sakuma
- 8 The Lanczos Method
Erik Koch
- 9 Hirsch-Fye Quantum Monte Carlo Method for Dynamical Mean-Field Theory
Nils Blümer
- 10 Continuous-Time Impurity Solvers
Philipp Werner
- 11 Non-Local Correlation Effects in Solids: Beyond DMFT
A. I. Lichtenstein and H. Hafermann
- 12 Multiple-Scattering Formalism for Correlated Systems: a KKR-DMFT Approach
H. Ebert, J. Minár, and D. Ködderitzsch
- 13 Hedin Equations, GW , GW +DMFT, and All That
K. Held, C. Taranto, G. Rohringer, and A. Toschi
- 14 Challenges from Experiment
Hao Tjeng

Index

Preface

Soon after the discovery of the basic principles of quantum mechanics theorists set out to explain the properties of solids from a first-principles, atomistic perspective. However, it soon became clear that theoretical methods based on the calculation of fermionic many-particle wave functions are notoriously difficult to handle. A crucial step forward was density-functional theory (DFT) and its local-density approximation (LDA). The success of DFT in explaining the physical and chemical properties of solids is remarkable. Nevertheless, LDA and its generalizations fail for systems whose low-energy properties are dominated by electron-electron correlations, such as Mott-insulating transition-metal oxides, Kondo and heavy-fermion materials, organic crystals, and many others. The realistic description of these strongly correlated materials remains, to date, one of the grand challenges of condensed matter-physics.

During the last few years conventional band-structure calculations in the local density approximation (LDA) have been merged with a modern many-body approach, the dynamical mean-field theory (DMFT), into a novel computational method referred to as LDA+DMFT. This framework has proved to be a breakthrough for the realistic modeling of the electronic, magnetic, and structural properties of materials such as transition metals and their oxides. Nevertheless the LDA+DMFT approach still needs to be considerably advanced to be able to treat increasingly complex systems. This requires, for example, an improvement of the interface between the band structure and many-body constituents of the approach, the refinement and integration of efficient impurity solvers, the realistic computation of free energies and forces, and the development of schemes to treat non-local correlations. For this purpose 25 researchers from 16 different institutions in the German-speaking part of Europe joined forces and established the Research Unit FOR 1346 on *Dynamical Mean-Field Approach with Predictive Power for Strongly Correlated Materials*, which is funded by the Deutsche Forschungsgemeinschaft since July 2010. It is the goal of this Research Unit to develop the LDA+DMFT framework into a comprehensive *ab initio* approach which will be able to describe, and eventually even predict, the properties of complex correlated materials.

By organizing the 2011 Autumn School *Hands-on LDA+DMFT* the researchers of the DFG Research Unit FOR 1346 offer a practical introduction into the LDA+DMFT approach for graduate students and young researchers in this novel branch of condensed matter physics.

The school covers the following topics

- Electronic correlations
- Basic DFT
- Model Hamiltonians
- Wannier functions
- DMFT
- LDA+DMFT
- Lanczos
- Hirsch-Fye quantum Monte Carlo
- Continuous-time quantum Monte Carlo
- Cluster DMFT and dual-Fermion approach
- KKR+DMFT
- GW+DMFT
- Challenges from experiments

We are grateful to all contributors for the time and effort spent in preparing the lectures as well as for their spontaneous help and support for this new school:

- Ferdi Aryasetiawan, Chiba University, Japan
- Peter Blöchl, Clausthal University of Technology
- Nils Blümer, Mainz University
- Karsten Held, Vienna Technical University, Austria
- Hubert Ebert, LMU Munich
- Marcus Kollar, Augsburg University
- Jan Kuneš, Institute of Physics, Academy of Sciences, Praha, Czech Republic
- Frank Lechermann, Hamburg University
- Hao Tjeng, Max Planck Institute for Chemical Physics of Solids, Dresden
- Philipp Werner, ETH Zurich, Switzerland

We are especially grateful to the contributors that prepared in addition tutorials, without which the school would have not be complete

- Ferdi Aryasetiawan, Chiba University
- Nils Blümer, Mainz University
- Jan Kuneš, Institute of Physics, Academy of Sciences, Praha
- Philipp Werner, ETH Zurich

We heartily thank all the colleagues and collaborators that helped us in proofreading the manuscripts, even on short notice

- Fabio Baruffa, GRS Jülich
- Michael Baumgärtel, GRS Jülich
- Andreas Flesch, IAS Jülich
- Christoph Friedrich, IAS Jülich
- Evgeny Gorelov, IAS Jülich
- Robert Jones, GRS Jülich
- Amin Kiani Shaikhabadi, IAS Jülich
- Joaquin Miranda, GRS Jülich
- German Ulm, GRS Jülich
- Rudolf Zeller, IAS Jülich
- Guoren Zhang, IAS Jülich

We are very grateful for the financial and organisational support we have received in organizing this school. The DFG Forschergruppe FOR 1346 provided the impetus for the school and a large part of the financial support. The Institut für Festkörperforschung and the German Research School for Simulation Sciences at the Forschungszentrum Jülich were vital for the local organization and the production of this book. The International Institute for Complex Adaptive Matter (ICAM) provided travel support.

The nature of a School requires the lecture-notes to be available during the lectures. In this way the participant have the chance to work through the lectures thoroughly while they are given. We are very grateful that all lecturers provided their manuscripts in time for the production of this book. We foresee that the lecture notes collected here will serve as a future standard entry point into the LDA+DMFT approach to strongly correlated materials. We thank Mrs. H. Lexis of the Forschungszentrum Jülich Verlag and Mr. D. Laufenberg of the Graphische Betriebe for providing their expert support in producing the present volume on a tight schedule and for making even seemingly impossible requests possible.

Finally, our special thanks go to Dipl.-Ing. R. Hölzle for his invaluable advice on all questions concerning the organization of such a school and to Mrs. L. Snyders for expertly handling all practical issues.

Eva Pavarini, Erik Koch, Dieter Vollhardt, and Alexander Lichtenstein

September 2011

I Introduction

Eva Pavarini and Erik Koch
Institute for Advanced Simulation
German Research School for Simulation Sciences
Forschungszentrum Jülich GmbH

Contents

1 Emergent behavior	2
2 Electronic Structure	5
3 The LDA+DMFT approach	7
4 Overview of the school	8

1 Emergent behavior

Understanding the laws that govern the universe is dream as old as mankind. Its motivation is the quest to reduce complexity to simplicity, succinctly expressed by Einstein [1]

The supreme test of the physicist is to arrive at those universal laws from which the cosmos can be built up by deduction.

This reductionist program achieved remarkable successes, starting with Newton's understanding that the gravitational pull on an apple is the same that holds the moon on its orbit around earth. Maxwell managed to describe the quite different phenomena of electricity and magnetism on the same footing, developing the theory of electromagnetism. The electromagnetic force could later be unified with the weak interaction into the electro-weak force, one of the pillars of the *standard model*. But there were also problems along the way. Theories became more and more complex as a whole zoo of elementary particles was discovered. Then, with the introduction of quarks, today's elementary particles became tomorrow's compound objects, while the new elementary particles ceased to have a meaning as independent objects. Interestingly, the considerable changes in what was believed to be the fundamental *Theory of Everything* had remarkably little influence on our understanding of the physics at lower energy scales: The quark-dynamics has hardly any relevance for understanding chemical bonding. When vastly different energy scales are involved, a clear-cut separation of the physical description into almost independent layers takes place.

The relevant elementary particles for describing matter at the energy scale of our everyday experience are the atomic nuclei, usually considered as point-charges, and, of course, the electrons, giving rise to *electronic structure*. The fundamental laws governing these are given by the innocent looking eigenvalue problem

$$H|\Psi\rangle = E|\Psi\rangle \quad (1)$$

where the Hamiltonian for a set of atomic nuclei $\{\alpha\}$ with atomic numbers $\{Z_\alpha\}$ and nuclear masses $\{M_\alpha\}$ and their accompanying $\{i\}$ electrons is given, in atomic units, by

$$H = -\frac{1}{2} \sum_i \nabla_i^2 - \sum_\alpha \frac{1}{2M_\alpha} \nabla_\alpha^2 - \sum_{\alpha,i} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \frac{1}{2} \sum_{ii'} \frac{1}{|\mathbf{r}_i - \mathbf{r}_{i'}|} + \frac{1}{2} \sum_{\alpha\alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{R}_\alpha - \mathbf{R}_{\alpha'}|}.$$

This equation, augmented by gravitational potentials and lowest-order relativistic corrections (spin) as the microscopic basis of magnetism, account for the phenomena of our everyday experience, i.e., they are a *Theory of almost Everything* [3,4]. Therefore, already shortly after the formulation of the Schrödinger equation, Dirac [5] remarked that the theory behind atomic and condensed-matter physics, as well as chemistry is completely known

The underlying laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that exact applications of these laws lead to equations which are too complicated to be soluble. It therefore becomes desirable that approximate practical

methods of applying quantum mechanics should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation.

From the reductionist point of view, this closes the case on chemistry and condensed-matter physics. Solving (1) is an exercise in applied mathematics, merely a practical problem. However, because of the quantum many-body nature of the problem, finding the solution, though possible in theory, is impossible in practice. To understand why, we consider a canonical example, a single atom of iron. Having 26 electrons, its wavefunction $\Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_{26})$ is a function of 78 coordinates. What does it take to store such a wave function? If we record Ψ at merely ten values for each coordinate, we would have to record 10^{78} values. There is not enough matter in the visible universe for storing even such a ridiculously crude representation of the wave function of a single iron-atom. This complexity of the wave function is the essence of the *many-body problem*. Already Laplace [6] realized that indeed the grand goal of simulating the world, even if it was possible in theory, in practice can never be reached

An intelligent being, who, at a given moment, knows all the forces that cause nature to move and the positions of the objects that it is made from, if also it is powerful enough to analyze his data, would have described in the same formula the movements of the largest bodies of the universe and those of the lightest atoms. [...] Although scientific research steadily approaches the abilities of this intelligent being, complete prediction will always remain infinitely far away.

But would it really be desirable to know the full wave function of a solid, even if it was possible? On the one hand, yes, because from the wave function we could readily calculate all expectation values. Thus we would be able to make reliable predictions of the properties of any given material. But *predictive power* does not just mean that calculations agree with experiment. Predictive power also means that we know what problems are worth looking at, i.e., for what materials interesting properties are to be expected. For this, knowing the full wave function would be of little help; it would just lead to information-overload. The physics would be buried in the masses of data, and the life of the universe would be too short to analyze it, let alone to understand it. Remarkably, Wigner and Seitz recognized this very early on [7], when electronic-structure calculations were still mainly done by hand, i.e., by human computers [8]

If one had a great calculating machine, one might apply it to the problem of solving the Schrödinger equation for each metal and obtain thereby the interesting physical quantities, such as the cohesive energy, the lattice constant, and similar parameters. It is not clear, however, that a great deal would be gained by this. Presumably the results would agree with the experimentally determined quantities and nothing vastly new would be learned from the calculation. It would be preferable instead to have a vivid picture of the behavior of the wave functions, a simple description of the essence of the factors which determine the cohesion and an understanding of the origin of variation in properties [...]

To understand Nature, we have to identify the *mechanisms* behind the phenomena. We want to know why Nature behaves the way she does, not just, by theory, reproduce some experiment. Why are metals shiny? Why do some materials superconduct while others do not? What opens the gap in a Mott insulator? What drives the ordering of orbitals? What is the mechanism behind colossal magneto-resistance? These properties do not descent in any obvious way from the Theory of almost Everything (1). They are examples of *emergent behavior*. When a system is made of many interacting components, no matter how simple the interactions are, complex many-body phenomena arise. We meet examples of emergence everyday, maybe the most astonishing being the human brain itself.

The limitations of the reductionist approach were probably most strongly voiced by P. W. Anderson in his famous article *More is different* [9]

The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe. [...]

The behavior of large and complex aggregates of elementary particles, it turns out, is not to be understood in terms of a simple extrapolation of the properties of a few particles.

Instead of a reduction to some specific Theory of Everything, science should rather be seen as a hierarchy of structures

Thus, with increasing complication at each stage, we go up the hierarchy of the sciences. We expect to encounter fascinating and, I believe, very fundamental questions at each stage in fitting together less complicated pieces into the more complicated system and understanding basically new types of behavior which can result.

The stages in this hierarchy are best defined, when the energy scales differ significantly. High energy physics deals with the interactions among elementary particles. On the next stage the elementary particles condense into bound states, the subject of nuclear physics. Going up in the hierarchy, at energy-scales of everyday-life, it is only the interaction between atoms and molecules that matters. This is the realm of the chemical bond. Each stage is well separated from the others. This means that new discoveries at a lower stage do not fundamentally change the description at the higher stage. On the one hand, this effective decoupling is what makes science possible without knowing the ultimate *Theory of Everything*. On the other hand, the same decoupling implies that describing the complexity at a higher stage in terms of the theory at the lower stage is a practical impossibility. At each level in the hierarchy fundamentally new properties emerge, which are largely independent of the details on the previous level. Moving up in the hierarchy we need to devise *effective theories* describing the physics at that level, which are derived from physical understanding rather than from an exact mathematical treatment of the underlying microscopic dynamics.

2 Electronic Structure

As we have seen, the fundamental problem in condensed-matter physics is solving (1). The most fascinating properties emerge from its many-body nature. This makes, except in the simplest cases, an exact solution a practical impossibility. Dirac's *approximate practical methods* for its solution therefore have to focus on specific aspects of the problem.

The most immediate many-body effect is Fermi-Dirac statistics. It gives us the *Aufbauprinzip*, letting us understand the structure of the periodic table and the basis of chemical bonding. This is well described in the Hartree-Fock approach. An even more practical approach is density-functional theory (DFT) [10]. It rephrases the many-body problem as the problem of minimizing a functional of the electron density – a tremendous simplification from a problem involving N -body wave-functions with $3N$ -coordinates to densities $n(\vec{r})$ with merely 3 variables. In practice, however, only approximations to the universal density-functional are known. The most successful route to constructing approximate density-functionals is via a mapping of the interacting system onto a non-interacting reference system. In this Kohn-Sham framework already very simple approaches like the local-density approximation (LDA) result in functionals that are surprisingly accurate for vast classes of materials. The basis for their success, again, is that the *Aufbauprinzip* is firmly built into the functional via the filling of the single-electron levels of the reference system.

Because of the single-electron nature of the Kohn-Sham reference system, practical density-functional calculations work well for materials with an electronic structure that can be understood in terms of the energy levels of individual electrons. In these cases density-functional calculations not only provide accurate results, but also contribute to our understanding, e.g., of bonding in materials in terms of molecular orbitals, or Bloch waves and Wannier functions in solids. The approach has proven so successful that its main creator, Walter Kohn, was recognized with the 1998 Nobel Prize in Chemistry.

The successes of the density-functional approach make it easy to forget that it was far from clear whether working density-functional approximations could ever be found. The electron density is a remarkably featureless quantity. As figure 1 exemplifies, it is almost impossible to distinguish the charge densities of a solid from the superposition of its constituent atoms or even to tell a metal from an insulator. This fact is the basis for the unexpected usefulness of the non self-consistent Harris functional [11]. It also lets us appreciate that finding good approximate density functionals is a highly nontrivial task. While the simple local-density approximation works surprisingly well for many materials, it has problems describing magnetism. In this case, expressing the functional in terms of spin-densities greatly simplifies the task of writing down a good functional. While all necessary information is included, in principle, in the density alone, only treating the spin-densities as separate variables gives a simple way, the local spin-density approximation [12], of ensuring that states with large spin tend to have lower energy.

This situation is typical for the art of devising approximate practical methods. We need to identify which details of the problem should be included. If we manage to identify the relevant details, finding a good approximation is significantly simplified. If we fail to identify them,

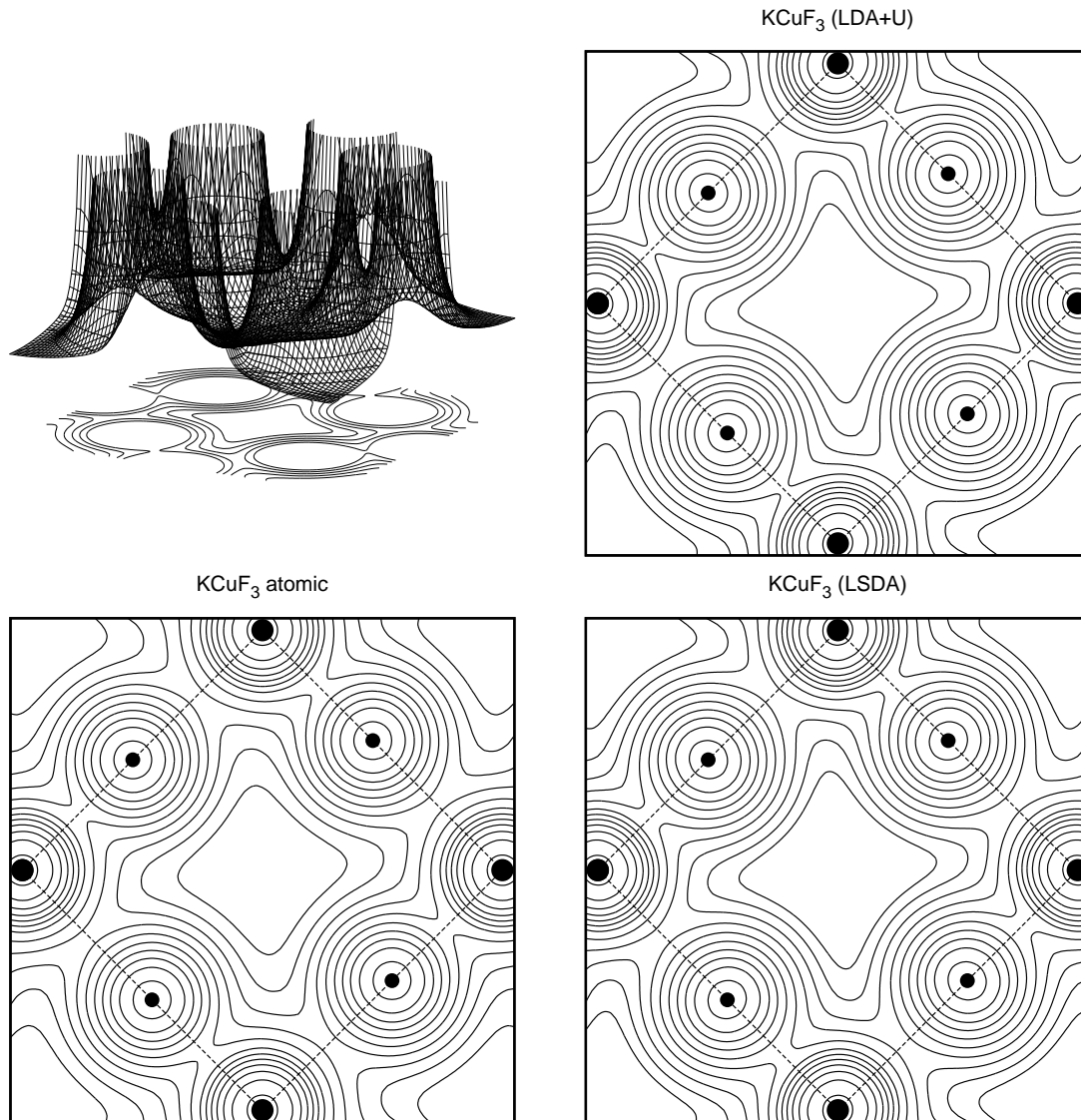


Fig. 1: Charge density for a cut through KCuF_3 . The charge density contours (on a logarithmic scale) for the insulating solid calculated with LDA+U hardly differs from the charge density of the metal obtained in LSDA. Both hardly differ from the superposition of atomic charge densities. Large black circles: position of Cu atoms, small black circles: F atoms.

finding a good approximation is virtually impossible.

By the nature of the Kohn-Sham approach, density-functional calculations are largely confined to materials, for which the picture of individual electrons is adequate, and *Fermi-liquid theory*, which models weakly interacting quasi-particles, applies. There is, however, a remarkable variety of *strongly correlated* materials for which this standard model of electronic structure theory breaks down. The hallmark of these materials is that some of their electrons are neither perfectly localized, nor fully itinerant. These electrons can no longer be considered individually. The resulting behavior presents some of the deepest intellectual challenges in physics. At the same time interest in strongly correlated materials is fueled by the astounding possibilities for technological applications. Prominent examples are the transition-metal oxides, e.g., the high-temperature superconductors, and molecular crystals [13].

3 The LDA+DMFT approach

When dealing with strongly-correlated electrons we have to confront the many-body problem. For a long time the only available strategy was to come up with minimal models, tailored to describe a specific phenomenon. Even these simple many-body problems (Kondo model, Anderson model, Hubbard model, periodic Anderson model, . . .), were still very hard to solve. Including the complexity of specific materials was out of the question. Most effort went into developing methods to solve these model problems. For single-impurity models, like the Kondo or the Anderson model, exact solutions were found, but lattice problems (Hubbard model, periodic Anderson model) remained unsolved, with the exception of special limits, typically one-dimensional systems. The formidable difficulty of finding approximate non-perturbative techniques to solve the Hubbard model and similar Hamiltonians lead to a decoupling of many-body physics from chemistry and electronic-structure. On the other hand, the hope of calculating model parameters from first-principles lead to largely independent developments, such as the LDA+ U method, in which correlation effects are fitted into DFT calculations even if only at the crude Hartree-Fock level.

In recent years, significant progress was driven by the development of dynamical mean-field theory (DMFT). The key insight was that in the limit of infinite dimensions the self energy becomes local [14, 15]. That suggested that, in finite but not too low dimension, the infinite lattice could be mapped approximately onto an impurity problem, which has to be solved self-consistently [15]. From the many-body point of view, the techniques developed for impurity models could now be used for the Hubbard model. From the materials point of view, the LDA+ U experience suggested that it would be possible to solve many-body models with parameters calculated *ab-initio*. These two lines of research culminated in the development to the LDA+DMFT method, today state-of-the-art approach for strongly-correlated systems.

While enormously reducing the cost of the simulation, non-perturbative calculations are still limited to quite simple model Hamiltonians [16]. It is therefore crucial to construct models that are as small as possible, while still capturing the essential chemistry of the real material. As we already saw in the development of approximate density-functionals, the art of modeling is to make the model as simple as possible, but not simpler. Ideally a model is simple enough that a simulation is well feasible, but detailed enough to be material-specific.

The LDA+DMFT approach works in two steps. In the first step, *ab-initio* calculations, based on density-functional theory, are used to obtain the one-electron part of the Hamiltonian. Next, the high-energy states are integrated out, so that only the low-energy partially filled (d or f) bands are retained, and a basis of first-principles Wannier functions is constructed. These Wannier functions, by construction, carry the information on the lattice and the chemistry; furthermore, they are localized, so that the correlation part of the Coulomb repulsion is very short range in this basis. In the second step, the material-specific few-bands many-body Hamiltonians, constructed in terms of these Wannier functions, are solved in the dynamical mean-field approximation. This two steps approach has been used very successfully, e.g., to understand the role of subtle crystal-field splittings for the metal-insulator transition in $3d^n$ transition-metal oxides [17].

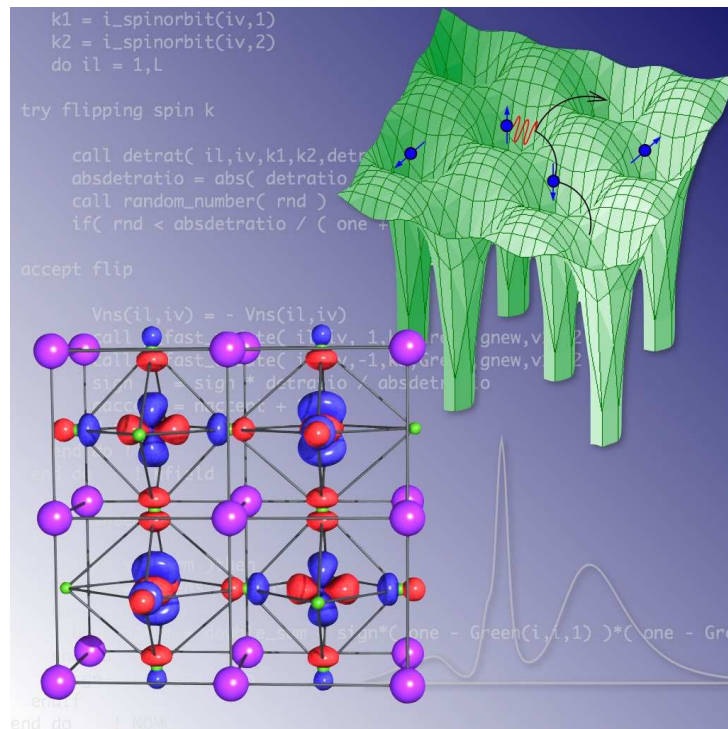


Fig. 2: Logo showing the essential ingredients of the LDA+DMFT approach: Electrons in localized orbitals are strongly correlated as their Coulomb interaction cannot be described by a simple static mean field. The background shows a section of a quantum Monte Carlo based impurity solver which is used to calculate, e.g., spectral functions.

Still, even low-energy few-bands models can be solved only thanks to high performance computers. The task of solving the full many-body problem in a realistic setting will remain the main challenge in condensed matter for years to come. Bridging the high and low energy electronic degrees of freedom is not only one of the deepest problem in contemporary physics but should also provide a wealth of exciting materials for novel technologies.

4 Overview of the school

The historical development and the current state-of-the-art of the LDA+DMFT approach is reflected in the organization of the school and the present collection of lecture notes.

To set the stage, Dieter Vollhardt will introduce the problem of electronic correlations and discuss the limit of infinite dimensions, which is the basis of the dynamical mean-field approach. The lecture of Peter Blöchl will then introduce the other pillar of LDA+DMFT, density-functional theory.

The second group of lectures is devoted to the art of model-building. First, Frank Lechermann will discuss some of the most important types of models. The construction of realistic models is based on Wannier functions from *ab-initio* calculations. The determination of Wannier functions is discussed in the lecture of Jan Kuneš. By construction these functions give a good representation of the kinetic part of the model Hamiltonian. In contrast, the calculation of

screened Coulomb integrals is much less developed. This challenging problem will be covered in the lecture of Ferdi Aryasetiawan. To reflect the central importance of model-building, two tutorials will cover the practical aspects of constructing Wannier functions and of calculating screened Coulomb parameters.

The dynamical mean-field approach is the central theme of the book. The foundations of this many-body method and the applications to simple models will be explained in the lecture of Marcus Kollar. To use it in a realistic contest, DMFT has to be combined with *ab-initio* techniques. The LDA+DMFT approach will be presented by Eva Pavarini.

An essential step of DMFT is to solve the auxiliary quantum-impurity problem. Three lectures are thus devoted to introducing some of the most important quantum-impurity solvers: The Lanczos method (Erik Koch), the Hirsch-Fye quantum Monte Carlo (Nils Blümer), and the continuous-time quantum Monte Carlo (Philipp Werner). Computational aspects of the quantum Monte Carlo approaches will be covered in two tutorials.

Further lectures are devoted to advanced topics. Hubert Ebert will introduce the KKR+DMFT approach, which tightly integrates DMFT in a band-structure method. Sasha Lichtenstein will discuss important extensions of DMFT to include spatial fluctuations. Finally, Karsten Held will discuss the GW+DMFT, which is based on Hedin's *GW* approach rather than DFT.

Each lecture will illustrate examples of the successes of LDA+DMFT. However, the ultimate word on a theory goes to experiments. Thus the book ends with the lecture of Hao Tjeng, maybe the most important of all. It will present challenges for theory from experiments. Facing the limits of a method is the essential step to improve a theory, and to make progress towards predictive power.

References

- [1] A. Einstein, *Principles of Research*, in *Essays in Science* **4** (1934)
- [2] P.M.A. Dirac, *Proceedings of the Royal Society (London)* **A 123**, 714 (1929)
- [3] D. M. Ceperley, *Rev. Mod. Phys.* **71**, 438 (1999)
- [4] R.B. Laughlin, and D. Pines,
Proceedings of the National Academy of Sciences (USA) **97**, 28 (2000)
- [5] P.M.A. Dirac, *Proceedings of the Royal Society (London)* **A 123**, 714 (1929)
- [6] P.S. Laplace, *Theorie Analytique des Probabilités*, Courier, Paris (1820)
- [7] E.P. Wigner and F. Seitz, *Solid State Physics* **1**, 97 (1955)
- [8] F. Hermann, *Physics Today*, June 1984, p. 56
- [9] P.W. Anderson, *Science* **177**, 393 (1972)
- [10] W. Kohn, *Rev. Mod. Phys.* **71**, 1253 (1999)
- [11] J. Harris, *Phys. Rev. B* **31**, 1770 (1985)
- [12] R.O. Jones and O. Gunnarsson, *Rev. Mod. Phys.* **61**, 689 (1989)
- [13] I.S. Osborne, *Science (Special Issue: Correlated Electron Systems)* **288**, 461 (2000)
- [14] W. Metzner and D. Vollhardt, *Phys. Rev. Lett.* **62**, 324 (1989)
- [15] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [16] A. Dolfen, E. Pavarini, and E. Koch,
Innovatives Supercomputing in Deutschland **4**, 16 (2006)
- [17] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, and O.K. Andersen,
Phys. Rev. Lett. **92**, 176403 (2004)

1 Dynamical Mean-Field Approach for Strongly Correlated Materials

Dieter Vollhardt

Center for Electronic Correlations and Magnetism

University of Augsburg

Contents

1	Introduction	2
1.1	Electronic correlations	2
1.2	Hubbard model	4
2	Mean-field approaches for many-body systems	5
2.1	Construction of mean-field theories	5
2.2	Weiss mean-field theory for the Ising model	6
3	Lattice fermions in high dimensions	7
3.1	Simplifications of perturbation theory	8
4	Dynamical mean-field theory for correlated lattice fermions	10
4.1	Solving the DMFT self-consistency equations	12
5	Mott-Hubbard metal-insulator transition	13
5.1	The characteristic structure of the spectral function	14
6	Electronic correlations in materials	17
6.1	LDA+DMFT	17
6.2	Single-particle spectrum of correlated electron materials	20
7	Summary and outlook	22

1 Introduction

1.1 Electronic correlations

In physics the average or expectation value of a product of quantities usually differs from the product of the averages of the individual quantities:

$$\langle AB \rangle \neq \langle A \rangle \langle B \rangle. \quad (1)$$

This is attributed to *correlations*. For example, in an interacting system a particle at position \mathbf{r} will, in general, influence other particles at positions \mathbf{r}' . Therefore the density-density correlation function of this system does not factorize

$$\langle n(\mathbf{r})n(\mathbf{r}') \rangle \neq \langle n(\mathbf{r}) \rangle \langle n(\mathbf{r}') \rangle = n^2, \quad (2)$$

i.e., is not given by the square of the average density n . (For quantum particles the unequal sign holds even in the non-interacting case, since the quantum statistics by itself already leads to a spatial dependence). Correlations are thus defined as effects which go beyond factorization approximations such as Hartree or Hartree-Fock theory.

Correlations in space and time are by no means abstract notions, but occur frequently in everyday life. Persons in an elevator or in a car are strongly correlated both in space and time, and it would be quite inadequate to describe the situation of a person in such a case within a factorization approximation where the influence of the other person(s) is described only by a static mean-field, i.e., a structureless cloud.

As in the case of two persons riding together on an elevator, two electrons with different spin direction occupying the same narrow d or f orbital in a real material are also correlated. Here the degree of correlation can be estimated in a very simplified picture as follows. Assuming the correlated electrons (or rather the quasiparticles, i.e., excitations) to have a well-defined dispersion $\epsilon_{\mathbf{k}}$, their velocity is given by $v_{\mathbf{k}} = \frac{1}{\hbar} |\nabla_{\mathbf{k}} \epsilon_{\mathbf{k}}|$. The typical velocity is given by $v_{\mathbf{k}} \sim \frac{a}{\tau}$, where a is the lattice spacing and τ is the average time spent on an atom. The derivative can be estimated as $\frac{1}{\hbar} |\nabla_{\mathbf{k}} \epsilon_{\mathbf{k}}| \sim \frac{1}{\hbar} aW$ since $|\nabla_{\mathbf{k}}| \sim 1/k \sim a$ and $|\epsilon_{\mathbf{k}}|$ corresponds to the band overlap t and hence to the band width W . Altogether this means that

$$\tau \sim \frac{\hbar}{W}. \quad (3)$$

The narrower a band, the longer an electron therefore resides on an atom and thereby feels the presence of other electrons. Hence a narrow band width implies strong electronic correlations. This is the case for many elements in the periodic table. Namely, in many materials with partially filled d and f electron shells, such as the transition metals V, Fe, and Ni and their oxides, or rare-earth metals such as Ce, electrons occupy narrow orbitals. This spatial confinement enhances the effect of the Coulomb interaction between the electrons, making them “strongly correlated”. Correlations give rise to profound quantitative and qualitative changes of the physical properties of electronic systems as compared to non-interacting particles. Indeed, already in 1937, at the outset of modern solid state physics, de Boer and Verwey [1] drew attention

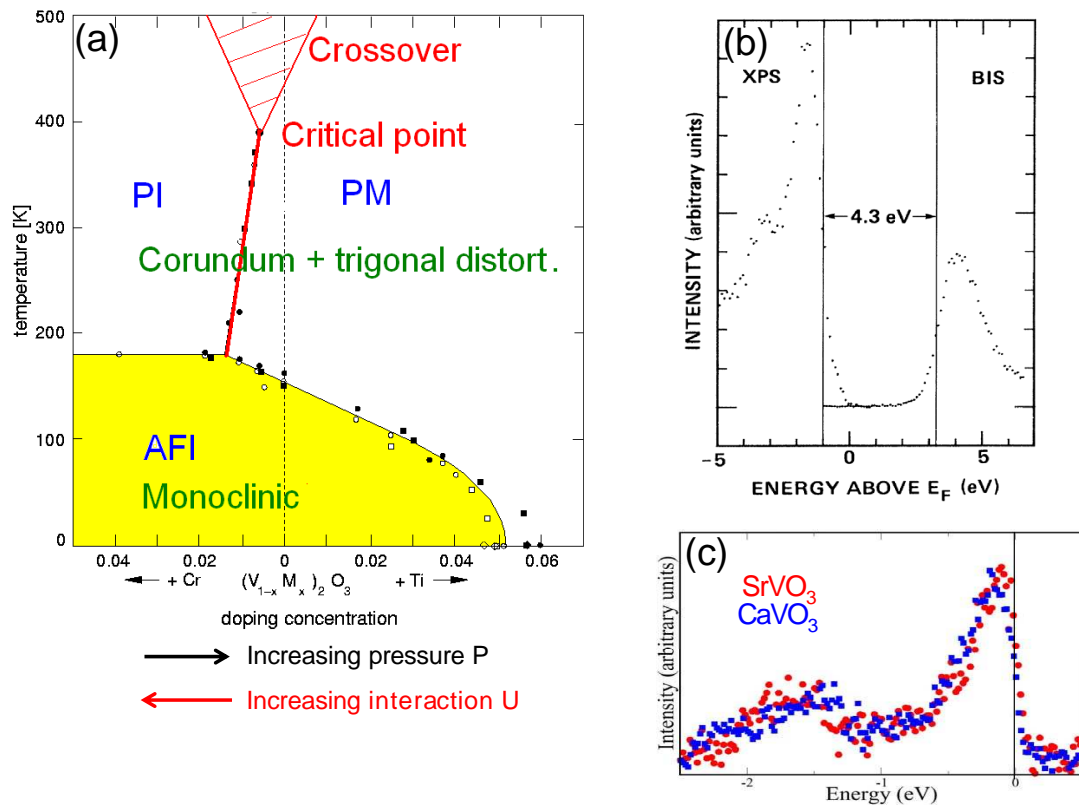


Fig. 1: Typical correlation effects in solids: (a) Mott- Hubbard metal-insulator transition in the paramagnetic phase of V_2O_3 doped with Cr (after [3]); (b) insulating energy gap at the Fermi surface of NiO (after [4]); (c) lower Hubbard band in $SrVO_3$ and $CaVO_3$ due to transfer of spectral weight from the Fermi energy to energies around -1.7 eV (after [5])

to the surprising properties of materials with incompletely filled $3d$ -bands, such as NiO. This observation prompted Mott and Peierls [2] to consider the interaction between the electrons.

Correlations may, for example, lead to a transition from metallic to insulating behavior as in V_2O_3 or NiO (see Fig. 1 (a),(b)). In particular, correlated materials often respond very strongly to changes in external parameters. This is expressed by large renormalizations of the response functions of the system, e.g., of the spin susceptibility and the charge compressibility, and by a strong transfer of spectral weight (see Fig. 1 (c)). Electronic correlations also play an essential role in high temperature superconductivity. In particular, the interplay between the spin, charge and orbital degrees of freedom of the correlated d and f electrons and with the lattice degrees of freedom leads to a wealth of unusual phenomena at low temperatures [6]. These properties cannot be explained within conventional mean-field theories, e.g., Hartree-Fock theory, since they describe the interaction only in an average way.

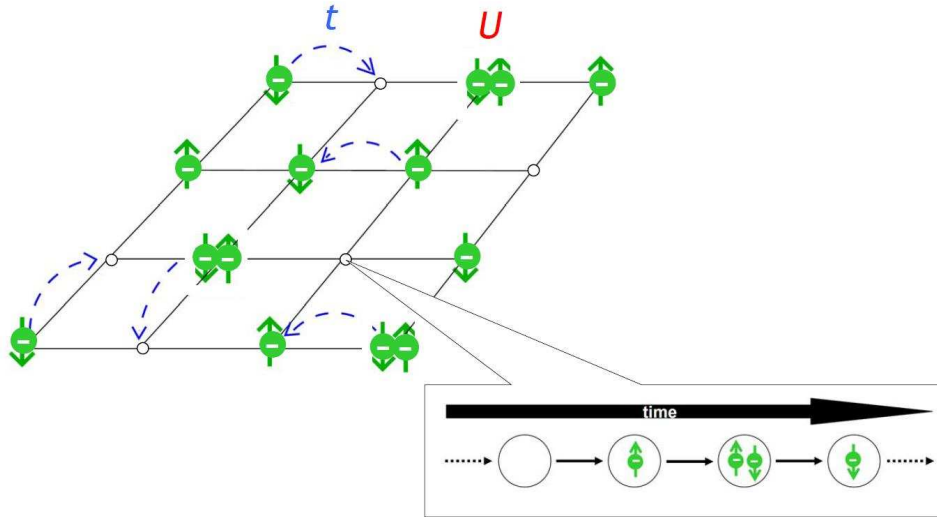


Fig. 2: Schematic illustration of interacting electrons in a solid in terms of the Hubbard model. The ions appear only as a rigid lattice (here represented as a square lattice). The electrons, which have a mass, a negative charge, and a spin (\uparrow or \downarrow), move from one lattice site to the next with a hopping amplitude t . The quantum dynamics thus leads to fluctuations in the occupation of the lattice sites as indicated by the time sequence. When two electrons meet on a lattice site (which is only possible if they have opposite spin because of the Pauli exclusion principle) they encounter an interaction U . A lattice site can either be unoccupied, singly occupied (\uparrow or \downarrow), or doubly occupied.

1.2 Hubbard model

The simplest model describing interacting electrons in a solid is the one-band, spin-1/2 Hubbard model [7–9] where the interaction between the electrons is assumed to be so strongly screened that it is taken as purely local. The Hamiltonian consists of two terms, the kinetic energy \hat{H}_0 and the interaction energy \hat{H}_I (here and in the following operators are denoted by a hat):

$$\hat{H} = \hat{H}_0 + \hat{H}_I \quad (4a)$$

$$\hat{H}_0 = \sum_{\mathbf{R}_i, \mathbf{R}_j} \sum_{\sigma} t_{ij} \hat{c}_{i\sigma}^{\dagger} \hat{c}_{j\sigma} = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} \hat{n}_{\mathbf{k}\sigma} \quad (4b)$$

$$\hat{H}_I = U \sum_{\mathbf{R}_i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}, \quad (4c)$$

where $\hat{c}_{i\sigma}^{\dagger}$ ($\hat{c}_{i\sigma}$) are creation (annihilation) operators of electrons with spin σ at site \mathbf{R}_i , and $\hat{n}_{i\sigma} = \hat{c}_{i\sigma}^{\dagger} \hat{c}_{i\sigma}$. The Fourier transform of the kinetic energy in (4b), where t_{ij} is the hopping amplitude, involves the dispersion $\epsilon_{\mathbf{k}}$ and the momentum distribution operator $\hat{n}_{\mathbf{k}\sigma}$.

A schematic picture of the Hubbard model is shown in Fig. 2. When we look only at a single site of this lattice model, this site will sometimes be empty, singly occupied or doubly occupied. In particular, for strong repulsion U double occupations are energetically very unfavorable and are therefore strongly suppressed. In this situation $\langle \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \rangle$ must not be factorized

since $\langle \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \rangle \neq \langle \hat{n}_{i\uparrow} \rangle \langle \hat{n}_{i\downarrow} \rangle$. Otherwise, correlation phenomena such as the Mott-Hubbard metal-insulator transition are eliminated from the very beginning. This explains why Hartree-Fock-type mean-field theories are generally insufficient to explain the physics of electrons in the paramagnetic phase for strong interactions.

The Hubbard model looks very simple. However, the competition between the kinetic energy and the interaction leads to a complicated many-body problem.

2 Mean-field approaches for many-body systems

2.1 Construction of mean-field theories

It is well known that theoretical investigations of quantum-mechanical many-body systems are faced with severe technical problems, particularly in those dimensions which are most interesting to us, i. e., $d = 2, 3$. This is due to the complicated dynamics and, in the case of fermions, the non-trivial algebra introduced by the Pauli exclusion principle. In the absence of exact methods there is clearly a great need for reliable, controlled approximation schemes.

In the statistical theory of classical and quantum-mechanical systems a rough, overall description of the properties of a model is often provided by a *mean-field theory*. Although the term is frequently used it is rather vague since there is no unique construction scheme.

There does exist a well-established route to mean-field theories which makes use of the simplifications that occur when some parameter is taken to be large (in fact, infinite), e.g., the length of the spins S , the spin degeneracy N , or the coordination number Z (the number of nearest neighbors of a lattice site¹). Investigations in this limit, supplemented, if at all possible, by an expansion in the inverse of the large parameter, often provide valuable insight into the fundamental properties of a system even when this parameter is not large.

One of the best-known theories of this kind is the Weiss molecular-field theory for the Ising model [10]. It is a prototypical *single-site mean-field theory* which becomes exact for infinite-range interaction, as well as in the limit of the coordination number $Z \rightarrow \infty$ or² the spatial dimension $d \rightarrow \infty$. In the latter case $1/Z$ or $1/d$ is a small parameter which can sometimes be used to improve the mean-field theory systematically. This mean-field theory contains no

¹In three dimensions ($d = 3$) one has $Z = 6$ for a simple cubic lattice ($Z = 2d$ for a hypercubic lattice in general dimensions d), $Z = 8$ for a bcc lattice and $Z = 12$ for an fcc-lattice. Since $Z \sim \mathcal{O}(10)$ is already quite large in $d = 3$, such that $1/Z$ is rather small, it is only natural and in the general spirit of theoretical physics to consider the extreme limit $Z \rightarrow \infty$ to simplify the problem. Later, if possible, one can try to improve the result by expanding in the small parameter $1/Z$. It turns out that several standard approximation schemes which are commonly used to explain experimental results in dimension $d = 3$, are exact only in $d = \infty$ (for a more detailed discussion see Ref. [11]).

²For regular lattices, e.g., Bravais-lattices, both a dimension d and a coordination number Z can be defined. In this case either d or Z can be used alternatively as an expansion parameter. However, there exist other lattices (or rather graphs) which cannot be associated with a physical dimension d although a coordination number Z is well-defined. The best-known example is the Bethe lattice, an infinitely extended Cayley tree [10, 12], which is not a regular lattice because it does not have loops. The coordination number Z is therefore a very useful parameter for theoretical investigations of lattice models, although the dimension d is the more general physical parameter. In the following discussion we mostly use both d and Z in parallel.

unphysical singularities and is applicable for all values of the input parameters, i.e., coupling parameters, magnetic field, and temperature. It is also diagrammatically controlled [13]. Insofar it is a very respectable approximation which sets very high standards for other mean-field theories.

2.2 Weiss mean-field theory for the Ising model

The Ising model with nearest-neighbor (NN) coupling is defined by

$$H = -\frac{1}{2}J \sum_{\langle \mathbf{R}_i, \mathbf{R}_j \rangle} S_i S_j, \quad (5)$$

where we assume ferromagnetic coupling ($J > 0$). Every spin S_i interacts with a local field h_i , produced by its nearest neighbors at site \mathbf{R}_i . In the Weiss mean-field approach the two-spin interaction in (5) is decoupled, i. e., H is replaced by a mean-field Hamiltonian

$$H^{\text{MF}} = -h_{\text{MF}} \sum_{\mathbf{R}_i} S_i + E_{\text{shift}}. \quad (6a)$$

Now a spin S_i interacts only with a global (“molecular”) field

$$h_{\text{MF}} = J \sum_{\mathbf{R}_j}^{(i)} \langle S_j \rangle \quad (6b)$$

$$\equiv J Z \langle S \rangle. \quad (6c)$$

Here $\langle \rangle$ indicates the thermal average, $E_{\text{shift}} = \frac{1}{2} L J Z \langle S \rangle^2$ is a constant energy shift with L as the number of lattice sites, and the superscript (i) implies summation over NN-sites of \mathbf{R}_i . This corresponds to the factorization

$$\langle [S_i - \langle S \rangle][S_j - \langle S \rangle] \rangle \equiv 0, \quad (7)$$

whereby correlated fluctuations of spins at sites \mathbf{R}_i and \mathbf{R}_j are neglected. In the limit $Z \rightarrow \infty$ the coupling constant J has to be rescaled as

$$J \rightarrow \frac{J^*}{Z}, \quad J^* = \text{const} \quad (8)$$

for h_{MF} to remain finite. In this limit the factorization procedure (7), and hence the replacement of (5), by the mean-field Hamiltonian (6a), becomes exact [14, 15].

Eq. (6a) implies that in the limit $Z \rightarrow \infty$ fluctuations in the bath of surrounding neighbors become unimportant, such that the surrounding of any site is completely described by a single mean-field parameter h_{MF} (see Fig. 3). Hence the Hamiltonian becomes purely local

$$H^{\text{MF}} = \sum_{\mathbf{R}_i} H_i + E_{\text{shift}} \quad (9)$$

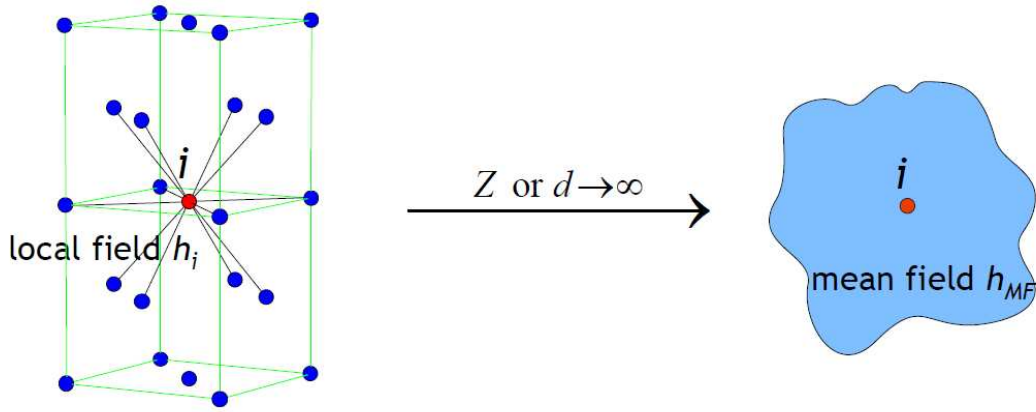


Fig. 3: Already in three dimensions ($d = 3$) the coordination number Z of a lattice can be quite high, as in the face-centered cubic lattice where $Z = 12$. In the limit $Z \rightarrow \infty$, or equivalently $d \rightarrow \infty$, the Ising model effectively reduces to a single-site problem where the local field h_i is replaced by a global mean (“molecular”) field h_{MF} .

$$H_i = -h_{MF}S_i. \quad (10)$$

Thereby the problem reduces to an effective *single-site problem*. The value of $\langle S \rangle$ is determined by the self-consistent equation

$$\langle S \rangle = \tanh(\beta J^* \langle S \rangle), \quad (11)$$

where $\beta = 1/T$ (here $k_B = 1$).

3 Lattice fermions in high dimensions

It is natural to ask whether the limit $d \rightarrow \infty$ may also be useful in the investigation of lattice models with itinerant quantum-mechanical degrees of freedom and, in particular, in the case of the Hubbard model. Following Ref. [16] we take a look at the kinetic energy term (4b), since the interaction term is purely local and is thereby completely independent of the lattice structure and the dimension. For nearest-neighbor hopping on a d -dimensional hypercubic lattice (where $Z = 2d$) $\epsilon_{\mathbf{k}}$ is given by³

$$\epsilon_{\mathbf{k}} = -2t \sum_{i=1}^d \cos k_i. \quad (12)$$

The density of states (DOS) corresponding to $\epsilon_{\mathbf{k}}$ is

$$N_d(\omega) = \sum_{\mathbf{k}} \delta(\hbar\omega - \epsilon_{\mathbf{k}}). \quad (13)$$

This is simply the probability density for finding $\omega = \epsilon_{\mathbf{k}}$ for a random choice of $\mathbf{k} = (k_1, \dots, k_d)$. If the k_i are chosen randomly, $\epsilon_{\mathbf{k}}$ in (12) is the sum of (independent) random numbers $-2t \cos k_i$.

³In the following we set Planck’s constant \hbar , Boltzmann’s constant k_B , and the lattice spacing equal to unity.

The central limit theorem then implies that in the limit $d \rightarrow \infty$ the DOS is given by a Gaussian

$$N_d(\omega) \xrightarrow{d \rightarrow \infty} \frac{1}{2t\sqrt{\pi d}} \exp \left[- \left(\frac{\omega}{2t\sqrt{d}} \right)^2 \right]. \quad (14)$$

Unless t is scaled properly with d this DOS will become arbitrarily broad and featureless for $d \rightarrow \infty$. Clearly only the scaling of the hopping amplitude

$$t \rightarrow \frac{t^*}{\sqrt{d}}, \quad t^* = \text{const.}, \quad (15)$$

yields a non-trivial DOS [17, 16]:

$$N_\infty(\omega) = \frac{1}{\sqrt{2\pi t^*}} \exp \left[- \frac{1}{2} \left(\frac{\omega}{t^*} \right)^2 \right]. \quad (16)$$

By contrast, the interaction term in (4) is purely local and independent of the surrounding. Hence it is independent of the spatial dimension of the system. Consequently, the on-site interaction U need not be scaled. So we see that the scaled Hubbard Hamiltonian

$$\hat{H} = - \frac{t^*}{\sqrt{Z}} \sum_{\langle \mathbf{R}_i, \mathbf{R}_j \rangle} \sum_{\sigma} \hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + U \sum_{\mathbf{R}_i} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \quad (17)$$

has a nontrivial $Z \rightarrow \infty$ limit, where both terms, the kinetic energy and the interaction, are of the same order of magnitude and are thereby able to compete. It is this competition between the two terms which leads to interesting many-body physics.⁴

The scaling (15) was determined within a \mathbf{k} -space formulation. We will now derive the same result within a position-space formulation.

3.1 Simplifications of perturbation theory

The most important consequence of the scaling (15) is the fact that it leads to significant simplifications in the investigation of Hubbard-type lattice models [16, 18–22]; for details see Ref. [11]. To understand this point better we take a look at the perturbation theory in terms of U . At $T = 0$ and $U = 0$ the kinetic energy of the Hubbard model may be written as

$$E_{\text{kin}}^0 = -t \sum_{\langle \mathbf{R}_i, \mathbf{R}_j \rangle} \sum_{\sigma} g_{ij,\sigma}^0, \quad (18)$$

⁴To obtain a physically meaningful mean-field theory for a model its internal or free energy has to remain finite in the limit d or $Z \rightarrow \infty$. While in the case of the Ising model the scaling $J \rightarrow \tilde{J}/Z$, $\tilde{J} = \text{const.}$, was rather obvious this is not so for more complicated models. Namely, quantum many-particle systems are usually described by a Hamiltonian consisting of several non-commuting terms, e.g., a kinetic energy and an interaction, each of which is associated with a coupling parameter, usually a hopping amplitude and an interaction, respectively. In such a case the question of how to scale these parameters has no unique answer since this depends on the physical effects one wishes to explore. In any case, the scaling should be performed such that the model remains non-trivial and that its internal or free energy stays finite in the $Z \rightarrow \infty$ limit. By “non-trivial” we mean that not only $\langle \hat{H}_0 \rangle$ and $\langle \hat{H}_{\text{int}} \rangle$, but also the *competition* between these terms, expressed by $\langle [\hat{H}_0, \hat{H}_{\text{int}}] \rangle$, should remain finite. In the case of the Hubbard model it would be possible to employ the scaling $t \rightarrow t^*/Z$, $t^* = \text{const.}$, but then the kinetic energy would be reduced to zero in the limit $d \rightarrow \infty$, making the resulting model uninteresting (but not unphysical) for most purposes.

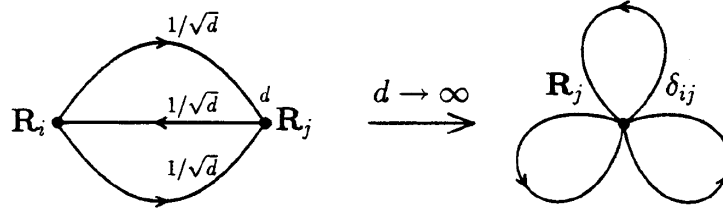


Fig. 4: Contribution to the irreducible self-energy for the Hubbard model in second-order perturbation theory in U , and its collapse in the limit $d \rightarrow \infty$.

where $g_{ij,\sigma}^0 = \langle \hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} \rangle_0$ is the one-particle density matrix. This quantity can also be interpreted as the amplitude for transitions between site \mathbf{R}_i and \mathbf{R}_j , whose square is proportional to the probability for a particle to hop from \mathbf{R}_j to \mathbf{R}_i , i.e., $|g_{ij,\sigma}^0|^2 \sim 1/Z \sim 1/d$ since \mathbf{R}_j has $\mathcal{O}(d)$ nearest neighbors \mathbf{R}_i . Thus the sum of $|g_{ij,\sigma}^0|^2$ over all nearest neighbors must yield a constant. In the limit $d \rightarrow \infty$ we then have

$$g_{ij,\sigma}^0 \sim \mathcal{O}\left(\frac{1}{\sqrt{d}}\right), \quad \mathbf{R}_i \text{ NN of } \mathbf{R}_j. \quad (19)$$

Since the sum over the NN-sites \mathbf{R}_i in (18) is of $\mathcal{O}(d)$ the NN-hopping amplitude t must obviously be scaled according to (15) for E_{kin}^0 to remain finite in the limit $d, Z \rightarrow \infty$. Hence, as expected, a real-space formulation yields the same results for the required scaling of the hopping amplitude.

The one-particle Green function (“propagator”) $G_{ij,\sigma}^0(\omega)$ of the non-interacting system obeys the same scaling as $g_{ij,\sigma}^0$. This follows directly from its definition

$$G_{ij,\sigma}^0(t) \equiv - \left\langle T \hat{c}_{i\sigma}(t) \hat{c}_{j\sigma}^\dagger(0) \right\rangle_0, \quad (20)$$

where T is the time ordering operator, and the time evolution of the operators is provided by the Heisenberg representation. The one-particle density matrix is obtained as $g_{ij,\sigma}^0 = \lim_{t \rightarrow 0^-} G_{ij,\sigma}^0(t)$. If $g_{ij,\sigma}^0$ obeys (19) the one-particle Green function $G_{ij,\sigma}^0(t)$ must follow the same scaling at all times since this property does not depend on the time evolution and the quantum mechanical representation. The Fourier transform $G_{ij,\sigma}^0(\omega)$ also preserves this property.

Although the propagator $G_{ij,\sigma}^0 \sim 1/\sqrt{d}$ vanishes for $d \rightarrow \infty$, the particles are not localized in this limit. Namely, even in the limit $d \rightarrow \infty$ the off-diagonal elements of $G_{ij,\sigma}^0$ contribute, since a particle may hop to d nearest neighbors with amplitude $t^*/\sqrt{2d}$. For general i, j one finds [23, 19]

$$G_{ij,\sigma}^0 \sim \mathcal{O}\left(1/d^{\|\mathbf{R}_i - \mathbf{R}_j\|/2}\right), \quad (21)$$

where $\|\mathbf{R}\| = \sum_{n=1}^d |R_n|$ is the length of \mathbf{R} in the so-called “Manhattan metric”.

It is the property (21) which is the origin of all simplifications arising in the limit $d \rightarrow \infty$. In particular, it implies the collapse of all connected, irreducible perturbation theory diagrams in position space [16, 18, 19]. This is illustrated in Fig. 4, where a contribution in second-order perturbation theory to the irreducible self-energy, $\Sigma_{ij}^{(2)}$, is shown. As a consequence the full,

irreducible self-energy becomes a purely local quantity [16, 18]:

$$\Sigma_{ij,\sigma}(\omega) \stackrel{d \rightarrow \infty}{\equiv} \Sigma_{ii,\sigma}(\omega) \delta_{ij}. \quad (22a)$$

In the paramagnetic phase we may write $\Sigma_{ii,\sigma}(\omega) \equiv \Sigma(\omega)$. The Fourier transform of $\Sigma_{ij,\sigma}$ is seen to become momentum-independent

$$\Sigma_{\sigma}(\mathbf{k}, \omega) \stackrel{d \rightarrow \infty}{\equiv} \Sigma_{\sigma}(\omega). \quad (22b)$$

This leads to tremendous simplifications in all many-body calculations for the Hubbard model and related models.

Due to the simplifications caused by (22), the most important obstacle for actual diagrammatic calculations in finite dimensions $d \geq 1$, namely the integration over intermediate momenta, is removed in $d = \infty$. While in finite dimensions these integrations lead to untractable technical problems, they become simple in $d = \infty$. In spite of the simplifications in position (or momentum) space the problem retains its full dynamics in $d = \infty$.

4 Dynamical mean-field theory for correlated lattice fermions

Itinerant quantum mechanical models such as the Hubbard model and its generalizations are much more complicated than classical, Ising-type models. Generally there do not even exist semiclassical approximations for such models that might serve as a starting point for further investigations. Under such circumstances the construction of a mean-field theory with the comprehensive properties of the Weiss molecular field theory for the Ising model will necessarily be much more complicated, too. As discussed above there do exist well-known mean-field approximation schemes, e. g. Hartree-Fock, random-phase approximation, saddle-point evaluations of path integrals, decoupling of operators. However, these approximations do not provide mean-field theories in the spirit of statistical mechanics, since they are not able to give a global description of a given model (e.g., the phase diagram, thermodynamics, etc.) in the entire range of input parameters.

Here the limit of high spatial dimensions d or coordination number Z has again been extremely useful [16]. It provides the basis for the construction of a comprehensive mean-field theory for lattice fermions which is diagrammatically controlled and whose free energy has no unphysical singularities. The construction is based on the scaled Hamiltonian (26) and the simplifications in the many-body perturbation theory discussed in Sec. 3.2. There we saw that the local propagator $G(\omega)$, i.e., the amplitude for an electron to return to a lattice site, and the local but dynamical self-energy $\Sigma(\omega)$ are the two central quantities in such a theory. Since the self-energy is a dynamical variable (in contrast to Hartree-Fock theory where it is merely a static potential) the resulting mean-field theory will also be dynamical and can thus describe genuine correlation effects.

The self-consistency equations of this *dynamical mean-field theory* (DMFT) [24–30] for correlated lattice fermions can be derived in different ways; for a discussion see Chapter 4 of

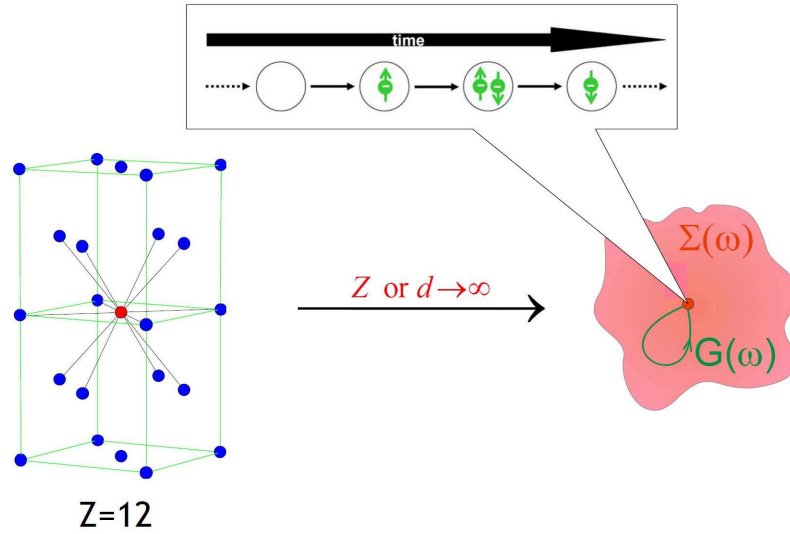


Fig. 5: In the limit $Z \rightarrow \infty$ the Hubbard model effectively reduces to a dynamical single-site problem, which may be viewed as a lattice site embedded in a dynamical mean field. Electrons may hop from the mean field onto this site and back, and interact on the site as in the original Hubbard model (see Fig. 2). The local propagator $G(\omega)$ (i.e., the return amplitude) and the dynamical self-energy $\Sigma(\omega)$ of the surrounding mean field play the main role in this limit. The quantum dynamics of the interacting electrons is still described exactly.

Ref. [31]. However, all derivations make use of the fact that in the limit of high spatial dimensions Hubbard-type models reduce to a dynamical single-site problem, where the d -dimensional lattice model is effectively described by the dynamics of the correlated fermions on a single site which is embedded in a bath provided by the other particles.

Today's standard derivation is based on the mapping of the lattice problem onto a self-consistent single-impurity Anderson model; for details see Ref. [29]. As such it makes contact with the theory of Anderson impurities and the Kondo problem — a well-understood branch of many-body physics, for whose solution efficient numerical codes have been developed already in the 1980's, in particular by making use of the quantum Monte-Carlo (QMC) method [32]. The self-consistent DMFT equations are given by

(i) the equation for the *local propagator* $G_\sigma(i\omega_n)$ which is expressed by a functional integral as

$$G_\sigma(i\omega_n) = -\frac{1}{Z} \int \prod_\sigma Dc_\sigma^* Dc_\sigma [c_\sigma(i\omega_n) c_\sigma^*(i\omega_n)] \exp[-S_{\text{loc}}] \quad (23)$$

with the partition function

$$Z = \int \prod_\sigma Dc_\sigma^* Dc_\sigma \exp[-S_{\text{loc}}], \quad (24)$$

and the local action

$$S_{\text{loc}} = - \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \sum_\sigma c_\sigma^*(\tau_1) \mathcal{G}_\sigma^{-1}(\tau_1 - \tau_2) c_\sigma(\tau_2) + U \int_0^\beta d\tau c_\uparrow^*(\tau) c_\uparrow(\tau) c_\downarrow^*(\tau) c_\downarrow(\tau), \quad (25)$$

where \mathcal{G}_σ is the effective local propagator (also called “bath Green function”, or “Weiss mean field”⁵) defined by a Dyson-type equation

$$\mathcal{G}_\sigma(i\omega_n) = [[G_\sigma(i\omega_n)]^{-1} + \Sigma_\sigma(i\omega_n)]^{-1}, \quad (26)$$

(ii) and the expression for the *lattice Green function* $G_{\mathbf{k}\sigma}(i\omega_n)$ given by

$$G_{\mathbf{k}\sigma}(i\omega_n) = \frac{1}{i\omega_n - \epsilon_{\mathbf{k}} + \mu - \Sigma_\sigma(i\omega_n)}, \quad (27)$$

from which, after performing the lattice Hilbert transform, one obtains the local Green function

$$G_\sigma(i\omega_n) = \sum_{\mathbf{k}} G_{\mathbf{k}\sigma}(i\omega_n) \quad (28)$$

$$= \int_{-\infty}^{\infty} d\epsilon \frac{N(\omega)}{i\omega_n - \epsilon + \mu - \Sigma_\sigma(i\omega_n)}, \quad (29)$$

which is equal to the local propagator (23). The ionic lattice on which the electrons move and its structure are seen to enter only through the DOS of the non-interacting electrons.

The self-consistent equations can be solved iteratively: Starting with an initial value for the self-energy $\Sigma_\sigma(i\omega_n)$ one obtains the local propagator $G_\sigma(i\omega_n)$ from (29) and thereby the bath Green function $\mathcal{G}_\sigma(i\omega_n)$ from (26). This determines the local action (25) which is needed to compute a new value for the local propagator $G_\sigma(i\omega_n)$ and, by employing the old self-energy, a new bath Green function, and so on.

4.1 Solving the DMFT self-consistency equations

The dynamics of the full Hubbard model (4) remains complicated even in the limit $d \rightarrow \infty$ because of the purely local nature of the interaction. Hence an exact, analytic evaluation of the self-consistent set of equations for the local propagator G_σ or the effective propagator \mathcal{G}_σ is not possible. A valuable semi-analytic approximation is provided by the “iterated perturbation theory” (IPT) [26,33,29]. Exact evaluations are only feasible when there is no coupling between the frequencies. This is the case, for example, in the Falicov-Kimball model [22,34].

Solutions of the general DMFT self-consistency equations require extensive numerical methods, in particular quantum Monte Carlo techniques [27,35,36] (for reviews see Refs. [29,37,38], the numerical renormalization group [39,40], exact diagonalization [41–43] and other techniques, which will be discussed later in this school.

⁵In principle, the local functions $\mathcal{G}_\sigma(i\omega_n)$ and $\Sigma_\sigma(i\omega_n)$ can both be viewed as a “dynamical mean field” acting on particles on a site, since they all appear in the bilinear term of the local action (25).

It quickly turned out that the DMFT is a powerful tool for the investigation of electronic systems with strong correlations. It provides a non-perturbative and thermodynamically consistent approximation scheme for finite-dimensional systems which is particularly valuable for the study of intermediate-coupling problems where perturbative techniques fail [29, 30].

In the remaining part of these lecture notes I shall discuss several applications of the DMFT to problems involving electronic correlations. In particular, I will address the Mott-Hubbard metal-insulator transition, and explain the connection of the DMFT with band-structure methods — the LDA+DMFT scheme — which is the first comprehensive framework for the *ab initio* investigation of correlated electron materials.

5 Mott-Hubbard metal-insulator transition

The correlation induced transition between a paramagnetic metal and a paramagnetic insulator, referred to as “Mott-Hubbard metal-insulator transition” (MIT), is one of the most intriguing phenomena in condensed matter physics [44–46]. This transition is a consequence of the competition between the kinetic energy of the electrons and their local interaction U . Namely, the kinetic energy prefers the electrons to move (a wave effect) which leads to doubly occupied sites and thereby to interactions between the electrons (a particle effect). For large values of U the doubly occupied sites become energetically very costly. The system may reduce its total energy by localizing the electrons. Hence the Mott transition is a localization-delocalization transition, demonstrating the particle-wave duality of electrons.

Mott-Hubbard MITs are, for example, found in transition metal oxides with partially filled bands near the Fermi level [6]. For such systems band theory typically predicts metallic behavior. The most famous example is V_2O_3 doped with Cr [3,47,48]. In particular, in $(V_{0.96}Cr_{0.04})_2O_3$ the metal-insulator transition is of first order below $T = 380$ K [3], with discontinuities in the lattice parameters and in the conductivity. However, the two phases remain isostructural.

Making use of the half-filled, single-band Hubbard model (4) the Mott-Hubbard MIT was studied intensively in the past [44–46]. Important early results were obtained by Hubbard [8, 49] within a Green function decoupling scheme, and by Brinkman and Rice [50] using the Gutzwiller variational method [7, 51], both at zero temperature.⁶ Hubbard’s approach yields a continuous splitting of the band into a lower and upper Hubbard band, but cannot describe quasiparticle features. By contrast, the Gutzwiller-Brinkman-Rice approach (for a review see Ref. [53]) gives a good description of the low-energy, quasiparticle behavior, but cannot reproduce the upper and lower Hubbard bands. In the latter approach the MIT is signalled by the disappearance of the quasiparticle peak.

To solve this problem the DMFT has been extremely valuable since it provided detailed insights into the nature of the Mott-Hubbard MIT for all values of the interaction U and temperature T [29, 54, 30].

⁶The Gutzwiller variational method [7,51] consists of the choice of a simple projected variational wave function (“Gutzwiller wave function”) and a semi-classical evaluation of expectation values in terms of this wave function (“Gutzwiller approximation”). The Gutzwiller approximation becomes exact in the limit $d \rightarrow \infty$ [52].

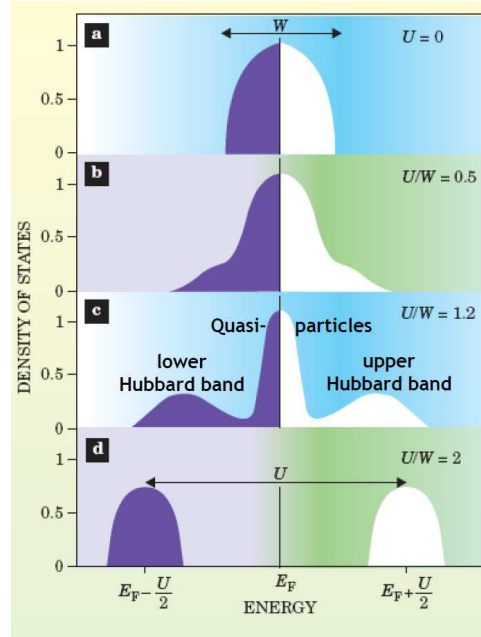


Fig. 6: Evolution of the spectral function (“density of states”) of the Hubbard model in the paramagnetic phase at half-filling. a) non-interacting case, b) for weak interactions there is only little transfer of spectral weight away from the Fermi energy, c) for strong interactions a typical three-peak structure consisting of coherent quasiparticle excitations close to the Fermi energy and incoherent lower and upper Hubbard bands is clearly seen, d) above a critical interaction the quasiparticle peak vanishes and the system is insulating, with two well-separated Hubbard bands remaining; after Ref. [30].

5.1 The characteristic structure of the spectral function

The Mott-Hubbard MIT is monitored by the spectral function $A(\omega) = -\frac{1}{\pi}\text{Im}G(\omega + i0^+)$ of the correlated electrons;⁷ here we follow the discussion of Refs. [55, 30]. The change of $A(\omega)$ obtained within the DMFT for the one-band Hubbard model (4) at $T = 0$ and half filling ($n = 1$) as a function of the Coulomb repulsion U (measured in units of the bandwidth W of non-interacting electrons) is shown in Figs. 6 and 7. While Fig. 6 is a schematic picture of the evolution of the spectrum when the interaction is increased, Fig. 7 shows actual numerical results obtained by the NRG [39, 56]. Here magnetic order is assumed to be suppressed (“frustrated”).

While at small U the system can be described by coherent quasiparticles whose DOS still resembles that of the free electrons, the spectrum in the Mott insulator state consists of two separate incoherent “Hubbard bands” whose centers are separated approximately by the energy U . The latter originate from atomic-like excitations at the energies $\pm U/2$ broadened by the hopping of electrons away from the atom. At intermediate values of U the spectrum then has a characteristic three-peak structure as in the single-impurity Anderson model, which includes both the atomic features (i.e., Hubbard bands) and the narrow quasiparticle peak at low excitation energies, near $\omega = 0$. This corresponds to a strongly correlated metal. The structure of the

⁷In the following we only consider the paramagnetic phase.

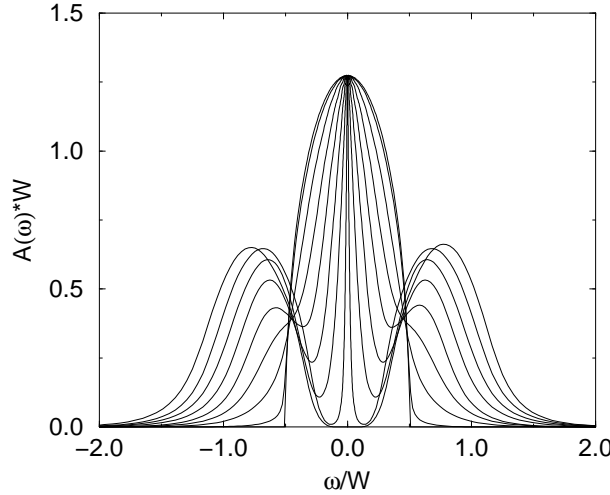


Fig. 7: Evolution of the $T = 0$ spectral function of the one-band Hubbard model with a semi-elliptic (“Bethe”) DOS for interaction values $U/W = 0, 0.2, 0.4, \dots, 1.6$ (W : band width) calculated with the numerical renormalization group. At the critical interaction $U_{c2}/W \simeq 1.47$ the metallic solution disappears and the Mott gap opens; from Ref. [56].

spectrum (lower Hubbard band, quasiparticle peak, upper Hubbard band) is quite insensitive to the specific form of the DOS of the non-interacting electrons.

The width of the quasiparticle peak vanishes for $U \rightarrow U_{c2}(T)$. The “Luttinger pinning” at $\omega = 0$ [20] is clearly observed. On decreasing U , the transition from the insulator to the metal occurs at a lower critical value U_{c1} , where the gap vanishes.

It is important to note that the three-peak spectrum originates from a lattice model with only *one* type of electrons. This is in contrast to the single-impurity Anderson model whose spectrum shows very similar features, but is due to *two* types of electrons, namely the localized orbital at the impurity site and the free conduction band. Therefore the screening of the magnetic moment which gives rise to the Kondo effect in impurity systems has a different origin in lattice systems. Namely, as explained by the DMFT, the same electrons provide both the local moments and the electrons which screen these moments [29].

The evolution of the spectral function of the half-filled frustrated Hubbard model at finite temperatures, $T = 0.0276 W$, is shown in Fig. 8. This temperature is above the temperature of the critical point so that there is no real transition but only a crossover from a metallic-like to an insulating-like solution. The height of the quasiparticle peak at the Fermi energy is no longer fixed at its zero temperature value. This is due to a finite value of the imaginary part of the self-energy. The spectral weight of the quasiparticle peak is seen to be gradually redistributed and shifted to the upper (lower) edge of the lower (upper) Hubbard band. The inset of Fig. 8 shows the U -dependence of the value of the spectral function at zero frequency $A(\omega = 0)$. For higher values of U the spectral density at the Fermi level is still finite and vanishes only in the limit $U \rightarrow \infty$ (or for $T \rightarrow 0$, provided that $U > U_{c2}(T = 0)$).

For the insulating phase DMFT predicts the filling of the Mott-Hubbard gap with increasing temperature. This is due to the fact that the insulator and the metal are not distinct phases in

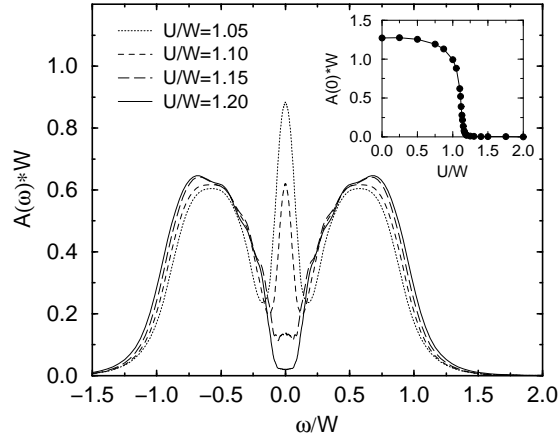


Fig. 8: Spectral function for the half-filled Hubbard model for various values of U at $T = 0.0276 W$ in the crossover region. The crossover from the metal to the insulator occurs via a gradual suppression of the quasiparticle peak at $\omega = 0$. The inset shows the U dependence of $A(\omega = 0)$, in particular the rapid decrease for $U \approx 1.1 W$; from Ref. [55].

the crossover regime, implying that the insulator has a finite spectral weight at the Fermi level. Altogether, the thermodynamic transition line $U_c(T)$ corresponding to the Mott-Hubbard MIT is found to be of first order at finite temperatures, being associated with a hysteresis region in the interaction range $U_{c1} < U < U_{c2}$ where U_{c1} and U_{c2} are the values at which the insulating and metallic solution, respectively, vanishes [29, 39, 57, 55, 58, 54]. The state-of-the-art MIT phase diagram [54] is shown in Fig. 9. The hysteresis region terminates at a critical point. For higher temperatures the transition changes into a smooth crossover from a bad metal to a bad insulator.

It is interesting to note that the slope of the phase transition line is negative down to $T = 0$, which implies that for constant interaction U the metallic phase can be reached from the insulator by decreasing the temperature T , i.e., by *cooling*. This anomalous behavior (which corresponds to the Pomeranchuk effect [59] in ^3He , if we associate solid ^3He with the insulator and liquid ^3He with the metal) can be easily understood from the Clausius-Clapeyron equation $dU/dT = \Delta S/\Delta D$. Here ΔS is the difference between the entropy in the metal and in the insulator, and ΔD is the difference between the number of doubly occupied sites in the two phases. Within the single-site DMFT there is no exchange coupling J between the spins of the electrons in the insulator, since the scaling (15) implies $J \propto -t^2/U \propto 1/d \rightarrow 0$ for $d \rightarrow \infty$. Hence the entropy of the macroscopically degenerate insulating state is $S_{\text{ins}} = k_B \ln 2$ per electron down to $T = 0$. This is larger than the entropy $S_{\text{met}} \propto T$ per electron in the Landau Fermi-liquid describing the metal, i.e., $\Delta S = S_{\text{met}} - S_{\text{ins}} < 0$. At the same time the number of doubly occupied sites is lower in the insulator than in the metal, i.e., $\Delta D = D_{\text{met}} - D_{\text{ins}} > 0$. The Clausius-Clapeyron equation then implies that the phase-transition line T vs. U has a negative slope down to $T = 0$. However, this is an artifact of the single-site DMFT. Namely, there will always exist an exchange coupling between the electrons leading to a vanishing entropy of the insulator at $T = 0$. Since the entropy of the insulator vanishes faster

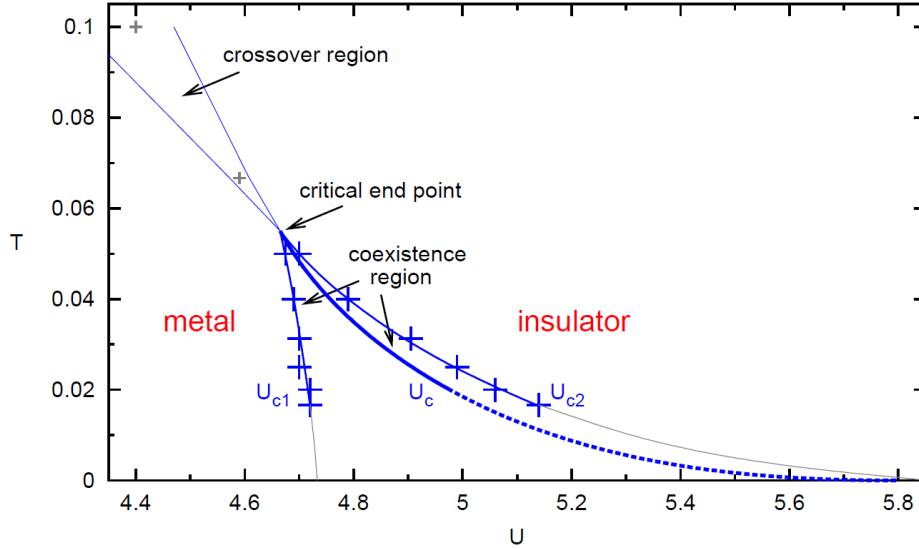


Fig. 9: Mott-Hubbard MIT phase diagram showing the metallic phase and the insulating phase, respectively, at temperatures below the critical end point, as well as a coexistence region; from Ref. [54].

than linearly with the temperature, the difference $\Delta S = S_{\text{met}} - S_{\text{ins}}$ eventually becomes positive, whereby the slope also becomes positive at lower temperatures;⁸ this is indeed observed in cluster DMFT calculations [60]. Since $\Delta S = 0$ at $T = 0$ the phase boundary must terminate at $T = 0$ with infinite slope.

At half filling and for bipartite lattices in dimensions $d > 2$ (in $d = 2$ only at $T = 0$), the paramagnetic phase is unstable against antiferromagnetic long-range order. The metal-insulator transition is then completely hidden by the antiferromagnetic insulating phase, as shown in Fig. 10.

6 Electronic correlations in materials

6.1 LDA+DMFT

Although the Hubbard model is able to explain basic features of the phase diagram of correlated electrons it cannot explain the physics of real materials in any detail. Clearly, realistic theories must take into account the explicit electronic and lattice structure of the systems.

Until recently the electronic properties of solids were investigated by two essentially separate communities, one using model Hamiltonians in conjunction with many-body techniques, the other employing density functional theory (DFT) [62, 63]. DFT and its local density approximation (LDA) have the advantage of being *ab initio* approaches which do not require empirical

⁸Here we assume for simplicity that the metal remains a Fermi liquid, and the insulator stays paramagnetic, down to the lowest temperatures. In fact, a Cooper pair instability will eventually occur in the metal, and the insulator will become long-range ordered, too. In this case the slope dU/dT can change sign several times depending on the value of the entropy of the two phases across the phase transition.

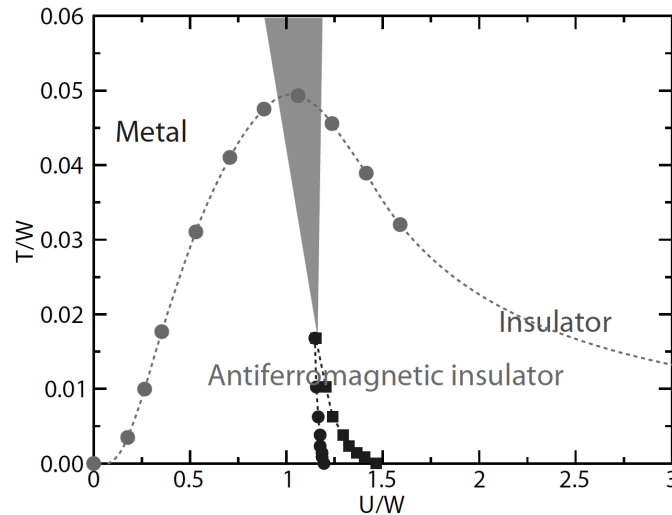


Fig. 10: On bipartite lattices and for half filling ($n = 1$) the paramagnetic phase is unstable against antiferromagnetism. The metal-insulator transition is then completely hidden by the antiferromagnetic insulating phase; from Ref. [61].

parameters as input. Indeed, they are highly successful techniques for the calculation of the electronic structure of real materials [64]. However, in practice DFT/LDA is seriously restricted in its ability to describe strongly correlated materials where the on-site Coulomb interaction is comparable with the band width. Here, the model Hamiltonian approach is more general and powerful since there exist systematic theoretical techniques to investigate the many-electron problem with increasing accuracy. Nevertheless, the uncertainty in the choice of the model parameters and the technical complexity of the correlation problem itself prevent the model Hamiltonian approach from being a flexible or reliable enough tool for studying real materials. The two approaches are therefore complementary. In view of the individual power of DFT/LDA and the model Hamiltonian approach, respectively, it had always been clear that a combination of these techniques would be highly desirable for *ab initio* investigations of real materials, including, e.g., *f*-electron systems and Mott insulators. One of the first successful attempts in this direction was the LDA+U method [65, 66], which combines LDA with a basically static, i.e., Hartree-Fock-like, mean-field approximation for a multi-band Anderson lattice model (with interacting and non-interacting orbitals). This method proved to be a very useful tool in the study of long-range ordered, insulating states of transition metals and rare-earth compounds. However, the paramagnetic metallic phase of correlated electron systems such as high-temperature superconductors and heavy-fermion systems clearly requires a treatment that goes beyond a static mean-field approximation and includes dynamical effects, e.g., the frequency dependence of the self-energy.

Here the recently developed LDA+DMFT method — a new computational scheme which merges electronic band structure calculations and the dynamical mean-field theory [67–76, 30] — has proved to be a breakthrough. Starting from conventional band structure calculations in the local density approximation (LDA) the correlations are taken into account by the Hubbard interaction

and a Hund's rule coupling term. The resulting DMFT equations are solved numerically with a quantum Monte-Carlo (QMC) algorithm. By construction, LDA+DMFT includes the correct quasiparticle physics and the corresponding energetics. It also reproduces the LDA results in the limit of weak Coulomb interaction U . More importantly, LDA+DMFT correctly describes the correlation induced dynamics near a Mott-Hubbard MIT and beyond. Thus, LDA+DMFT and related approaches [77, 78] are able to account for the physics at all values of the Coulomb interaction and doping level.

In the LDA+DMFT approach the LDA band structure is expressed by a one-particle Hamiltonian \hat{H}_{LDA}^0 , and is then supplemented by the local Coulomb repulsion U and Hund's rule exchange J (here we follow the presentation of Held *et al.* [72]). This leads to a material specific generalization of the one-band model Hamiltonian

$$\hat{H} = \hat{H}_{\text{LDA}}^0 + U \sum_m \sum_i \hat{n}_{im\uparrow} \hat{n}_{im\downarrow} + \sum_{i,m \neq m', \sigma, \sigma'} (V - \delta_{\sigma\sigma'} J) \hat{n}_{im\sigma} \hat{n}_{im'\sigma'}. \quad (30)$$

Here m and m' enumerate the three interacting t_{2g} orbitals of the transition metal ion or the $4f$ orbitals in the case of rare earth elements. The interaction parameters are related by $V = U - 2J$ which holds exactly for degenerate orbitals and is a good approximation for the t_{2g} . The actual values for U and V can be obtained from an averaged Coulomb parameter \bar{U} and Hund's exchange J , which can be calculated by constrained LDA.

In the one-particle part of the Hamiltonian

$$\hat{H}_{\text{LDA}}^0 = \hat{H}_{\text{LDA}} - \sum_i \sum_{m\sigma} \Delta\epsilon_d \hat{n}_{im\sigma}. \quad (31)$$

the energy term containing $\Delta\epsilon_d$ is a shift of the one-particle potential of the interacting orbitals. It cancels the Coulomb contribution to the LDA results ("double-counting correction") and can also be calculated by constrained LDA [72].

Within the LDA+DMFT scheme the self-consistency condition connecting the self-energy Σ and the Green function G at frequency ω reads:

$$G_{qm,q'm'}(\omega) = \frac{1}{V_B} \int d^3k \left([\omega \mathbf{1} + \mu \mathbf{1} - H_{\text{LDA}}^0(\mathbf{k}) - \Sigma(\omega)]^{-1} \right)_{qm,q'm'}. \quad (32)$$

Here, $\mathbf{1}$ is the unit matrix, μ the chemical potential, $H_{\text{LDA}}^0(\mathbf{k})$ is the orbital matrix of the LDA Hamiltonian derived, for example, in a linearized muffin-tin orbital (LMTO) basis, $\Sigma(\omega)$ denotes the self-energy matrix which is nonzero only between the interacting orbitals, and $[\dots]^{-1}$ implies the inversion of the matrix with elements n ($=qm$), n' ($=q'm'$), where q and m are the indices of the atom in the primitive cell and of the orbital, respectively. The integration extends over the Brillouin zone with volume V_B (we note that \hat{H}_{LDA}^0 may include additional non-interacting orbitals).

For cubic transition metal oxides Eq. (32) can be simplified to

$$G(\omega) = G^0(\omega - \Sigma(\omega)) = \int d\epsilon \frac{N^0(\epsilon)}{\omega - \Sigma(\omega) - \epsilon} \quad (33)$$

if the degenerate t_{2g} orbitals crossing the Fermi level are well separated from the other orbitals. For non-cubic systems the degeneracy is lifted. In this case we employ Eq. (33) as an approximation, using different $\Sigma_m(\omega)$, $N_m^0(\epsilon)$ and $G_m(\omega)$ for the three non-degenerate t_{2g} orbitals. The Hamiltonian (30) is solved within the DMFT using standard quantum Monte-Carlo (QMC) techniques [32] to solve the self-consistency equations. From the imaginary time QMC Green function one can calculate the physical (real frequency) spectral function with the maximum entropy method [79].

6.2 Single-particle spectrum of correlated electron materials

Transition metal oxides are an ideal laboratory for the study of electronic correlations in solids. Among these materials, cubic perovskites have the simplest crystal structure and thus may be viewed as a starting point for understanding the electronic properties of more complex systems. Typically, the $3d$ states in those materials form comparatively narrow bands with width $W \sim 2-3$ eV, which leads to strong Coulomb correlations between the electrons.

Photoemission spectra provide a direct experimental tool to study the electronic structure and spectral properties of electronically correlated materials. In particular, spectroscopic studies of strongly correlated $3d^1$ transition metal oxides [80, 6, 5, 81–83] find a pronounced lower Hubbard band in the photoemission spectra which cannot be explained by conventional band structure theory. These are typical correlation effects which will now be illustrated by results obtained with the LDA+DMFT approach for the simple $3d^1$ transition metal compounds SrVO_3 and CaVO_3 .

The main effect of the substitution of Sr ions in SrVO_3 by the isovalent, but smaller, Ca ions is to decrease the V-O-V angle from $\theta = 180^\circ$ in SrVO_3 to $\theta \approx 162^\circ$ in the orthorhombically distorted structure of CaVO_3 . However, this rather strong bond bending results only in a 4% decrease of the one-particle bandwidth W and thus in a correspondingly small increase of the ratio U/W as one moves from SrVO_3 to CaVO_3 .

LDA+DMFT(QMC) spectra of SrVO_3 and CaVO_3 were calculated [5] by starting from the respective LDA DOS of the two materials; they are shown in Fig. 11. These spectra show genuine correlation effects, i.e., the formation of lower Hubbard bands at about 1.5 eV and upper Hubbard bands at about 2.5 eV, with well-pronounced quasiparticle peaks at the Fermi energy. Therefore both SrVO_3 and CaVO_3 are strongly correlated metals [5, 84]. The DOS of the two systems shown in Fig. 11 are quite similar. In fact, SrVO_3 is slightly less correlated than CaVO_3 , in accord with their different LDA bandwidths. The inset of Fig. 11 shows that the effect of temperature on the spectrum is small for $T \lesssim 700$ K.

Since the three t_{2g} orbitals of this simple $3d^1$ material are (almost) degenerate the spectral function has the same three-peak structure as that of the one-band Hubbard model shown in Fig. 8. The temperature induced decrease of the quasiparticle peak height is also clearly seen. As noted in Sec. 5 the actual form of the spectrum no longer resembles the input (LDA) DOS, i.e., it essentially depends only on the first three energy moments of the LDA DOS (electron density, average energy, band width).

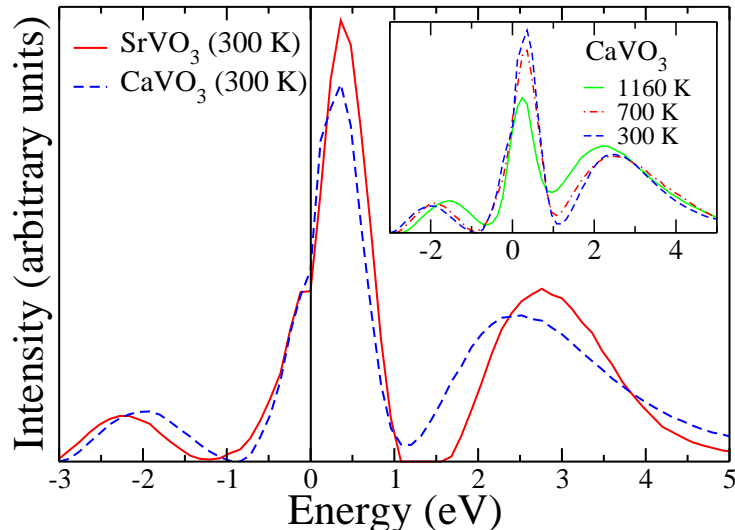


Fig. 11: LDA+DMFT(QMC) spectrum of SrVO_3 (solid line) and CaVO_3 (dashed line) calculated at $T=300$ K; inset: effect of temperature in the case of CaVO_3 ; after Ref. [5].

In the left panel of Fig. 12 the LDA+DMFT(QMC) spectra at 300K are compared with experimental high-resolution bulk PES; for details see ref. [85]. The quasiparticle peaks in theory and experiment are seen to be in very good agreement. In particular, their height and width are almost identical for both SrVO_3 and CaVO_3 . The difference in the positions of the lower Hubbard bands may be partly due to (i) the subtraction of the (estimated) oxygen contribution which might also remove some $3d$ spectral weight below -2 eV, and (ii) uncertainties in the *ab initio* calculation of the local Coulomb interaction strength. In the right panel of Fig. 12 comparison is made with XAS data [86]. Again, the overall agreement of the weights and positions of the quasiparticle and upper t_{2g} Hubbard band is good, including the tendencies when going from SrVO_3 to CaVO_3 ($\text{Ca}_{0.9}\text{Sr}_{0.1}\text{VO}_3$ in the experiment). For CaVO_3 the weight of the quasiparticle peak is somewhat lower than in the experiment. In contrast to one-band Hubbard model calculations, the material specific results reproduce the strong asymmetry around the Fermi energy with respect to weights and bandwidths. The slight differences in the quasiparticle peaks (see Fig. 11) lead to different effective masses, namely $m^*/m = 2.1$ for SrVO_3 and $m^*/m = 2.4$ for CaVO_3 . These theoretical values agree with $m^*/m = 2 - 3$ for SrVO_3 and CaVO_3 as obtained from de Haas-van Alphen experiments and thermodynamics [87].

The experimentally determined spectra of SrVO_3 and CaVO_3 and the good agreement with parameter-free LDA+DMFT calculations confirm the existence of a pronounced three-peak structure in a correlated bulk material. Although the DMFT had predicted such a behavior for the Hubbard model (see Sec. 5.1.) it was not clear whether the DMFT result would really be able to describe real materials in three dimensions. Now it has been confirmed that the three-peak structure not only occurs in single-impurity Anderson models but also in three-dimensional correlated bulk matter.

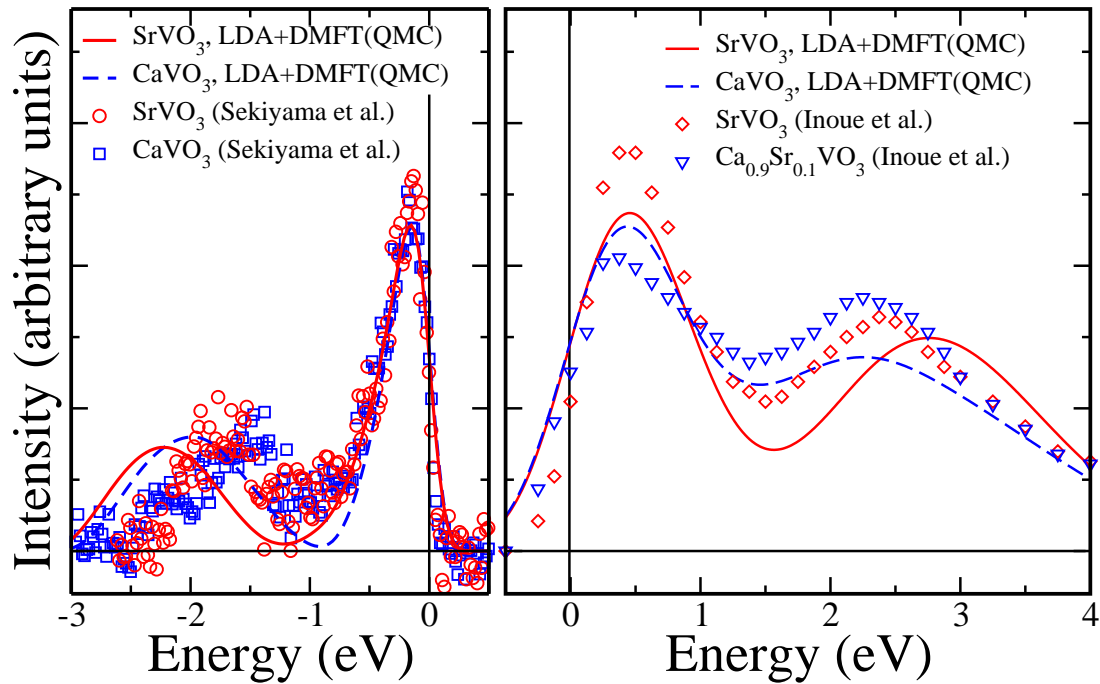


Fig. 12: Comparison of the calculated, parameter-free LDA+DMFT(QMC) spectra of SrVO_3 (solid line) and CaVO_3 (dashed line) with experiment. Left: Bulk-sensitive high-resolution PES (SrVO_3 : circles; CaVO_3 : rectangles). Right: XAS for SrVO_3 (diamonds) and $\text{Ca}_{0.9}\text{Sr}_{0.1}\text{VO}_3$ (triangles) [86]. Horizontal line: experimental subtraction of the background intensity; after Ref. [85].

7 Summary and outlook

Due to the intensive international research over the last two decades the DMFT has quickly developed into a powerful method for the investigation of electronic systems with strong correlations. It provides a comprehensive, non-perturbative and thermodynamically consistent approximation scheme for the investigation of finite-dimensional systems (in particular for dimension $d = 3$), and is particularly useful for the study of problems where perturbative approaches are inapplicable. For this reason the DMFT has now become the standard mean-field theory for fermionic correlation problems, including cold atoms in optical lattices [88–92]. The study of models in non-equilibrium within a suitable generalization of the DMFT has become yet another fascinating new research area [93–101].

Until a few years ago research into correlated electron systems concentrated on homogeneous bulk systems. DMFT investigations of systems with internal or external inhomogeneities such as thin films and multi-layered nanostructures are still very new [102–107]. They are particularly important in view of the novel types of functionalities of such systems, which may have important applications in electronic devices. Here the DMFT and its generalizations will certainly be very useful.

In particular, the development of the *ab initio* band-structure calculation technique referred to as LDA+DMFT has proved to be a breakthrough in the investigation of electronically correlated

materials. It has already provided important insights into the spectral and magnetic properties of correlated electron materials, e.g., transition metals and their oxides. Clearly, this approach has a great potential for further developments. It is the goal of the DFG Research Unit on *Dynamical Mean-Field Approach with Predictive Power for Strongly Correlated Materials* which organizes this Autumn School “Hands-on LDA+DMFT” to develop the LDA+DMFT framework into a comprehensive *ab initio* approach which will be able to describe, and even predict, the properties of complex correlated materials.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] J.H. de Boer and E.J.W. Verwey, Proc. Phys. Soc. **49**, No. **4S**, 59 (1937)
- [2] N.F. Mott and R. Peierls, Proc. Phys. Soc. A **49**, 72 (1937)
- [3] D.B. McWhan, A. Menth, J.P. Remeika, W.F. Brinkman, and T.M. Rice, Phys. Rev. B **B2**, 1920 (1973)
- [4] G.A. Sawatzky and J.W. Allen, Phys. Rev. Lett. **53**, 2339 (1984)
- [5] A. Sekiyama, H. Fujiwara, S. Imada, S. Suga, H. Eisaki, S.I. Uchida, K. Takegahara, H. Harima, Y. Saitoh, I.A. Nekrasov, G. Keller, D. E. Kondakov, A.V. Kozhevnikov, T. Pruschke, K. Held, D. Vollhardt, and V. I. Anisimov, Phys. Rev. Lett. **93**, 156402 (2004)
- [6] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)
- [7] M.C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963)
- [8] J. Hubbard, Proc. Roy. Soc. London **A276**, 238 (1963)
- [9] J. Kanamori, Prog. Theor. Phys. **30**, 275 (1963)
- [10] R.J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982)
- [11] D. Vollhardt, in *Correlated Electron Systems*, edited by V.J. Emery (World Scientific, Singapore, 1993) p. 57; http://www.physik.uni-augsburg.de/theo3/Research/research_jerusalem.vollha.en.shtml
- [12] M. Eckstein, M. Kollar, K. Byczuk and D. Vollhardt, Phys. Rev. B **71**, 235119 (2005)
- [13] C. Itzykson and J.-M. Drouffe, *Statistical Field Theory* (Cambridge University Press, Cambridge, 1989)
- [14] R. Brout, Phys. Rev. **118**, 1009 (1960)
- [15] C.J. Thompson, Commun. Math. Phys. **36**, 255 (1974)
- [16] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)
- [17] U. Wolff, Nucl. Phys. B **225**, 391 (1983)
- [18] E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989)
- [19] W. Metzner, Z. Phys. B **77**, 253 (1989)
- [20] E. Müller-Hartmann, Z. Phys. B **76**, 211 (1989)
- [21] H. Schweitzer and G. Czycholl, Solid State Comm. **69**, 171 (1989)

- [22] U. Brandt and C. Mielsch, *Z. Phys.* **B 75**, 365 (1989)
- [23] P.G.J. van Dongen, F. Gebhard, and D. Vollhardt, *Z. Phys.* **76**, 199 (1989)
- [24] V. Janiš, *Z. Phys. B* **83**, 227 (1991)
- [25] V. Janiš and D. Vollhardt, *Int. J. Mod. Phys. B* **6**, 731 (1992)
- [26] A. Georges and G. Kotliar, *Phys. Rev. B* **45**, 6479 (1992)
- [27] M. Jarrell, *Phys. Rev. Lett.* **69**, 168 (1992)
- [28] Th. Pruschke, M. Jarrell, and J.K. Freericks, *Adv. Phys.* **44**, 187 (1995)
- [29] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [30] G. Kotliar and D. Vollhardt, *Physics Today* **3**, 53 (2004)
- [31] D. Vollhardt, in *Lectures on the Physics of Strongly Correlated Systems XIV, AIP Conference Proceedings*, vol. 1297, ed. by A. Avella, F. Mancini (American Institute of Physics, Melville, 2010), p. 339; arXiv:1004.5069v3
- [32] J.E. Hirsch and R.M. Fye, *Phys. Rev. Lett.* **56**, 2521 (1986)
- [33] X.Y. Zhang, M.J. Rozenberg, and G. Kotliar, *Phys. Rev. Lett.* **70**, 1666 (1993)
- [34] J.K. Freericks and V. Zlatić, *Rev. Mod. Phys.* **75**, 1333 (2003)
- [35] M.J. Rozenberg, X.Y. Zhang, and G. Kotliar, *Phys. Rev. Lett.* **69**, 1236 (1992)
- [36] A. Georges and W. Krauth, *Phys. Rev. Lett.* **69**, 1240 (1992)
- [37] Th. Maier, M. Jarrell, Th. Pruschke, and M.H. Hettler, *Rev. Mod. Phys.* **77**, 1027 (2005)
- [38] E. Gull, A. J. Millis, A. I. Lichtenstein, A. N. Rubtsov, M. Troyer, P. Werner, *Rev. Mod. Phys.* **83**, 349 (2011)
- [39] R. Bulla, *Phys. Rev. Lett.* **83**, 136 (1999)
- [40] R. Bulla, T.A. Costi, and Th. Pruschke, *Rev. Mod. Phys.* **80**, 395 (2008)
- [41] M. Caffarel and W. Krauth, *Phys. Rev. Lett.* **72**, 1545 (1994)
- [42] Q. Si, M.J. Rozenberg, G. Kotliar, and A.E. Ruckenstein, *Phys. Rev. Lett.* **72**, 2761 (1994)
- [43] M.J. Rozenberg, G. Moeller, and G. Kotliar, *Mod. Phys. Lett. B* **8**, 535 (1994)
- [44] N.F. Mott, *Rev. Mod. Phys.* **40**, 677 (1968)
- [45] N.F. Mott, *Metal–Insulator Transitions* (Taylor and Francis, London, 1990), 2 edition

-
- [46] F. Gebhard, *The Mott Metal-Insulator Transition*, Springer, Berlin, 1997
- [47] D.B. McWhan and J.P. Remeika, Phys. Rev. B **B2**, 3734 (1970)
- [48] T.M. Rice and D.B. McWhan, IBM J. Res. Develop. **251** (May 1970)
- [49] J. Hubbard, Proc. Roy. Soc. London **A281**, 401 (1964)
- [50] W.F. Brinkman and T.M. Rice, Phys. Rev. B **2**, 4302 (1970)
- [51] M.C. Gutzwiller, Phys. Rev. **137**, A1726 (1965)
- [52] W. Metzner and D. Vollhardt, Phys. Rev. B **37**, 7382 (1988)
- [53] D. Vollhardt, Rev. Mod. Phys. **56**, 99 (1984)
- [54] N. Blümer, *Metal-Insulator Transition and Optical Conductivity in High Dimensions* (Shaker Verlag, Aachen, 2003)
- [55] R. Bulla, T.A. Costi, and D. Vollhardt, Phys. Rev. B **64**, 045103 (2001)
- [56] D. Vollhardt, K. Held, G. Keller, R. Bulla, Th. Pruschke, I.A. Nekrasov, and V.I. Anisimov, J. Phys. Soc. Jpn. **74**, 136 (2005)
- [57] M.J. Rozenberg, R. Chitra, and G. Kotliar, Phys. Rev. Lett. **83**, 3498 (1999)
- [58] J. Joo and V. Oudovenko, Phys. Rev. B **64**, 193102 (2001)
- [59] D. Vollhardt and P. Wölfle, *The Superfluid Phases of Helium 3* (Taylor and Francis, London, 1990)
- [60] H. Park, K. Haule, and G. Kotliar, Phys. Rev. Lett. **101**, 186403 (2008)
- [61] Th. Pruschke, Prog. Theor. Phys. Suppl. **160**, 274 (2005)
- [62] P. Hohenberg and W. Kohn, Phys. Rev. **136B**, 864 (1964).
- [63] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965)
- [64] R.O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989)
- [65] V.I. Anisimov, J. Zaanen, and O.K. Andersen, Phys. Rev. B **44**, 943 (1991)
- [66] V.I. Anisimov, F. Aryasetiawan, and A.I. Lichtenstein, J. Phys. Cond. Matter **9**, 767 (1997)
- [67] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, and G. Kotliar, J. Phys.: Cond. Matt. **9**, 7359 (1997)
- [68] A.I. Lichtenstein and M.I. Katsnelson, Phys. Rev. B **57**, 6884 (1998)

- [69] I.A. Nekrasov, K. Held, N. Blümer, A.I. Poteryaev, V.I. Anisimov, and D. Vollhardt, *Eur. Phys. J. B* **18**, 55 (2000)
- [70] K. Held, I.A. Nekrasov, N. Blümer, V.I. Anisimov, and D. Vollhardt, *Int. J. Mod. Phys. B* **15**, 2611 (2001)
- [71] A.I. Lichtenstein, M.I. Katsnelson, and G. Kotliar, in *Electron Correlations and Materials Properties*, edited by A. Gonis, N. Kioussis, and M. Ciftan (Kluwer Academic/Plenum, New York, 2002), p. 428
- [72] K. Held, I.A. Nekrasov, G. Keller, V. Eyert, N. Blümer, A.K. McMahan, R.T. Scalettar, T. Pruschke, V.I. Anisimov, and D. Vollhardt, *Psi-k Newsletter* **56**, 65 (2003); reprinted in *Phys. Status Solidi B* **243**, 2599 (2006)
- [73] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, C. A. Marianetti, *Rev. Mod. Phys.* **78**, 865 (2006)
- [74] K. Held, *Adv. Phys.* **56**, 829 (2007)
- [75] M.I. Katsnelson, V.Yu. Irkhin, L. Chioncel, A.I. Lichtenstein, and R.A. de Groot, *Rev. Mod. Phys.* **80**, 315 (2008)
- [76] J. Kuneš, I. Leonov, M. Kollar, K. Byczuk, V.I. Anisimov, and D. Vollhardt, *Eur. Phys. J. Special Topics* **180**, 5 (2010)
- [77] S. Biermann, F. Aryasetiawan, and A. Georges, *Phys. Rev. Lett.* **90**, 086402 (2003)
- [78] J. Minár, L. Chioncel, A. Perlov, H. Ebert, M.I. Katsnelson, and A.I. Lichtenstein, *Phys. Rev. B* **72**, 45125 (2005)
- [79] M. Jarrell and J.E. Gubernatis, *Phys. Rep.* **269**, 133 (1996)
- [80] A. Fujimori, I. Hase, H. Namatame, Y. Fujishima, Y. Tokura, H. Eisaki, S. Uchida, K. Takegahara, and F.M.F. de Groot, *Phys. Rev. Lett.* **69**, 1796 (1992)
- [81] S.-K. Mo, H.-D. Kim, J.W. Allen, G.H. Gweon, J.D. Denlinger, J.-H. Park, A. Sekiyama, A. Yamasaki, S. Suga, P. Metcalf, and K. Held, *Phys. Rev. Lett.* **93**, 076404 (2004)
- [82] D. Schrupp, M. Sing, M. Tsunekawa, H. Fujiwara, S. Kasai, A. Sekiyama, S. Suga, T. Muro, V.A.M. Brabers, and R. Claessen, *Europhys. Lett.* **70**, 789 (2005)
- [83] T. C. Koethe, Z. Hu, M.W. Haverkort, C. Schüßler-Langeheine, F. Venturini, N.B. Brookes, O. Tjernberg, W. Reichelt, H.H. Hsieh, H.-J. Lin, C.T. Chen, and L.H. Tjeng, *Phys. Rev. Lett.* **97**, 116402 (2006)
- [84] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges, and O.K. Andersen, *Phys. Rev. Lett.* **92**, 176403 (2004)

-
- [85] I.A. Nekrasov, G. Keller, D.E. Kondakov, A.V. Kozhevnikov, Th. Pruschke, K. Held, D. Vollhardt, and V.I. Anisimov, *Phys. Rev. B* **72**, 155106 (2005)
- [86] I.H. Inoue, I. Hase, Y. Aiura, A. Fujimori, K. Morikawa, T. Mizokawa, Y. Haruyama, T. Maruyama, and Y. Nishihara, *Physica C* **235-240**, 1007 (1994)
- [87] I.H. Inoue, C. Bergemann, I. Hase, and S.R. Julian, *Phys. Rev. Lett.* **88**, 236403 (2002)
- [88] A. Rapp, G. Zarand, C. Honerkamp, W. Hofstetter, *Phys. Rev. Lett.* **98**, 160405 (2007)
- [89] M. Snoek, I. Titvinidze, C. Toke, K. Byczuk, and W. Hofstetter, *New J. Phys.* **10**, 093008 (2008)
- [90] R. Jördens, N. Strohmaier, K. Günter, H. Moritz, and T. Esslinger, *Nature* **455**, 204 (2008)
- [91] U. Schneider, L. Hackermüller, S. Will, Th. Best, I. Bloch, T.A. Costi, R.W. Helmes, D. Rasch, A. Rosch, *Science* **322**, 1520 (2008)
- [92] I. Bloch, J. Dalibard, and W. Zwerger, *Rev. Mod. Phys.* **80**, 885 (2008)
- [93] V. Turkowski and J.K. Freericks, *Phys. Rev. B* **71**, 085104 (2005)
- [94] J.K. Freericks, V.M. Turkowski, and V. Zlatić, *Phys. Rev. Lett.* **97** 266408 (2006)
- [95] J.K. Freericks, *Phys. Rev. B* **77**, 075109 (2008)
- [96] N. Tsuji, T. Oka, and H. Aoki, *Phys. Rev. B* **78**, 235124 (2008)
- [97] M. Eckstein, and M. Kollar, *Phys. Rev. Lett.* **100**, 120404 (2008)
- [98] M. Eckstein, and M. Kollar, *Phys. Rev. B* **78**, 205119 (2008)
- [99] J.K. Freericks, H.R. Krishnamurthy, and Th. Pruschke, *Phys. Rev. Lett.* **102**, 136401 (2009)
- [100] M. Eckstein, M. Kollar, and P. Werner, *Phys. Rev. Lett.* **103**, 056403 (2009)
- [101] M. Eckstein and P. Werner, *Phys. Rev. B* **82**, 115115 (2010)
- [102] M. Potthoff and W. Nolting, *Phys. Rev. B* **59**, 2549 (1999)
- [103] J.K. Freericks, *Transport in multilayered nanostructures — The dynamical mean-field approach* (Imperial College Press, London, 2006)
- [104] M. Takizawa, H. Wadati, K. Tanaka, M. Hashimoto, T. Yoshida, A. Fujimori, A. Chikamtsu, H. Kumigashira, M. Oshima, K. Shibuya, T. Mihara, T. Ohnishi, M. Lippmaa, M. Kawasaki, H. Koinuma, S. Okamoto, and A.J. Millis, *Phys. Rev. Lett.* **97**, 057601 (2006)

- [105] L. Chen and J.K. Freericks, *Phys. Rev. B* **75**, 1251141 (2007)
- [106] K. Byczuk, in *Condensed Matter Physics in the Prime of the 21st Century: Phenomena, Materials, Ideas, Methods*, ed. J. Jedrzejewski (World Scientific, Singapore, 2008), p. 1
- [107] R.W. Helmes, T.A. Costi, and A. Rosch, *Phys. Rev. Lett.* **100**, 056403 (2008)

2 Theory and Practice of Density-Functional Theory

Peter E. Blöchl

Institute for Theoretical Physics

Clausthal University of Technology

Contents

1	Introduction	2
2	Basics of density-functional theory	3
3	Jacob's ladder of density functionals	12
4	Benchmarks, successes and failures	16
5	Electronic structure methods	17
6	Augmented wave methods	19
7	Pseudopotentials	21
8	Projector augmented-wave method	25
A	Model exchange-correlation energy	37
B	Large-gradient limit of the enhancement factor	37

1 Introduction

On the nanoscale, materials around us have surprisingly simple structures: The standard model of solid state physics and chemistry only knows of two types of particles, namely the nuclei making up the periodic table and the electrons. Only one kind of interaction between them needs to be considered, namely the electrostatic interaction. Even magnetic forces are important only in rare occasions. All other fundamental particles and interactions are irrelevant for chemistry. The behavior of these particles can be described by the Schrödinger equation (or better the relativistic Dirac equation), which is easily written down. However, the attempt to solve this equation for any system of interest fails miserably due to what Walter Kohn termed the exponential wall [1].

To obtain an impression of the powers of the exponential wall, imagine the wave function of a N_2 molecule, having two nuclei and fourteen electrons. For N particles, the Schrödinger equation is a partial differential equation in $3N$ dimensions. Let us express the wave function on a grid with about 100 points along each spatial direction and let us consider two spin states for each electron. Such a wave function is represented by $2^{14} 100^{3 \times 16} \approx 10^{100}$ complex numbers. A data server for this amount of data, made of current terabyte hard disks, would occupy a volume with a diameter of 10^{10} light years!

Treating the nuclei as classical particles turned out to be a good approximation, but the quantum nature of the electrons cannot be ignored. A great simplification is to describe electrons as non-interacting quasi particles. Instead of one wave function in $3N$ dimensions, one only needs to describe N wave functions in three dimensions each, a dramatic simplification from 10^{100} to 10^7 numbers.

While the independent-particle model is very intuitive, and while it forms the basis of most text books on solid-state physics, materials physics, and chemistry, the Coulomb interaction between electrons is clearly not negligible.

Here, density-functional theory [2,3] comes to our rescue: it provides a rigorous mapping from interacting electrons onto a system of non-interacting electrons. Unfortunately, the exact mapping is utterly complicated, and this is where all the complexity goes. Luckily, there are simple approximations that are both intuitive and surprisingly accurate. Furthermore, with the help of clever algorithms, density-functional calculations can be performed on current computers for large systems with several hundred atoms in a unit cell or a molecule. The microscopic insight gained from density functional calculations is a major source of progress in solid state physics, chemistry, material science, and biology.

In the first part of this article, I will try to familiarize the novice reader with the basics of density-functional theory, provide some guidance into common approximations and give an idea of the type of problems that can be studied with density functional theory.

Beyond this article, I recommend the insightful review articles on density functional theory by Jones and Gunnarsson [4], Baerends [5], von Barth [6], Perdew [7], Yang [8], and their collaborators.

Solving the one-particle Schrödinger equation, which results from density-functional theory,

for real materials is a considerable challenge. Several avenues have been developed to their solution. This is the field of electronic structure methods, which will be discussed in the second part of this article. This part is taken from earlier versions by Clemens Först, Johannes Kästner and myself [9, 10].

2 Basics of density-functional theory

The dynamics of the electron wave function is governed by the Schrödinger equation $i\hbar\partial_t|\Psi\rangle = \hat{H}|\Psi\rangle$ with the N -particle Hamiltonian \hat{H} .

$$\hat{H} = \sum_{j=1}^N \left(\frac{-\hbar^2}{2m_e} \nabla_j^2 + v_{ext}(\vec{r}_j) \right) + \frac{1}{2} \sum_{i \neq j}^N \frac{e^2}{4\pi\epsilon_0 |\vec{r}_i - \vec{r}_j|}. \quad (1)$$

With m_e we denote the electron mass, with ϵ_0 the vacuum permittivity, e is the elementary charge and \hbar is the Planck quantum divided by 2π . The Coulomb potentials of the nuclei have been combined into an external potential $v_{ext}(\vec{r})$.

All N -electron wave functions $\Psi(\vec{x}_1, \dots, \vec{x}_N)$ obey the Pauli principle, that is they change their sign, when two of its particle coordinates are exchanged.

We use a notation that combines the position vector $\vec{r} \in \mathbb{R}^3$ of an electron with its discrete spin coordinate $\sigma \in \{\uparrow, \downarrow\}$ into a single vector $\vec{x} := (\vec{r}, \sigma)$. Similarly, we use the notation of a four-dimensional integral $\int d^4x := \sum_{\sigma} \int d^3r$ for the sum over spin indices and the integral over the position. With the generalized symbol $\delta(\vec{x} - \vec{x}') := \delta_{\sigma, \sigma'} \delta(\vec{r} - \vec{r}')$ we denote the product of Kronecker delta of the spin coordinates and Dirac's delta function for the positions. While, at first sight, it seems awkward to combine continuous and discrete numbers, this notation is less error prone than the notation that treats the spin coordinates as indices, where they can be confused with quantum numbers. During the first reading, the novice can ignore the complexity of the spin coordinates, treating \vec{x} like a coordinate. During careful study, he will nevertheless have the complete and concise expressions.

One-particle reduced density matrix and two-particle density

In order to obtain the ground state energy $E = \langle \Psi | \hat{H} | \Psi \rangle$ we need to perform 2^N integrations in $3N$ dimensions each, i.e.

$$E = \int d^4x_1 \cdots \int d^4x_N \Psi^*(\vec{x}_1, \dots, \vec{x}_N) \hat{H} \Psi(\vec{x}_1, \dots, \vec{x}_N). \quad (2)$$

However, only two different types of integrals occur in the expression for the energy, so that most of these integrations can be performed beforehand leading to two quantities of physical significance.

- One of these quantities is the one-particle reduced density matrix $\rho^{(1)}(\vec{x}, \vec{x}')$, which allows one to evaluate all expectation values of one-particle operators such as the kinetic energy

and the external potential,

$$\rho^{(1)}(\vec{x}, \vec{x}') := N \int d^4x_2 \dots \int d^4x_N \Psi(\vec{x}, \vec{x}_2, \dots, \vec{x}_N) \Psi^*(\vec{x}', \vec{x}_2, \dots, \vec{x}_N). \quad (3)$$

- The other one is the two-particle density $n^{(2)}(\vec{r}, \vec{r}')$, which allows to determine the interaction between the electrons,

$$n^{(2)}(\vec{r}, \vec{r}') := N(N-1) \sum_{\sigma, \sigma'} \int d^4x_3 \dots \int d^4x_N |\Psi(\vec{x}, \vec{x}', \vec{x}_3, \dots, \vec{x}_N)|^2. \quad (4)$$

If it is confusing that there are two different quantities depending on two particle coordinates, note that the one-particle reduced density matrix $\rho^{(1)}$ depends on two \vec{x} -arguments of the same particle, while the two-particle density $n^{(2)}$ depends on the positions of two different particles.

With these quantities the total energy is

$$\begin{aligned} E = & \int d^4x' \int d^4x \delta(\vec{x}' - \vec{x}) \left(\frac{-\hbar^2}{2m_e} \vec{\nabla}^2 + v_{ext}(\vec{r}) \right) \rho^{(1)}(\vec{x}, \vec{x}') \\ & + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 n^{(2)}(\vec{r}, \vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|}, \end{aligned} \quad (5)$$

where the gradient of the kinetic energy operates on the first argument \vec{r} of the density matrix.

One-particle reduced density matrix and natural orbitals

In order to make oneself familiar with the one-particle reduced density matrix, it is convenient to diagonalize it. The eigenstates $\varphi_n(\vec{r})$ are called natural orbitals [11] and the eigenvalues \bar{f}_n are their occupations. The index n labeling the natural orbitals may stand for a set of quantum numbers.

The density matrix can be written in the form

$$\rho^{(1)}(\vec{x}, \vec{x}') = \sum_n \bar{f}_n \varphi_n(\vec{x}) \varphi_n^*(\vec{x}'). \quad (6)$$

The natural orbitals are orthonormal one-particle orbitals, i.e.

$$\int d^4x \varphi_m^*(\vec{x}) \varphi_n(\vec{x}) = \delta_{m,n}. \quad (7)$$

Due to the Pauli principle, occupations are non-negative and never larger than one [12]. The natural orbitals already point the way to the world of effectively non-interacting electrons.

The one-particle density matrix provides us with the electron density

$$n^{(1)}(\vec{r}) = \sum_{\sigma} \rho^{(1)}(\vec{x}, \vec{x}) = \sum_{\sigma} \sum_n \bar{f}_n \varphi_n^*(\vec{x}) \varphi_n(\vec{x}). \quad (8)$$

With the natural orbitals, the total energy Eq. 5 obtains the form

$$\begin{aligned} E = & \sum_n \bar{f}_n \int d^4x \varphi_n^*(\vec{x}) \frac{-\hbar^2}{2m} \vec{\nabla}^2 \varphi_n(\vec{x}) + \int d^3r v_{ext}(\vec{r}) n^{(1)}(\vec{r}) \\ & + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 n^{(2)}(\vec{r}, \vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|}. \end{aligned} \quad (9)$$

Two-particle density and exchange-correlation hole

The physical meaning of the two-particle density $n^{(2)}(\vec{r}, \vec{r}')$ is the following: For particles that are completely uncorrelated, meaning that they do not even experience the Pauli principle, the two particle density would be¹ the product of one-particle densities, i.e. $n^{(2)}(\vec{r}, \vec{r}') = n^{(1)}(\vec{r})n^{(1)}(\vec{r}')$. If one particle is at position \vec{r}_0 , the density of the remaining $N - 1$ particles is the conditional density

$$\frac{n^{(2)}(\vec{r}_0, \vec{r})}{n^{(1)}(\vec{r}_0)}.$$

The conditional density is the electron density seen by one of the electrons at \vec{r}_0 . This observer electron obviously only sees the remaining $N - 1$ electrons.

It is convenient to express the two-particle density by the hole function $h(\vec{r}, \vec{r}')$, i.e.

$$n^{(2)}(\vec{r}, \vec{r}') = n^{(1)}(\vec{r}) \left[n^{(1)}(\vec{r}') + h(\vec{r}, \vec{r}') \right]. \quad (10)$$

One electron at position \vec{r} does not “see” the total electron density $n^{(1)}$ with N electrons, but only the density of the $N - 1$ other electrons, because it does not see itself. The hole function $h(\vec{r}_0, \vec{r})$ is simply the difference of the total electron density and the electron density seen by the observer electron at \vec{r}_0 .

The division of the two-particle density in Eq. 10 suggests that we split the electron-electron interaction into the so-called Hartree energy

$$E_H \stackrel{\text{def}}{=} \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 n^{(1)}(\vec{r}) n^{(1)}(\vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} \quad (11)$$

and the *potential energy of exchange and correlation*

$$U_{xc} \stackrel{\text{def}}{=} \int d^3r n^{(1)}(\vec{r}) \frac{1}{2} \int d^3r' \frac{e^2 h(\vec{r}, \vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|}. \quad (12)$$

Keep in mind that U_{xc} is *not* the exchange correlation energy. The difference is a kinetic energy correction that will be discussed later in Eq. 19.

The hole function has a physical meaning: An electron sees the total density minus the electrons accounted for by the hole. Thus each electron not only experiences the electrostatic potential of the total electron density $n^{(1)}(\vec{r})$, but also the attractive potential of its own exchange correlation hole $h(\vec{r}_0, \vec{r})$.

A few facts for this hole density are apparent:

1. Because each electron of a N -electron system sees $N - 1$ other electrons, the hole function integrates to exactly minus one electron

$$\int d^3r h(\vec{r}_0, \vec{r}) = -1 \quad (13)$$

irrespective of the position \vec{r}_0 of the observing electron.

¹This is correct only up to a term that vanishes in the limit of infinite particle number.

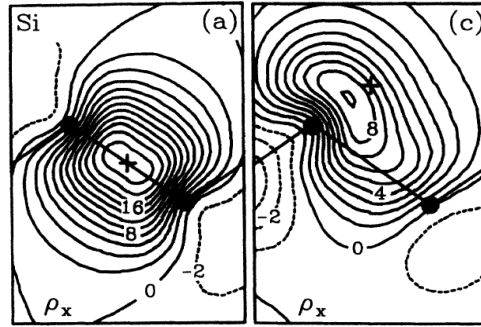


Fig. 1: Exchange hole in silicon. The cross indicates the position of the observer electron. The black spheres and the lines indicate the atomic positions and bonds in the (110) plane. Reprinted figure with permission from Mark S. Hybertsen and Steven G. Louie, *Physical Review B* 34, 5390 (1986). Copyright 1986 by the American Physical Society.

2. The density of the remaining $N - 1$ electrons can not be larger than the total electron density. This implies

$$h(\vec{r}_0, \vec{r}) \geq -n^{(1)}(\vec{r}) . \quad (14)$$

3. Due to the Pauli principle, no other electron with the same spin as the observer electron can be at the position \vec{r}_0 . Thus the on-top hole $h(\vec{r}_0, \vec{r}_0)$ obeys the limits [13]

$$-\frac{1}{2}n^{(1)}(\vec{r}_0) \geq h(\vec{r}_0, \vec{r}_0) \geq -n^{(1)}(\vec{r}_0) . \quad (15)$$

4. Assuming locality, the hole function vanishes at large distances from the observer electron at \vec{r}_0 , i.e.

$$h(\vec{r}_0, \vec{r}) \rightarrow 0 \quad \text{for} \quad |\vec{r} - \vec{r}_0| \rightarrow \infty . \quad (16)$$

With locality I mean that the density does not depend on the position or the presence of an observer electron, if the latter is very far away.

A selfmade functional

It is fairly simple to make our own density functional²: For a given density, we choose a simple shape for the hole function, such as a spherical box. Then we scale the value and the radius such that the hole function integrates to -1 , and that its value is opposite equal to the spin density at its center. The electrostatic potential of this hole density at its center is the exchange-correlation energy for the observer electron. Our model has an exchange correlation energy³ of

$$U_{xc}[n^{(1)}] \approx -\frac{1}{2} \int d^3r n^{(1)}(\vec{r}) \left(\frac{3}{4} \frac{e^2}{4\pi\epsilon_0} \sqrt[3]{\frac{2\pi}{3}} \left(n^{(1)}(\vec{r}) \right)^{\frac{1}{3}} \right) \sim \int d^3r \left(n^{(1)}(\vec{r}) \right)^{\frac{4}{3}} . \quad (17)$$

²A functional $F[y]$ maps a function $y(x)$ to a number F . It is a generalization of the function $F(\vec{y})$ of a vector \vec{y} , where the vector index of \vec{y} is turned into a continuous argument x .

³For this model we do not distinguish between the energy of exchange and correlation and its potential energy contribution

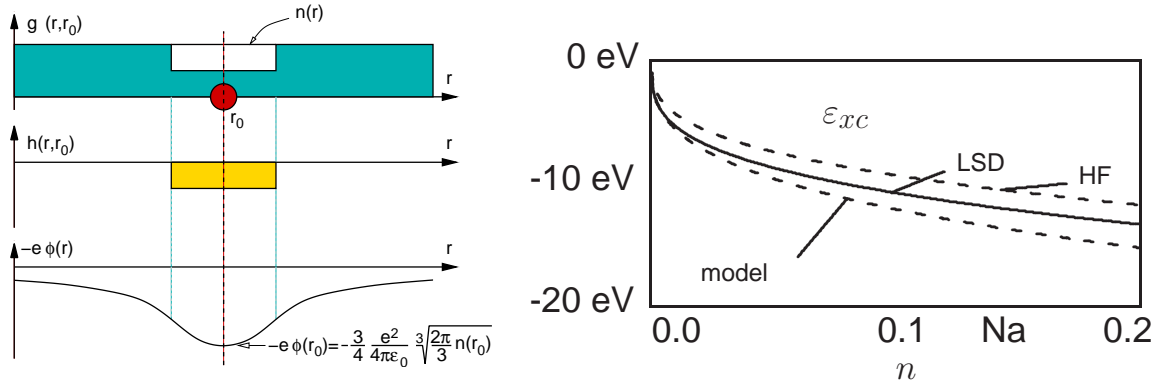


Fig. 2: Left: Scheme to demonstrate the construction of the exchange correlation energy from a simple model. Right: exchange correlation energy per electron ϵ_{xc} as function of electron density from our model, Hartree-Fock approximation and the exact result. The symbol “Na” indicates the density of Sodium.

The derivation is an elementary exercise and is given in the appendix. The resulting energy per electron ϵ_{xc} is given on the right-hand side of Fig. 2 indicated as “model” and compared with the exact result indicated as “LSD” and the Hartree-Fock result indicated as “HF” for a homogeneous electron gas.

The agreement with the correct result, which is surprisingly good for such a crude model, provides an idea of how robust the density-functional theory is with respect to approximations. While this model has been stripped to the bones, it demonstrates the way physical insight enters the construction of density functionals. Modern density functionals are far more sophisticated and exploit much more information [14], but the basic method of construction is similar.

Kinetic energy

While the expression for the kinetic energy in Eq. 9 seems familiar, there is a catch to it. In order to know the natural orbitals and the occupations we need access to the many-particle wave function or at least to its reduced density matrix.

A good approximation for the kinetic energy of the interacting electrons is the kinetic energy functional $T_s[n^{(1)}]$ of the ground state of non-interacting electrons with the same density as the true system. It is defined by

$$\begin{aligned}
 T_s[n^{(1)}] = \min_{\{f_n \in [0,1], |\psi_n\rangle\}} & \left\{ \sum_n f_n \int d^4x \psi_n^*(\vec{x}) \frac{-\hbar^2 \nabla^2}{2m} \psi_n(\vec{x}) \right. \\
 & + \int d^3r v_{eff}(\vec{r}) \left(\left[\sum_n f_n \sum_\sigma \psi_n^*(\vec{x}) \psi_n(\vec{x}) \right] - n^{(1)}(\vec{r}) \right) \\
 & \left. - \sum_{n,m} A_{m,n} \left(\langle \psi_n | \psi_m \rangle - \delta_{n,m} \right) \right\}. \quad (18)
 \end{aligned}$$

Note that $f_n \neq \bar{f}_n$ and that the so-called Kohn-Sham orbitals $\psi_n(\vec{x})$ differ⁴ from the natural

⁴To be precise, Kohn-Sham orbitals are the natural orbitals for non-interacting electrons of a given density.

orbitals $\varphi_n(\vec{x})$. Natural orbitals and Kohn-Sham wave functions are fairly similar, while the occupations f_n of Kohn-Sham orbitals differ considerably from those \bar{f}_m of the natural orbitals. The effective potential $v_{eff}(\vec{r})$ is the Lagrange multiplier for the density constraint. $\Lambda_{m,n}$ is the Lagrange multiplier for the orthonormality. Diagonalization of \mathbf{A} yields a diagonal matrix with the one-particle energies on the diagonal.

This kinetic energy $T_s[n^{(1)}]$ is a unique functional of the density, which is the first sign that we are approaching a density-functional theory. Also it is the introduction of this kinetic energy, where we made for the first time a reference to a ground state. Density functional theory as described here is inherently a ground-state theory.

Why does the true kinetic energy of the interacting system differ from that of the non-interacting energy? Consider the hole function of a non-interacting electron gas. When inserted into Eq. 12 for U_{xc} the potential energy of exchange and correlation, we obtain a contribution to the total energy that is called exchange energy. The interaction leads to a second energy contribution that is called correlation energy. Namely, when the interaction is switched on, the wave function is deformed in such a way that the Coulomb repulsion between the electrons is reduced. This makes the hole function more compact. However, there is a price to pay when the wave functions adjust to reduce the Coulomb repulsion between the electrons, namely an increase of the kinetic energy: Pushing electrons away from the neighborhood of the reference electrons requires that work be performed against the kinetic pressure of the electron gas, which raises the kinetic energy. Thus, the system has to find a compromise between minimizing the electrostatic repulsion of the electrons and increasing its kinetic energy. As a result, the correlation energy has a potential-energy contribution and a kinetic-energy contribution.

This tradeoff can be observed in Fig. 2. The correct exchange correlation energy is close to our model at low densities, while it becomes closer to the Hartree-Fock result at high densities. This is consistent with the fact that the electron gas can easily be deformed at low densities, while the deformation becomes increasingly costly at high densities due to the larger pressure of the electron gas.

The difference between T_s and the true kinetic energy is combined with the potential energy of exchange and correlation U_{xc} from Eq. 12 into the exchange correlation energy E_{xc} , i.e.

$$E_{xc} = U_{xc} + \sum_n \bar{f}_n \int d^4x \varphi_n^*(\vec{x}) \frac{-\hbar^2}{2m} \nabla^2 \varphi_n(\vec{x}) - T_s[n^{(1)}]. \quad (19)$$

Note, that the $\phi_n(\vec{x})$ and the \bar{f}_n are natural orbitals and occupations of the interacting electron gas, and that they differ from the Kohn-Sham orbitals $\psi_n(\vec{x})$ and occupations f_n .

They are however different from the natural orbitals of interacting electrons at the same density.

Total energy

The total energy obtains the form

$$\begin{aligned}
E = & \min_{|\Phi\rangle, \{|\psi_n\rangle, f_n \in [0,1]\}} \left\{ \sum_n f_n \int d^4x \psi_n^*(\vec{x}) \frac{-\hbar^2}{2m} \vec{\nabla}^2 \psi_n(\vec{x}) \right. \\
& + \int d^3r v_{eff}(\vec{r}) \left(\left[\sum_n f_n \sum_\sigma \psi_n^*(\vec{x}) \psi_n(\vec{x}) \right] - n(\vec{r}) \right) + \int d^3r v_{ext}(\vec{r}) n^{(1)}(\vec{r}) \\
& \left. + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 n^{(1)}(\vec{r}) n^{(1)}(\vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + E_{xc} - \sum_{n,m} \Lambda_{m,n} \left(\langle \psi_n | \psi_m \rangle - \delta_{n,m} \right) \right\}. \quad (20)
\end{aligned}$$

In order to evaluate the total energy with Eq. 20, we still have to start from the many-particle wave function $|\Phi\rangle$. Only the many-particle wave function allows us to evaluate the one-particle density $n^{(1)}(\vec{r})$ and the exchange correlation energy E_{xc} . Kohn-Sham orbitals $|\psi_n\rangle$ and occupations f_n are obtained by an independent minimization for each density.

If, however, we were able to express the exchange-correlation energy E_{xc} as a functional of the density alone, there would be no need for the many-particle wave function at all and the terrors of the exponential wall would be banned. We could minimize Eq. 20 with respect to the density, Kohn-Sham orbitals and their occupations.

Let us, for the time being, simply assume that $E_{xc}[n^{(1)}]$ is a functional of the electron density and explore the consequences of this assumption. Later, I will show that this assumption is actually valid.

The minimization in Eq. 20 with respect to the one-particle wave functions yields the Kohn-Sham equations

$$\left[\frac{-\hbar^2}{2m_e} \vec{\nabla}^2 + v_{eff}(\vec{r}) - \epsilon_n \right] \psi_n(\vec{x}) = 0 \quad \text{with} \quad \int d^4x \psi_m(\vec{x}) \psi_n(\vec{x}) = \delta_{m,n}. \quad (21)$$

The Kohn-Sham energies ϵ_n are the diagonal elements of the Lagrange multiplier Λ , when the latter is forced to be diagonal.

The requirement that the derivative of the total energy Eq. 20 with respect to the density vanishes, yields an expression for the effective potential

$$v_{eff}(\vec{r}) = v_{ext}(\vec{r}) + \int d^3r' \frac{e^2 n^{(1)}(\vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + \frac{\delta E_{xc}[n^{(1)}]}{\delta n^{(1)}(\vec{r})}. \quad (22)$$

Both equations, together with the density constraint

$$n^{(1)}(\vec{r}) = \sum_n f_n \sum_\sigma \psi_n^*(\vec{x}) \psi_n(\vec{x}), \quad (23)$$

form a set of coupled equations, that determine the electron density and the total energy. This set of coupled equations, Eqs. 21, 22, and 23, is what is solved in the so-called self-consistency loop. Once the set of self-consistent equations has been solved, we obtain the electron density and we can evaluate the total energy.

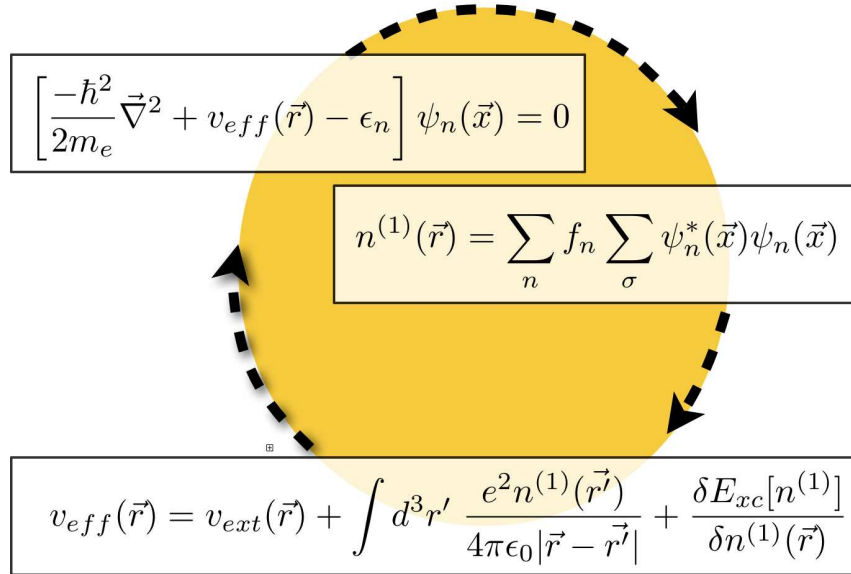


Fig. 3: *Self-consistency cycle.*

In practice, one often makes the assumption that the non-interacting electrons in the effective potential closely resemble the true interacting electrons, and extracts a wealth of other physical properties from the Kohn-Sham wave functions $|\psi_n\rangle$ and the Kohn-Sham energies ϵ_n . However, there is little theoretical backing for this approach and, if it fails, one should not blame density functional theory!

Is there a density functional?

The argument leading to the self-consistent equations, Eqs. 21, 22, and 23, relied entirely on the hope that exchange correlation functional can be expressed as a functional of the electron density. In fact, this can easily be shown, if we restrict us to ground state densities. The proof goes back to the seminal paper by Levy [15, 16].

Imagine that one could construct all fermionic many-particle wave functions. For each of these wave functions, we can determine in a unique way the electron density

$$n^{(1)}(\vec{r}) = N \sum_\sigma \int d^3x_2 \dots \int d^3x_N |\Psi(\vec{x}, \vec{x}_2, \dots, \vec{x}_N)|^2. \quad (24)$$

Having the electron densities, we sort the wave functions according to their density. For each density, I get a mug $M[n^{(1)}]$ that holds all wave functions with that density, which is written on the label of the mug.

Now we turn to each mug $M[n^{(1)}]$ in sequence and determine for each the wave function with the lowest energy. Because the external potential energy is the same for all wave functions with the same density, we need to consider only the kinetic energy operator \hat{T} and the operator \hat{W} of

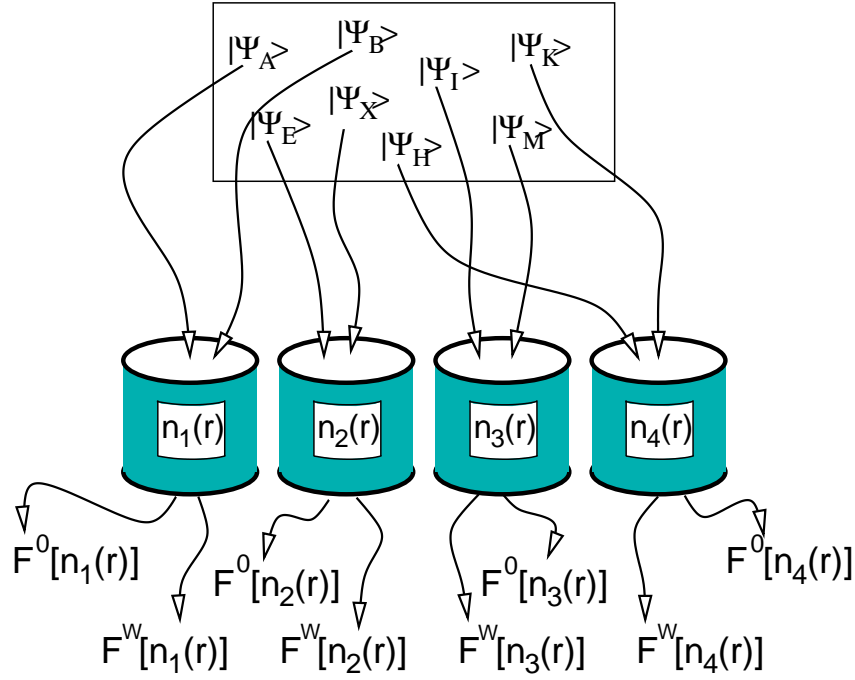


Fig. 4: Illustration for Levy's proof that there exists a density functional.

the electron-electron interaction, and we do not need to consider the external potential.

$$F^{\hat{W}}[n^{(1)}] = \min_{|\psi\rangle \in M[n^{(1)}]} \langle \psi | \hat{T} + \hat{W} | \psi \rangle \quad (25)$$

$F^{\hat{W}}[n^{(1)}]$ is the universal density functional. It is universal in the sense that it is an intrinsic property of the electron gas and absolutely independent of the external potential.

Next, we repeat the same construction as that for a universal density functional, but now we leave out the interaction \hat{W} and consider only the kinetic energy \hat{T} .

$$F^0[n^{(1)}] = \min_{|\psi\rangle \in M[n^{(1)}]} \langle \psi | \hat{T} | \psi \rangle \quad (26)$$

The resulting functional $F^0[n^{(1)}]$ is nothing but the kinetic energy of non-interacting electrons $T_s[n^{(1)}]$.

Now we can write down the total energy as functional of the density

$$E[n^{(1)}] = F^{\hat{W}}[n^{(1)}] + \int d^3r v_{ext}(\vec{r}) n^{(1)}(\vec{r}) \quad (27)$$

When we compare Eq. 27 with Eq. 20, we obtain an expression for the exchange correlation energy.

$$E_{xc}[n^{(1)}] = F^{\hat{W}}[n^{(1)}(\vec{r})] - F^0[n^{(1)}(\vec{r})] - \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 n^{(1)}(\vec{r}) n^{(1)}(\vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} \quad (28)$$

This completes the proof that the exchange correlation energy is a functional of the electron density. The latter was the assumption for the derivation of the set of self-consistent equations, Eqs. 21, 22, and 23 for the Kohn-Sham wave functions $\psi_n(\vec{x})$.

With this, I finish the description of the theoretical basis of density-functional theory. We have seen that the total energy can rigorously be expressed as a functional of the density or, in practice, as a functional of a set of one-particle wave functions, the Kohn-Sham wave functions and their occupations. Density functional theory per se is not an approximation and, in contrast to common belief, it is not a mean-field approximation. Nevertheless, we need to introduce approximations to make density functional theory work. This is because the exchange correlation energy $E_{xc}[n^{(1)}]$ is not completely known. These approximations will be discussed in the next section.

3 Jacob's ladder of density functionals

The development of density functionals is driven by mathematical analysis of the exact exchange correlation hole [14, 7], physical insight and numerical benchmark calculations on real systems. The functionals evolved in steps from one functional form to another, with several parameterizations at each level. Perdew pictured this development by Jacob's ladder leading up to heaven [17, 7]. In his analogy the different rungs of the ladder represent the different levels of density functionals leading to the unreachable, ultimately correct functional.

LDA, the big surprise

The first density functionals used in practice were based on the local-density approximation (LDA). The hole function for an electron at position \vec{r} has been approximated by the one of a homogeneous electron gas with the same density as $n^{(1)}(\vec{r})$. The exchange correlation energy for the homogeneous electron gas has been obtained by quantum Monte Carlo calculations [18] and analytic calculations [19]. The local density approximation has been generalized early to local spin-density approximation (LSD) [20].

Truly surprising was how well the theory worked for real systems. Atomic distances could be determined within a few percent of the bond length and energy differences in solids were surprisingly good.

This was unexpected, because the density in real materials is far from homogeneous. Gunnarsson and Lundquist [21] explained this finding with sumrules that are obeyed by the local density approximation: Firstly, the exchange correlation energy depends only on the spherical average of the exchange correlation hole. Of the radial hole density only the first moment contributes, while the second moment is fixed by the sum-rule that the electron density of the hole integrates to -1 . Thus we can use

$$\int d^3r \frac{e^2 h(\vec{r}_0, \vec{r})}{4\pi\epsilon_0 |\vec{r} - \vec{r}_0|} = -\frac{e^2}{4\pi\epsilon_0} \frac{\int_0^\infty dr r \langle h(\vec{r}_0, \vec{r}') \rangle_{|\vec{r}' - \vec{r}_0| = r}}{\int_0^\infty dr r^2 \langle h(\vec{r}_0, \vec{r}') \rangle_{|\vec{r}' - \vec{r}_0| = r}} \quad (29)$$

where the angular brackets imply the angular average of $\vec{r}' - \vec{r}_0$. This dependence on the hole density is rather insensitive to small changes of the hole density. Even for an atom, the *spherically averaged* exchange hole closely resembles that of the homogeneous electron gas [4].

The main deficiency of the LDA was the strong overbinding with bond energies in error by about one electron volt. On the one hand, this rendered LDA useless for most applications in chemistry. On the other hand, the problem was hardly visible in solid state physics where bonds are rarely broken, but rearranged so that errors cancelled.

GGA, entering chemistry

Being concerned about the large density variations in real materials, one tried to include the first terms of a Taylor expansion in the density gradients. These attempts failed miserably. The culprit has been a violation of the basic sum rules as pointed out by Perdew [22]. The cure was a cutoff for the gradient contributions at high gradients, which lead to the class of generalized gradient approximations (GGA) [23].

Becke [24] provides an intuitive description for the workings of GGA's, which I will sketch here in a simplified manner: Becke uses an ansatz $E_{xc} = \int d^3r A(n(\vec{r}))F(x(\vec{r}))$ for the exchange-correlation energy where $n(\vec{r})$ is the local density and $x = |\vec{\nabla}n|/n^{4/3}$ is a dimensionless reduced gradient. Do not confuse this symbol with the combined position-and-spin coordinate \vec{x} . The function A is simply the LDA expression and $F(x)$ is the so-called enhancement factor. The large-gradient limit of $F(x)$ is obtained from a simple physical argument:

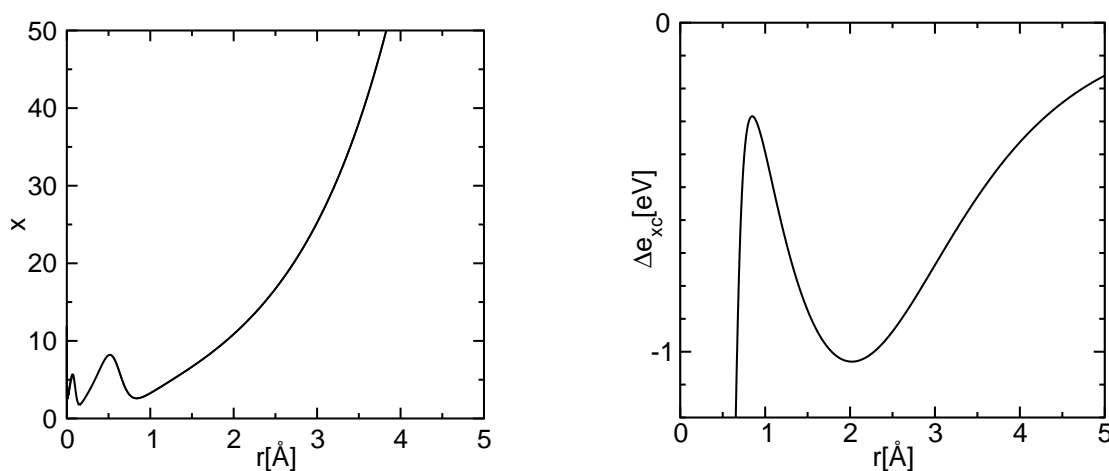


Fig. 5: Left figure: reduced density gradient $x = |\vec{\nabla}n|/n^{4/3}$ of a silicon atom as function of distance from the nucleus demonstrating that the largest reduced gradients occur in the exponential tails. Right figure: additional contribution from the gradient correction (PBE versus PW91 LDA) of the exchange correlation energy per electron. The figure demonstrates that the gradient correction stabilizes the tails of the wave function. The covalent radius of silicon is at 1.11 Å.

Somewhat surprisingly, the reduced gradient is largest not near the nucleus but in the exponentially decaying charge-density tails as shown in Fig. 5. For an electron that is far from an atom, the hole is on the atom, because a hole can only be dug where electrons are. Thus the Coulomb interaction energy of the electron with its hole is $-\frac{e^2}{4\pi\epsilon_0 r}$, where r is the distance of the reference

electron from the atom. As shown in appendix B, the enhancement factor can now be obtained by enforcing this behavior for exponentially decaying densities.

As a result, the exchange and correlation energy per electron in the tail region of the electron density falls off with the inverse distance in GGA, while it has a much faster, exponential decay in the LDA. Thus, the tail region is stabilized by GGA. This contribution acts like a negative “surface energy”.

When a bond between two atoms is broken, the surface is increased. In GGA this bond-breaking process is more favorable than in LDA, and, hence, the bond is weakened. Thus the GGA cures the overbinding error of the LDA.

These gradient corrections greatly improved the bond energies and made density functional theory useful also for chemists. The most widely distributed GGA functional is the Perdew-Burke-Ernzerhof (PBE) functional [25].

Meta GGA's

The next level of density functionals are the so-called meta GGA's [26–28] that include not only the gradient of the density, but also the second derivatives of the density. These functionals can be reformulated so that the additional parameter is the kinetic energy density instead of the second density derivatives. Perdew recommends his TPSS functional [29].

Hybrid functionals

Another generation of functionals are hybrid functionals [30, 31], which replace some of the exchange energy by the exact exchange

$$E_X^{HF} = -\frac{1}{2} \sum_{m,n} \bar{f}_m \bar{f}_n \int d^4x \int d^4x' \frac{e^2 \psi_m^*(\vec{x}) \psi_n(\vec{x}) \psi_n^*(\vec{x}') \psi_m(\vec{x}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} \quad (30)$$

where \bar{f}_n and the $\psi_n(\vec{x})$ are the Kohn-Sham occupations and wave functions, respectively.

The motivation for this approach goes back to the adiabatic connection formula [32, 33, 21]

$$E_{xc}[n(\vec{r})] = \int_0^1 d\lambda U_{xc}^{\lambda\hat{W}}[n(\vec{r})] = \int d^3r n(\vec{r}) \int_0^1 d\lambda \frac{1}{2} \int d^3r' \frac{h_\lambda(\vec{r}, \vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} \quad (31)$$

which expresses the exchange correlation energy as an integral of the potential energy of exchange and correlation over the interaction strength. Here the interaction in the Hamiltonian is scaled by a factor λ , leading to a λ -dependent universal functional $F^{\lambda\hat{W}}[n^{(1)}]$. The interaction energy can be expressed by

$$\begin{aligned} F^{\hat{W}}[n] &= F^0[n] + \int_0^1 d\lambda \frac{d}{d\lambda} F^{\lambda\hat{W}}[n] \\ &= T_s[n] + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 n(\vec{r}) n(\vec{r}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + \int_0^1 d\lambda U_{xc}^{\lambda\hat{W}}[n] \end{aligned} \quad (32)$$

which leads via Eq. 28 to Eq. 31. Using perturbation theory, the derivative of $F^{\lambda\hat{W}}[n]$ simplifies to the expectation value of the interaction $\langle\Psi(\lambda)|\hat{W}|\Psi(\lambda)\rangle$, which is the potential energy of exchange and correlation evaluated for a many-particle wave function obtained for the specified given interaction strength.

The underlying idea of the hybrid functionals is to interpolate the integrand between the end points. In the non-interacting limit, i.e. for $\lambda = 0$ the integrand $U_{xc}^{\lambda\hat{W}}$ is exactly given by the exact exchange energy of Eq. 30. For the full interaction, on the other hand, the LDA or GGA functionals are considered correctly. Thus a linear interpolation would yield

$$E_{xc} = \frac{1}{2}\left(U_{xc}^0 + U_{xc}^{\hat{W}}\right) = \frac{1}{2}\left(E_X^{HF} + U_{xc}^{\hat{W}}\right) = E_{xc}^{GGA} + \frac{1}{2}\left(E_X^{HF} - E_X^{GGA}\right). \quad (33)$$

Depending on whether the λ -dependence is a straight line or whether it is convex, the weight factor may be equal or smaller than $\frac{1}{2}$. Perdew [34] has given arguments that a factor $\frac{1}{4}$ would actually be better than a factor $\frac{1}{2}$.

Hybrid functionals perform substantially better than GGA functionals regarding binding energies, band gaps and reaction energies. However, they are flawed for the description of solids. The reason is that the exact exchange hole in a solid is very extended. These long-range tails are screened away quickly when the interaction is turned on, because they are cancelled by the correlation. Effectively, we should use a smaller mixing factor for the long range part of the exchange hole. This can be taken into account, by cutting off the long-range part of the interaction for the calculation of the Hartree-Fock exchange [35]. This approach improves the results for band gaps while reducing the computational effort [36].

The effective cancellation of the long-ranged contribution of exchange with a similar contribution from correlation, which is also considered properly already in the LDA, is one of the explanation for the superiority of the LDA over the Hartree-Fock approximation.

The most widely used hybrid functional is the B3LYP functional [37], which is, however, obtained from a parameter fit to a database of simple molecules. The functional PBE0 [38, 39] is born out of the famous PBE GGA functional and is a widely distributed parameter-free functional.

LDA+U and local hybrid functionals

Starting from a completely different context, Anisimov et. al. [40] introduced the so-called LDA+U method, which, as described below, has some similarities to the hybrid functionals above.

The main goal was to arrive at a proper description of transition metal oxides, which tend to be Mott insulators, while GGA calculations predict them often to be metals. The remedy was to add a correlation term⁵ [41] borrowed from the Hubbard model and to correct the resulting

⁵The expression given here looks unusually simple. This is due to the notation of spin orbitals, which takes care of the spin indices.

double counting of the interactions by E_{dc} .

$$E = E^{GGA} + \frac{1}{2} \sum_R \sum_{\alpha, \beta, \gamma, \delta \in \mathcal{C}_R} U_{\alpha, \beta, \gamma, \delta} \left(\rho_{\gamma, \alpha} \rho_{\delta, \beta} - \rho_{\delta, \alpha} \rho_{\gamma, \beta} \right) - E_{dc} \quad (34)$$

$$U_{\alpha, \beta, \gamma, \delta} = \int d^4x \int d^4x' \frac{e^2 \chi_\alpha^*(\vec{x}) \chi_\beta^*(\vec{x}') \chi_\gamma(\vec{x}) \chi_\delta(\vec{x}')}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} \quad (35)$$

$$\rho_{\alpha, \beta} = \langle \pi_\alpha | \psi_n \rangle f_n \langle \psi_n | \pi_\beta \rangle, \quad (36)$$

where $|\chi_\alpha\rangle$ are atomic tight-binding orbitals and $|\pi_\alpha\rangle$ are their projector functions.⁶ The additional energy is a Hartree-Fock exchange energy, that only considers the exchange for specified sets of local orbitals. The exchange term does only consider a subset of orbitals \mathcal{C}_R for each atom R and it ignores the contribution involving orbitals centered on different atoms.

Novak et al. [42] made the connection to the hybrid functionals explicit and restricted the exact exchange contribution of a hybrid functional to only a shell of orbitals. While in the LDA+U method the bare Coulomb matrix elements are reduced by a screening factor, in the hybrid functionals it is the mixing factor that effectively plays the same role. Both LDA+U and the local hybrid method have in common that they radically remove the contribution of off-site matrix elements of the interaction. Tran et al. [43] applied this method to transition metal oxides and found results that are similar to those of the full implementation of hybrid functionals.

Van der Waals interactions

One of the major difficulties for density functionals is the description of van der Waals forces, because it is due to the quantum mechanical synchronization of charge fluctuations on distinct molecules. I refer the reader to the work made in the group of Lundqvist [44–46].

4 Benchmarks, successes and failures

The development of density functionals has profited enormously from careful benchmark studies. The precondition is a data set of test cases for which reliable and accurate experimental data exist. The most famous data sets are the G1 and G2 databases [47–50] that have been set up to benchmark quantum-chemistry codes. Becke [51, 52, 31, 53, 54] set a trend by using these large sets of test cases for systematic studies of density functionals. In order to separate out the accuracy of the density functionals, it is vital to perform these calculations on extremely accurate numerical methods. Becke used basis set free calculations that were limited to small molecules, while being extremely accurate. Paier et. al. [55–57, 36] have later performed careful comparisons of two methods, Gaussian and the projector augmented-wave method, to single out the error of the electronic structure method.

⁶Projector functions obey the biorthogonality condition $\langle \chi_\alpha | \pi_\beta \rangle = \delta_{\alpha, \beta}$. Within the sub-Hilbert space of the tight-binding orbitals, i.e. for wave functions of the form $|\psi\rangle = \sum_\alpha |\chi_\alpha\rangle c_\alpha$, the projector functions decompose the wave function into tight binding orbitals, i.e. $|\psi\rangle = \sum_\alpha |\chi_\alpha\rangle \langle \pi_\alpha | \psi \rangle$. A similar projection is used extensively in the projector augmented-wave method described later.

Overall, the available density functionals predict molecular structures very well. Bond distances agree with the experiment often within one percent. Bond angles come out within a few degrees. The quality of total energies depends strongly on the level of functionals used. On the LDA level bonds are overestimated in the 1 eV range, on the GGA level these errors are reduced to a about 0.3 eV, and hybrid functionals reduce the error by another factor of 2. The ultimate goal is to reach chemical accuracy, which is about 0.05 eV. Such an accuracy allows to predict reaction rates at room temperature within a factor of 10.

Band gaps are predicted to be too small with LDA and GGA. The so-called band gap problem has been one of the major issues during the development of density functionals. Hybrid functionals clearly improve the situation. A problem is the description of materials with strong electron correlations. For LDA and GGA many insulating transition metal oxides are described as metals. This changes again for the hybrid functionals, which turns them into antiferromagnetic insulators, which is a dramatic improvement.

5 Electronic structure methods

In this second part of my lecture notes, I will address the problem of how to solve the Kohn-Sham equations and how to obtain the total energy and other observables. It is convenient to use a slightly different notation: Instead of treating the nuclei via an external potential, we combine all electrostatic interactions into a single double integral.

This brings the total energy into the form

$$E[\{\psi_n(\vec{r})\}, \{\vec{R}_R\}] = \sum_n f_n \langle \psi_n | \frac{\hat{p}^2}{2m_e} | \psi_n \rangle + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 (n(\vec{r}) + Z(\vec{r})) (n(\vec{r}') + Z(\vec{r}'))}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + E_{xc}[n], \quad (37)$$

where $Z(\vec{r}) = -\sum_R Z_R \delta(\vec{r} - \vec{R}_R)$ is the nuclear charge density expressed in electron charges. Z_R is the atomic number of a nucleus at position \vec{R}_R .

The electronic ground state is determined by minimizing the total energy functional $E[\Psi_n]$ of Eq. 37 at a fixed ionic geometry. The one-particle wave functions have to be orthogonal. This constraint is implemented with the method of Lagrange multipliers. We obtain the ground-state wave functions from the extremum condition for

$$Y[\{\psi_n\}, \Lambda] = E[\{\psi_n\}] - \sum_{n,m} [\langle \psi_n | \psi_m \rangle - \delta_{n,m}] \Lambda_{m,n} \quad (38)$$

with respect to the wavefunctions and the Lagrange multipliers $\Lambda_{m,n}$. The extremum condition for the wavefunctions has the form

$$\hat{H}|\psi_n\rangle f_n = \sum_m |\psi_m\rangle \Lambda_{m,n}, \quad (39)$$

where $\hat{H} = \frac{1}{2m_e}\hat{p}^2 + \hat{v}_{\text{eff}}$ is the effective one-particle Hamilton operator.

The corresponding effective potential depends itself on the electron density via

$$v_{\text{eff}}(\vec{r}) = \int d^3r' \frac{e^2 \left(n(\vec{r}') + Z(\vec{r}') \right)}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + \mu_{xc}(\vec{r}), \quad (40)$$

where $\mu_{xc}(\vec{r}) = \frac{\delta E_{xc}[n(\vec{r})]}{\delta n(\vec{r})}$ is the functional derivative of the exchange and correlation functional. After a unitary transformation that diagonalizes the matrix of Lagrange multipliers Λ , we obtain the Kohn-Sham equations

$$\hat{H}|\psi_n\rangle = |\psi_n\rangle\epsilon_n. \quad (41)$$

The one-particle energies ϵ_n are the eigenvalues of the matrix with the elements $\Lambda_{n,m}(f_n + f_m)/(2f_n f_m)$ [58].

The one-electron Schrödinger equations, namely the Kohn-Sham equations given in Eq. 21, still pose substantial numerical difficulties: (1) in the atomic region near the nucleus, the kinetic energy of the electrons is large, resulting in rapid oscillations of the wavefunction that require fine grids for an accurate numerical representation. On the other hand, the large kinetic energy makes the Schrödinger equation stiff, so that a change of the chemical environment has little effect on the shape of the wavefunction. Therefore, the wavefunction in the atomic region can be represented well already by a small basis set. (2) In the bonding region between the atoms the situation is opposite. The kinetic energy is small and the wavefunction is smooth. However, the wavefunction is flexible and responds strongly to the environment. This requires large and nearly complete basis sets.

Combining these different requirements is non-trivial and various strategies have been developed.

- The atomic point of view has been most appealing to quantum chemists. Basis functions are chosen that resemble atomic orbitals. This choice exploits that the wavefunction in the atomic region can be described by a few basis functions, while the chemical bond is described by the overlapping tails of these atomic orbitals. Most techniques in this class are a compromise of, on the one hand, a well adapted basis set, where the basis functions are difficult to handle, and, on the other hand, numerically convenient basis functions such as Gaussians, where the inadequacies are compensated by larger basis sets.
- Pseudopotentials regard an atom as a perturbation of the free electron gas. The most natural basis functions for the free electron gas are plane waves. Plane-wave basis sets are in principle complete and suitable for sufficiently smooth wavefunctions. The disadvantage of the comparably large basis sets required is offset by their extreme numerical simplicity. Finite plane-wave expansions are, however, absolutely inadequate to describe the strong oscillations of the wavefunctions near the nucleus. In the pseudopotential approach the Pauli repulsion by the core electrons is therefore described by an effective potential that expels the valence electrons from the core region. The resulting wavefunctions are smooth

and can be represented well by plane waves. The price to pay is that all information on the charge density and wavefunctions near the nucleus is lost.

- Augmented-wave methods compose their basis functions from atom-like wavefunctions in the atomic regions and a set of functions, called envelope functions, appropriate for the bonding in between. Space is divided accordingly into atom-centered spheres, defining the atomic regions, and an interstitial region in between. The partial solutions of the different regions are matched with value and derivative at the interface between atomic and interstitial regions.

The projector augmented-wave method is an extension of augmented wave methods and the pseudopotential approach, which combines their traditions into a unified electronic structure method.

After describing the underlying ideas of the various approaches, let us briefly review the history of augmented wave methods and the pseudopotential approach. We do not discuss the atomic-orbital based methods, because our focus is the PAW method and its ancestors.

6 Augmented wave methods

The augmented wave methods have been introduced in 1937 by Slater [59]. His method was called augmented plane-wave (APW) method. Later Korringa [60], Kohn and Rostoker [61] modified the idea, which lead to the so-called KKR method. The basic idea behind the augmented wave methods has been to consider the electronic structure as a scattered-electron problem: Consider an electron beam, represented by a plane wave, traveling through a solid. It undergoes multiple scattering at the atoms. If, for some energy, the outgoing scattered waves interfere destructively, so that the electrons can not escape, a bound state has been determined. This approach can be translated into a basis-set method with energy- and potential-dependent basis functions. In order to make the scattered wave problem tractable, a model potential had to be chosen: The so-called muffin-tin potential approximates the true potential by a potential that is spherically symmetric in the atomic regions and constant in between.

Augmented-wave methods reached adulthood in the 1970s: O. K. Andersen [62] showed that the energy dependent basis set of Slater's APW method can be mapped onto one with energy independent basis functions by linearizing the partial waves for the atomic regions with respect to their energy. In the original APW approach, one had to determine the zeros of the determinant of an energy dependent matrix, a nearly intractable numerical problem for complex systems. With the new energy independent basis functions, however, the problem is reduced to the much simpler generalized eigenvalue problem, which can be solved using efficient numerical techniques. Furthermore, the introduction of well defined basis sets paved the way for full-potential calculations [63]. In that case, the muffin-tin approximation is used solely to define the basis set $|\chi_i\rangle$, while the matrix elements $\langle\chi_i|H|\chi_j\rangle$ of the Hamiltonian are evaluated with the full potential.

In the augmented wave methods one constructs the basis set for the atomic region by solving the radial Schrödinger equation for the spherically averaged effective potential

$$\left[\frac{-\hbar^2}{2m_e} \vec{\nabla}^2 + v_{eff}(\vec{r}) - \epsilon \right] \phi_{\ell,m}(\epsilon, \vec{r}) = 0$$

as function of the energy. Note that a partial wave $\phi_{\ell,m}(\epsilon, \vec{r})$ is an angular-momentum eigenstate and can be expressed as a product of a radial function and a spherical harmonic. The energy-dependent partial wave is expanded in a Taylor expansion about some reference energy $\epsilon_{\nu,\ell}$

$$\phi_{\ell,m}(\epsilon, \vec{r}) = \phi_{\nu,\ell,m}(\vec{r}) + (\epsilon - \epsilon_{\nu,\ell}) \dot{\phi}_{\nu,\ell,m}(\vec{r}) + O((\epsilon - \epsilon_{\nu,\ell})^2),$$

where $\phi_{\nu,\ell,m}(\vec{r}) = \phi_{\ell,m}(\epsilon_{\nu,\ell}, \vec{r})$. The energy derivative of the partial wave $\dot{\phi}_{\nu}(\vec{r}) = \left. \frac{\partial \phi(\epsilon, \vec{r})}{\partial \epsilon} \right|_{\epsilon_{\nu,\ell}}$ is obtained from the energy derivative of the Schrödinger equation

$$\left[\frac{-\hbar^2}{2m_e} \vec{\nabla}^2 + v_{eff}(\vec{r}) - \epsilon_{\nu,\ell} \right] \dot{\phi}_{\nu,\ell,m}(\vec{r}) = \phi_{\nu,\ell,m}(\vec{r}).$$

Next, one starts from a regular basis set, such as plane waves, Gaussians or Hankel functions. These basis functions are called envelope functions $|\tilde{\chi}_i\rangle$. Within the atomic region they are replaced by the partial waves and their energy derivatives, such that the resulting wavefunction $\chi_i(\vec{r})$ is continuous and differentiable. The augmented envelope function has the form

$$\chi_i(\vec{r}) = \tilde{\chi}_i(\vec{r}) - \sum_R \theta_R(\vec{r}) \tilde{\chi}_i(\vec{r}) + \sum_{R,\ell,m} \theta_R(\vec{r}) \left[\phi_{\nu,R,\ell,m}(\vec{r}) a_{R,\ell,m,i} + \dot{\phi}_{\nu,R,\ell,m}(\vec{r}) b_{R,\ell,m,i} \right]. \quad (42)$$

$\theta_R(\vec{r})$ is a step function that is unity within the augmentation sphere centered at \vec{R}_R and zero elsewhere. The augmentation sphere is atom-centered and has a radius about equal to the covalent radius. This radius is called the muffin-tin radius, if the spheres of neighboring atoms touch. These basis functions describe only the valence states; the core states are localized within the augmentation sphere and are obtained directly by a radial integration of the Schrödinger equation within the augmentation sphere.

The coefficients $a_{R,\ell,m,i}$ and $b_{R,\ell,m,i}$ are obtained for each $|\tilde{\chi}_i\rangle$ as follows: The envelope function is decomposed around each atomic site into spherical harmonics multiplied by radial functions

$$\tilde{\chi}_i(\vec{r}) = \sum_{\ell,m} u_{R,\ell,m,i}(|\vec{r} - \vec{R}_R|) Y_{\ell,m}(\vec{r} - \vec{R}_R). \quad (43)$$

Analytical expansions for plane waves, Hankel functions or Gaussians exist. The radial parts of the partial waves $\phi_{\nu,R,\ell,m}$ and $\dot{\phi}_{\nu,R,\ell,m}$ are matched with value and derivative to $u_{R,\ell,m,i}(|\vec{r}|)$, which yields the expansion coefficients $a_{R,\ell,m,i}$ and $b_{R,\ell,m,i}$.

If the envelope functions are plane waves, the resulting method is called the linear augmented plane-wave (LAPW) method. If the envelope functions are Hankel functions, the method is called linear muffin-tin orbital (LMTO) method.

A good review of the LAPW method [62] has been given by Singh [64]. Let us now briefly mention the major developments of the LAPW method: Soler [65] introduced the idea of additive

augmentation: While augmented plane waves are discontinuous at the surface of the augmentation sphere if the expansion in spherical harmonics in Eq. 42 is truncated, Soler replaced the second term in Eq. 42 by an expansion of the plane wave with the same angular momentum truncation as in the third term. This dramatically improved the convergence of the angular momentum expansion. Singh [66] introduced so-called local orbitals, which are non-zero only within a muffin-tin sphere, where they are superpositions of ϕ and $\dot{\phi}$ functions from different expansion energies. Local orbitals substantially increase the energy transferability. Sjöstedt [67] relaxed the condition that the basis functions are differentiable at the sphere radius. In addition she introduced local orbitals, which are confined inside the sphere, and that also have a kink at the sphere boundary. Due to the large energy cost of kinks, they will cancel, once the total energy is minimized. The increased variational degree of freedom in the basis leads to a dramatically improved plane-wave convergence [68].

The second variant of the linear methods is the LMTO method [62]. A good introduction into the LMTO method is the book by Skriver [69]. The LMTO method uses Hankel functions as envelope functions. The atomic spheres approximation (ASA) provides a particularly simple and efficient approach to the electronic structure of very large systems. In the ASA the augmentation spheres are blown up so that the sum of their volumes is equal to the total volume. Then, the first two terms in Eq. 42 are ignored. The main deficiency of the LMTO-ASA method is the limitation to structures that can be converted into a closed packed arrangement of atomic and empty spheres. Furthermore, energy differences due to structural distortions are often qualitatively incorrect. Full potential versions of the LMTO method, that avoid these deficiencies of the ASA have been developed. The construction of tight binding orbitals as superposition of muffin-tin orbitals [70] showed the underlying principles of the empirical tight-binding method and prepared the ground for electronic structure methods that scale linearly instead of with the third power of the number of atoms. The third generation LMTO [71] allows to construct true minimal basis sets, which require only one orbital per electron pair for insulators. In addition, they can be made arbitrarily accurate in the valence band region, so that a matrix diagonalization becomes unnecessary. The first steps towards a full-potential implementation, that promises good accuracy, while maintaining the simplicity of the LMTO-ASA method, are currently under way. Through the minimal basis-set construction the LMTO method offers unrivaled tools for the analysis of the electronic structure and has been extensively used in hybrid methods combining density functional theory with model Hamiltonians for materials with strong electron correlations [72].

7 Pseudopotentials

Pseudopotentials have been introduced to (1) avoid describing the core electrons explicitly and (2) to avoid the rapid oscillations of the wavefunction near the nucleus, which normally require either complicated or large basis sets.

The pseudopotential approach can be traced back to 1940 when C. Herring invented the orthogonalized plane-wave method [73]. Later, Phillips [74] and Antončik [75] replaced the or-

thogonality condition by an effective potential, which mimics the Pauli repulsion by the core electrons and thus compensates the electrostatic attraction by the nucleus. In practice, the potential was modified, for example, by cutting off the singular potential of the nucleus at a certain value. This was done with a few parameters that have been adjusted to reproduce the measured electronic band structure of the corresponding solid.

Hamann, Schlüter and Chiang [76] showed in 1979 how pseudopotentials can be constructed in such a way that their scattering properties are identical to that of an atom to first order in energy. These first-principles pseudopotentials relieved the calculations from the restrictions of empirical parameters. Highly accurate calculations have become possible especially for semiconductors and simple metals. An alternative approach towards first-principles pseudopotentials by Zunger and Cohen [77] even preceded the one mentioned above.

The idea behind the pseudopotential construction

In order to construct a first-principles pseudopotential, one starts out with an all-electron density-functional calculation for a spherical atom. Such calculations can be performed efficiently on radial grids. They yield the atomic potential and wavefunctions $\phi_{\ell,m}(\vec{r})$. Due to the spherical symmetry, the radial parts of the wavefunctions for different magnetic quantum numbers m are identical.

For the valence wavefunctions one constructs pseudo wavefunctions $|\tilde{\phi}_{\ell,m}\rangle$. There are numerous ways [78–81] to construct those pseudo wavefunctions: Pseudo wave functions are identical to the true wave functions outside the augmentation region, which is called core region in the context of the pseudopotential approach. Inside the augmentation region the pseudo wavefunction should be nodeless and have the same norm as the true wavefunctions, that is $\langle \tilde{\phi}_{\ell,m} | \tilde{\phi}_{\ell,m} \rangle = \langle \phi_{\ell,m} | \phi_{\ell,m} \rangle$ (compare Figure 6).

From the pseudo wavefunction, a potential $u_{\ell}(\vec{r})$ can be reconstructed by inverting the respective Schrödinger equation, i.e.

$$\left[-\frac{\hbar^2}{2m_e} \vec{\nabla}^2 + u_{\ell}(\vec{r}) - \epsilon_{\ell,m} \right] \tilde{\phi}_{\ell,m}(\vec{r}) = 0 \quad \Rightarrow \quad u_{\ell}(\vec{r}) = \epsilon + \frac{1}{\tilde{\phi}_{\ell,m}(\vec{r})} \cdot \frac{\hbar^2}{2m_e} \vec{\nabla}^2 \tilde{\phi}_{\ell,m}(\vec{r}) .$$

This potential $u_{\ell}(\vec{r})$ (compare Figure 6), which is also spherically symmetric, differs from one main angular momentum ℓ to the other. Note, that this *inversion of the Schrödinger equation* works only if the wave functions are nodeless.

Next we define an effective pseudo Hamiltonian

$$\hat{H}_{\ell} = -\frac{\hbar^2}{2m_e} \vec{\nabla}^2 + v_{\ell}^{ps}(\vec{r}) + \int d^3r' \frac{e^2 (\tilde{n}(\vec{r}') + \tilde{Z}(\vec{r}'))}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + \mu_{xc}([n(\vec{r})], \vec{r}) , \quad (44)$$

where $\mu_{xc}(\vec{r}) = \delta E_{xc}[n]/\delta n(\vec{r})$ is the functional derivative of the exchange and correlation energy with respect to the electron density. Then, we determine the pseudopotentials v_{ℓ}^{ps} such that the pseudo Hamiltonian produces the pseudo wavefunctions, that is

$$v_{\ell}^{ps}(\vec{r}) = u_{\ell}(\vec{r}) - \int d^3r' \frac{e^2 (\tilde{n}(\vec{r}') + \tilde{Z}(\vec{r}'))}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} - \mu_{xc}([\tilde{n}(\vec{r})], \vec{r}) . \quad (45)$$

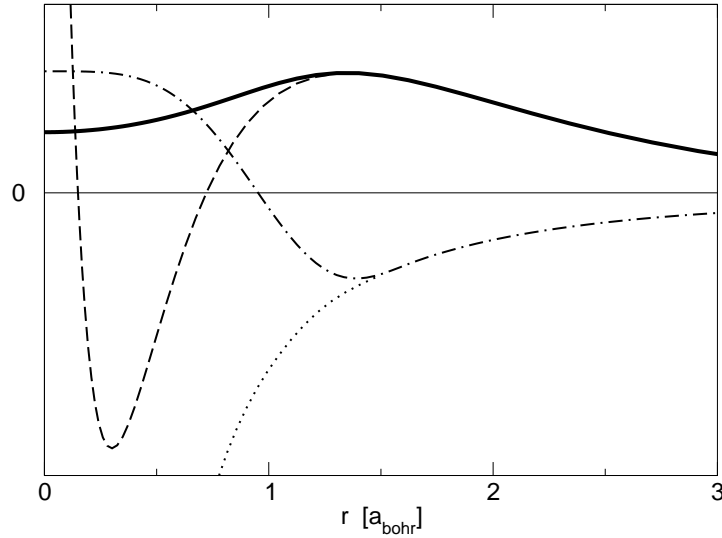


Fig. 6: Illustration of the pseudopotential concept for the example of the 3s wavefunction of Si. The solid line shows the radial part of the pseudo wavefunction $\tilde{\phi}_{\ell,m}$. The dashed line corresponds to the all-electron wavefunction $\phi_{\ell,m}$, which exhibits strong oscillations at small radii. The angular momentum dependent pseudopotential u_{ℓ} (dash-dotted line) deviates from the all-electron potential v_{eff} (dotted line) inside the augmentation region. The data are generated by the fhi98PP code [82].

This process is called “unscreening”.

$\tilde{Z}(\vec{r})$ mimics the charge density of the nucleus and the core electrons. It is usually an atom-centered, spherical Gaussian that is normalized to the charge of nucleus and core of that atom. In the pseudopotential approach, $\tilde{Z}_R(\vec{r})$ does not change with the potential. The pseudo density $\tilde{n}(\vec{r}) = \sum_n f_n \tilde{\psi}_n^*(\vec{r}) \tilde{\psi}_n(\vec{r})$ is constructed from the pseudo wavefunctions.

In this way, we obtain a different potential for each angular momentum channel. In order to apply these potentials to a given wavefunction, the wavefunction must first be decomposed into angular momenta. Then each component is applied to the pseudopotential v_{ℓ}^{ps} for the corresponding angular momentum.

The pseudopotential defined in this way can be expressed in a semi-local form

$$v^{ps}(\vec{r}, \vec{r}') = \bar{v}(\vec{r}) \delta(\vec{r} - \vec{r}') + \sum_{\ell,m} \left[Y_{\ell,m}(\vec{r}) [v_{\ell}^{ps}(\vec{r}) - \bar{v}(\vec{r})] \frac{\delta(|\vec{r}| - |\vec{r}'|)}{|\vec{r}|^2} Y_{\ell,m}^*(\vec{r}') \right]. \quad (46)$$

The local potential $\bar{v}(\vec{r})$ only acts on those angular momentum components that are not already considered explicitly in the non-local, angular-momentum dependent pseudopotentials v_{ℓ}^{ps} . Typically it is chosen to cancel the most expensive nonlocal terms, the one corresponding to the highest physically relevant angular momentum.

The pseudopotential $v^{ps}(\vec{r}, \vec{r}')$ is non-local as it depends on two position arguments, \vec{r} and \vec{r}' . The expectation values are evaluated as a double integral

$$\langle \tilde{\psi} | \hat{v}_{ps} | \tilde{\psi} \rangle = \int d^3r \int d^3r' \tilde{\psi}^*(\vec{r}) v^{ps}(\vec{r}, \vec{r}') \tilde{\psi}(\vec{r}') \quad (47)$$

The semi-local form of the pseudopotential given in Eq. 46 is computationally expensive. Therefore, in practice one uses a separable form of the pseudopotential [83–85]

$$\hat{v}^{ps} \approx \sum_{i,j} \hat{v}^{ps} |\tilde{\phi}_i\rangle \left[\langle \tilde{\phi}_j | \hat{v}^{ps} | \tilde{\phi}_i \rangle \right]_{i,j}^{-1} \langle \tilde{\phi}_j | \hat{v}^{ps} . \quad (48)$$

Thus, the projection onto spherical harmonics used in the semi-local form of Eq. 46 is replaced by a projection onto angular momentum dependent functions $\hat{v}^{ps} |\tilde{\phi}_i\rangle$.

The indices i and j are composite indices containing the atomic-site index R , the angular momentum quantum numbers ℓ, m and an additional index α . The index α distinguishes partial waves with otherwise identical indices R, ℓ, m when more than one partial wave per site and angular momentum is allowed. The partial waves may be constructed as eigenstates of the pseudopotential \hat{v}_ℓ^{ps} for a set of energies.

One can show that the identity of Eq. 48 holds by applying a wavefunction $|\tilde{\psi}\rangle = \sum_i |\tilde{\phi}_i\rangle c_i$ to both sides. If the set of pseudo partial waves $|\tilde{\phi}_i\rangle$ in Eq. 48 is complete, the identity is exact. The advantage of the separable form is that $\langle \tilde{\phi} | \hat{v}^{ps}$ is treated as one function, so that expectation values are reduced to combinations of simple scalar products $\langle \tilde{\phi}_i | \hat{v}^{ps} | \tilde{\psi} \rangle$.

The total energy of the pseudopotential method can be written in the form

$$E = \sum_n f_n \langle \tilde{\psi}_n | \frac{\hat{p}^2}{2m_e} | \tilde{\psi}_n \rangle + E_{self} + \sum_n f_n \langle \tilde{\psi}_n | \hat{v}_{ps} | \tilde{\psi}_n \rangle + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 (\tilde{n}(\vec{r}) + \tilde{Z}(\vec{r})) (\tilde{n}(\vec{r}') + \tilde{Z}(\vec{r}'))}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + E_{xc}[\tilde{n}(\vec{r})] . \quad (49)$$

The constant E_{self} is adjusted such that the total energy of the atom is the same for an all-electron calculation and the pseudopotential calculation.

For the atom, from which it has been constructed, this construction guarantees that the pseudopotential method produces the correct one-particle energies for the valence states and that the wave functions have the desired shape.

While pseudopotentials have proven to be accurate for a large variety of systems, there is no strict guarantee that they produce the same results as an all-electron calculation, if they are used in a molecule or solid. The error sources can be divided into two classes:

- Energy transferability problems: Even for the potential of the reference atom, the scattering properties are accurate only in given energy window.
- Charge transferability problems: In a molecule or crystal, the potential differs from that of the isolated atom. The pseudopotential, however, is strictly valid only for the isolated atom.

The plane-wave basis set for the pseudo wavefunctions is defined by the shortest wave length $\lambda = 2\pi/|\vec{G}|$, where \vec{G} is the wave vector, via the so-called plane-wave cutoff $E_{PW} = \frac{\hbar^2 G_{max}^2}{2m_e}$ with $G_{max} = \max\{|\vec{G}|\}$. It is often specified in Rydberg (1 Ry = $\frac{1}{2}$ H \approx 13.6 eV). The plane-wave cutoff is the highest kinetic energy of all basis functions. The basis-set convergence can systematically be controlled by increasing the plane-wave cutoff.

The charge transferability is substantially improved by including a nonlinear core correction [86] into the exchange-correlation term of Eq. 49. Hamann [87] showed how to construct pseudopotentials also from unbound wavefunctions. Vanderbilt [88, 89] generalized the pseudopotential method to non-normconserving pseudopotentials, so-called ultra-soft pseudopotentials, which dramatically improve the basis-set convergence. The formulation of ultra-soft pseudopotentials has already many similarities with the projector augmented-wave method. Truncated separable pseudopotentials suffer sometimes from so-called ghost states. These are unphysical core-like states, which render the pseudopotential useless. These problems have been discussed by Gonze [90]. Quantities such as hyperfine parameters that depend on the full wavefunctions near the nucleus, can be extracted approximately [91]. Good reviews of pseudopotential methodology have been written by Payne *et al.* [92] and Singh [64].

In 1985 R. Car and M. Parrinello published the ab-initio molecular dynamics method [93]. Simulations of the atomic motion have become possible on the basis of state-of-the-art electronic structure methods. Besides making dynamical phenomena and finite temperature effects accessible to electronic structure calculations, the ab-initio molecular dynamics method also introduced a radically new way of thinking into electronic structure methods. Diagonalization of a Hamilton matrix has been replaced by classical equations of motion for the wavefunction coefficients. If one applies friction, the system is quenched to the ground state. Without friction truly dynamical simulations of the atomic structure are performed. By using thermostats [94–97], simulations at constant temperature can be performed. The Car-Parrinello method treats electronic wavefunctions and atomic positions on an equal footing.

8 Projector augmented-wave method

The Car-Parrinello method had been implemented first for the pseudopotential approach. There seemed to be insurmountable barriers against combining the new technique with augmented wave methods. The main problem was related to the potential dependent basis set used in augmented wave methods: the Car-Parrinello method requires a well defined and unique total energy functional of atomic positions and basis set coefficients. Furthermore the analytic evaluation of the first partial derivatives of the total energy with respect to wave functions, $\frac{\partial E}{\partial \langle \psi_n |} = \hat{H} | \psi_n \rangle f_n$, and atomic positions, the forces $\vec{F}_j = -\vec{\nabla}_j E$, must be possible. Therefore, it was one of the main goals of the PAW method to introduce energy and potential independent basis sets, which were as accurate as the previously used augmented basis sets. Other requirements have been: (1) The method should at least match the efficiency of the pseudopotential approach for Car-Parrinello simulations. (2) It should become an exact theory when converged and (3) its convergence should be easily controlled. We believe that these criteria have been met, which explains why the use of the PAW method has become widespread today.

Transformation theory

At the root of the PAW method lies a transformation that maps the true wavefunctions with their complete nodal structure onto auxiliary wavefunctions that are numerically convenient. We aim for smooth auxiliary wavefunctions, which have a rapidly convergent plane-wave expansion. With such a transformation we can expand the auxiliary wave functions into a convenient basis set such as plane waves, and evaluate all physical properties after reconstructing the related physical (true) wavefunctions.

Let us denote the physical one-particle wavefunctions as $|\psi_n\rangle$ and the auxiliary wavefunctions as $|\tilde{\psi}_n\rangle$. Note that the tilde refers to the representation of smooth auxiliary wavefunctions and n is the label for a one-particle state and contains a band index, a k -point and a spin index. The transformation from the auxiliary to the physical wave functions is denoted by $\hat{\mathcal{T}}$, i.e.

$$|\psi_n\rangle = \hat{\mathcal{T}}|\tilde{\psi}_n\rangle. \quad (50)$$

Now we express the constrained density functional F of Eq. 38 in terms of our auxiliary wavefunctions

$$F[\{\hat{\mathcal{T}}|\tilde{\psi}_n\rangle\}, \{A_{m,n}\}] = E[\{\hat{\mathcal{T}}|\tilde{\psi}_n\rangle\}] - \sum_{n,m} [\langle \tilde{\psi}_n | \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}} | \tilde{\psi}_m \rangle - \delta_{n,m}] A_{m,n}. \quad (51)$$

The variational principle with respect to the auxiliary wavefunctions yields

$$\hat{\mathcal{T}}^\dagger \hat{H} \hat{\mathcal{T}} |\tilde{\psi}_n\rangle = \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}} |\tilde{\psi}_n\rangle \epsilon_n. \quad (52)$$

Again, we obtain a Schrödinger-like equation (see derivation of Eq. 41), but now the Hamilton operator has a different form, $\hat{\hat{H}} = \hat{\mathcal{T}}^\dagger \hat{H} \hat{\mathcal{T}}$, an overlap operator $\hat{\hat{O}} = \hat{\mathcal{T}}^\dagger \hat{\mathcal{T}}$ occurs, and the resulting auxiliary wavefunctions are smooth.

When we evaluate physical quantities, we need to evaluate expectation values of an operator \hat{A} , which can be expressed in terms of either the true or the auxiliary wavefunctions, i.e.

$$\langle \hat{A} \rangle = \sum_n f_n \langle \psi_n | \hat{A} | \psi_n \rangle = \sum_n f_n \langle \tilde{\psi}_n | \hat{\mathcal{T}}^\dagger \hat{A} \hat{\mathcal{T}} | \tilde{\psi}_n \rangle. \quad (53)$$

In the representation of auxiliary wavefunctions we need to use transformed operators $\hat{\hat{A}} = \hat{\mathcal{T}}^\dagger \hat{A} \hat{\mathcal{T}}$. As it is, this equation only holds for the valence electrons. The core electrons are treated differently, as will be shown below.

The transformation takes us conceptionally from the world of pseudopotentials to that of augmented wave methods, which deal with the full wavefunctions. We will see that our auxiliary wavefunctions, which are simply the plane-wave parts of the full wavefunctions, translate into the wavefunctions of the pseudopotential approach. In the PAW method the auxiliary wavefunctions are used to construct the true wavefunctions and the total energy functional is evaluated from the latter. Thus it provides the missing link between augmented wave methods and the pseudopotential method, which can be derived as a well-defined approximation of the PAW method.

In the original paper [58], the auxiliary wavefunctions were termed pseudo wavefunctions and the true wavefunctions were termed all-electron wavefunctions, in order to make the connection more evident. We avoid this notation here, because it resulted in confusion in cases where the correspondence is not clear-cut.

Transformation operator

So far we have described how we can determine the auxiliary wave functions of the ground state and how to obtain physical information from them. What is missing is a definition of the transformation operator $\hat{\mathcal{T}}$.

The operator $\hat{\mathcal{T}}$ has to modify the smooth auxiliary wave function in each atomic region, so that the resulting wavefunction has the correct nodal structure. Therefore, it makes sense to write the transformation as identity plus a sum of atomic contributions $\hat{\mathcal{S}}_R$

$$\hat{\mathcal{T}} = \hat{1} + \sum_R \hat{\mathcal{S}}_R. \quad (54)$$

For every atom, $\hat{\mathcal{S}}_R$ adds the difference between the true and the auxiliary wavefunction.

The local terms $\hat{\mathcal{S}}_R$ are defined in terms of solutions $|\phi_i\rangle$ of the Schrödinger equation for the isolated atoms. This set of partial waves $|\phi_i\rangle$ will serve as a basis set so that, near the nucleus, all relevant valence wavefunctions can be expressed as superposition of the partial waves with yet unknown coefficients as

$$\psi(\vec{r}) = \sum_{i \in R} \phi_i(\vec{r}) c_i \quad \text{for } |\vec{r} - \vec{R}_R| < r_{c,R}. \quad (55)$$

With $i \in R$ we indicate those partial waves that belong to site R .

Since the core wavefunctions do not spread out into the neighboring atoms, we will treat them differently. Currently we use the frozen-core approximation, which imports the density and the energy of the core electrons from the corresponding isolated atoms. The transformation $\hat{\mathcal{T}}$ shall produce only wavefunctions orthogonal to the core electrons, while the core electrons are treated separately. Therefore, the set of atomic partial waves $|\phi_i\rangle$ includes only valence states that are orthogonal to the core wavefunctions of the atom.

For each of the partial waves we choose an auxiliary partial wave $|\tilde{\phi}_i\rangle$. The identity

$$\begin{aligned} |\phi_i\rangle &= (\hat{1} + \hat{\mathcal{S}}_R) |\tilde{\phi}_i\rangle \quad \text{for } i \in R \\ \hat{\mathcal{S}}_R |\tilde{\phi}_i\rangle &= |\phi_i\rangle - |\tilde{\phi}_i\rangle \end{aligned} \quad (56)$$

defines the local contribution $\hat{\mathcal{S}}_R$ to the transformation operator. Since $\hat{1} + \hat{\mathcal{S}}_R$ should change the wavefunction only locally, we require that the partial waves $|\phi_i\rangle$ and their auxiliary counterparts $|\tilde{\phi}_i\rangle$ are pairwise identical beyond a certain radius $r_{c,R}$.

$$\phi_i(\vec{r}) = \tilde{\phi}_i(\vec{r}) \quad \text{for } i \in R \quad \text{and} \quad |\vec{r} - \vec{R}_R| > r_{c,R} \quad (57)$$

Note that the partial waves are not necessarily bound states and are therefore not normalizable unless we truncate them beyond a certain radius $r_{c,R}$. The PAW method is formulated such that

the final results do not depend on the location where the partial waves are truncated, as long as this is not done too close to the nucleus and identical for auxiliary and all-electron partial waves.

In order to be able to apply the transformation operator to an arbitrary auxiliary wavefunction, we need to be able to expand the auxiliary wavefunction locally into the auxiliary partial waves

$$\tilde{\psi}(\vec{r}) = \sum_{i \in R} \tilde{\phi}_i(\vec{r}) c_i = \sum_{i \in R} \tilde{\phi}_i(\vec{r}) \langle \tilde{p}_i | \tilde{\psi} \rangle \quad \text{for } |\vec{r} - \vec{R}_R| < r_{c,R}, \quad (58)$$

which defines the projector functions $|\tilde{p}_i\rangle$. The projector functions probe the local character of the auxiliary wave function in the atomic region, and examples are shown in Figure 7. From Eq. 58 we can derive $\sum_{i \in R} |\tilde{\phi}_i\rangle \langle \tilde{p}_i| = 1$, which is valid within $r_{c,R}$. It can be shown by insertion, that the identity Eq. 58 holds for any auxiliary wavefunction $|\tilde{\psi}\rangle$ that can be expanded locally into auxiliary partial waves $|\tilde{\phi}_i\rangle$, if

$$\langle \tilde{p}_i | \tilde{\phi}_j \rangle = \delta_{i,j} \quad \text{for } i, j \in R. \quad (59)$$

Note that neither the projector functions nor the partial waves need to be mutually orthogonal. The projector functions are fully determined with the above conditions and a closure relation that is related to the unscreening of the pseudopotentials (see Eq. 90 in [58]).

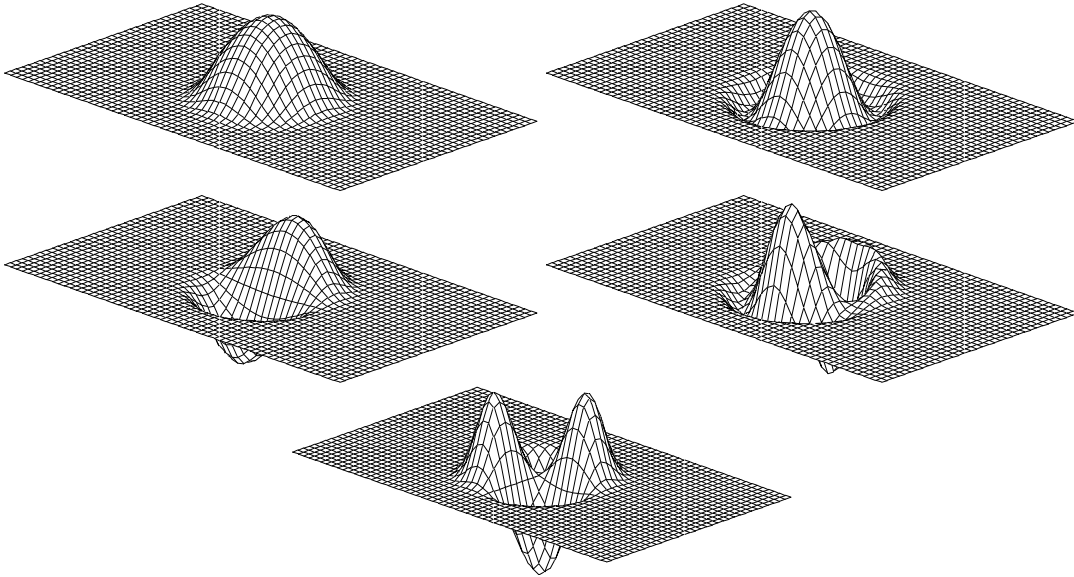


Fig. 7: Projector functions of the chlorine atom. Top: two s-type projector functions, middle: p-type, bottom: d-type.

By combining Eq. 56 and Eq. 58, we can apply \hat{S}_R to any auxiliary wavefunction.

$$\hat{S}_R |\tilde{\psi}\rangle = \sum_{i \in R} \hat{S}_R |\tilde{\phi}_i\rangle \langle \tilde{p}_i | \tilde{\psi} \rangle = \sum_{i \in R} (|\phi_i\rangle - |\tilde{\phi}_i\rangle) \langle \tilde{p}_i | \tilde{\psi} \rangle. \quad (60)$$

Hence, the transformation operator is

$$\hat{\mathcal{T}} = \hat{1} + \sum_i (|\phi_i\rangle - |\tilde{\phi}_i\rangle) \langle \tilde{p}_i |, \quad (61)$$

where the sum runs over all partial waves of all atoms. The true wave function can be expressed as

$$|\psi\rangle = |\tilde{\psi}\rangle + \sum_i (|\phi_i\rangle - |\tilde{\phi}_i\rangle) \langle \tilde{p}_i | \tilde{\psi} \rangle = |\tilde{\psi}\rangle + \sum_R (|\psi_R^1\rangle - |\tilde{\psi}_R^1\rangle) \quad (62)$$

with

$$|\psi_R^1\rangle = \sum_{i \in R} |\phi_i\rangle \langle \tilde{p}_i | \tilde{\psi} \rangle \quad (63)$$

$$|\tilde{\psi}_R^1\rangle = \sum_{i \in R} |\tilde{\phi}_i\rangle \langle \tilde{p}_i | \tilde{\psi} \rangle. \quad (64)$$

In Fig. 8 the decomposition of Eq. 62 is shown for the example of the bonding p - σ state of the Cl_2 molecule.

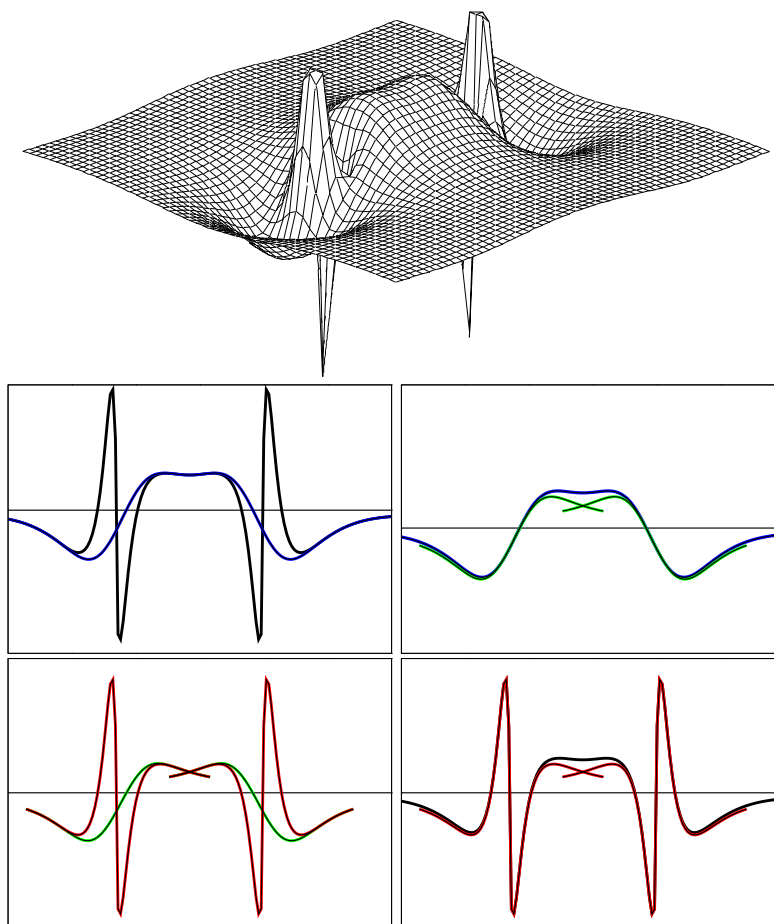


Fig. 8: Bonding p - σ orbital of the Cl_2 molecule and its decomposition into auxiliary wavefunction and the two one-center expansions. Top-left: True and auxiliary wave function; top-right: auxiliary wavefunction and its partial wave expansion; bottom-left: the two partial wave expansions; bottom-right: true wavefunction and its partial wave expansion.

To understand the expression Eq. 62 for the true wave function, let us concentrate on different regions in space. (1) Far from the atoms, the partial waves are, according to Eq. 57, pairwise

identical so that the auxiliary wavefunction is identical to the true wavefunction, that is $\psi(\vec{r}) = \tilde{\psi}(\vec{r})$. (2) Close to an atom R , however, the auxiliary wavefunction is, according to Eq. 58, identical to its one-center expansion, that is $\tilde{\psi}(\vec{r}) = \tilde{\psi}_R^1(\vec{r})$. Hence the true wavefunction $\psi(\vec{r})$ is identical to $\psi_R^1(\vec{r})$, which is built up from partial waves that contain the proper nodal structure. In practice, the partial wave expansions are truncated. Therefore, the identity of Eq. 58 does not hold strictly. As a result, the plane waves also contribute to the true wavefunction inside the atomic region. This has the advantage that the missing terms in a truncated partial wave expansion are partly accounted for by plane waves. This explains the rapid convergence of the partial wave expansions. This idea is related to the additive augmentation of the LAPW method of Soler [65].

Frequently, the question comes up, whether the transformation Eq. 61 of the auxiliary wavefunctions indeed provides the true wavefunction. The transformation should be considered merely as a change of representation analogous to a coordinate transform. If the total energy functional is transformed consistently, its minimum will yield auxiliary wavefunctions that produce the correct wave functions $|\psi\rangle$.

Expectation values

Expectation values can be obtained either from the reconstructed true wavefunctions or directly from the auxiliary wave functions

$$\begin{aligned} \langle \hat{A} \rangle &= \sum_n f_n \langle \psi_n | \hat{A} | \psi_n \rangle + \sum_{n=1}^{N_c} \langle \phi_n^c | \hat{A} | \phi_n^c \rangle \\ &= \sum_n f_n \langle \tilde{\psi}_n | \hat{T}^\dagger \hat{A} \hat{T} | \tilde{\psi}_n \rangle + \sum_{n=1}^{N_c} \langle \phi_n^c | \hat{A} | \phi_n^c \rangle, \end{aligned} \quad (65)$$

where f_n are the occupations of the valence states and N_c is the number of core states. The first sum runs over the valence states, and second over the core states $|\phi_n^c\rangle$.

Now we can decompose the matrix element for a wavefunction ψ into its individual contributions according to Eq. 62.

$$\begin{aligned} \langle \psi | \hat{A} | \psi \rangle &= \langle \tilde{\psi} + \sum_R (\psi_R^1 - \tilde{\psi}_R^1) | \hat{A} | \tilde{\psi} + \sum_{R'} (\psi_{R'}^1 - \tilde{\psi}_{R'}^1) \rangle \\ &= \underbrace{\langle \tilde{\psi} | \hat{A} | \tilde{\psi} \rangle + \sum_R \left(\langle \psi_R^1 | \hat{A} | \psi_R^1 \rangle - \langle \tilde{\psi}_R^1 | \hat{A} | \tilde{\psi}_R^1 \rangle \right)}_{\text{part 1}} \\ &\quad + \underbrace{\sum_R \left(\langle \psi_R^1 - \tilde{\psi}_R^1 | \hat{A} | \tilde{\psi} - \tilde{\psi}_R^1 \rangle + \langle \tilde{\psi} - \tilde{\psi}_R^1 | \hat{A} | \psi_R^1 - \tilde{\psi}_R^1 \rangle \right)}_{\text{part 2}} \\ &\quad + \underbrace{\sum_{R \neq R'} \langle \psi_R^1 - \tilde{\psi}_R^1 | \hat{A} | \psi_{R'}^1 - \tilde{\psi}_{R'}^1 \rangle}_{\text{part 3}} \end{aligned} \quad (66)$$

Only the first part of Eq. 66 is evaluated explicitly, while the second and third parts of Eq. 66 are neglected, because they vanish for sufficiently local operators as long as the partial wave expansion is converged: The function $\psi_R^1 - \tilde{\psi}_R^1$ vanishes per construction beyond its augmentation region, because the partial waves are pairwise identical beyond that region. The function $\tilde{\psi} - \tilde{\psi}_R^1$ vanishes inside its augmentation region if the partial wave expansion is sufficiently converged. In no region of space are both functions $\psi_R^1 - \tilde{\psi}_R^1$ and $\tilde{\psi} - \tilde{\psi}_R^1$ simultaneously nonzero. Similarly the functions $\psi_R^1 - \tilde{\psi}_R^1$ from different sites are never non-zero in the same region in space. Hence, the second and third parts of Eq. 66 vanish for operators such as the kinetic energy $\frac{-\hbar^2}{2m_e} \vec{\nabla}^2$ and the real space projection operator $|r\rangle\langle r|$, which produces the electron density. For truly nonlocal operators the parts 2 and 3 of Eq. 66 would have to be considered explicitly. The expression, Eq. 65, for the expectation value can therefore be written with the help of Eq. 66 as

$$\begin{aligned}
\langle \hat{A} \rangle &= \sum_n f_n \left(\langle \tilde{\psi}_n | \hat{A} | \tilde{\psi}_n \rangle + \langle \psi_n^1 | \hat{A} | \psi_n^1 \rangle - \langle \tilde{\psi}_n^1 | \hat{A} | \tilde{\psi}_n^1 \rangle \right) + \sum_{n=1}^{N_c} \langle \phi_n^c | \hat{A} | \phi_n^c \rangle \\
&= \sum_n f_n \langle \tilde{\psi}_n | \hat{A} | \tilde{\psi}_n \rangle + \sum_{n=1}^{N_c} \langle \tilde{\phi}_n^c | \hat{A} | \tilde{\phi}_n^c \rangle \\
&+ \sum_R \left(\sum_{i,j \in R} D_{i,j} \langle \phi_j | \hat{A} | \phi_i \rangle + \sum_{n \in R}^{N_{c,R}} \langle \phi_n^c | \hat{A} | \phi_n^c \rangle \right) \\
&- \sum_R \left(\sum_{i,j \in R} D_{i,j} \langle \tilde{\phi}_j | \hat{A} | \tilde{\phi}_i \rangle + \sum_{n \in R}^{N_{c,R}} \langle \tilde{\phi}_n^c | \hat{A} | \tilde{\phi}_n^c \rangle \right), \tag{67}
\end{aligned}$$

where \mathbf{D} is the one-center density matrix defined as

$$D_{i,j} = \sum_n f_n \langle \tilde{\psi}_n | \tilde{p}_j \rangle \langle \tilde{p}_i | \tilde{\psi}_n \rangle = \sum_n \langle \tilde{p}_i | \tilde{\psi}_n \rangle f_n \langle \tilde{\psi}_n | \tilde{p}_j \rangle. \tag{68}$$

The auxiliary core states, $|\tilde{\phi}_n^c\rangle$ allow us to incorporate the tails of the core wavefunction into the plane-wave part, and therefore assure that the integrations of partial wave contributions cancel exactly beyond r_c . They are identical to the true core states in the tails, but are a smooth continuation inside the atomic sphere. It is not required that the auxiliary wave functions are normalized.

Following this scheme, the electron density is given by

$$\begin{aligned}
n(\vec{r}) &= \tilde{n}(\vec{r}) + \sum_R \left(n_R^1(\vec{r}) - \tilde{n}_R^1(\vec{r}) \right) \tag{69} \\
\tilde{n}(\vec{r}) &= \sum_n f_n \tilde{\psi}_n^*(\vec{r}) \tilde{\psi}_n(\vec{r}) + \tilde{n}_c(\vec{r}) \\
n_R^1(\vec{r}) &= \sum_{i,j \in R} D_{i,j} \phi_j^*(\vec{r}) \phi_i(\vec{r}) + n_{c,R}(\vec{r}) \\
\tilde{n}_R^1(\vec{r}) &= \sum_{i,j \in R} D_{i,j} \tilde{\phi}_j^*(\vec{r}) \tilde{\phi}_i(\vec{r}) + \tilde{n}_{c,R}(\vec{r}), \tag{70}
\end{aligned}$$

where $n_{c,R}$ is the core density of the corresponding atom and $\tilde{n}_{c,R}$ is the auxiliary core density, which is identical to $n_{c,R}$ outside the atomic region, but smooth inside.

Before we continue, let us discuss a special point: The matrix elements of a general operator with the auxiliary wavefunctions may be slowly converging with the plane-wave expansion, because the operator \hat{A} may not be well behaved. An example of such an operator is the singular electrostatic potential of a nucleus. This problem can be alleviated by adding an “intelligent zero”: If an operator \hat{B} is purely localized within an atomic region, we can use the identity between the auxiliary wavefunction and its own partial wave expansion

$$0 = \langle \tilde{\psi}_n | \hat{B} | \tilde{\psi}_n \rangle - \langle \tilde{\psi}_n^1 | \hat{B} | \tilde{\psi}_n^1 \rangle. \quad (71)$$

Now we choose an operator \hat{B} so that it cancels the problematic behavior of the operator \hat{A} , but is localized in a single atomic region. By adding \hat{B} to the plane-wave part and the matrix elements with its one-center expansions, the plane-wave convergence can be improved without affecting the converged result. A term of this type, namely \hat{v} will be introduced in the next section to cancel the Coulomb singularity of the potential at the nucleus.

Total energy

As with wavefunctions and expectation values, the total energy can be divided into three parts.

$$E[\{\tilde{\psi}_n\}, \{R_R\}] = \tilde{E} + \sum_R (E_R^1 - \tilde{E}_R^1) \quad (72)$$

The plane-wave part \tilde{E} involves only smooth functions and is evaluated on equi-spaced grids in real and reciprocal space. This part is computationally most demanding, and is similar to the expressions in the pseudopotential approach.

$$\begin{aligned} \tilde{E} = & \sum_n \langle \tilde{\psi}_n | \frac{\hat{p}^2}{2m_e} | \tilde{\psi}_n \rangle + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 (\tilde{n}(\vec{r}) + \tilde{Z}(\vec{r})) (\tilde{n}(\vec{r}') + \tilde{Z}(\vec{r}'))}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} \\ & + \int d^3r \bar{v}(\vec{r}) \tilde{n}(\vec{r}) + E_{xc}[\tilde{n}] \end{aligned} \quad (73)$$

$\tilde{Z}(\mathbf{r})$ is an angular-momentum dependent core-like density that will be described in detail below. The remaining parts can be evaluated on radial grids in a spherical-harmonics expansion. The nodal structure of the wavefunctions can be properly described on a logarithmic radial grid that

becomes very fine near the nucleus,

$$E_R^1 = \sum_{i,j \in R} D_{i,j} \langle \phi_j | \frac{\hat{p}^2}{2m_e} | \phi_i \rangle + \sum_{n \in R}^{N_{c,R}} \langle \phi_n^c | \frac{\hat{p}^2}{2m_e} | \phi_n^c \rangle + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 (n^1(\vec{r}) + Z(\vec{r})) (n^1(\vec{r}') + Z(\vec{r}'))}{|\vec{r} - \vec{r}'|} + E_{xc}[n^1] \quad (74)$$

$$\tilde{E}_R^1 = \sum_{i,j \in R} D_{i,j} \langle \tilde{\phi}_j | \frac{\hat{p}^2}{2m_e} | \tilde{\phi}_i \rangle + \frac{1}{2} \int d^3r \int d^3r' \frac{e^2 (\tilde{n}^1(\vec{r}) + \tilde{Z}(\vec{r})) (\tilde{n}^1(\vec{r}') + \tilde{Z}(\vec{r}'))}{4\pi\epsilon_0 |\vec{r} - \vec{r}'|} + \int d^3r \bar{v}(\vec{r}) \tilde{n}^1(\vec{r}) + E_{xc}[\tilde{n}^1]. \quad (75)$$

The compensation charge density $\tilde{Z}(\vec{r}) = \sum_R \tilde{Z}_R(\vec{r})$ is given as a sum of angular momentum dependent Gauss functions, which have an analytical plane-wave expansion. A similar term occurs also in the pseudopotential approach. In contrast to the norm-conserving pseudopotential approach, however, the compensation charge of an atom \tilde{Z}_R is non-spherical and constantly adapts instantaneously to the environment. It is constructed such that

$$n_R^1(\vec{r}) + Z_R(\vec{r}) - \tilde{n}_R^1(\vec{r}) - \tilde{Z}_R(\vec{r}) \quad (76)$$

has vanishing electrostatic multipole moments for each atomic site. With this choice, the electrostatic potentials of the augmentation densities vanish outside their spheres. This is the reason why there is no electrostatic interaction of the one-center parts between different sites.

The compensation charge density as given here is still localized within the atomic regions. A technique similar to an Ewald summation, however, allows it to be replaced by a very extended charge density. Thus we can achieve that the plane-wave convergence of the total energy is not affected by the auxiliary density.

The potential $\bar{v} = \sum_R \bar{v}_R$, which occurs in Eqs. 73 and 75 enters the total energy in the form of “intelligent zeros” described in Eq. 71

$$0 = \sum_n f_n \left(\langle \tilde{\psi}_n | \bar{v}_R | \tilde{\psi}_n \rangle - \langle \tilde{\psi}_n^1 | \bar{v}_R | \tilde{\psi}_n^1 \rangle \right) = \sum_n f_n \langle \tilde{\psi}_n | \bar{v}_R | \tilde{\psi}_n \rangle - \sum_{i,j \in R} D_{i,j} \langle \tilde{\phi}_i | \bar{v}_R | \tilde{\phi}_j \rangle. \quad (77)$$

The main reason for introducing this potential is to cancel the Coulomb singularity of the potential in the plane-wave part. The potential \bar{v} allows us to influence the plane-wave convergence beneficially, without changing the converged result. \bar{v} must be localized within the augmentation region, where Eq. 58 holds.

Approximations

Once the total energy functional provided in the previous section has been defined, everything else follows: Forces are partial derivatives with respect to atomic positions. The potential is the derivative of the non-kinetic energy contributions to the total energy with respect to the density,

and the auxiliary Hamiltonian follows from derivatives $\tilde{H}|\tilde{\psi}_n\rangle$ with respect to auxiliary wave functions. The fictitious Lagrangian approach of Car and Parrinello [98] does not allow any freedom in the way these derivatives are obtained. Anything else than analytic derivatives will violate energy conservation in a dynamical simulation. Since the expressions are straightforward, even though rather involved, we will not discuss them here.

All approximations are incorporated already in the total energy functional of the PAW method. What are those approximations?

- Firstly we use the frozen-core approximation. In principle this approximation can be overcome.
- The plane-wave expansion for the auxiliary wavefunctions must be complete. The plane-wave expansion is controlled easily by increasing the plane-wave cutoff defined as $E_{PW} = \frac{1}{2}\hbar^2 G_{max}^2$. Typically we use a plane-wave cutoff of 30 Ry.
- The partial wave expansions must be converged. Typically we use one or two partial waves per angular momentum (ℓ, m) and site. It should be noted that the partial wave expansion is not variational, because it changes the total energy functional and not the basis set for the auxiliary wavefunctions.

We do not discuss here numerical approximations such as the choice of the radial grid, since those are easily controlled.

Relation to pseudopotentials

We mentioned earlier that the pseudopotential approach can be derived as a well defined approximation from the PAW method: The augmentation part of the total energy $\Delta E = E^1 - \tilde{E}^1$ for one atom is a functional of the one-center density matrix \mathbf{D} defined in Eq. 68. The pseudopotential approach can be recovered if we truncate a Taylor expansion of ΔE about the atomic density matrix after the linear term. The term linear in \mathbf{D} is the energy related to the nonlocal pseudopotential.

$$\begin{aligned} \Delta E(\mathbf{D}) &= \Delta E(\mathbf{D}^{at}) + \sum_{i,j} \left. \frac{\partial \Delta E}{\partial D_{i,j}} \right|_{\mathbf{D}^{at}} (D_{i,j} - D_{i,j}^{at}) + O(\mathbf{D} - \mathbf{D}^{at})^2 \\ &= E_{self} + \sum_n f_n \langle \tilde{\psi}_n | \hat{v}^{ps} | \tilde{\psi}_n \rangle - \int d^3r \bar{v}(\vec{r}) \tilde{n}(\vec{r}) + O(\mathbf{D} - \mathbf{D})^2, \end{aligned} \quad (78)$$

which can directly be compared with the total energy expression Eq. 49 of the pseudopotential method. The local potential $\bar{v}(\vec{r})$ of the pseudopotential approach is identical to the corresponding potential of the projector augmented-wave method. The remaining contributions in the PAW total energy, namely \tilde{E} , differ from the corresponding terms in Eq. 49 only in two features: our auxiliary density also contains an auxiliary core density, reflecting the nonlinear core correction of the pseudopotential approach, and the compensation density $\tilde{Z}(\vec{r})$ is non-spherical and depends on the wave function. Thus we can look at the PAW method also as a pseudopotential

method with a pseudopotential that adapts instantaneously to the electronic environment. In the PAW method, the explicit nonlinear dependence of the total energy on the one-center density matrix is properly taken into account.

What are the main advantages of the PAW method compared with the pseudopotential approach? Firstly all errors can be systematically controlled, so that there are no transferability errors. As shown by Watson [99] and Kresse [100], most pseudopotentials fail for high spin atoms such as Cr. While it is probably true that pseudopotentials can be constructed that cope even with this situation, a failure can not be known beforehand, so that some empiricism remains in practice: A pseudopotential constructed from an isolated atom is not guaranteed to be accurate for a molecule. In contrast, the converged results of the PAW method do not depend on a reference system such as an isolated atom, because PAW uses the full density and potential.

Like other all-electron methods, the PAW method provides access to the full charge and spin density, which is relevant, for example, for hyperfine parameters. Hyperfine parameters are sensitive probes of the electron density near the nucleus. In many situations they are the only information available that allows us to deduce atomic structure and chemical environment of an atom from experiment.

The plane-wave convergence is more rapid than in norm-conserving pseudopotentials and should in principle be equivalent to that of ultra-soft pseudopotentials [88]. Compared to the ultra-soft pseudopotentials, however, the PAW method has the advantage that the total energy expression is less complex and can therefore be expected to be more efficient.

The construction of pseudopotentials requires us to determine a number of parameters. As they influence the results, their choice is critical. Also the PAW methods provides some flexibility in the choice of auxiliary partial waves. However, this choice does not influence the converged results.

Recent developments

Since the first implementation of the PAW method in the CP-PAW code [58], a number of groups have adopted the PAW method. The second implementation was done by the group of Holzwarth [101], and the resulting PWPAAW code is freely available [102]. This code is also used as a basis for the PAW implementation in the ABINIT project [104]. An independent PAW code has been developed by Valiev and Weare [103]. This implementation has entered the NWChem code [107]. An independent implementation of the PAW method is that of the VASP code [100]. The PAW method has also been implemented by W. Kromen [106] into the EStCoMPP code of Blügel and Schroeder. Another implementation is in the Quantum Espresso code [105]. A real-space-grid based version of the PAW method is the code GPAW developed by Mortensen et al. [108].

Another branch of methods uses the reconstruction of the PAW method, without taking into account the full wavefunctions in the energy minimization. Following chemists' notation, this approach could be termed "post-pseudopotential PAW". This development began with the evaluation for hyperfine parameters from a pseudopotential calculation using the PAW reconstruc-

tion operator [109] and is now used in the pseudopotential approach to calculate properties that require the correct wavefunctions such as hyperfine parameters.

The implementation of the PAW method by Kresse and Joubert [100] has been particularly useful as they had an implementation of PAW in the same code as the ultra-soft pseudopotentials, so that they could critically compare the two approaches. Their conclusion is that both methods compare well in most cases, but they found that magnetic energies are seriously – by a factor two – in error in the pseudopotential approach, while the results of the PAW method were in line with other all-electron calculations using the linear augmented plane-wave method. As an aside, Kresse and Joubert claim incorrectly that their implementation is superior as it includes a term that is analogous to the non-linear core correction of pseudopotentials [110]: this term, however, is already included in the original version in the form of the pseudized core density.

Several extensions of the PAW have been done in the recent years: For applications in chemistry truly isolated systems are often of great interest. As any plane-wave based method introduces periodic images, the electrostatic interaction between these images can cause serious errors. The problem has been solved by mapping the charge density onto a point charge model, so that the electrostatic interaction could be subtracted out in a self-consistent manner [111]. In order to include the influence of the environment, the latter was simulated by simpler force fields using the quantum-mechanics–molecular-mechanics (QM-MM) approach [112].

In order to overcome the limitations of the density functional theory several extensions have been performed. Bengone [113] implemented the LDA+U approach into our CP-PAW code. Soon after this, Arnaud [114] accomplished the implementation of the GW approximation into our CP-PAW code. The VASP-version of PAW [115] and our CP-PAW code have now been extended to include a non-collinear description of the magnetic moments. In a non-collinear description, the Schrödinger equation is replaced by the Pauli equation with two-component spinor wavefunctions.

The PAW method has proven useful to evaluate electric field gradients [116] and magnetic hyperfine parameters with high accuracy [117]. Invaluable will be the prediction of NMR chemical shifts using the GIPAW method of Pickard and Mauri [118], which is based on their earlier work [119]. While the GIPAW is implemented in a post-pseudopotential manner, the extension to a self-consistent PAW calculation should be straightforward. An post-pseudopotential approach has also been used to evaluate core level spectra [120] and momentum matrix elements [121].

Acknowledgment

Part of this article has been written together with Clemens Först and Johannes Kästner. I am also grateful for the careful reading and helpful suggestions by Rolf Fader, Johannes Kirschner, Jürgen Noffke and Philipp Seichter. Financial support by the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

Appendices

A Model exchange-correlation energy

We consider a model with a constant density and a hole function that describes a situation where all electrons of the same spin are repelled completely from a sphere centered at the reference electron

The hole function has the form

$$h(\vec{r}, \vec{r}_0) = \begin{cases} -\frac{1}{2}n(\vec{r}_0) & \text{for } |\vec{r} - \vec{r}_0| < r_h \\ 0 & \text{otherwise} \end{cases}$$

where $n(\vec{r})$ is the electron density and the hole radius $r_h = \sqrt[3]{\frac{2}{4\pi n}}$ is the radius of the sphere, which is determined such that the exchange correlation hole integrates to -1 , i.e. $\frac{4\pi}{3}r_h^3(\frac{1}{2}n) = 1$.

The potential of a homogeneously charged sphere with radius r_h and one positive charge is

$$v(r) = \frac{e^2}{4\pi\epsilon_0} \begin{cases} -\frac{3}{2r_h} + \frac{1}{2r_h} \left(\frac{r}{r_h}\right)^2 & \text{for } r \leq r_h \\ -\frac{1}{r} & \text{for } r > r_h \end{cases}$$

where $r = |\vec{r} - \vec{r}_0|$.

With Eq. 12 we obtain for the potential contribution of the exchange correlation energy

$$U_{xc} = - \int d^3r n(\vec{r})v(r=0) = - \int d^3r \frac{e^2}{4\pi\epsilon_0} \frac{3}{4} \sqrt[3]{\frac{2\pi}{3}} \cdot n^{\frac{4}{3}}$$

B Large-gradient limit of the enhancement factor

An exponentially decaying density

$$n(r) = \exp(-\lambda r) \quad (79)$$

has a reduced gradient

$$x := \frac{|\vec{\nabla}n|}{n^{\frac{4}{3}}} = \lambda \exp(+\frac{1}{3}\lambda r) \quad (80)$$

We make the following ansatz for the exchange correlation energy per electron

$$\epsilon_{xc}(n, x) = -Cn^{\frac{1}{3}}F(x) \quad (81)$$

where only the local exchange has been used and C is a constant.

Enforcing the long-distance limit of the exchange correlation energy per electron for exponentially decaying densities

$$\epsilon_{xc}((n(r), x(r))) = -\frac{1}{2} \frac{e^2}{4\pi\epsilon_0 r} \quad (82)$$

yields

$$F(x) = \frac{e^2}{4\pi\epsilon_0 r(x) 2C n^{\frac{1}{3}}(r(x))} \quad (83)$$

Using Eqs. 79 and 80, we express the radius and the density by the reduced gradient, i.e.

$$r(x) = -\frac{3}{\lambda} \left(\ln[\lambda] - \ln[x] \right) \quad (84)$$

$$n(x) = n(r(x)) = \lambda^3 x^{-3}, \quad (85)$$

and obtain

$$F(x) = \frac{e^2}{4\pi\epsilon_0 \left[-\frac{3}{\lambda} \left(\ln[\lambda] - \ln[x] \right) \right] \left[2C \lambda x^{-1} \right]} = \left(\frac{e^2}{4\pi\epsilon_0 \cdot 6C} \right) \frac{x^2}{x \ln(\lambda) - x \ln(x)}$$

$$\xrightarrow{x \rightarrow \infty} - \left(\frac{e^2}{4\pi\epsilon_0 \cdot 6C} \right) \frac{x^2}{x \ln(x)} \quad (86)$$

Now we need to ensure that $F(0) = 1$, so that the gradient correction vanishes for the homogeneous electron gas, and that $F(x) = F(-x)$ to enforce spin reversal symmetry. There are several possible interpolations for these requirements, but the simplest is

$$F(x) = 1 - \frac{\beta x^2}{1 + \frac{4\pi\epsilon_0}{e^2} \cdot 6C \beta x \cdot \text{asinh}(x)} \quad (87)$$

This is the enhancement factor for exchange used by Becke in his B88 functional [24].

References

- [1] W. Kohn, *Nobel Lectures, Chemistry, 1996-2000* (World Scientific, Singapore, 2003), Ch.: Electronic Structure of Matter - Wave functions and Density Functionals, p. 213
- [2] P. Hohenberg, W. Kohn, *Phys. Rev.* **136**, B864 (1964)
- [3] W. Kohn, L. Sham, *Phys. Rev.* **140**, A1133 (1965)
- [4] R.O. Jones, O. Gunnarsson, *Rev. Mod. Phys.* **61**, 689 (1989)
- [5] E.J. Baerends, O.V. Gritsenko, *J. Phys. Chem. A* **101**, 5383 (1997)
- [6] U. von Barth, *Physica Scripta* **109**, 9 (2004)
- [7] J. Perdew, *et al.*, *J. Chem. Phys.* **123**, 62201 (2005)
- [8] A.J. Cohen, P. Mori-Sánchez, W. Yang, *Science* **321**, 792 (2008)
- [9] P.E. Blöchl, C.J. Först, J. Kästner, *Handbook for Materials Modeling* (Springer, 2005), chap. Electronic Structure Methods: Augmented Waves, Pseudopotentials and the Projector augmented wave method, p. 93
- [10] P.E. Blöchl, C.J. Först, J. Schimpl, *Bull. Mater. Sci.* **26**, 33 (2003)
- [11] P.-O. Löwdin, *Phys. Rev.* **97**, 1474 (1955)
- [12] A. Coleman, *Rev. Mod. Phys.* **35**, 668 (1963)
- [13] K. Burke, J.P. Perdew, M. Ernzerhof, *J. Chem. Phys.* **109**, 3760 (1998)
- [14] J.P. Perdew, K. Burke, *Int. J. Quant. Chem.* **57**, 309 (1996)
- [15] M. Levy, *Proc. Nat'l Acad. Sci. USA* **76**, 6062 (1979)
- [16] E. Lieb, *Int. J. Quant. Chem.* **24**, 243 (1983)
- [17] J.P. Perdew, K. Schmidt, *AIP Conf. Proc.* **577**, 1 (2001)
- [18] D.M. Ceperley, B. Alder, *Phys. Rev. Lett.* **45**, 566 (1980)
- [19] S.K. Ma, K. Brueckner, *Phys. Rev.* **165**, 18 (1968)
- [20] U. von Barth, L. Hedin, *J. Phys. C: Solid State Phys.* **5**, 1629 (1972)
- [21] O. Gunnarsson, B.I. Lundquist, *Phys. Rev. B* **13**, 4274 (1976)
- [22] J.P. Perdew, *Phys. Rev. Lett.* **55**, 1665 (1985)
- [23] D.C. Langreth, M. Mehl, *Phys. Rev. B* **28**, 1809 (1983)

- [24] A.D. Becke, *Phys. Rev. A* **38**, 3098 (1988)
- [25] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996)
- [26] E. Proynov, E. Ruiz, A. Vela, D.R. Salahub, *Int. J. Quant. Chem.* **56**, S29, 61 (1995)
- [27] T.V. Voorhis, G. Scuseria, *J. Chem. Phys.* **109**, 400 (1998)
- [28] A.D. Becke, *J. Chem. Phys.* **109**, 2092 (1998)
- [29] J. Tao, J.P. Perdew, V.N. Staroverov, G.E. Scuseria, *Phys. Rev. Lett.* **91**, 146401 (2003)
- [30] A.D. Becke, *J. Chem. Phys.* **98**, 1372 (1993)
- [31] A.D. Becke, *J. Chem. Phys.* **98**, 5648 (1993)
- [32] J. Harris, R.O. Jones, *J. Phys. F: Met. Phys.* **4**, 1170 (1974)
- [33] D.C. Langreth, J.P. Perdew, *Sol. St. Commun.* **17**, 1425 (1975)
- [34] J.P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **105**, 9982 (1996)
- [35] J. Heyd, G. Scuseria, M. Ernzerhof, *J. Chem. Phys.* **118**, 8207 (2003)
- [36] M. Marsman, J. Paier, A. Stroppa, G. Kresse, *J. Phys. Chem.* **20**, 64201 (2008)
- [37] P. Stephens, F. Devlin, C. Chabalowski, M. Frisch, *J. Phys. Chem.* **98**, 11623 (1994)
- [38] M. Ernzerhof, G. Scuseria, *J. Chem. Phys.* **110**, 5029 (99)
- [39] C. Adamo, V. Barone, *J. Chem. Phys.* **110**, 6158 (1999)
- [40] V.I. Anisimov, J. Zaanen, O.K. Andersen, *Phys. Rev. B* **44**, 943 (1991)
- [41] A.I. Lichtenstein, V.I. Anisimov, J. Zaanen, *Phys. Rev. B* **52**, 5467 (1995)
- [42] P. Novak, J. Kunes, L. Chaput, W.E. Pickett, *Phys. Stat. Sol. B* **243**, 563 (2006)
- [43] F. Tran, P. Blaha, K. Schwarz, P. Novák, *Phys. Rev. B* **74**, 155108 (2006)
- [44] M. Dion, H. Rydberg, E. Schröder, D.C. Langreth, B. I. Lundqvist, *Phys. Rev. Lett.* **92**, 246401 (2004)
- [45] T. Thonhauser, V.R. C.S. Li, A. Puzder, P. Hyldgaard, D. Langreth, *Phys. Rev. B* **76**, 125112 (2007)
- [46] K. Lee, E.D. Murray, L. Kong, B.I. Lundqvist, D.C. Langreth, *Phys. Rev. B* **82**, 81101 (2010)
- [47] J.A. Pople, M. Head-Gordon, D.J. Fox, K. Raghavachari, L.A. Curtiss, *J. Chem. Phys.* **90**, 5622 (1989)

- [48] L.A. Curtiss, C. Jones, G.W. Trucks, K. Raghavachari, J.A. Pople, *J. Chem. Phys.* **93**, 2537 (1990)
- [49] L.A. Curtiss, K. Raghavachari, P.C. Redfern, J.A. Pople, *J. Chem. Phys.* **106**, 1063 (1997)
- [50] L.A. Curtiss, P.C. Redfern, K. Raghavachari, J.A. Pople, *J. Chem. Phys.* **109**, 42 (1998)
- [51] A.D. Becke, *J. Chem. Phys.* **96**, 2155 (1992)
- [52] A.D. Becke, *J. Chem. Phys.* **97**, 9173 (1992)
- [53] A.D. Becke, *J. Chem. Phys.* **104**, 1040 (1996)
- [54] A.D. Becke, *J. Chem. Phys.* **107**, 8554 (1997)
- [55] J. Paier, R. Hirschl, M. Marsman, G. Kresse, *J. Chem. Phys.* **122**, 234102 (2005)
- [56] J. Paier, *et al.*, *J. Chem. Phys.* **124**, 154709 (2006)
- [57] J. Paier, *et al.*, *J. Chem. Phys.* **125**, 249901 (2006)
- [58] P.E. Blöchl, *Phys. Rev. B* **50**, 17953 (1994)
- [59] J.C. Slater, *Phys. Rev.* **51**, 846 (1937)
- [60] J. Koringa, *Physica* **13**, 392 (1947)
- [61] W. Kohn, N. Rostoker, *Phys. Rev.* **94**, 1111 (1954)
- [62] O.K. Andersen, *Phys. Rev. B* **12**, 3060 (1975)
- [63] H. Krakauer, M. Posternak, A.J. Freeman, *Phys. Rev. B* **19**, 1706 (1979)
- [64] D.J. Singh, *Planewaves, pseudopotentials and the LAPW method* (Kluwer, 1994)
- [65] J. Soler, A.R. Williams, *Phys. Rev. B* **40**, 1560 (1989)
- [66] D.J. Singh, *Phys. Rev. B* **43**, 6388 (1991)
- [67] E. Sjöstedt, L. Nordström, D.J. Singh, *Solid State Commun.* **113**, 15 (2000)
- [68] G.K. H. Madsen, P. Blaha, K. Schwarz, E. Sjoestedt, L. Nordstroem, *Phys. Rev. B* **64**, 195134 (2001)
- [69] H.L. Skriver, *The LMTO method: muffin-tin orbitals and electronic structure* (Springer, 1984)
- [70] O.K. Andersen, O. Jepsen, *Phys. Rev. Lett.* **53**, 2571 (1984)

-
- [71] O.K. Andersen, T. Saha-Dasgupta, S. Ezhof, *Bull. Mater. Sci.* **26**, 19 (2003)
- [72] K. Held, *et al.*, *NIC Series (John von Neumann Institute for Computing)* **10**, 175 (2002)
- [73] C. Herring, *Phys. Rev.* **57**, 1169 (1940)
- [74] J. Phillips, L. Kleinman, *Phys. Rev.* **116**, 287 (1959)
- [75] E. Antončík, *J. Phys. Chem. Solids* **10**, 314 (1959)
- [76] D.R. Hamann, M. Schlüter, C. Chiang, *Phys. Rev. Lett.* **43**, 1494 (1979)
- [77] A. Zunger, M.L. Cohen, *Phys. Rev. B* **18**, 5449 (1978)
- [78] G. Kerker, *J. Phys. C: Solid State Phys.* **13**, L189 (1980)
- [79] G.B. Bachelet, D.R. Hamann, M. Schlüter, *Phys. Rev. B* **26**, 4199 (1982)
- [80] N. Troullier, J. Martins, *Phys. Rev. B* **43**, 1993 (1991)
- [81] J. Lin, A. Qteish, M. Payne, V. Heine, *Phys. Rev. B* **47**, 4174 (1993)
- [82] M. Fuchs, M. Scheffler, *Comp. Phys. Comm.* **119**, 67 (1998)
- [83] L. Kleinman, D.M. Bylander, *Phys. Rev. Lett.* **48**, 1425 (1982)
- [84] P.E. Blöchl, *Phys. Rev. B* **41**, 5414 (1990)
- [85] D. Vanderbilt, *Phys. Rev. B* **41**, 7892 (1990)
- [86] S.G. Louie, S. Froyen, M.L. Cohen, *Phys. Rev. B* **26**, 1738 (1982)
- [87] D.R. Hamann, *Phys. Rev. B* **40**, 2980 (1989)
- [88] D. Vanderbilt, *Phys. Rev. B* **41**, 17892 (1990)
- [89] K. Laasonen, A. Pasquarello, R. Car, C. Lee, D. Vanderbilt, *Phys. Rev. B* **47**, 110142 (1993)
- [90] X. Gonze, R. Stumpf, M. Scheffler, *Phys. Rev. B* **44**, 8503 (1991)
- [91] C.V. de Walle, P.E. Blöchl, *Phys. Rev. B* **47**, 4244 (1993)
- [92] M.C. Payne, M.P. Teter, D.C. Allan, T.A. Arias, J.D. Joannopoulos, *Rev. Mod. Phys* **64**, 11045 (1992)
- [93] R. Car, M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985)
- [94] S. Nose, *J. Chem. Phys.* **81**, 511 (1984)
- [95] W.G. Hoover, *Phys. Rev. A* **31**, 1695 (1985)

- [96] P.E. Blöchl, M. Parrinello, *Phys. Rev. B* **45**, 9413 (1992)
- [97] P.E. Blöchl, *Phys. Rev. B* **65**, 104303 (2002)
- [98] R. Car, M. Parrinello, *Phys. Rev. Lett.* **55**, 2471 (1985)
- [99] S.C. Watson, E.A. Carter, *Phys. Rev. B* **58**, R13309 (1998)
- [100] G. Kresse, J. Joubert, *Phys. Rev. B* **59**, 1758 (1999)
- [101] N.A.W. Holzwarth, G.E. Matthews, R.B. Dunning, A.R. Tackett, Y. Zheng, *Phys. Rev. B* **55**, 2005 (1997)
- [102] A.R. Tackett, N.A.W. Holzwarth, G.E. Matthews, *Computer Physics Communications* **135**, 329 (2001)
- [103] M. Valiev, J.H. Weare, *J. Phys. Chem. A* **103**, 10588 (1999)
- [104] M. Torrent, F. Jollet, F. Bottin, G. Zerah and X. Gonze, *Comp. Mat. Sci.* **42**, 337 (2008)
- [105] P. Gianozzi *et al.*, *J. Phys C: Condens. Matt.* 21,395502 (2009)
- [106] W. Kromen, *Die Projector Augmented Wave-Methode: Ein schnelles Allelektronenverfahren für die ab-initio-Molekulardynamik*, (PhD thesis, RWTH Aachen, 2001)
- [107] T.L. Windus, E.J. Bylaska, M. Dupuis, S. Hirata and L. Pollack, et al. *Lecture Notes in Computer Science*, 2003, Volume 2660, Computational Science ICCS 2003, 168 (2003)
- [108] J. J. Mortensen, L. B. Hansen and K.W. Jacobsen, *Phys. Rev. B* **71**, 35109 (2005)
- [109] C.G. Van de Walle, P.E. Blöchl, *Phys. Rev. B* **47**, 4244 (1993)
- [110] S.G. Louie, S. Froyen, M.L. Cohen, *Phys. Rev. B* **26**, 1738 (1982)
- [111] P.E. Blöchl, *J. Chem. Phys.* **103**, 7422 (1995)
- [112] T.K. Woo, P.M. Margl, P.E. Blöchl, T. Ziegler, *J. Phys. Chem. B* **101**, 7877 (1997)
- [113] O. Bengone, M. Alouani, P.E. Blöchl, J. Hugel, *Phys. Rev. B* **62**, 16392 (2000)
- [114] B. Arnaud, M. Alouani, *Phys. Rev. B.* **62**, 4464 (2000)
- [115] D. Hobbs, G. Kresse, J. Hafner, *Phys. Rev. B* **62**, 11556 (2000)
- [116] H.M. Petrilli, P.E. Blöchl, P. Blaha, K. Schwarz, *Phys. Rev. B* **57**, 14690 (1998)
- [117] P.E. Blöchl, *Phys. Rev. B* **62**, 6158 (2000)
- [118] C.J. Pickard, F. Mauri, *Phys. Rev. B.* **63**, 245101 (2001)

- [119] F. Mauri, B.G. Pfrommer, S.G. Louie, *Phys. Rev. Lett.* **77**, 5300 (1996)
- [120] D.N. Jayawardane, C.J. Pickard, L.M. Brown, M.C. Payne,
Phys. Rev. B **64**, 115107 (2001)
- [121] H. Kageshima, K. Shiraishi, *Phys. Rev. B* **56**, 14985 (1997)

3 Model Hamiltonians and Basic Techniques

Frank Lechermann

I. Institut für Theoretische Physik

Universität Hamburg

Contents

1	Motivation	2
2	Introduction to the single-band Hubbard model	3
2.1	Competition between itinerancy and localization	3
2.2	Heuristic derivation	4
2.3	Some model limits and basic excitations	6
3	Relatives and extensions of the Hubbard model	9
3.1	Anderson Hamiltonian	9
3.2	Multi-band Hubbard Hamiltonians	11
3.3	More interactions	14
4	Basic approaches to the single-band Hubbard model	15
4.1	Hartree-Fock	15
4.2	Hubbard I	18
5	Slave-boson approach	20
5.1	Infinite- U limit	20
5.2	Kotliar-Ruckenstein representation	23
6	Final remarks	25

1 Motivation

The problem of interacting electrons in realistic systems is a difficult one. Although conventional Bloch band theory with its roots dating back to the 1920s (see Ref. [1] for a historical review) is very successful in describing many so-called weakly correlated materials, the intricate phenomenology of electronic systems that display signatures of strong correlations are basically outside that scope. First-principles or *ab-initio* approaches to tackle such systems on a similar level of sophistication as their weakly correlated neighbors is highly non-trivial and the central topic of this school. In this context, model Hamiltonians play an essential role by, loosely speaking, bridging the gap between the possible to the (nearly) impossible. Their justification can in principle be categorized more concretely by

- (i) the tremendous complexity of materials systems on a microscopic level because of the large numbers of various degrees of freedom in often low-symmetry environments that asks for simplifications to discuss the dominant physics of interest.
- (ii) the condensation of different robust physical mechanisms in simplified mathematical terms in order to look for new physical processes via their fascinating mutual interplay.

Both theoretical perspectives allow for a predictive character in the use of models, yet the first one is somewhat more directly associated with given materials problems. Therefore we want to concentrate thereon, namely on model Hamiltonians geared to simulate the key physics of notoriously complicated complete Hamiltonians of large-scale interacting systems. Note that a model approach not always has to cover only low-energy scales, but that often also high-energy excitations are of vital interest. In addition, here we are aiming at the full quantum nature of the problem and leave possible classical approximations (e.g., Ising model in zero field [2], etc.), though often also interesting, aside.

Lets indeed start from the complete Hamiltonian \mathcal{H} of a condensed matter system with N_e electrons and N_K nuclei on a lattice with position vectors \mathbf{R}_α . In first quantization the problem reads

$$\mathcal{H} = \underbrace{-\sum_{\alpha}^{N_K} \frac{\hbar^2 \Delta_{\alpha}}{2M_{\alpha}} + \frac{1}{2} \sum_{\substack{\alpha\alpha' \\ \alpha \neq \alpha'}} \frac{Z_{\alpha} Z_{\alpha'} e^2}{|\mathbf{R}_{\alpha} - \mathbf{R}_{\alpha'}|}}_{=:\mathcal{T}_K + \mathcal{V}_{KK} =: \mathcal{H}_K} \underbrace{- \sum_{\alpha\mu} \frac{Z_{\alpha} e^2}{|\mathbf{R}_{\alpha} - \mathbf{r}_{\mu}|}}_{=:\mathcal{V}_{Ke}} \underbrace{- \sum_{\mu}^{N_e} \frac{\hbar^2 \Delta_{\mu}}{2m} + \frac{1}{2} \sum_{\substack{\mu\mu' \\ \mu \neq \mu'}} \frac{e^2}{|\mathbf{r}_{\mu} - \mathbf{r}_{\mu'}|}}_{=:\mathcal{T}_e + \mathcal{V}_{ee} =: \mathcal{H}_e} \quad , \quad (1)$$

composed of a pure part \mathcal{H}_K for the nuclei, a pure electronic part \mathcal{H}_e as well as the coupling \mathcal{V}_{Ke} thereof. The Born-Oppenheimer approximation (BOA) [3] allows a decoupling of the problem for the nuclei from the electrons, ending up with an electronic Hamiltonian of the form

$$H_e = H_e(\{\mathbf{R}\}) := \mathcal{H}_e + \mathcal{V}_{Ke} \quad , \quad (2)$$

where the set $\{\mathbf{R}\}$ of lattice points is a mere fixed parameter set for the given electronic problem. The periodic lattice potential \mathcal{V}_{Ke} is then often denoted as *external potential* v_{ext} , to highlight

the electronic-system character. The remaining lattice part of \mathcal{H} eventually leads to phonon excitations and their interactions, which we will not cover in this overview. Only in section 3.3 a brief discussion of modeled electron-phonon coupling, i.e., physics beyond the BOA, will mildly touch this matter.

The sobering news is that after the decoupling from the lattice degrees of freedom the solid-state problem of interacting electrons within an external potential is still much too complicated. In band theory it is assumed that H_e may be written as a sum of effective single-particle terms, i.e., the electron-electron interaction \mathcal{V}_{ee} as a two-particle operator is transformed into a one-particle form. Although for ground-state properties such an exact transformation *in principle* exists (see the chapter on density functional theory), so far, however, any practical approximation leads to severe failures for correlated solid-state systems. Not only is the exact analytical treatment of the true H_e impossible, also tough numerical methods surrender to the explosion of the associated Hilbert space in this technically very demanding potential landscape of a realistic solid. Unlike in quantum chemistry, where these numerical techniques for atoms and smaller molecules are often still feasible (however nowadays also with a huge amount of labor, see e.g. [4] for an introduction), condensed matter theory is in need of model approaches.

Thus in the following sections we will deal with ways of simplifying H_e in order to squeeze some interesting physics out of it. While sections 2 and 3 cope with the definition of the Hubbard model and its friends, sections 4 and 5 deal with first simple and not so simple tools for the actual solutions. In fact, albeit we just emphasized the complexity of realistic systems, solving, once derived, model Hamiltonians is very far from being easy and a whole community in theoretical condensed matter physics is devoted to this. In the end it is mainly about changing an almost utopian problem into a less utopian one.

2 Introduction to the single-band Hubbard model

Since there is a whole zoo of model Hamiltonians available, we have to limit the discussion to a certain subset. As our main interest in this school is in problems where the interplay between kinetic energy gain and cost of Coulomb interaction is central, it is meaningful to circle the discussion around the Hubbard model. Excellent accounts of pure spin models can be found elsewhere (see e.g. the books [5–7]).

2.1 Competition between itinerancy and localization

Before presenting a heuristic derivation of the basic Hubbard model it is useful to provide an intuitive view on the main driving forces that govern interacting electron systems in a solid.

When placing atoms in regular periodic arrays, the first obvious deviation to atomic physics is due to symmetry, namely the splitting of states because of the emerging crystal field. The latter competes usually with the exchange splittings responsible for Hund's rules as well as with the spin-orbit interaction. In the following we assume the general understanding that for *spd* systems intermediate crystal-field strengths apply. Thus Hund's first and second rule are vital

and the still smaller crystal field dominates over the spin-orbit interaction.

Because of the Heisenberg uncertainty principle the electrons are in favor of minimizing their kinetic energy through dislocation, i.e., hopping processes between different atoms are preferable, leading eventually to crystal bonding. However as in atomic physics, whenever electrons come close the Coulomb energy raises due to the mutual interaction. Hence the overall movements within the electronic system are quite intricate and usually highly correlated. But there are two limits when everything becomes much simpler. If the effective Coulomb interaction in the system is very well screened, in a simple picture one may assume that the electrons are rather free to optimize their kinetic energy irrespective of the restrictions imposed by interaction. Note that one truly has to invoke the quasiparticle concept of Landau Fermi-liquid theory (see e.g. [8]) to fully justify such a view, since the screening is itself mediated by the electronic system. An even simpler limit is provided by the absence of sufficient screening processes in a commensurable filling scenario. The latter means an integer ratio between the number of electrons N_e and the number of lattice sites N_l . Then the electrons' tendency to leave their lattice site and hop around the lattice is approaching zero and an insulating state of matter results. This electron localization in real space, termed *Mott insulator*, is quite different from the commonly known band-insulating state. In a band insulator the absence of electrical conductivity is based on the complete filling of a band of effective single-particle Bloch states and the existence of an HOMO-LUMO gap. A Mott insulator is a quite different beast with a charge gap that is not governed by (renormalized) hybridization effects but which originates from the (renormalized) Coulomb interaction between the electrons.

In summary, we expect the competition between the itinerant and the localized character of the electrons as the vital ingredients to get a hold on the key physics of many correlated electron systems. In the following the concentration will be therefore on the simplified coherent modeling of this rivalry in mathematical terms, leaving other more specific mechanisms aside. The very basic model Hamiltonian that is tailored to serve this goal is the so-called single-band Hubbard Hamiltonian.

2.2 Heuristic derivation

Instead of approximating H_e of eq. (2) as a sum of single-particle terms, another viewpoint is taken now. The aim is to keep the explicit many-particle structure of the full Hamiltonian, but to chop off those terms that are not vital for the basic modeling we outlined in the last subsection. Since best suited for many-body systems, we therefore start by writing H_e in second quantization¹ as

$$H_e = - \sum_{\alpha\beta ab\sigma} t_{\mathbf{R}_\alpha\mathbf{R}_\beta}^{L_a L_b} c_{\mathbf{R}_\alpha L_a \sigma}^\dagger c_{\mathbf{R}_\beta L_b \sigma} + \frac{1}{2} \sum_{\substack{\alpha\beta\gamma\delta \\ abcd\sigma\sigma'}} V_{ee}(\{\mathbf{R}, L\}) c_{\mathbf{R}_\alpha L_a \sigma}^\dagger c_{\mathbf{R}_\beta L_b \sigma'}^\dagger c_{\mathbf{R}_\delta L_d \sigma'} c_{\mathbf{R}_\gamma L_c \sigma} . \quad (3)$$

Here the electron creation and annihilation operator $c^{(\dagger)}$ is represented in a localized Wannier basis $\varphi(\mathbf{r})$ in real space, marked by lattice site \mathbf{R} , orbital character L and spin projection σ .

¹In the following hermiticity of the hopping term is assumed to be enforced.

While the first sum over single-particle terms carries the kinetic energy as well as the interaction with the periodic lattice potential, the second sum describes the electron-electron Coulomb interaction. Note that we do not allow for spin-dependent hopping processes, as here magnetic effects should emerge from the interacting part. The respective matrix elements read

$$t_{\mathbf{R}_\alpha \mathbf{R}_\beta}^{L_a L_b} = \int d\mathbf{r} \varphi_{\mathbf{R}_\alpha L_a}^*(\mathbf{r}) \left\{ \frac{\hbar^2 \Delta}{2m} - v_{\text{ext}}(\mathbf{r}) \right\} \varphi_{\mathbf{R}_\beta L_b}(\mathbf{r}) \quad , \quad (4)$$

$$V_{ee}(\{\mathbf{R}, L\}) = \int d\mathbf{r} d\mathbf{r}' \varphi_{\mathbf{R}_\alpha L_a \sigma}^*(\mathbf{r}) \varphi_{\mathbf{R}_\beta L_b \sigma'}^*(\mathbf{r}') \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \varphi_{\mathbf{R}_\gamma L_c \sigma'}(\mathbf{r}') \varphi_{\mathbf{R}_\delta L_d \sigma}(\mathbf{r}) \quad . \quad (5)$$

Everybody with a basic training in quantum mechanics immediately realizes that its impossible to work out the solution for such a problem with so many degrees of freedom. Whereas the single-particle term is feasible, resulting in a simple form of band theory, the two-particle part is the tough one. From a formal theoretical point of view, one proper method for simplification would be to integrate out high energy degrees of freedom within the spirit of the renormalization-group theory. However for general materials systems this proves unmanageable. Moreover, since we want to study metallic and gapped systems on an equal footing such a recipe is inadequate at the present level. The natural way of dealing with it is by applying two key observations:

- (1) In order to get a very first understanding of competing itinerancy and localization, a full multi-orbital structure is not necessary. Furthermore, one can limit the discussion to only nearest-neighbor (NN) hopping of the electrons.
- (2) The Coulomb interaction between two electrons is usually strongest if both come close in the same localized Wannier orbital. With distance the interaction rapidly decays for many systems also due to efficient screening processes of other interfering electrons.

We therefore take a rather radical approach and assume a model system with NN hoppings between only one Wannier orbital per identical sites, e.g. a model lattice of hydrogen atoms, and keeping solely the on-site Coulomb matrix element $V_{ee}(\mathbf{R}, \mathbf{R}, \mathbf{R}, \mathbf{R}) =: U$. In what follows, to keep notations simple it is customary to label lattice sites with (i, j, \dots) and introduce the particle-number operator $n = c^\dagger c$. With the given dramatic simplifications the complicated Hamiltonian (3) may then be cast into

$$H_{\text{hub}} = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \varepsilon_0 \sum_{i\sigma} n_{i\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \quad . \quad (6)$$

This is the famous single-band Hubbard model, named after J. Hubbard due to its seminal work² [9]. It has only two explicit relevant parameters, namely the hopping amplitude t and the on-site Coulomb interaction, the so-called Hubbard U . Note that the second term scaling by the local single-particle level energy ε_0 (stemming from $t_{\mathbf{R}_\alpha \mathbf{R}_\alpha}$ in eq. (3)) is not strictly mandatory.

²Note however that many authors independently came up with very similar Hamiltonians at around the same time, including P. W. Anderson, M. C. Gutzwiller and J. Kanamori.

Because it only amounts to a shift in energy, one could easily choose $\varepsilon_0=0$ without changing the model properties. The braces in the first sum denote it only runs over NN sites.

Its innocent appearance compared to the full form (3) should not mask the fact that solving the Hubbard model is a very tough job. There is only an exact analytical solution in one dimension (1D) and a numerical exact approach for the infinite-dimensional case which is just the dynamical mean-field theory (DMFT) this school is to a large part devoted to. Sadly enough, the two- and three-dimensional cases, closest to many strong correlation problems in nature, are the really hard nuts to crack. The difficulty arises from the difference of the summation parts. While the first and second sum may easily be diagonalized in *reciprocal space*, the same easy diagonalization can be performed for the third sum, however, on every lattice site in *real space*. In other words, the first sums pose a standard band-theory problem, whereas the third sum may be interpreted as a standard quantum-chemical one. Diagonalizing both parts simultaneously in 2D and 3D appears to be impossible.

Besides t and U there are other implicit “parameters” for the model, such as the lattice type (e.g., square, triangular, fcc, etc.) and the filling $n=N_e/N_l$ (not to be confused with the particle-number operator). A very important setup in this respect is the so-called half-filled case, where $n=1$, i.e., there is one electron per lattice site (and hence per orbital). In that integer filling regime the system is in principle susceptible to a Mott-insulating state, whereas the rather trivial integer fillings $n=0,2$ denote band insulators. Its important to realize that at any other filling no insulating state is reachable for finite t in this model. Since then there will be always double occupations of lattice sites coexisting with single occupations and thus the possibility to lower the total energy via hopping processes.

There is a huge literature on the Hubbard model which in this small review we do not dare to approach. Very detailed overviews can e.g. be found in the books of Fazekas [6] and Gebhard [10].

2.3 Some model limits and basic excitations

When trying to understand a given model it is always a good idea to first examine the limiting regimes. Restricting the discussion to the interesting half-filled $n=1$ case, it is obvious that the ratio U/t serves as a perfect marker. For a certain lattice type one would rather use the bandwidth $W \sim t$ in the denominator, yet we try to keep things simple and do not choose a specific lattice. In order to render the discussion a bit more quantitative, let us focus on computing the zero-temperature *spectral function* $A_\sigma(\mathbf{k}, \omega)$ given by the following expression^{3,4}

$$A_\sigma(\mathbf{k}, \omega) = \begin{cases} A_\sigma^+(\mathbf{k}, \omega) = \sum_m |\langle \Psi_m^{(N_e+1)} | c_{\mathbf{k}\sigma}^\dagger | \Psi_0 \rangle|^2 \delta(\omega - \omega_{m0}) & \text{for } \omega \geq 0 \\ A_\sigma^-(\mathbf{k}, \omega) = \sum_m |\langle \Psi_m^{(N_e-1)} | c_{\mathbf{k}\sigma} | \Psi_0 \rangle|^2 \delta(\omega - \omega_{0m}) & \text{for } \omega < 0 \end{cases}, \quad (7)$$

³This form is exact in the thermodynamic limit, where the energy differences upon particle addition and removal are identical (see e.g. [11]).

⁴The convention $\hbar=1$ is used.

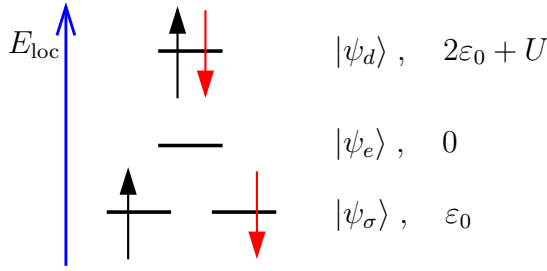


Fig. 1: Level diagram for a single lattice site with one orbital.

with $c_{\mathbf{k}\sigma}^{(\dagger)}$ now creating/annihilating electrons with wave vector \mathbf{k} and spin projection σ . This function provides information about the system's energy distribution via summation over processes after adding and removing an electron to the ground-state wave function $|\Psi_0\rangle$. Thereby the $|\Psi_m\rangle$ denote excited $(N_e \pm 1)$ -particle eigenstates with eigenenergies ω_{m0}, ω_{0m} with respect to the ground state. Of course, these states are unbelievably complicated for generic U/t , however everything becomes rather simple for two limiting cases:

Fermi-gas limit $U=0$. The problem reduces to the simplest band-theoretical form since the Hubbard model loses its interaction term. Hence \mathbf{k} is a good quantum number and creating/annihilating electrons according to (7) already produces eigenstates of the system, the well-known Bloch states. Since $\langle \Psi_{(m0)} | c_{\mathbf{k}\sigma}^{(\dagger)} | \Psi_{(m0)} \rangle$ equals one for the associated eigenstate and zero otherwise, the \mathbf{k} -integrated spectral function $\rho_\sigma(\omega)$ reduces to⁵

$$\rho_\sigma(\omega) = \sum_{\mathbf{k}} A_\sigma(\mathbf{k}, \omega) = \sum_{\mathbf{k}} \delta(\omega - \varepsilon_{\mathbf{k}}) \quad , \quad (8)$$

which is the familiar electronic density of states (DOS) from single-particle theory for the band dispersion $\varepsilon_{\mathbf{k}}$. So one recovers the good old band-theoretical result for a metallic state based on a simple NN hopping. Because interactions among the electrons are completely absent, one usually refers to this solution also as the Fermi gas. This means that in real space the local site occupations, i.e., empty, single and double, are according to the statistics of a non-interacting lattice gas.

Atomic limit $t=0$. This is just the opposite case, no hopping allowed and therefore insulating by default. But now we are facing a truly interacting problem. Albeit a purely local one since obviously the whole job is easily separable into a sum of N_l isolated atomic problems. There are four states per site, which are summarized with their energetics in Fig. 1. For the computation of $\rho_\sigma(\omega)$ one first needs to apply the Fourier-transformation rule $\sqrt{N_l} c_{\mathbf{k}\sigma}^\dagger = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} c_{\mathbf{R}\sigma}^\dagger$ (and c.c. for $c_{\mathbf{k}\sigma}$) to eq. (7) and the problem can be examined for the two branches $\omega \lesseqgtr 0$ as

$$\rho_\sigma(\omega) = \begin{cases} \sum_{\mathbf{R}} \langle n_{\mathbf{R}\bar{\sigma}} \rangle |\langle \psi_d | c_{\mathbf{R}\sigma}^\dagger | \psi_{\bar{\sigma}} \rangle|^2 \delta(\omega - \omega_{d\bar{\sigma}}) = \sum_{\mathbf{R}} \langle n_{\mathbf{R}\bar{\sigma}} \rangle \delta(\omega - (\varepsilon_0 + U)) \\ \sum_{\mathbf{R}} \langle n_{\mathbf{R}\sigma} \rangle |\langle \psi_e | c_{\mathbf{R}\sigma} | \psi_\sigma \rangle|^2 \delta(\omega - \omega_{e\sigma}) = \sum_{\mathbf{R}} \langle n_{\mathbf{R}\sigma} \rangle \delta(\omega - \varepsilon_0) \end{cases} \quad , \quad (9)$$

since for $n=1$ the degenerate $|\psi_{\sigma,\sigma'}\rangle$ forms the local ground state. Thus there are only δ -Peak excitations at ε_0 and $\varepsilon_0 + U$ as expected for an atomic problem. Contrary to the Fermi-gas limit

⁵Throughout the text the proper normalization of \mathbf{k} - and \mathbf{R} -sums to the number of \mathbf{k}/\mathbf{R} -points is assumed.

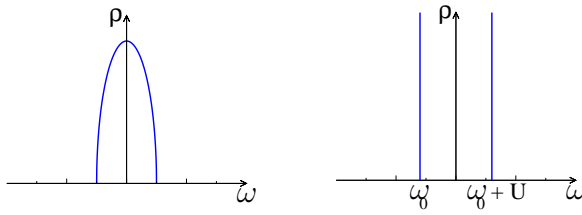


Fig. 2: Schematic local spectral function for the Fermi-gas limit $U=0$ (left) and the atomic limit $t=0$ (right) of the Hubbard model at half filling.

now only single occupied sites appear in the real space ($n=1$)-ground state, because double occupations cost energy.

The local spectral function $\rho_\sigma(\omega)$ for the two limits is depicted in Fig. 2. Note again the very different nature of excitations: whereas for $U=0$ there are ordinary Bloch excitations in k -space forming a rather broad band, for $t=0$ the system shows local multiplet excitations in r -space. Now what happens away from these limits? A simplistic first guess would be to more or less overlay the limiting spectra to mediate between them, whereby one has to take into account that any finite t will broaden spectral peaks in energy. The pictures that emerge presumably look like the ones shown in Fig. 3. Turning on U when starting in the Fermi gas has two main effects. First it leads to *band narrowing* (or the Brinkman-Rice effect [12]), because hopping processes are now somewhat suppressed due to the cost of energy when there are necessarily double occupations formed which are penalized by U . Secondly, multiplet excitations gradually emerge at higher energies already in the metal, since spectral weight that is now missing in the band-like excitations is transferred into the former. This spectral-weight transfer is a hallmark signature of strongly correlated electron systems.

One refers to the renormalized band part at low energy, i.e., close to the Fermi level ε_F , as *quasiparticle excitations* and to the smeared multiplet excitations at higher energies to *Hubbard excitations* or *Hubbard bands*. At very large U/t the system finally undergoes a Mott-insulating transition (MIT). The quasiparticles that remain close to ε_F until the Mott insulator sets in still may be marked with wave vector \mathbf{k} , but now have finite lifetime due to the fact that $c_{\mathbf{k}\sigma}^{(\dagger)}$ do not anymore create/annihilate eigenstates of the solid. In other words, Bloch's theorem does not hold on the interacting lattice. The broader Hubbard bands are doomed with even shorter lifetime, since the incoherent multiplets do not propagate well on the lattice.

The limit $U/t \gg 1$ deserves indeed further discussion. It is important to realize that this case is of course very different from the atomic limit of decoupled lattice sites. Therefore the Mott-insulating state at half filling is still an unique state of condensed matter. In order to approach this very strongly correlated limit for arbitrary filling theoretically, a systematic perturbative

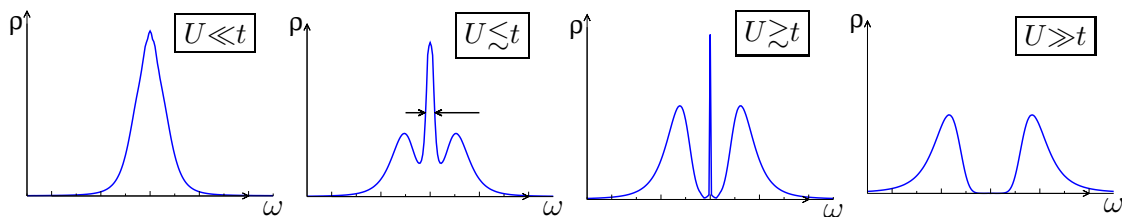


Fig. 3: Local spectral function for the Hubbard model on a Bethe lattice with increasing U at half filling $n=1$. The graphs were obtained within DMFT.

expansion of the Hubbard model in t/U may be performed (for details see [5, 6, 10]). For convenience setting $\varepsilon_0=0$, this leads to the t - J model

$$H_{\text{ij}} = P \left[-t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} - \frac{t^2}{U} \sum_{\langle ijk \rangle} \left(c_{i\uparrow}^\dagger c_{j\downarrow}^\dagger - c_{i\downarrow}^\dagger c_{j\uparrow}^\dagger \right) \left(c_{j\downarrow} c_{k\uparrow} - c_{j\uparrow} c_{k\downarrow} \right) \right] P \quad . \quad (10)$$

The projection operator P restricts the whole configuration space to only those configurations with empty and single occupied sites. Some theorists believe that this model is a good starting point to enter the physics of copper-oxide high-temperature superconductors [13]. In the Mott insulator at half filling only spin excitations appear at low-energy and this model further reduces to the *antiferromagnetic quantum Heisenberg model*⁶

$$H_{\text{heis}} = \frac{2t^2}{U} \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j \quad , \quad (11)$$

with the spin operators defined as $S_i^\mu = 1/2 \sum_{\sigma\sigma'} c_{i\sigma}^\dagger \tau_{\sigma\sigma'}^\mu c_{i\sigma'}$, where $\tau_{\sigma\sigma'}^\mu$ are the elements of the ($\mu=x, y, z$) Pauli matrices. The coupling constant is also most often written via the superexchange parameter $J=4t^2/U > 0$. So in the end the well-known Heisenberg model is contained in the minimal Hubbard model. For a deeper discussion of the t - J and the Heisenberg model we again refer to [4–6, 10]) and references therein.

3 Relatives and extensions of the Hubbard model

In the following sections we briefly want to discuss further Hamiltonians related to the Hubbard model and also extensions of the latter. This is important because many materials problems obviously just do not easily boil down to a simple Hubbard-model form. For instance already the assumption of only NN hopping is rather restrictive and many system display more sophisticated hopping paths. However it is easy to lift this restriction by taking more distant hoppings (e.g., obtained from a Slater-Koster parametrization [14]) into account and writing

$$H_{\text{hub}} = - \sum_{ij, \sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \varepsilon_0 \sum_{i\sigma} n_{i\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \quad . \quad (12)$$

The allowance of more than one orbital per site needs some more thinking and will be discussed in section 3.2. But before we want to have a look on a natural companion of the Hubbard Hamiltonian that has also relevance for the DMFT construction.

3.1 Anderson Hamiltonian

Instead of considering an ensemble of fully correlated lattice sites, it is also very instructive to investigate so-called *impurity models*. There only a few down to one correlated lattice site(s) exist within a given host lattice that supports otherwise rather weakly correlated sites. Since we

⁶Constant energy shifts are neglected.

are here mainly interested in the competition between localization and itinerancy, it is assumed that the electrons stemming from the weakly correlated sites form a simple metallic state. Motivated by experiments on the local-moment behavior of correlated impurity atoms in metals (such as e.g. Fe atoms in copper) and after preliminary theoretical work, P. W. Anderson set up an interacting model Hamiltonian for the single-impurity problem [15], which reads

$$H_{\text{and}} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \varepsilon_d \sum_{\sigma} n_{d\sigma} + U n_{d\uparrow} n_{d\downarrow} + \sum_{\mathbf{k}\sigma} \left(V_{\mathbf{k}d} c_{\mathbf{k}\sigma}^\dagger d_{\sigma} + V_{\mathbf{k}d}^* d_{\sigma}^\dagger c_{\mathbf{k}\sigma} \right) . \quad (13)$$

This Anderson Hamiltonian has in principle similar components as the Hubbard Hamiltonian, but note that there is now only one correlated impurity site with atomic level ε_d and Hubbard interaction U , embedded in a Fermi sea of band electrons with dispersion $\varepsilon_{\mathbf{k}}$. The important ingredient is now of course the coupling or hybridization between these two parts, marked by the matrix element $V_{\mathbf{k}d}$. This model is usually suited to describe isolated rather well-localized d - or f -levels within a metallic host. A more stringent introduction is e.g. provided in reference [16], here we only try to sketch the main features of the Hamiltonian.

The atomic states of the isolated impurity are the same as the ones depicted in Fig. 1. Let us assume again a local filling $n=1$ of the impurity ion. Upon hybridization with the conduction sea, the impurity level broadens by Δ . When turning on U one again faces a competing situation: for large U/Δ the single broadened level will split into two with energy separation U , giving rise to a mean-field local moment $m=n_{d\uparrow}-n_{d\downarrow}$. The critical value U_c for this to happen can be calculated in mean field as $U_c=\pi\Delta$. Hence the appearance of local moments for impurity atoms in a host metal depends on the strength of the screened on-site impurity Coulomb interaction as well as on the impurity-host hybridization. The latter may often be drawn already from the DOS of the host system.

Beyond mean field, the Anderson Hamiltonian paves the road towards even deeper physics. Namely, the splitting of the impurity level is such that a resonance at the Fermi level remains due to adiabaticity, in resemblance to what we have seen in the Hubbard model on the lattice in Fig. 3. What we identified there as the quasiparticle excitation on the lattice is now called the *Kondo resonance* and is a bit harder to grasp. The intuitive idea behind it is that when the local moment forms at large U , at some point local charge fluctuations become only virtual and the excitations that remain due to the impurity-host coupling are solely of spin nature. In that sense the Kondo resonance is based on a highly correlated quantum-fluctuating many-body state. In fact it can be shown that there is a well-defined theoretical formalism, the so-called Schrieffer-Wolff transformation (see e.g. [16] for details), that allows to reduce the Anderson Hamiltonian at very large U/Δ to the following spin-interacting form⁷

$$H_{\text{kondo}} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + J \mathbf{s}_{\text{host}} \cdot \mathbf{S}_d , \quad (14)$$

where \mathbf{s}_{host} denotes the spin density due to the conduction electrons at the impurity site. This is the Kondo Hamiltonian named after J. Kondo due to its famous analysis of the relation be-

⁷The formal \mathbf{k} dependence of J is here neglected.

tween local-moment scattering and the resistivity minimum in such metallic impurity-host systems [17]. Note that the spin interaction is antiferromagnetic, i.e., $J > 0$. Unfortunately there is no space in the present scope to discuss the fascinating physics of the Kondo effect in more detail and so we just refer to Coleman's discussion [16] or even more extended surveys like e.g. the book by Hewson [18].

3.2 Multi-band Hubbard Hamiltonians

Let us now come back to the problem of correlated sites building up a periodic lattice (in fact, there are also Kondo lattices, but that's a bit a different story (see, e.g., [6])). There are only a few cases of materials-specific problems where focussing the theoretical discussion on sole single-band properties is then truly sufficient. For instance, many correlated materials studies deal with the $3d$ -shell of transition-metal ions in a certain crystal-field environment. There, as already noted, one expects the splitting of the levels according to symmetry (see Fig. 4). Though it is often legitimate then not to invoke the full five-orbital shell as correlated subspace, but to focus on the three-orbital t_{2g} or the two-orbital e_g manifold. But reducing it to only one effective orbital/band is often too much asked. Although the single band Hubbard model is very important in condensed matter physics, in most cases the hybridizations among orbitals in realistic solids are too strong to single out only one dominating band.

That being said, the extension of the single-band Hubbard model to the multi-band case is in order. As discussed in the last section, the hopping part is not hard to generalize. One only needs to enlarge the hopping matrix by further intra- and inter-orbital entries in the form of eq. (4). On the other hand, the interacting part is not quite that trivial to modify, since it amounts to generalize the quantum-chemical problem to a multi-orbital one in a given crystal field. The first thing to realize is that even when restricting to sole on-site interactions, according to eqs. (3,5) more Coulomb integrals and associated interactions now come into play. Immediately the question arises how to choose/compute the additional Coulomb parameters/functions in a certain crystal field. Since a deeper discussion of this (especially when accounting for the various symmetry issues) is rather involved and because many issues are still current line of research, we choose to discuss only briefly the model cases of e_g , t_{2g} and full p -, d - or f -shell within a cu-

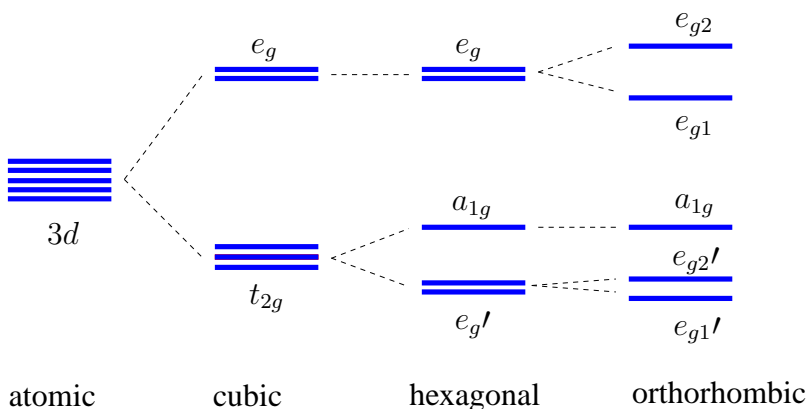


Fig. 4: Splitting of the d -states in different crystal fields.

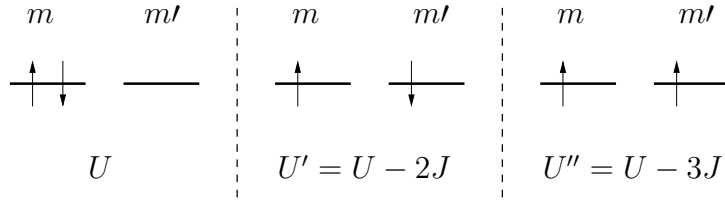


Fig. 5: Local Coulomb interaction between two orbitals with enforcing Hund's first rule.

bic crystal field. Some concrete results for the individual models will thereby be postponed to section 5 where we discuss the slave-boson approach to the numerical solution.

Lets assume first a cubic crystal-field splitting such that the t_{2g} states are significantly lower in energy than the e_g states (cf. Fig. 4). Moreover, a low-spin scenario shall hold, i.e., the lower-lying t_{2g} states first become filled according to Hund's first rule and only after that the e_g levels start to accumulate electrons. Then the filling of eight electrons (e.g. Ni^{2+}), because of the full t_{2g} subshell, poses in a good approximation the problem of two interacting electrons in the two e_g states. In a different case where the filling amounts to only four electrons (e.g. Ru^{4+}), one deals with a three-orbital problem composed of the t_{2g} manifold. The appropriate interacting lattice problem may be formulated in both cases as a generalized rotationally invariant multi-band Hubbard Hamiltonian through [20–22]

$$\begin{aligned}
 H_{\text{cub}} = & \sum_{ij,mm',\sigma} t_{ij}^{mm'} c_{im\sigma}^\dagger c_{jm'\sigma} + U \sum_{im} n_{im\uparrow} n_{im\downarrow} + \frac{1}{2} \sum_{i,m \neq m',\sigma} [U' n_{im\sigma} n_{im'\bar{\sigma}} + U'' n_{im\sigma} n_{im'\sigma}] \\
 & + \frac{1}{2} \sum_{m \neq m',\sigma} \left[J c_{im\sigma}^\dagger c_{im'\bar{\sigma}}^\dagger c_{im\bar{\sigma}} c_{im'\sigma} + J_C c_{im\sigma}^\dagger c_{im'\bar{\sigma}}^\dagger c_{im'\bar{\sigma}} c_{im'\sigma} \right] . \quad (15)
 \end{aligned}$$

It is seen that the complexity has quite increased compared to the single-band form (6), mainly due to the necessity of including the on-site exchange integral J . At this point its instructive to compare the exact expression for U and J derived from the general form (5).

$$U_{mm'} = \int d\mathbf{r} d\mathbf{r}' \varphi_{\mathbf{R}m\sigma}^*(\mathbf{r}) \varphi_{\mathbf{R}m'\sigma'}^*(\mathbf{r}') \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \varphi_{\mathbf{R}m'\sigma'}(\mathbf{r}') \varphi_{\mathbf{R}m\sigma}(\mathbf{r}) , \quad (16)$$

$$J_{mm'} = \int d\mathbf{r} d\mathbf{r}' \varphi_{\mathbf{R}m\sigma}^*(\mathbf{r}) \varphi_{\mathbf{R}m'\sigma'}^*(\mathbf{r}') \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \varphi_{\mathbf{R}m\sigma'}(\mathbf{r}') \varphi_{\mathbf{R}m'\sigma}(\mathbf{r}) . \quad (17)$$

Thus in principle these two Coulomb integrals are matrices that depend on the orbital indices m, m' . In a model spirit it proves however sufficient for the present symmetry to perform an orbital-independent parametrization, yet differentiating between intra- and inter-orbital terms. The intra-orbital Coulomb interaction shall be given by U and the inter-orbital Coulomb interactions by $U'=U-2J$ for different spin and $U''=U-3J$ for identical spin. Obviously, therewith Hund's first rule is correctly incorporated (see Fig. 5). In eq. (15) the terms in U , U' and U'' are *density-density interactions*, i.e., they can be written as products of particle-number operators. The remaining two terms in the second line of eq. (15) may not be written in this form, but are important to establish the full orbital rotational invariance of the Hamiltonian. These interactions are related to spin-flip and pair-hopping processes as can be verified by inspection. In

principle, the pair-hopping Coulomb integral J_C may be different than the one for the spin-flip term. However usually both interaction integrals are chosen to be equally identified with the exchange integral J . Remember again that the form of the model Coulomb integrals in eq. (15) is usually tailored to cubic symmetry. However in practice this structure for the multi-band Hamiltonian is also successfully applied in lower-symmetry cases (e.g. [23]).

Note that the given multi-band Hamiltonian may for the three-orbital t_{2g} problem also be written in a different way (but mathematical equivalent form) as

$$H_{t_{2g}} = \sum_{ij,mm',\sigma} t_{ij}^{mm'} c_{im\sigma}^\dagger c_{jm'\sigma} + \frac{(U-3J)}{2} N(N-1) + \frac{5}{2} JN - 2JS^2 - \frac{1}{2} JL^2 \quad , \quad (18)$$

with the total operators for particle number N , spin S and angular momentum L . This form is furthermore most often used in the context of a correlated p shell.

In the case of a full d - or f -shell there are even more complicated terms appearing in the derivation of the atomic Hamiltonian. In order to again establish complete rotational invariance, one best relies on the general form for local interactions

$$H_d = \sum_{ij,mm',\sigma} t_{ij}^{mm'} c_{im\sigma}^\dagger c_{jm'\sigma} + \frac{1}{2} \sum_{i,mm'm''m''',\sigma\sigma'} U_{mm'm''m'''} c_{im\sigma}^\dagger c_{im'\sigma'}^\dagger c_{im''\sigma''} c_{im'''\sigma'''} \quad . \quad (19)$$

The interaction matrix element can then be evaluated through a multipole expansion into effective Slater integrals F_k via

$$U_{mm'm''m'''} = \langle mm' | V_{ee}^{\text{loc}} | m''m''' \rangle = \sum_{k=0} a_k(m, m', m'', m''') F_k \quad . \quad (20)$$

In a spherical approximation, only a finite number of Slater-integral terms form the sum and those are related to U and J . Then $U=F_0$ always holds and the further relations for the different ($l>0$)-shells are usually chosen as follows

$$\begin{aligned} l = 1 & : J = \frac{1}{5} F_2 \quad , \\ l = 2 & : J = \frac{1}{14} (F_2 + F_4) \quad , \quad F_4 = 0.625 F_2 \quad , \\ l = 3 & : J = \frac{1}{6435} (286 F_2 + 195 F_4 + 250 F_6) \quad , \quad F_4 = 0.668 F_2 \quad , \quad F_6 = 0.494 F_2 \quad . \end{aligned} \quad (21)$$

Note that in current research there is nowadays the option to directly compute the necessary screened Coulomb integrals for a certain material system from approximated first-principles schemes, but this is beyond the scope of the present chapter.

Without going into the details of the multi-orbital models properties, lets just note that in the fully degenerate case the simple Mott scenario still applies, i.e., at integer filling the multi-band system can undergo a Mott transition. This means that e.g. for a three-band problem one may have an MIT for $n=1,2$.

3.3 More interactions

So far we only dealt with local Coulomb interactions in our model Hamiltonians, which may be indeed sufficient in the context of this school and as a good starting point. However of course, there are also limits to this kind of interaction, even when working in a multi-orbital scenario. Since nature does not care so much about the taste and wellness of theoretical physicists, eventually there are materials problems that ask for more.

For instance, one may easily think of problems where inter-site Coulomb interactions are not negligible anymore. In low-dimensional systems, especially quasi-1D compounds, the NN Coulomb integral V is often relevant at specific fillings to drive e.g. charge-density-wave (CDW) or spin-density-wave (SDW) instabilities [24]. The *extended Hubbard Hamiltonians* that are used for such modelings look like

$$H_{\text{hubex}} = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \varepsilon_0 \sum_{i\sigma} n_{i\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} + \sum_{ij\sigma\sigma'} V_{ij} n_{i\sigma} n_{j\sigma'} . \quad (22)$$

Most often the density-density interaction associated with V_{ij} is restricted to NN, but in principle the treatment of long-range Coulomb terms is possible.

Another action, retrieved from the collection of all possible pair interactions (see e.g. [6]) that may be relevant in some materials cases [25] is the so-called *correlated hopping* process with the interaction integral X . The associated Hamiltonian term has the following basic structure

$$H_{\text{ch}} = \sum_{ij\sigma} X_{ij} \left(c_{i\sigma}^\dagger c_{j\sigma} + c_{j\sigma}^\dagger c_{i\sigma} \right) (n_{i\bar{\sigma}} + n_{j\bar{\sigma}}) . \quad (23)$$

Since the term within the first braces has the form of a bond-charge operator when summed over σ , the Hamiltonian H_{ch} may be interpreted as a bond-charge-site-charge repulsion [6].

Finally, the reserach area of model Hamiltonians also opens a route to go beyond the initial Born-Oppenheimer approximation. Remember that the BOA was introduced to decouple electronic and lattice degrees of freedom. However in many materials (e.g. manganites [26]) the electron-phonon coupling is significant and the BOA breaks down. Then there is the possibility to formulate a simplified Fröhlich Hamiltonian in real space using further only Einstein phonons, the so-called *Holstein Hamiltonian* [27]

$$H_{\text{hol}} = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} - g \sum_{i\sigma} n_{i\sigma} (b_i^\dagger + b_i) + \omega \sum_i b_i^\dagger b_i , \quad (24)$$

where $b^{(\dagger)}$ are bosonic creation/annihilation operators for the phonons, ω the Einstein-mode frequency and g the electron-phonon coupling strength.

There are many more model Hamiltonian approaches to specific microscopic processes. Note that for instance spin-orbit coupling was always excluded in our considerations, though the interplay of this phenomenon with Coulomb correlations is becoming a very active line of theoretical research (e.g. [19]). However this would here go beyond the scope of an elementary introduction to the field. We therefore now end the first part of this chapter of introducing basic model hamiltonians and turn to an, again brief, account of simpler theoretical methodolgies to actually put us in position to compute some of the discussed models properties.

4 Basic approaches to the single-band Hubbard model

In this first section of techniques to approach Hubbard-like models we deal with two traditional methodologies. We first discuss the standard mean-field framework of Hartree-Fock in the present model context. The second subsection introduces the simplest Green's-function based ansatz to the problem of locally interacting electrons, namely the so-called Hubbard-I approach. There are numerous excellent reviews of these basic approaches, e.g. [4,6,10], and it's somehow a thankless task to add something to that in this short overview. Thus we will not challenge to confuse the reader by trying something terribly fancy but instead provide the necessary information in a nutshell along the lines of the already existing literature.

4.1 Hartree-Fock

When facing an interacting problem it is usually a very good idea to start with a mean-field approach, since it is simple but, importantly, non-trivial. Albeit nowadays many rather sophisticated techniques are available, one should never forget about the power and successes of mean-field (MF) theory.

Remember your quantum-mechanics class on many-particle wave functions in first quantization. Back then the simplest idea was to assume the full wave function may be decoupled and represented as a product of single-particle wave functions. After inserting in the Schrödinger equation, in the end every individual particle is moving in the *mean-field* build up by the other ones. Let us try to translate and apply this idea to the single-band Hubbard model in the form

$$H = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \quad . \quad (25)$$

Instead of applying a decoupling on the *states*, it is in the present context more efficient to decouple already on the *operator* level. Therefore we write the particle-number operator as

$$n_{i\sigma} = \langle n_{i\sigma} \rangle + \delta n_{i\sigma} \quad , \quad (26)$$

which means that there is a bulk part of $n_{i\sigma}$, its expectation value and hence a *c*-number, that accounts for the major observable physics. And a smaller part $\delta n_{i\sigma}$ that carries the still essential quantum-fluctuating nature around that. Neglecting the latter would be a bad idea, because this would lead to rather trivial results. Better write the interaction kernel therewith now as

$$n_{i\uparrow} n_{i\downarrow} = \langle n_{i\uparrow} \rangle \langle n_{i\downarrow} \rangle + \langle n_{i\downarrow} \rangle \delta n_{i\uparrow} + \langle n_{i\uparrow} \rangle \delta n_{i\downarrow} + \delta n_{i\uparrow} \delta n_{i\downarrow} =: A(n_{i\uparrow} n_{i\downarrow}) + \delta n_{i\uparrow} \delta n_{i\downarrow} \quad . \quad (27)$$

A similar analytical structure is obtained, i.e., the kernel separates into a, supposingly, bulky part and a smaller product of the fluctuations. Now we perform the approximation and neglect the *correlation of fluctuations* $\delta n_{i\uparrow} \delta n_{i\downarrow}$, which is at the heart of MF theory on the microscopic level. By eliminating $\delta n_{i\sigma}$ the Hubbard term hence reads

$$U \sum_i n_{i\uparrow} n_{i\downarrow} \approx U \sum_i A(n_{i\uparrow} n_{i\downarrow}) = U \sum_i (n_{i\uparrow} \langle n_{i\downarrow} \rangle + n_{i\downarrow} \langle n_{i\uparrow} \rangle - \langle n_{i\uparrow} \rangle \langle n_{i\downarrow} \rangle) \quad . \quad (28)$$

So within this approximation the tough interaction part looks rather harmless, merely equal to the particle-number operator with a site- and spin-dependent mean field. However before even examining details of this MF approach, one easily sees that something dramatic has happened to symmetry. The original spin rotational invariance of the Hubbard model is gone, since the MF interacting part can be rewritten using only the z component of the spin operator, i.e., $S_i^z = 1/2(n_{i\uparrow} - n_{i\downarrow})$. It is in fact a very important issue to understand that even if one uses the (\uparrow, \downarrow) spin-projection basis in Fock space for the operator representation, the Hubbard term is still spin-rotational invariant. In some sense, our present MF approach shows the Hartree term but lacks its Fock companion.

The problem is that in eq. (27) we have unconsciously made a particular choice for decoupling the Hubbard term into a sum of simpler terms by singling out the particle-number operators. But the Hubbard model is more clever than that. In fact based on Wick's theorem a certain composition of operators can always be decoupled by forming all possible creation-annihilation pairs. Thus here

$$\begin{aligned} n_{i\uparrow}n_{i\downarrow} = c_{i\uparrow}^\dagger c_{i\uparrow} c_{i\downarrow}^\dagger c_{i\downarrow} = -c_{i\uparrow}^\dagger c_{i\downarrow}^\dagger c_{i\uparrow} c_{i\downarrow} \rightarrow & -\langle c_{i\uparrow}^\dagger c_{i\downarrow} \rangle c_{i\downarrow}^\dagger c_{i\uparrow} - \langle c_{i\downarrow}^\dagger c_{i\uparrow} \rangle c_{i\uparrow}^\dagger c_{i\downarrow} \\ & + \langle c_{i\uparrow}^\dagger c_{i\uparrow} \rangle c_{i\downarrow}^\dagger c_{i\downarrow} + \langle c_{i\downarrow}^\dagger c_{i\downarrow} \rangle c_{i\uparrow}^\dagger c_{i\uparrow} \\ & + \langle c_{i\uparrow}^\dagger c_{i\downarrow} \rangle \langle c_{i\downarrow}^\dagger c_{i\uparrow} \rangle - \langle c_{i\uparrow}^\dagger c_{i\uparrow} \rangle \langle c_{i\downarrow}^\dagger c_{i\downarrow} \rangle \quad . \end{aligned} \quad (29)$$

We easily identify the bulky A part of eq. (27) in there, but there is obviously a similar part stemming from the combinations $c_{i\downarrow}^\dagger c_{i\uparrow} =: S_i^+$ and $c_{i\downarrow}^\dagger c_{i\uparrow} =: S_i^-$. However this is just what we were looking for, as the spin-ladder operators should be capable to restore the spin-rotational invariance. Indeed bringing the former Hartree and this new Fock part together, the Hartree-Fock (HF) approximation to the Hubbard model can be cast into [6]

$$H_{\text{hub}}^{\text{HF}} = -t \sum_{\langle ij \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + \frac{U}{2} \sum_i \left\{ n_i \langle n_i \rangle - 4 \mathbf{S}_i \langle \mathbf{S}_i \rangle - \frac{1}{2} \langle n_i \rangle^2 - 2 \langle \mathbf{S}_i \rangle^2 \right\} \quad , \quad (30)$$

with $n_i = \sum_\sigma n_{i\sigma}$. At a first glance the interaction part looks more complicated, but note that it is only a single-particle term with a somewhat more sophisticated mean field. The approximate dispersion $\varepsilon_{\mathbf{k}\sigma}^{\text{HF}}$ with interaction is readily computed from (30) by e.g. choosing \hat{e}_z as the quantization axis, transforming to \mathbf{k} -space and using the identity $\sum_i n_{i\sigma} = \sum_{\mathbf{k}} n_{\mathbf{k}\sigma}$, which eventually leads to

$$\varepsilon_{\mathbf{k}\sigma}^{\text{HF}} := \frac{\partial E_{\text{hub}}^{\text{HF}}}{\partial \langle n_{\mathbf{k}\sigma} \rangle} = \frac{\partial \langle H_{\text{hub}}^{\text{HF}} \rangle}{\partial \langle n_{\mathbf{k}\sigma} \rangle} = \begin{cases} \varepsilon_{\mathbf{k}} + U \left(\frac{n}{2} - m \right) & \text{for } \sigma = \uparrow \\ \varepsilon_{\mathbf{k}} + U \left(\frac{n}{2} + m \right) & \text{for } \sigma = \downarrow \end{cases} \quad , \quad (31)$$

with $m = 1/2(\langle n_{i\uparrow} \rangle - \langle n_{i\downarrow} \rangle)$. Thus Hartree-Fock reveals the expected *exchange splitting* between spin-polarized bands in the ferromagnetic phase. Though HF is a weak-coupling approach to the Coulomb-interacting problem⁸, due to the lack of screening it is ill-defined for metals (see e.g. [11]). In fact in some sense it is suited for *long-range ordered* Mott insulators, when

⁸HF only accounts for the first-order terms in the proper many-body diagrammatic perturbation theory for the interacting electron system (see e.g. [33]).

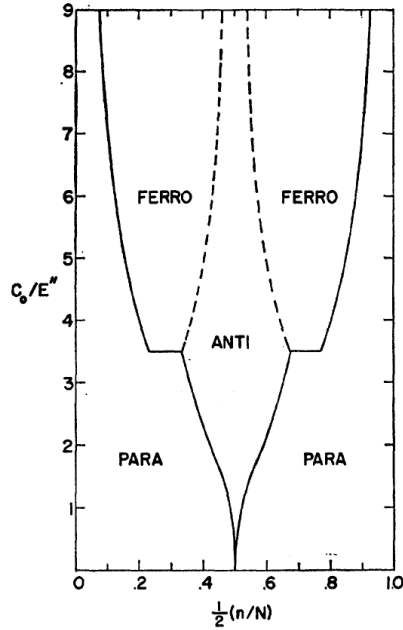


Fig. 6: Normalized U - n Hartree-Fock phase diagram of the 3D Hubbard model on a simple cubic lattice taking paramagnetic, ferromagnetic and antiferromagnetic phases into account. Solid/Dashed lines indicate continuous/first order phase transitions. From [29].

the charge gap can be described in the HF approximation via an U -driven exchange gap (cf. eq. (31)). The so-called *LDA+ U method* [28] makes explicit use of that in the context of first-principles calculations.

Equipped with this approximation one can nonetheless then start to study the competition between different phases like CDW, SDW, etc. But it is important to realize that the HF approximation may not account for explicit many-body effects such as e.g. band narrowing, finite lifetimes, or even a clear notion of Hubbard bands in the spectral function.

A magnetic phase diagram with taking NN collinear orderings into account is shown in Fig. 6. Non-surprisingly increasing U seems to trigger magnetic order. As expected close to half filling and larger U an antiferromagnetic phase is stabilized due to superexchange. However a closer inspection shows that for $U \rightarrow \infty$ there will be competition also with the ferromagnetic phase. This is condensed in the Nagaoka theorem [30], stating that on many lattices at $U \rightarrow \infty$ the state with a single hole is indeed ferromagnetic. The whole issue of stabilizing ferromagnetic order in the Hubbard model is in fact far from being trivial (see e.g. [31, 6] for a discussion). Note that since especially close to phase transitions the MF approximation of neglecting the correlation of fluctuations can be dangerous, one should keep in mind that there may be various changes to the HF phase diagram when using a more elaborate technique (as e.g. in [32]). Furthermore, incommensurate magnetic phases not displayed in Fig. 6 also have to be considered.

At this point some readers may still show some scepticism about the uniqueness of the decoupling exhibited in eq. (29). For instance one knows from the solid-state physics lecture that a metal may be unstable against pairing, i.e., forming a superconducting state. Obviously the shown decoupling is not capable of supporting such a phase, yet there is no a priori reason why the full Hubbard model should display the same insufficiency. And indeed, there are still more ways of decoupling the inconspicuous Hubbard term, e.g. allowing also for a superconducting amplitude (see [6]). Just think of the possibility of transforming the electron operators into

other bases and then redoing the decoupling. It is in the end once more some kind of “magic” of quantum mechanics that the simple-looking Hubbard model bears so much physics in it.

4.2 Hubbard I

At the introduction of the Hartree-Fock method we made reference to the derivation in the wave-function picture. Besides directly manipulating the operator terms as in the last section, one can also make direct use of other kind of representations in many-body physics, namely *many-particle Green's functions*⁹. The formal definition of the one-particle Green's function is given by

$$G_\sigma(\mathbf{R}_\alpha, t; \mathbf{R}_{\alpha'}, t') = -i \langle \Psi_0 | c_{\mathbf{R}_\alpha \sigma}(t) c_{\mathbf{R}_{\alpha'} \sigma}^\dagger(t') | \Psi_0 \rangle \quad , \quad (32)$$

where the Heisenberg representation of the fermionic operators is important. The physical picture behind it is in principle simple. This function takes notes of the life of a created electron at spacetime $(\mathbf{R}_{\alpha'}, t')$ on top of the many-particle ground state $|\Psi_0\rangle$ until its death at spacetime (\mathbf{R}_α, t) , asking for the probability to end up in the very same $|\Psi_0\rangle$. In a periodic solid its of course useful to Fourier transform this function to (\mathbf{k}, ω) -space by invoking translational invariance and the explicit time independence of our Hamiltonian. The recorded information can then be read out in spectral form via the Lehmann representation [34]

$$G_\sigma(\mathbf{k}, \omega) = \sum_m \frac{|\langle \Psi_m^{(N_e+1)} | c_{\mathbf{k}\sigma}^\dagger | \Psi_0 \rangle|^2}{\omega + \mu - \omega_{m0} + i\eta} + \sum_m \frac{|\langle \Psi_m^{(N_e-1)} | c_{\mathbf{k}\sigma} | \Psi_0 \rangle|^2}{\omega + \mu - \omega_{0m} - i\eta} \quad , \quad (33)$$

where μ denotes the chemical potential. We can benefit from having introduced already the spectral function $A_\sigma(\mathbf{k}, \omega)$ in eq. (7) and write this in a more condensed form as an integral¹⁰

$$G_\sigma(\mathbf{k}, \omega) = \int_{-\infty}^{\infty} d\omega' \frac{A_\sigma(\mathbf{k}, \omega')}{\omega + \mu - \omega' + \text{sgn}(\omega')i0^+} \quad . \quad (34)$$

For instance, taking the Fermi-gas limit from section 2.3, one immediately obtains

$$G_\sigma^{\text{FG}}(\mathbf{k}, \omega) = \int_{-\infty}^{\infty} d\omega' \frac{\delta(\omega' - \varepsilon_{\mathbf{k}})}{\omega + \mu - \omega' + \text{sgn}(\omega')i0^+} = \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}} + i\eta_{\mathbf{k}}} \quad . \quad (35)$$

with $\eta_{\mathbf{k}} = \text{sgn}(|\mathbf{k}| - k_F)$. The power of the one-particle Green's function is build on the fact that it has complete record of the one-particle correlations of the system and allows, e.g., also to compute the expectation value of any single-particle operator, such as the total energy. In general it can be compactly rewritten as

$$G_\sigma(\mathbf{k}, \omega) = \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}} - \Sigma_\sigma(\mathbf{k}, \omega)} \quad , \quad (36)$$

where the *self-energy* $\Sigma_\sigma(\mathbf{k}, \omega)$ carries all deviations from the Fermi-gas limit due to Coulomb interactions¹¹. Note that the self-energy is a true complex function and accounts therefore also for lifetime effects (e.g. such as we pointed out in section 2.3).

⁹An excellent introduction to this framework of theoretical many-body physics is e.g. given in [33].

¹⁰Note that going to an integral representation is only allowed in the large- N_e limit since it bears deeper physics about the excitations (see appendix H of [33]).

¹¹The Pauli principle is usually enforced by hand in the Fermi gas, but Σ includes also the Fock term.

Lets try to compute also the Green's function in the atomic limit of the Hubbard model. Since the problem is separable we can limit the concentration on the local part of the Green's function, i.e., $G_{i\sigma}(\omega) = \sum_{\mathbf{k}} G_{\sigma}(\mathbf{k}, \omega)$ and insert in (34) the result (9) to reveal

$$\begin{aligned} G_{i\sigma}(\omega) &= \int_{-\infty}^0 d\omega' \frac{(1 - \langle n_{i\bar{\sigma}} \rangle) \delta(\omega' - \varepsilon_0)}{\omega + \mu - \omega' - i0^+} + \int_0^{\infty} d\omega' \frac{\langle n_{i\bar{\sigma}} \rangle \delta(\omega' - (\varepsilon_0 + U))}{\omega + \mu - \omega' + i0^+} \\ &= \frac{1 - \langle n_{i\bar{\sigma}} \rangle}{\omega + \mu - \varepsilon_0 - i0^+} + \frac{\langle n_{i\bar{\sigma}} \rangle}{\omega + \mu - \varepsilon_0 - U + i0^+} . \end{aligned} \quad (37)$$

Thus the former δ -Peak excitations are easily identified as the poles of the atomic Green's function. In order to extract the corresponding self-energy we have to bring it in the form of eq. (36). For simplicity we assume once again $\varepsilon_0=0$ and aim at the analytic structure

$$G_{i\sigma}(\omega) = \frac{1}{\omega + \mu - \Sigma_{i\sigma}^{\text{atom}}(\omega)} , \quad (38)$$

since no k-dependence should emerge. One readily realizes that the following form for the self-energy does the job:

$$\Sigma_{i\sigma}^{\text{atom}}(\omega) = U n_{i\bar{\sigma}} + U^2 \frac{n_{i\bar{\sigma}}(1 - n_{i\bar{\sigma}})}{\omega + \mu - U(1 - n_{i\bar{\sigma}})} . \quad (39)$$

Albeit somewhat artificial, this representation opens the door to a first true many-body approximation compared to the effective single-particle HF treatment in mean field discussed in the last section. Because we can now make the radical assumption that the *true* self-energy of the solid shall have the analytical form of the atomic self-energy (39). This defines the so-called Hubbard-I (HI) approximation to the lattice Green's function, introduced in Hubbard's original paper [9], written as

$$G_{\sigma}^{\text{HI}}(\mathbf{k}, \omega) = \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}} - \Sigma_{i\sigma}^{\text{atom}}(\omega)} . \quad (40)$$

As usual the poles of $G_{\sigma}^{\text{HI}}(\mathbf{k}, \omega)$ define the excitations. Inserting (39) in (40) and computing the roots of the denominator yields

$$\varepsilon_{\mathbf{k}\sigma}^{\text{HI}} = \frac{1}{2} \left\{ \varepsilon_{\mathbf{k}} + U \vee \pm \sqrt{(\varepsilon_{\mathbf{k}} + U)^2 + 4U \langle n_{i\bar{\sigma}} \rangle} \right\} . \quad (41)$$

There are two solutions for each spin projection since there is, as already discussed in section 2.3, an upper and a lower Hubbard band. Hence whereas HF somehow approximates (with severe deficiencies) around the Fermi-gas limit, Hubbard I approximates around the atomic limit. Note that in HI the explicit Hubbard-band dispersion has k-dependent spectral weight $A_{\sigma}(\mathbf{k}, \omega)$, meaning that $\langle \Psi_{(m0)} | c_{\mathbf{k}\sigma}^{(\dagger)} | \Psi_{(m0)} \rangle$ now differs from one. This can be seen by bringing $G_{\sigma}^{\text{HI}}(\mathbf{k}, \omega)$ in Lehmann form. But still, because the self-energy is a pure real object, no finite-lifetime effects are addressed in HI.

There are several further pros and cons of the HI approximation, for details see e.g. [10]. It is exact in the atomic limit as well as in the Fermi-gas limit (easily seen from eq. (41)), however it does not match the HF solution for small U . Furthermore perhaps the most important drawback

comes from the violation of particle-hole symmetry, which can be traced back to the use of *one* self-energy structure for the *two* Hubbard bands. In the end it is perhaps more the spirit of the approximation that renders the HI form rather relevant. One could think of a better approximate form of the local $\Sigma_{i\sigma}(\omega)$ in order to improve the method. This thought was indeed successful and lead finally to the development of DMFT.

5 Slave-boson approach

In the last section of this brief introduction we aim to improve on the HF approximation by studying the true quasiparticle excitations in the strongly correlated metallic regime. This shall be done within a still simple but rather efficient technique with the politically incorrect naming slave-boson method. That approach shares its roots with the so-called *Gutzwiller approach* and both formalisms may in many cases be transformed into each other. However in the present text we will restrict the discussion to the slave bosons and provide references to some common and related initial work in this field [35–38]. More details on the nature of the performed approximation to the correlated problem may be also found in the review by Vollhardt [39] and references therein. Once again, the following presentation will be short and far from complete. It should mainly give the reader a first taste of the method.

5.1 Infinite- U limit

The general idea behind the Gutzwiller and the slave-boson method is given by the fact that electronic correlations impose certain constraints on the Hilbert space of available states of the problem. For instance its quite clear that in the context of the Hubbard model, double occupations on the lattice sites are severely suppressed at large U . Use of this was already made when writing down the t - J Hamiltonian for the limit $U/t \gg 1$ in section 2.3, where we explicitly projected onto (empty, single) occupied sites. However instead of “hard-coding” this effect in a new Hamiltonian form for a certain limit, one can also implement this physics in a more flexible way, giving rise to a new methodology for solving the actual Hubbard Hamiltonian.

In order to understand the basic principles of the approach it is instructive to first have look at the dispersive part of the single-band Hubbard model at $U \rightarrow \infty$, which we can write as

$$H = -t \sum_{\langle ij \rangle \sigma} P c_{i\sigma}^\dagger c_{j\sigma} P \quad , \quad (42)$$

whereby the projection P excludes the double occupied states $|\uparrow\downarrow\rangle$, leaving only the states $\{|0\rangle, |\sigma\rangle\}$ locally available (cf. Fig. 1). The question of course arises how to actually enforce this projection over the whole lattice in the calculation. In principle one has to demand that $\sum_\sigma n_{i\sigma} < 2$ on each site i . Recalling the standard lecture on Lagrange multipliers we however know that a constraint is usually best imposed via an *equality relation*. This can be formally achieved by introducing new auxiliary quantum degrees of freedom $\phi_i^{(\dagger)}$ such that we can de-

compose the original operators via

$$c_{i\sigma}^\dagger = f_{i\sigma}^\dagger \phi_i \quad \wedge \quad c_{i\sigma} = f_{i\sigma} \phi_i^\dagger \quad . \quad (43)$$

Since the $f_{i\sigma}^{(\dagger)}$ operators are still carrying the fermionic character, quantum mechanics teaches us that the $\phi_i^{(\dagger)}$ have to be of bosonic kind. The new operators are defined by their action and the corresponding particle-number-site statistics, i.e.,

$$\begin{aligned} \phi_i^\dagger |\text{vac}\rangle &= |0_i\rangle \quad , & |0_i\rangle &: n_b = 1 \quad \wedge \quad n_{f\sigma} = 0 \quad , \\ f_{i\sigma}^\dagger |\text{vac}\rangle &= |\sigma_i\rangle & |\uparrow_i\rangle &: n_b = 0 \quad \wedge \quad n_{f\uparrow} = 1 \quad , \\ & & |\downarrow_i\rangle &: n_b = 0 \quad \wedge \quad n_{f\downarrow} = 1 \quad , \end{aligned} \quad (44)$$

where $n_{f,b}$ denote the respective site occupation numbers and $|\text{vac}\rangle$ marks the vacuum state. Hence we now can truly formulate the necessary constraint on each lattice site i in the limit $U \rightarrow \infty$ as

$$\sum_{\sigma} f_{i\sigma}^\dagger f_{i\sigma} + \phi_i^\dagger \phi_i = 1 =: \mathcal{Q} \quad (45)$$

and using the relations (43) we can directly choose to express the Hamiltonian form (42) in that limit via

$$H = -t \sum_{\langle ij \rangle \sigma} \phi_i \phi_j^\dagger f_{i\sigma}^\dagger f_{j\sigma} \quad . \quad (46)$$

In a simple picturing the decomposition of the physical electron operator $c_{i\sigma}^{(\dagger)}$ amounts here to a fragmentation into low-energy quasiparticle and high-energy Hubbard excitations on the operator level. The original operator takes care of both, the itinerant and the localized character of the electron, while $f_{i\sigma}^{(\dagger)}$ carries the sole quasiparticle part and the slave boson $\phi_i^{(\dagger)}$ only the high-energy remainings. Loosely speaking, the slave boson “releases” the electron from its high-energy excitations, which gives the political incorrectness a mild spin. However in detail things are truly a bit more complicated. As we have learned already in the HF subsection, such decouplings in quantum mechanics are seldom unique. In fact there is a gauge symmetry group associated with the redundancy of representing $c_{i\sigma}^{(\dagger)}$ via slave-boson techniques. The former manifests itself in the conservation of the pseudo charge \mathcal{Q} defined in the constraint (45), thereby generating invariance under the group of local $U(1)$ gauge transformations (see e.g. [40, 13] and references therein for further details). This issue is interesting but drives us at present a bit away from our simple goal of getting first concrete results for the Hubbard model, so let's get back to this.

Enforcing the constraint (45) on each lattice site individually is too tough, but in order to proceed it is a good idea to perform a MF approximation by condensing the bosons and averaging the constraint over all sites. We can then write

$$r := \langle \phi_i \rangle \quad , \quad \sum_{\sigma} \langle n_{f\sigma} \rangle + |r|^2 = 1 \quad , \quad H_{\text{eff}} = -|r|^2 t \sum_{\langle ij \rangle \sigma} f_{i\sigma}^\dagger f_{j\sigma} \quad , \quad (47)$$

with H_{eff} as the effective Hamiltonian in this slave-boson mean-field (SBMF) theory. In a functional-integral representation of the problem this approximation amounts to a saddle-point

approximation and therefore resembles the MF concept. Note that since $|r|^2=1-\langle n_f \rangle$ it also equals the doping δ away from half filling. Therewith one can define an effective hopping $t_{\text{eff}}=\delta t$ which provides readily the important Brinkman-Rice effect, namely that the quasiparticles become heavy (i.e., show small hopping amplitude) at small doping δ .

The ground state for finite chemical potential can then be found from minimizing the grand potential per lattice site

$$\Omega = \langle H_{\text{eff}} \rangle + \lambda \left(\sum_{\sigma} \langle n_{f\sigma} \rangle + |r|^2 - 1 \right) - \mu \sum_{\sigma} \langle n_{f\sigma} \rangle \quad (48)$$

with respect to the lagrange multiplier λ . In some sense the nature of the slave-boson technique is first to enlarge the Hilbert space by introducing additional bosonic degrees of freedom, which translates secondly in an enhancement of the variational freedom to select the actual physical states.

The quasiparticle dispersion is finally obtained from (48) via

$$\varepsilon_{\mathbf{k}\sigma}^{\text{SBMF}} := \frac{\partial E^{\text{SBMF}}}{\partial \langle n_{\mathbf{k}\sigma}^f \rangle} = |r|^2 \varepsilon_{\mathbf{k}\sigma} + \lambda \quad (49)$$

and the Green's function of the non-interacting fermionic quasiparticles hence reads

$$G_f(\mathbf{k}, \omega) = \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}\sigma}^{\text{SBMF}}} = \frac{1}{\omega + \mu - |r|^2 \varepsilon_{\mathbf{k}} - \lambda} \quad (50)$$

However the one-particle Green's function of the true physical electrons results from inserting (43) in the definition (32) and using the MF approximation (47), i.e.,

$$G^{\text{SBMF}}(\mathbf{k}, \omega) = |r|^2 G_f(\mathbf{k}, \omega) = \frac{|r|^2}{\omega + \mu - |r|^2 \varepsilon_{\mathbf{k}} - \lambda} =: \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}} - \Sigma^{\text{SBMF}}(\omega)} \quad (51)$$

The so defined local SBMF self-energy is then given by

$$\Sigma^{\text{SBMF}}(\omega) = \omega \left(1 - \frac{1}{|r|^2} \right) + \mu - \frac{\mu - \lambda}{|r|^2} \quad , \quad (52)$$

consisting of term linear in frequency ω and a static part. The latter accounts for a shift of the low-energy excitations and the former for the proper band renormalization. In fact $|r|^2$ is often named the quasiparticle weight Z , as for $Z=1$ one retrieves the Fermi-gas limit and $Z=0$ marks the vanishing of quasiparticle excitations. A value $0 < Z < 1$ therefore characterizes the so-called *Fermi-liquid regime*. As in HF, note that the Hubbard excitations do not appear in the associated spectral function, because we condensed the slave bosons. In principle Hubbard bands can be gained from treating fluctuations around the saddle-point, but this is a rather tricky issue (see e.g. [41]).

Although the slave-boson technique is approximate, we see that with already modest effort it can account for explicit many-body effects like the band-narrowing close to the Fermi level. But so far the method is not very useful for too many practical concerns, because we only dealt with the rather specific $U \rightarrow \infty$ case, mainly appropriate in the Kondo regime. However we would not have introduced the method here if there wasn't the possibility to extent it to the finite- U case, to be discussed next.

5.2 Kotliar-Ruckenstein representation

There are various options to extend the slave-boson method to finite U . Two prominent realizations are the Kotliar-Ruckenstein (KR) representation [42] and the slave-rotor formalism [43]. In the following we briefly sketch the former methodology.

Somehow it is quite clear what is needed, since at finite U double occupations are of course still accessible one has to the extent the number of bosonic degrees of freedom in order to variationally cope with the enlarged number of lattice states. To study this without getting immediately lost in too many details, let us concentrate first on the slave-boson treatment of the simple atomic problem of a single correlated orbital [44]. The Hamiltonian for this Hubbard atom reads

$$H_{\text{loc}} = \varepsilon_0 \sum_{\sigma} c_{\sigma}^{\dagger} c_{\sigma} + U c_{\uparrow}^{\dagger} c_{\uparrow} c_{\downarrow}^{\dagger} c_{\downarrow} \quad . \quad (53)$$

There are the already familiar four atomic states $\Gamma = \{|0\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\}$ available for occupation. In the KR formalism one simply associates with *each* state a specific pair of slave-boson operators $\phi^{(\dagger)}$, i.e.,

$$\{|0\rangle, |\uparrow\rangle, |\downarrow\rangle, |\uparrow\downarrow\rangle\} \rightarrow \{\phi_0^{(\dagger)}, \phi_{\uparrow}^{(\dagger)}, \phi_{\downarrow}^{(\dagger)}, \phi_{\uparrow\downarrow}^{(\dagger)}\} \quad . \quad (54)$$

Since we are at finite U there is now no unique constraint on the actual occupation. However there are *two* new types of constraints which ensure that by enlarging the Hilbert space we truly recover in the calculation the physical electronic states based on the original $c_{\sigma}^{(\dagger)}$ operators. As there is no Pauli principle for bosons, we first have to demand that we only care about states with only a single slave boson, namely

$$\sum_{\Gamma} \phi_{\Gamma}^{\dagger} \phi_{\Gamma} = 1 \quad . \quad (55)$$

Secondly, the fermionic and bosonic content have to match in order to possibly recombine both parts to the actual physical electron, thus

$$\sum_{\Gamma} n_{\sigma}^{\Gamma} \phi_{\Gamma}^{\dagger} \phi_{\Gamma} = f_{\sigma}^{\dagger} f_{\sigma} \quad , \quad (56)$$

whereby n_{σ}^{Γ} marks the number of σ -electrons in the state Γ . We then *choose* to write the effective slave-boson Hamiltonian with the constraints already included as

$$H_{\text{loc}}^{\text{SB}} = \varepsilon_0 \sum_{\sigma} f_{\sigma}^{\dagger} f_{\sigma} + U \phi_{\uparrow\downarrow}^{\dagger} \phi_{\uparrow\downarrow} + \lambda_0 \left(\sum_{\Gamma} \phi_{\Gamma}^{\dagger} \phi_{\Gamma} - 1 \right) + \sum_{\sigma} \lambda_{\sigma} \left(\sum_{\Gamma} n_{\sigma}^{\Gamma} \phi_{\Gamma}^{\dagger} \phi_{\Gamma} - f_{\sigma}^{\dagger} f_{\sigma} \right) \quad . \quad (57)$$

With the important choice of representing the single-particle part with the fermionic $f_{\sigma}^{(\dagger)}$ operators and the interacting term with the slave bosons $\phi_{\uparrow\downarrow}^{(\dagger)}$ we rendered the problem quadratic in the operators and therefore easily solvable. In order to do so, we do not even have to condense the bosons in our simple problem. It is possible to directly connect the finite-temperature

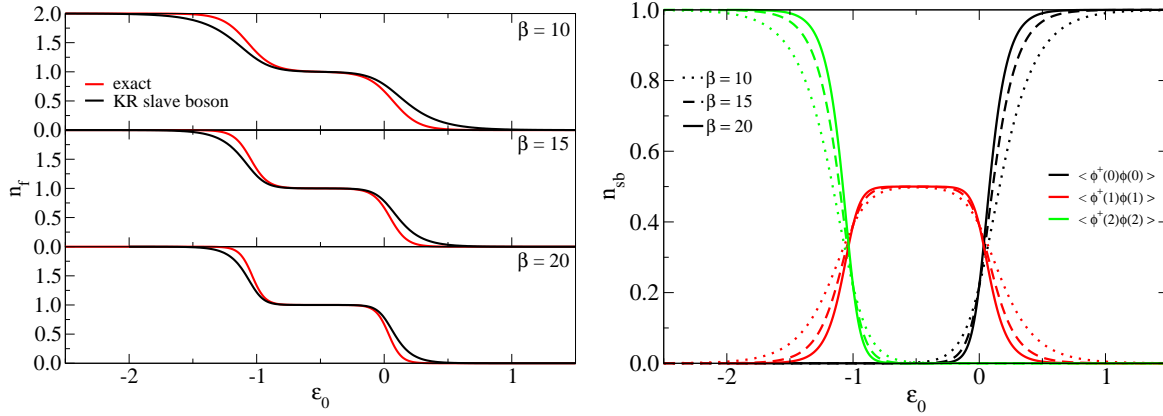


Fig. 7: Hubbard atom within KR slave-boson theory. Left: Comparison with the exact result for the filling with level energy ϵ_0 . Right: Individual level fillings from the slave-boson amplitudes.

expectation values to the respective fermionic/bosonic distribution, namely

$$\begin{aligned} \langle f_{\sigma}^{\dagger} f_{\sigma} \rangle &= n^{\text{F}}(\epsilon_0 - \lambda_{\sigma}) \quad , \quad \langle \phi_0^{\dagger} \phi_0 \rangle = n^{\text{B}}(\lambda_0) \\ \langle \phi_{\sigma}^{\dagger} \phi_{\sigma} \rangle &= n^{\text{B}}(\lambda_0 + \lambda_{\sigma}) \quad , \quad \langle \phi_{\uparrow\downarrow}^{\dagger} \phi_{\uparrow\downarrow} \rangle = n^{\text{B}}(U + \lambda_0 + \lambda_{\uparrow} + \lambda_{\downarrow}) \quad , \end{aligned} \quad (58)$$

with $n^{\text{B,F}}$ as the Bose-Einstein/Fermi-Dirac distribution functions. Therewith the constraints (55,56) read on average

$$n^{\text{B}}(\lambda_0) + \sum_{\sigma} n^{\text{B}}(\lambda_0 + \lambda_{\sigma}) + n^{\text{B}}(U + \lambda_0 + \lambda_{\uparrow} + \lambda_{\downarrow}) - 1 = 0 \quad , \quad (59)$$

$$n^{\text{B}}(U + \lambda_0 + \lambda_{\uparrow} + \lambda_{\downarrow}) + n^{\text{B}}(\lambda_0 + \lambda_{\sigma}) - n^{\text{F}}(\epsilon_0 - \lambda_{\sigma}) = 0 \quad . \quad (60)$$

Aiming for a spin-unpolarized solution with $\lambda_{\sigma} = \lambda$, the above system of two nonlinear equations can be easily solved for (λ_0, λ) . Note that the solution will be approximate, because the exact solution for the expectation value n_s for $s=0,1,2$ electrons on the model atom is of course given by

$$\langle n_s \rangle = \frac{\sum_{\Gamma} n_s^{\Gamma} e^{-\beta E_{\Gamma}}}{\sum_{\Gamma} e^{-\beta E_{\Gamma}}} \quad , \quad (61)$$

with E_{Γ} as the energy in state Γ according to the true Hamiltonian (53) and β the inverse temperature. Figure 7 shows the slave-boson solution along with the exact Coulomb staircase for different inverse temperatures.

The general application of the KR scheme is surely not concentrated on atomic problems, but of course on single- and multi-orbital Hubbard-like models on the lattice. In general its useful to increase formality and write the representation of the physical electron operator for a certain (site, orbital, spin projection) as (see e.g. [45] for details)

$$c_{im\sigma}^{\dagger(\text{SB})} := \mathcal{C}_{im\sigma}^{\dagger} = r_{im}^*(\{\phi\}) f_{im\sigma}^{\dagger} = \sum_{pq} \langle p_i | f_{im\sigma}^{\dagger} | q_i \rangle \phi_{p_i}^{\dagger} \phi_{q_i} f_{im\sigma}^{\dagger} \quad . \quad (62)$$

Here $|p\rangle, |q\rangle$ mark the quasiparticle Fock states $f_{im\sigma}^{\dagger}$ is acting on. Its convenient to express the original physical states in the enlarged product Hilbert space of the quasiparticles and bosons

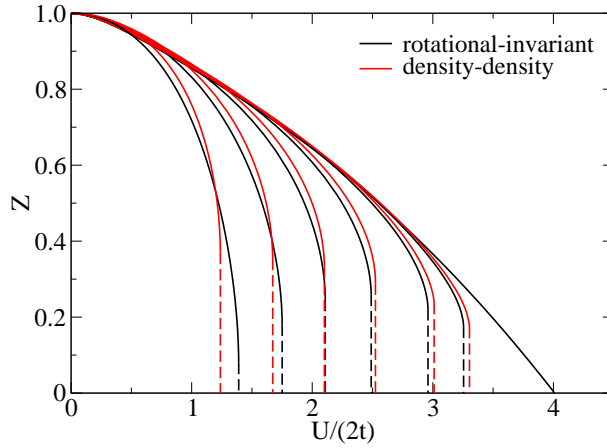


Fig. 8: Quasiparticle weight Z versus U for the two-band Hubbard model on a 3D simple-cubic lattice, NN hopping and no inter-band hopping. Influence of $J=J_C$ on the Mott transition in calculations with and without spin-flip and pair-hopping terms at half filling ($n=2$) for equal bandwidths (from [45]). Right to left: $J/U=0,0.01,0.02,0.05,0.10,0.20,0.45$.

therefore as

$$|p_i\rangle := \phi_{pi}^\dagger |\text{vac}_i\rangle \otimes |p_i\rangle_f \quad . \quad (63)$$

Note that we again only have one slave boson for the physical state, something that has again to be enforced via the constraint (55). The effective Hamiltonian for the single-band Hubbard model will then be given by

$$\underline{H}_{\text{hub}} = -t \sum_{\langle ij \rangle \sigma} r_i r_j^* f_{i\sigma}^\dagger f_{j\sigma} + U \sum_i \phi_{\uparrow i}^\dagger \phi_{\downarrow i} \quad . \quad (64)$$

Throughout this equations for the quasiparticle and slave-boson operators the site index i is surely lost in the MF approximation most often applied in the actual calculations on the lattice. In that sense the conventional KR-SBMF theory is a local approach to correlations, i.e., no inter-site self-energy terms are revealed.

A serious drawback of the standard KR scheme is given by the fact that it can only be applied to density-density interactions. Therefore general multi-orbital Hamiltonians like (15) with spin-flip and pair-hopping terms can not be adequately treated. One can overcome this problem by further extending the framework to a rotationally invariant scheme [46, 47, 45, 48, 49]. However the details of that topic are beyond our brief introduction and we therefore close this section with just showing the rotationally invariant slave-boson result for the general two-band case of the Hamiltonian (15) at half filling in Fig. 8.

6 Final remarks

The goal of this chapter was to give readers that are new to the field a basic introduction to the theoretical research on model Hamiltonians in condensed matter physics. Those who have already or who will have a look at some of the literature referenced here will readily see that we

only scarcely scratched the tip of the iceberg associated with this topic. Nonetheless, the local viewpoint on electronic correlations, taken here for most of the Hamiltonians, and the basic techniques appear to be an appropriate starting point for many phenomena.

Concerning the techniques, the reader should clearly see that there is much room for improvement. Especially treating quasiparticles and Hubbard excitations on an equal footing appears, from what we have seen so far, to be tough. In the chapters to come we will realize how dynamical mean-field theory exactly cures this problem within an again local representation of the electronic self-energy. That is, however, still not the end of the story, there are and will be challenges where one eventually has to go beyond this, just think of long-range fluctuations close to (quantum) critical points.

When discussing the justification for model Hamiltonians compared to pure first-principles approaches in the introductory section of this chapter we already mentioned the reasons. But let us remember in the end once more that such models are excellent tools to single out the essential physical mechanisms in given problems without getting lost in too many, often irrelevant, details. In order to finally deal with emergent phenomena in physics one has to remember this thinking and one should therefore be strongly motivated not to forget it when working with modern first-principles(-like) approaches, as for instance provided by LDA+DMFT.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] L. Hoddeson, G. Baym and M. Eckert, *Rev. Mod. Phys.* **59**, 287 (1987)
- [2] E. Ising, *Zeitschrift für Physik* **31**, 253 (1925)
- [3] M. Born and R. Oppenheimer, *Annalen der Physik* **389**, 457 (1927)
- [4] P. Fulde: *Electron Correlations in Molecules and Solids* (Springer, 1995)
- [5] A. Auerbach: *Interacting Electrons and Quantum Magnetism* (Springer, 1998)
- [6] P. Fazekas: *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, 1999)
- [7] H.T. Diep (ed.): *Frustrated spin systems* (World-Scientific, 2003)
- [8] E.M. Lifshitz and L. P. Pitaevskii: *Statistical Physics Pt. 2* (Butterworth-Heinemann, 1980)
- [9] J. Hubbard, *Proc. Roy. Soc. A* **276** 238 (1963)
- [10] F. Gebhard: *The Mott Metal-Insulator Transition* (Springer, 2000)
- [11] W. Jones and N.H. March: *Theoretical Solid State Physics* (Dover, 1973)
- [12] W.F. Brinkman and T.M. Rice, *Phys. Rev. B* **2**, 430 (1970)
- [13] P.A. Lee, N. Nagaosa and X.-G. Wen, *Rev. Mod. Phys.* **78**, 17 (2006)
- [14] J.C. Slater and G.F. Koster, *Phys. Rev.* **94**, 1498 (1954)
- [15] P.W. Anderson, *Phys. Rev.* **124**, 41 (1961)
P.W. Anderson, Nobel lecture (1977)
- [16] P. Coleman, in *Lectures on the Physics of Highly Correlated Systems VI*, 79 (2002),
arXiv:cond-mat/0206003
- [17] J. Kondo, *Prog. Theoret. Phys.* **32**, 37 (1964)
- [18] A.C. Hewson, *The Kondo Problem to Heavy Fermions* (Cambridge, 1997)
- [19] D. Pesin and L. Balents, *Nature Physics* **6**, 376 (2010)
- [20] C. Castellani, C.R. Natoli and J. Ranninger, *Phys. Rev. B* **18**, 4945 (1978)
- [21] R. Frésard and G. Kotliar, *Phys. Rev. B* **56**, 12909 (1997)
- [22] M. Imada, A. Fujimori and Y. Tokura, *Rev. Mod. Phys.* **70**, 1039 (1998).

-
- [23] D. Grieger, L. Boehnke and F. Lechermann, *J. of Phys.: Cond. Mat.* **22**, 275601 (2010)
- [24] J.E. Hirsch, *Phys. Rev. Lett.* **53**, 2327 (1984)
- [25] S. Kivelson, W.-P. Su, J.R. Schrieffer and A.J. Heeger, *Phys. Rev. Lett.* **58**, 1899 (1987)
- [26] D.W. Visser, A.P. Ramirez and M.A. Subramanian, *Phys. Rev. Lett.* **78**, 3947 (1997)
- [27] T. Holstein, *Ann. Phys.* **8**, 343 (1959).
- [28] V.I. Anisimov, J. Zaanen and O.K. Andersen, *Phys. Rev. B* **44**, 943 (1991)
- [29] D.R. Penn, *Phys. Rev.* **142**, 350 (1966)
- [30] Y. Nagaoka, *Solid State Commun.* **3**, 409 (1965)
D.J. Thouless, *Proc. Phys. Soc.* **86**, 893 (1965)
- [31] J. Wahle, N. Blümer, J. Schlipf, K. Held and D. Vollhardt, *Phys. Rev. B* **58**, 12749 (1998)
- [32] A.N. Tahvildar-Zadeh, J.K. Freericks and M. Jarrell, *Phys. Rev. B* **55**, 942 (1997)
- [33] R.D. Mattuck: *A Guide to Feynman Diagrams in the Many-Body Problem*
(McGraw-Hill/Dover 1976)
- [34] H. Lehmann, *Nuovo Cimento* **11**, 342 (1954)
- [35] M.C. Gutzwiller, *Phys. Rev. Lett.* **10** 159 (1963)
- [36] S.E. Barnes, *J. Phys. F: Met. Phys.* **6**, 1375 (1976)
- [37] N. Read and D.M. Newns, *J. Phys. C: Solid State Phys.* **16**, 3273 (1983)
- [38] P. Coleman, *Phys. Rev. B* **29**, 3035 (1984)
- [39] D. Vollhardt, *Rev. Mod. Phys.* **56**, 99 (1984)
- [40] R. Frésard and T. Kopp, *Nucl. Phys. B* **594** (2001) 769
R. Frésard, H. Ouerdane and T. Kopp, *Nucl. Phys. B* **785**, 286 (2007)
- [41] R. Raimondi and C. Castellani, *Physical Review B* **48**, R11453 (1993)
- [42] G. Kotliar and A.E. Ruckenstein, *Phys. Rev. Lett.* **57**, 1362 (1986)
- [43] S. Florens and A. Georges, *Phys. Rev. B* **66**, 165111 (2002)
- [44] F. Lechermann and A. Georges, unpublished.
- [45] F. Lechermann, A. Georges, G. Kotliar and O. Parcollet, *Phys. Rev. B* **76**, 155102 (2007).
- [46] T. Li, P. Wölfle and P.J. Hirschfeld, *Phys. Rev. B* **40**, 6817 (1989)

- [47] J. Bünnemann, W. Weber and F. Gebhard, Phys. Rev. B **57**, 6896 (1998)
- [48] A. Isidori and M. Capone, Phys. Rev. B **80**, 115120 (2009)
- [49] J. Bünnemann, phys. stat. sol. B **248**, 203 (2011)

4 Wannier Functions and Construction of Model Hamiltonians

Jan Kuneš

Institute of Physics, Academy of Sciences
Praha, Czech Republic

Contents

1	Introduction	2
1.1	Electron in periodic potential	2
1.2	Why localized basis?	3
2	Preliminaries	4
2.1	Orthogonal atomic orbitals	4
2.2	Asymptotic behavior of Fourier coefficients	5
3	Wannier functions	7
3.1	Basic definitions	7
3.2	Existence of exponentially localized Wannier functions – single band	9
3.3	Existence of exponentially localized Wannier functions – composite bands	11
3.4	Uniqueness	12
4	Numerical methods for construction of Wannier functions	13
4.1	Maximally localized Wannier functions	13
4.2	Projection method	15
4.3	Entangled bands	15
5	Examples of applications	16
5.1	Wide versus narrow energy window: SrVO_3	17
5.2	Unfolding band structures: LaOFeAs	18
5.3	Entangled bands	20
5.4	Spin-orbit coupling: Sr_2IrO_4	21
5.5	Wannier functions in dynamical mean-field theory	22
6	Summary	24

1 Introduction

Although introduced already in 1937, Wannier functions might be enjoying their 'golden age' right now, thanks to the rapid growth of methods linking first-principles band structure calculations with model theories based on the second quantization formalism, such as LDA+DMFT. While using Wannier functions as a computational tool is the likely goal of the reader, there is another fascinating aspect of Wannier functions, namely the connection between the spatial localization of Wannier functions and the topological properties of the corresponding Bloch states. It is this topology that gave the name to the topological insulators.

In the first part of these notes we establish the connection between the smoothness and periodicity of Bloch functions as functions of the electron quasi-momenta and the exponential localization of the corresponding Wannier functions. As this is a rather mathematical topic we do not attempt a comprehensive presentation. Our goal is to point out the general ideas and concepts and to direct the reader to the original literature. In the second part of the notes we present the commonly used computational methods for construction of the Wannier functions and examples of their application.

1.1 Electron in periodic potential

The introduction of density functional theory [1] started the era of *ab initio* calculations of electronic structure for real materials. The Kohn-Sham equation has the form of Schrödinger equation for non-interacting electrons

$$-\nabla^2\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = \epsilon\psi(\mathbf{r}), \quad (1)$$

where we chose atomic units with $\hbar = 1$, $m_e = 1/2$. With periodic boundary conditions (at infinity) and the requirement that $\psi(\mathbf{r})$ is normalizable, the equation represents an eigenvalue problem for a Hermitian Hamiltonian $H = -\nabla^2 + V(\mathbf{r})$. If the crystal potential $V(\mathbf{r})$ possesses translational symmetry

$$V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r}), \quad (2)$$

where \mathbf{R} is a vector of the crystal lattice, the Bloch theorem [2] allows a partial diagonalization. The eigenfunctions of H can be written in the form

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}}u_{n,\mathbf{k}}(\mathbf{r}), \quad (3)$$

where $u_{n,\mathbf{k}}(\mathbf{r})$ is an \mathbf{r} -periodic function, n is a discrete band index, and \mathbf{k} is a continuous index, a vector from the first Brillouin zone. The corresponding eigenvalues $\epsilon_{n,\mathbf{k}}$ are continuous and periodic functions of \mathbf{k} – they are said to form energy bands. The cell periodic functions $u_{n,\mathbf{k}}(\mathbf{r})$ obey the equation

$$(-i\nabla - \mathbf{k})^2u_{n,\mathbf{k}}(\mathbf{r}) + V(\mathbf{r})u_{n,\mathbf{k}}(\mathbf{r}) = \epsilon_{n,\mathbf{k}}u_{n,\mathbf{k}}(\mathbf{r}), \quad (4)$$

with periodic boundary conditions in the unit cell. Besides mathematical elegance, the Bloch theorem is crucial for numerical applications since it replaces the Hamiltonian H with continuous spectrum, by a (continuous) set of Hamiltonians $H_{\mathbf{k}} = (-i\nabla - \mathbf{k})^2 + V(\mathbf{r})$ with discrete spectra, i.e., a problem which is numerically tractable with standard methods of linear algebra. One of the key characteristics of any material is the presence or absence of an energy gap separating the ground state from the excitations, which translates into presence or absence of a direct gap between the occupied valence bands and empty conduction bands in the case of non-interacting electrons in a periodic solid. On the other hand, for the existence and other properties of the Wannier functions the band filling is irrelevant. Also the distinction between direct and indirect gaps does not matter for the properties of Wannier functions. We say that the n th band is isolated if $\epsilon_{n-1,\mathbf{k}} < \epsilon_{n,\mathbf{k}} < \epsilon_{n+1,\mathbf{k}}$ for each \mathbf{k} , i.e., the n th band is separated by a finite, possibly indirect, gap from the bands below and above. Similarly a composite band is isolated if $\epsilon_{n_{\min}-1,\mathbf{k}} < \epsilon_{n_{\min},\mathbf{k}} \leq \epsilon_{n_{\max},\mathbf{k}} < \epsilon_{n_{\max}+1,\mathbf{k}}$. Bands that are not isolated will be called entangled.

1.2 Why localized basis?

The reasons why the description of materials in terms of localized orbitals is attractive are both conceptual and technical. By conceptual we mean those features that do not simplify computations, but provide better insights into the physics. For example, the chemist's language of chemical bonds is 'difficult' to understand for a physicist speaking the 'Bloch wave' language. The path from a set of isolated atoms with the *localized* atomic orbitals to a periodic solid with *extended* Bloch states may be difficult to grasp, as it involves qualitative differences such as localized versus extended or discrete versus continuous. The Wannier functions provide a natural extension of the concept of atomic orbitals into solids and thus bridge this gap. Another example is provided by the theory of dielectric polarization [3]. While in terms of Bloch states the polarization is expressed through a rather abstract concept of Berry phase, the formulation in terms of Wannier functions uses a simple notion of the center of mass of the corresponding charge distribution.

Many important physical properties and phenomena involve spatially localized objects such as impurities or defects in the crystal structure, screened electron-electron interaction, composite excitations such as excitons or polarons. The technical advantages of localized orbitals in the context of local electronic correlations are evident. As we typically consider only the short-range part of the electron-electron interaction explicitly, expressing the interaction in terms of localized orbitals reduces the number of non-zero terms considerably (for a more detailed discussion of the Hubbard model see section 5.5).

While atomic orbitals appear to provide a natural framework there is a problem. As put by Wannier [4] in the context of semiconductor physics: "It would no doubt be more satisfactory for insulating crystals, to discuss the Hamiltonian using atomic functions rather than Bloch functions. But this line of attack has been hampered by the fact that atomic functions are not orthogonal". Obviously, the orthogonality of a basis set is a great advantage, in particular when the formalism of second quantization is to be used.

2 Preliminaries

2.1 Orthogonal atomic orbitals

There are various ways to introduce the Wannier functions (WFs). Let us follow Wannier [4] for a moment and consider the following warm-up problem. Assume that we have a solid formed by atoms sitting on lattice sites \mathbf{R} with a single valence orbital described by an atomic wave function $v(\mathbf{x})$, i.e., the energies of the other atomic orbitals are far away so that their effect can be neglected. The basis $\{v(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$ is in general non-orthogonal. How to orthogonalize it while preserving the explicit translational symmetry of the basis set? A straightforward Gram-Schmidt procedure does break the translational symmetry of the basis. Instead, the following construction may be used.

First, we form the Bloch sum (discrete Fourier transform) of the atomic orbitals

$$\phi(\mathbf{k}, \mathbf{x}) = \frac{A(\mathbf{k})}{N^{1/2}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} v(\mathbf{x} - \mathbf{R}), \quad (5)$$

where the summation runs over the lattice vectors \mathbf{R} . The normalization constant $A(\mathbf{k})$ is given by

$$\frac{1}{A(\mathbf{k})^2} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \int d\mathbf{x} v^*(\mathbf{x})v(\mathbf{x} - \mathbf{R}). \quad (6)$$

It is easy to verify that the basis $\{\phi(\mathbf{k}, \mathbf{x})\}_{\mathbf{k}}$, indexed by the reciprocal vectors \mathbf{k} , is orthonormal. Next, we perform the inverse Fourier transform

$$w(\mathbf{R}, \mathbf{x}) = \frac{1}{N^{1/2}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{R}} \phi(\mathbf{k}, \mathbf{x}) = \sum_{\mathbf{R}'} c(\mathbf{R} - \mathbf{R}') v(\mathbf{x} - \mathbf{R}'). \quad (7)$$

The new functions $w(\mathbf{R}, \mathbf{x})$ are combinations of the original atomic functions $v(\mathbf{x} - \mathbf{R}')$ centered at different lattice sites with coefficients $c(\mathbf{R})$ given by

$$c(\mathbf{R}) = \frac{1}{N} \sum_{\mathbf{k}} A(\mathbf{k}) e^{-i\mathbf{k}\cdot\mathbf{R}}. \quad (8)$$

Let us make a few observations about $w(\mathbf{R}, \mathbf{x})$. First, since the coefficients $c(\mathbf{R} - \mathbf{R}')$ in (7) depend only on the difference $\mathbf{R} - \mathbf{R}'$ we can write $w(\mathbf{R}, \mathbf{x}) \equiv w(\mathbf{x} - \mathbf{R})$, i.e. $w(\mathbf{R}, \mathbf{x})$ on different lattice sites are shifted images of the same functional form $w(\mathbf{x})$. Second, since (7) is a unitary transformation of an orthonormal basis $\{\phi(\mathbf{k}, \mathbf{x})\}_{\mathbf{k}}$ the basis $\{w(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$ is also orthonormal as can be verified by explicit calculation. In fact, we have performed forward and inverse Fourier transformation, but we did not get the original atomic functions. The 'magic' is in the normalization factor $A(\mathbf{k})$, which was introduced between the Fourier transforms. Let us look at the behavior of $A(\mathbf{k})$ in more detail. If the atomic functions $v(\mathbf{x} - \mathbf{R})$ do not overlap at all, expression (6) returns $A(\mathbf{k}) = 1$ and the inverse Fourier transform (7) recovers the original atomic functions. Once the overlap between $v(\mathbf{x} - \mathbf{R})$ on different sites is finite, the \mathbf{k} -dependence of $A(\mathbf{k})$ leads to admixture of atomic functions from neighboring sites in $w(\mathbf{x} - \mathbf{R})$. Generally, the larger the overlap of $v(\mathbf{x} - \mathbf{R})$'s the stronger the \mathbf{k} -dependence of

$A(\mathbf{k})$ and the less the $w(\mathbf{R} - \mathbf{x})$ resembles the starting atomic functions. For exponentially localized atomic functions the orthonormal $w(\mathbf{R} - \mathbf{x})$ are also exponentially localized [4].

Since the orbitals $\{v(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$ give rise to a single band we can say that we have constructed a localized, translationally invariant orthonormal basis $\{w(\mathbf{x} - \mathbf{R})\}_{\mathbf{R}}$, which spans the Hilbert space of this band.

Typically, in numerical computations we do not have an atomic basis for a given band to start with. Thus a question arises whether a similar construction is possible for an arbitrary band and how it can be realized if we know the Bloch eigenstates in some basis, as is usually the case.

2.2 Asymptotic behavior of Fourier coefficients

Localization, that is how fast a function decays away from its center of mass, is one of the central issues concerning Wannier orbitals. Since the Wannier functions are, vaguely speaking, Fourier transforms of the Bloch functions let us review some properties of the Fourier series [5]. We do not aspire to provide a full mathematical background of the complicated issue of localization. Our intention is to point out the intimate connection between smoothness of a function and the convergence (localization) of its Fourier series.

The Fourier coefficients of a function $f(x)$ of a real variable x integrable (L_1 integrable) on the interval $[0, 2\pi)$ are defined by

$$a_n = \frac{1}{2\pi} \int_0^{2\pi} dx e^{-inx} f(x) \quad (9)$$

and the corresponding Fourier series reads

$$F(x) = \sum_{n=-\infty}^{\infty} a_n e^{inx}. \quad (10)$$

Let us assume that $f(x)$ is periodic, $f(x) = f(x + 2\pi)$, and piecewise smooth up to the second derivative, i.e., there is a finite number of points α_i where the first derivative $f^{(1)}(x)$ is discontinuous and $f^{(1)}(x)$ and $f^{(2)}(x)$ are continuous and finite on every interval $[\alpha_i, \alpha_{i+1}]$. This is sufficient for the series (10) to converge uniformly towards $f(x)$ and for the Fourier coefficients to converge as $|a_n| < K/n^2$, where K is a finite constant. Differentiating (10) with respect to x and applying the above statement to higher derivatives of $f(x)$ it is easy to see that the smoother the function, the faster its Fourier coefficients decay. In general, if a discontinuity appears first in the k th-derivative, the Fourier coefficients decay as K/n^{k+1} . Therefore a necessary condition for an exponential decay $a_n \propto \exp(-Kn)$ is that $f(x)$ is smooth, i.e., has derivatives to an arbitrary order. However, this is not sufficient as is shown by an example at the end of this section.

To derive a sufficient condition for the exponential decay of a_n , let us assume that $f(x)$ can be analytically continued into the complex plane. Introducing a new variable $z = e^{ix}$, we map the real interval $[0, 2\pi)$ onto a unit circle and define a function $\tilde{f}(z)$ by $\tilde{f}(e^{ix}) = f(x)$. Now, let

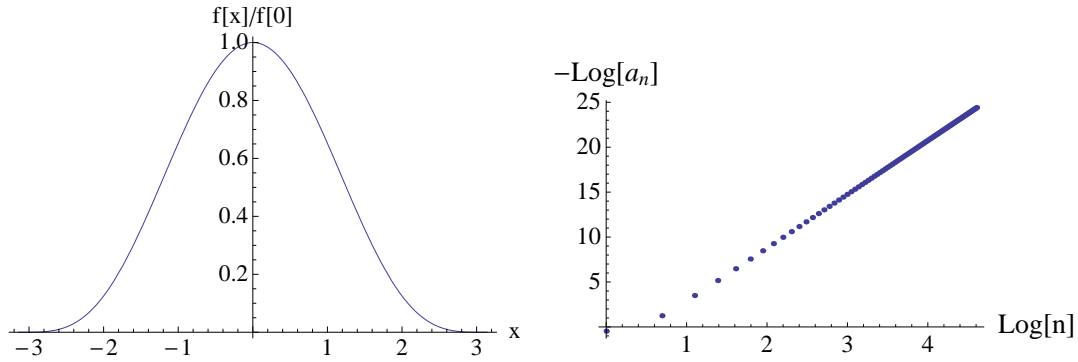


Fig. 1: The graph of $x \rightarrow \left(1 - \frac{x}{\pi}\right)^4 \left(1 + \frac{x}{\pi}\right)^4$ (left) and its Fourier coefficients a_n in a log-log plot (right).

us assume that $\tilde{f}(z)$ is analytic on a ring $r < 1 < R$ which implies that there exists a Laurent series

$$\tilde{f}(z) = \sum_{n=-\infty}^{\infty} b_n z^n, \quad (11)$$

where the coefficients b_n are given by an integral over a contour γ , which encircles zero and lies inside the ring $r < z < R$,

$$b_n = \frac{1}{2\pi i} \oint_{\gamma} \frac{\tilde{f}(z) dz}{z^{n+1}}. \quad (12)$$

Taking the unit circle for γ and comparing to (9) one sees that $a_n = b_n$. Let us define $\rho = \min(R, 1/r)$. The analyticity of $\tilde{f}(z)$ implies that the series (11) is absolutely convergent for any real x from the interval $1/\rho < x < \rho$, but diverges for either $x > \rho$, $x < 1/\rho$, or both. This means that $b_{|n|} x^{|n|} \rightarrow 0$ for $x < \rho$, but $b_{|n|} x^{|n|} \rightarrow \infty$ for $x > \rho$. This is the precise meaning of $b_{|n|} \sim \exp(-|n|h)$, where $h = \ln \rho$.

We close this section with a few examples of the convergence of Fourier series for functions defined on $[-\pi, \pi)$ and periodically repeated over the real axis. In Fig. 1 we start with $\left(1 - \frac{x}{\pi}\right)^4 \left(1 + \frac{x}{\pi}\right)^4$, which has a discontinuous 5th derivative at $x = \pm\pi$. Its Fourier coefficients exhibit a power law decay, which appears linear in the log-log plot, Fig. 1.

Function $(1 + \cos(x)) \sqrt{1 + \cos(x)}$ provides an example of an analytic function with an infinite Fourier series. The Fourier coefficients decay exponentially as shown in Fig. 2

Finally we study the Fourier series of a 'bump' function $\exp(-1/[1 - (x/\pi)^2])$, which is smooth, but not analytic. Its Fourier coefficients exhibit $|a_n| \sim |\pi n|^{-3/4} \exp(-\sqrt{|\pi n|})$ asymptotic behavior as can be shown by the saddle point method [6]. The decay of a_n is obviously faster than a power law, but slower than exponential. The actual Fourier coefficients obtained by numerical integration are shown in Fig. 3.

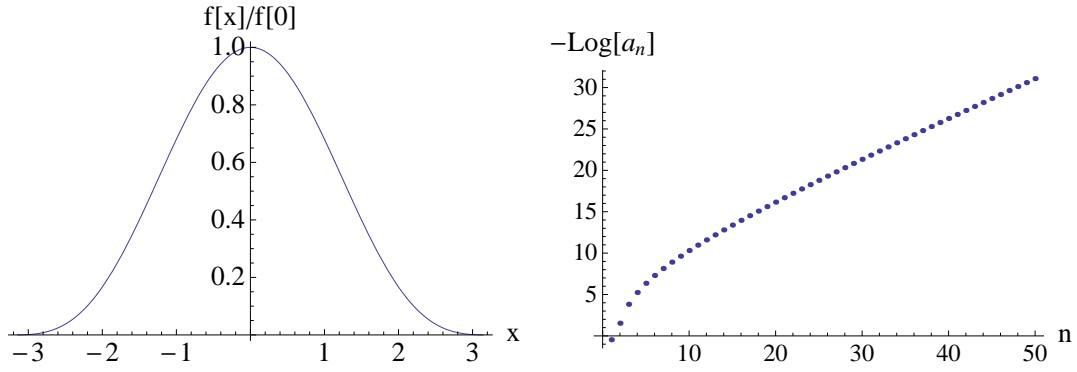


Fig. 2: The graph of $x \rightarrow (1 + \cos(x))\sqrt{1.1 + \cos(x)}$ (left) and its Fourier coefficients a_n in a log-lin plot (right).

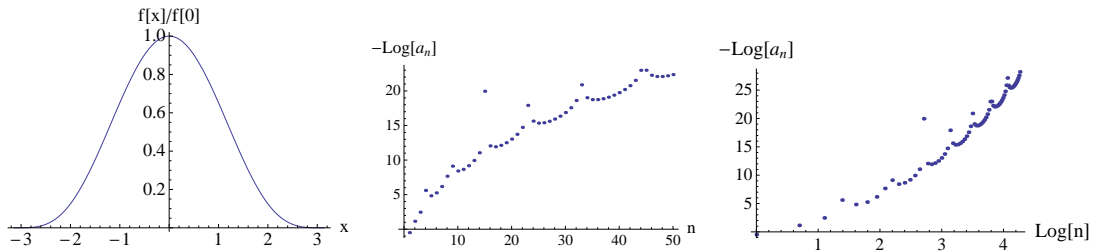


Fig. 3: The graph of $x \rightarrow \exp\left(-\frac{1}{1-(x/\pi)^2}\right)$ (left) and its Fourier coefficients a_n in a log-lin (center) and log-log (right) plot.

3 Wannier functions

3.1 Basic definitions

The Wannier function (WF) of an isolated band is defined as

$$w_n(\mathbf{r} - \mathbf{R}) = \frac{V}{(2\pi)^d} \int_{\text{BZ}} d\mathbf{k} e^{-i\mathbf{k}\cdot\mathbf{R}} \psi_{n,\mathbf{k}}(\mathbf{r}), \quad (13)$$

where V is the unit cell volume, d is the dimension, $\psi_{n,\mathbf{k}}(\mathbf{r})$ are the Bloch functions (3) corresponding to the n th band. The integration runs over the first Brillouin zone. The inverse transformation reads

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} w_n(\mathbf{r} - \mathbf{R}). \quad (14)$$

The transformation properties of the Bloch functions under lattice translation (3) ensure the mutual orthogonality of WFs centered in different unit cells. The definition (13) is not unique since the overall phase associated with the Bloch function is arbitrary. For reasons that become clear later we will consider only situations where $\psi_{n,\mathbf{k}}(\mathbf{r})$ is a smooth function of \mathbf{k} . A particular choice of the phases will be called a gauge and the transformation

$$\psi_{n,\mathbf{k}}(\mathbf{r}) \rightarrow e^{i\phi(\mathbf{k})} \psi_{n,\mathbf{k}}(\mathbf{r}), \quad (15)$$

between different gauges, where $\phi(\mathbf{k})$ is an analytic real-valued function, will be called a gauge transformation.

The generalization to the case of isolated composite bands is straightforward.

$$\mathbf{w}(\mathbf{r} - \mathbf{R}) = \frac{V}{(2\pi)^d} \int_{\text{BZ}} d\mathbf{k} \mathbf{U}(\mathbf{k}) \boldsymbol{\psi}_{\mathbf{k}}(\mathbf{r}), \quad (16)$$

where $\mathbf{w}(\mathbf{r})$ and $\boldsymbol{\psi}_{\mathbf{k}}(\mathbf{r})$ are vectors

$$\mathbf{w} = \begin{pmatrix} w_{n_{\min}} \\ \vdots \\ w_{n_{\max}} \end{pmatrix}, \quad \boldsymbol{\psi}_{\mathbf{k}} = \begin{pmatrix} \psi_{n_{\min},\mathbf{k}} \\ \vdots \\ \psi_{n_{\max},\mathbf{k}} \end{pmatrix} \quad (17)$$

in the band index spanning the composite band, and $\mathbf{U}(\mathbf{k})$ is a unitary matrix acting on the band index. We define a quasi-inverse transformation

$$\tilde{\boldsymbol{\psi}}_{\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \mathbf{w}(\mathbf{r} - \mathbf{R}), \quad (18)$$

where the quasi Bloch states $\tilde{\boldsymbol{\psi}}_{\mathbf{k}}(\mathbf{r})$ have the transformation property (3), but are not the eigenstates of the Hamiltonian. The gauge transformation (15) generalizes to a set of unitary transformations $\mathbf{U}(\mathbf{k})$ analytic and periodic in \mathbf{k} .

The construction of Wannier functions poses several non-trivial problems.

Existence. Given an isolated band, is it possible to span the corresponding Hilbert space with a basis of exponentially localized Wannier functions? As discussed below an equivalent question reads: Is it possible to find a gauge such that the resulting Bloch states are periodic and analytic functions of \mathbf{k} ? The requirement of periodicity has topological implications and it turns out that the existence or non-existence of exponentially localized Wannier function is a topological characteristic of a given band.

Multiple bands. How do the arguments about existence of the exponentially localized WFs generalize to the case of isolated composite bands? Mathematically the difference between single and multiple bands lies in the Abelian (commutative multiplication of numbers) or non-Abelian (non-commutative multiplication of matrices) character of the respective gauge transformations. As a result, the technique used originally for a single band cannot be simply generalized to a composite band. For example, the results for composite bands depend on the dimension of the lattice.

Entanglement. Often it is not possible to find an isolated (composite) band spanning the desired part of the Hilbert space. Nevertheless, we would like to have a localized WFs basis representing the bands of interest, if only approximately. What is the optimal procedure to obtain WFs with such property?

Uniqueness. Assuming we can construct exponentially localized WFs, what is the meaning of the remaining gauge freedom? Under which additional conditions are the WFs defined uniquely and what are their physical consequences?

Implementation. How to compute the Wannier functions numerically? The theory of Wannier functions operates with the concepts of analytic continuation or topology with respect to \mathbf{k} ,

which are inherently connected to the notion of continuity. How to perform practical calculations in the computer, which deal inevitably with discrete quantities?

In the following we will discuss these issues in some detail.

3.2 Existence of exponentially localized Wannier functions – single band

In the following we will sketch the conditions and prove of the existence of exponentially localized WFs for a single isolated band.

Let us assume that a given isolated band has eigenvalues $\{\epsilon_{\mathbf{k}}\}$ and the corresponding Bloch states $\{|\psi_{\mathbf{k}}\rangle\}$. The key ingredients for exponential localization are analyticity and periodicity as a function of \mathbf{k} (see 2.2). The existence of exponentially localized WFs is equivalent to the possibility of finding a set of Bloch states such that $\langle \mathbf{r} | \psi_{\mathbf{k}} \rangle$ is a periodic and analytic function of \mathbf{k} throughout the Brillouin zone. To appreciate the question of localization one has to realize that the eigenvalue problem does not specify $\langle \mathbf{r} | \psi_{\mathbf{k}} \rangle$ uniquely, but only up to an arbitrary multiplicative phase factor, which may differ from \mathbf{k} -point to \mathbf{k} -point. While the exponentially localized WFs exist under rather general conditions (see below), it is not true that they exist for every Hermitian periodic Hamiltonian.

The first partial answer to the question of existence of exponentially localized WFs in **1D** was given by Kohn [7] for non-relativistic electrons in a periodic potential with inversion symmetry. He proved the existence of exponentially localized WFs as well as their uniqueness upon the requirement of reality and definite parity. Kohn studied the solutions of the initial-value problem for the Schrödinger equation

$$-\frac{d^2}{dx^2}\phi_{1,2}(E, x) + V(x)\phi_{1,2}(E, x) = E\phi_{1,2}(E, x) \quad (19)$$

as a function of the energy E (the indices 1 and 2 stand for the two linearly independent solutions corresponding to different initial conditions). The solutions obtained this way are analytic functions of E . Kohn showed that Bloch functions, defined by the periodicity condition $\psi(E, x + a) = \lambda\psi(E, x)$, can be build as $\psi(E, x) = \alpha(E)\phi_1(E, x) + \beta(E)\phi_2(E, x)$ in such a way that analyticity as a function of E and in turn as a function of k is preserved. It turns out that while analyticity of $\langle x | \psi_{\mathbf{k}} \rangle$ can be satisfied relatively easily, the requirement of periodicity on top of it, is a rather stringent condition.

A straightforward generalization of Kohn's global analysis, based on the initial value problem to the Schrödinger equation, to **higher dimensions** is not possible. To analyze the dimensions two and higher des Cloizeaux [8] introduced the band projection operator $P(\mathbf{k})$

$$P(\mathbf{k}) = |\psi_{\mathbf{k}}\rangle\langle\psi_{\mathbf{k}}| \quad (20)$$

and proved the analyticity in \mathbf{k} of its \mathbf{r} -elements $\langle \mathbf{r} | P(\mathbf{k}) | \mathbf{r}' \rangle$ as a function of \mathbf{k} for an isolated as well as composite band. While \mathbf{k} -analyticity of Bloch states $\langle \mathbf{r} | \psi_{\mathbf{k}} \rangle$ implies the analyticity of $\langle \mathbf{r} | P(\mathbf{k}) | \mathbf{r}' \rangle$, the opposite is not true. To illustrate the problem we follow Ref. [9] and try to

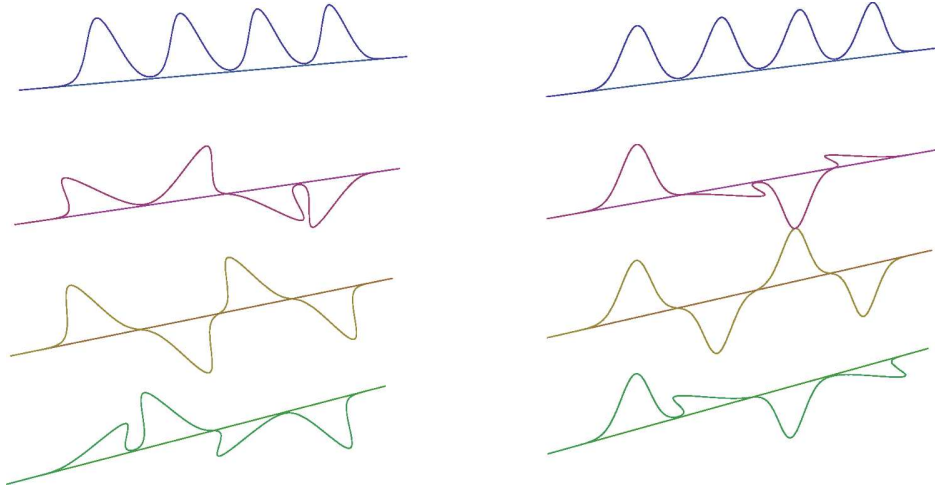


Fig. 4: *Cartoon depiction of the projection method. The Bloch functions of different \mathbf{k} with random overall phases (left) and the same functions with the aligned phase on the first site (right).*

construct the \mathbf{k} -analytic Bloch state $|\psi_{\mathbf{k}}\rangle$ as

$$\langle \mathbf{r} | \psi_{\mathbf{k}} \rangle = G^{-1/2}(\mathbf{k}) \langle \mathbf{r} | P(\mathbf{k}) | M_0 \rangle, \quad (21)$$

$$G(\mathbf{k}) = \langle M_0 | P(\mathbf{k}) | M_0 \rangle, \quad (22)$$

where $|M_0\rangle$ is an arbitrary trial orbital. The idea is to let the Bloch functions on the lattice site specified by the trial orbital $|M_0\rangle$ add up with the same sign (constructive interference) and to interfere destructively on the other lattice sites as depicted in Fig. 4. The Bloch function $|\psi_{\mathbf{k}}\rangle$ defined by (21) is analytic as long as the non-negative $G(\mathbf{k})$ is different from 0. Thus points $G(\mathbf{k}_0) = 0$ need special attention. As $G(\mathbf{k})$ is analytic, $G(\mathbf{k} - \mathbf{k}_0) \propto (\mathbf{k} - \mathbf{k}_0)^{2p}$ must be a quadratic (or even higher) form. Here again the 1D is distinct, as only in 1D a square root $G^{1/2}(\mathbf{k})$ analytic at \mathbf{k}_0 is guaranteed to exist, but can possibly change sign. Assuming time-reversal symmetry implies that zeros of $G(\mathbf{k})$ appear in pairs \mathbf{k}_0 and $-\mathbf{k}_0$ due to Kramers degeneracy (the cases $G(0) = 0$ and $G(\pi) = 0$ may be excluded by the choice of $|M_0\rangle$). Thus in a 1D time-reversal symmetric system a periodic and analytic $G^{1/2}(\mathbf{k})$ exists. It can be shown that in this case the zeros of $G^{1/2}(\mathbf{k})$ are canceled out by the corresponding zeros in $\langle \mathbf{r} | P(\mathbf{k}_0) | M_0 \rangle$ and $\langle \mathbf{r} | \psi_{\mathbf{k}} \rangle$ (21) is periodic and analytic.

Similar argument cannot be used in higher dimensions since singularities of $G^{1/2}(\mathbf{k})$ at zeros of $G(\mathbf{k})$ cannot be removed. The present technique allows to prove that exponentially localized Wannier functions exist in time-reversal invariant systems with inversion symmetry or systems that can be adiabatically connected (i.e., keeping the band isolated along the adiabatic path) to a system with localized WFs [9]. On the other hand, in 2D quantum Hall systems (i.e., no time-reversal symmetry) one can prove that a zero of $G(\mathbf{k})$ exists for any choice of $|M_0\rangle$ [10] and that exponentially localized WFs do not exist for bands carrying the Hall current. Using a different approach based on the band projection operator Nenciu proved the existence of exponentially

localized Wannier functions for an isolated band in systems with time reversal symmetry in arbitrary dimension [11].

3.3 Existence of exponentially localized Wannier functions – composite bands

The arguments from the previous section cannot be simply generalized to the case of composite bands. It is nevertheless possible to prove that time-reversal invariance of the Hamiltonian is sufficient for exponentially localized WFs to exist for spinless fermions in 2D and 3D [12]. The proof relies on the concepts of connection, curvature, and Chern numbers from differential geometry. The basic idea is that Brillouin zone provides a base manifold and the inner product in the Hilbert space of the cell-periodic functions $u_{n,\mathbf{k}}$ supplies a sense of parallelism between nearby \mathbf{k} -points, so called (Berry) connection

$$A_{mn}^\alpha(\mathbf{k}) = \left(u_{n,\mathbf{k}} \left| \frac{\partial}{\partial k_\alpha} u_{m,\mathbf{k}} \right. \right), \quad (23)$$

where the scalar product $(\cdot|\cdot)$ denotes an \mathbf{r} -integration over the unit cell. Although the connection is not gauge-invariant it allows for the calculation of gauge invariant quantities by standard techniques of differential geometry. Following [12, 13], we define the trace of the curvature corresponding to $A_{mn}^\alpha(\mathbf{k})$ by

$$B^{\alpha\beta}(\mathbf{k}) = \text{tr} \left(\frac{\partial \mathbf{A}^\beta}{\partial k_\alpha} - \frac{\partial \mathbf{A}^\alpha}{\partial k_\beta} - [\mathbf{A}^\alpha, \mathbf{A}^\beta] \right) \quad (24)$$

$$= 2 \text{Im} \sum_{n=n_{\min}}^{n_{\max}} \left(\frac{\partial}{\partial k_\alpha} u_{n,\mathbf{k}} \left| \frac{\partial}{\partial k_\beta} u_{n,\mathbf{k}} \right. \right) \quad (25)$$

$$= \text{Tr} \left(P(\mathbf{k}) \left[\frac{\partial}{\partial k_\alpha} P(\mathbf{k}), \frac{\partial}{\partial k_\beta} P(\mathbf{k}) \right] \right). \quad (26)$$

Here the first line is the definition of $B^{\alpha\beta}(\mathbf{k})$ in terms of the connection $A_{mn}^\alpha(\mathbf{k})$, with the trace and the commutator acting on the band indices. The second line is $B^{\alpha\beta}(\mathbf{k})$ for the connection (23). The third line shows explicitly the gauge invariance and analyticity by expressing $B^{\alpha\beta}(\mathbf{k})$ in terms of the generalized band projection operator $P(\mathbf{k})$, with the trace and commutator understood in the operator sense. The generalized band projection operator (20)

$$P(\mathbf{k}) = \sum_{n=n_{\min}}^{n_{\max}} |\psi_{n,\mathbf{k}}\rangle \langle \psi_{n,\mathbf{k}}| \quad (27)$$

is an analytic function of \mathbf{k} [8]. Panati [14] showed that the analytic and periodic quasi Bloch states exist only when all Chern numbers associated with the curvature $B^{\alpha\beta}(\mathbf{k})$ are zero.

Modern examples of systems with non-trivial band topology, which prohibits construction of exponentially localized WFs, are topological insulators [15]. We close this section by pointing out that the topological characteristics are the property of a given band or composite bands. The topological characteristics change when other bands are included in the composite bands and thus the construction of localized Wannier functions may become possible.

3.4 Uniqueness

Analytic functions are very rigid objects in the sense that, once the function is known on an arbitrarily small open interval Δk^d (where \mathbf{k} is viewed as a complex variable!), it is uniquely defined everywhere. Nevertheless it is clear that having an exponentially localized WF any gauge transformation (15) also produces exponentially localized WF.

Uniqueness of the WF center. Let us define the center of WF as $\langle w|\mathbf{r}|w\rangle$ (the center of mass). We will show that for an isolated band the WF center does not depend on the gauge.

$$\begin{aligned}\langle w|\mathbf{r}|w\rangle &= \int d\mathbf{r} \left(\frac{V}{(2\pi)^d} \right)^2 \int_{\text{BZ}} d\mathbf{k} d\mathbf{k}' \psi_{\mathbf{k}}^*(\mathbf{r}) \mathbf{r} \psi_{\mathbf{k}'}(\mathbf{r}) \\ &= -i \frac{V}{(2\pi)^d} \int_{\text{BZ}} d\mathbf{k} \langle u_{\mathbf{k}} | \nabla_{\mathbf{k}} u_{\mathbf{k}} \rangle,\end{aligned}\tag{28}$$

where we have used $\mathbf{r}\psi_{\mathbf{k}}(\mathbf{r}) = i\nabla_{\mathbf{k}}\psi_{\mathbf{k}}(\mathbf{r}) - ie^{-i\mathbf{k}\cdot\mathbf{r}}\nabla_{\mathbf{k}}u_{\mathbf{k}}(\mathbf{r})$ and the fact the $\psi_{\mathbf{k}}(\mathbf{r})$ is a periodic function of \mathbf{k} . Now, performing the gauge transformation (15) $\tilde{\psi} = e^{i\phi}\psi$ we get

$$\begin{aligned}\langle \tilde{w}|\mathbf{r}|\tilde{w}\rangle &= \langle w|\mathbf{r}|w\rangle + \frac{V}{(2\pi)^d} \int_{\text{BZ}} d\mathbf{k} \nabla_{\mathbf{k}}\phi(\mathbf{k}) \\ &= \langle w|\mathbf{r}|w\rangle + \mathbf{R},\end{aligned}\tag{29}$$

where \mathbf{R} is a lattice vector. We have used the fact that $\phi(\mathbf{k})$ is smooth and periodic modulo 2π . Therefore, we can conclude that the position of the WF center is unique modulo a lattice vector \mathbf{R} . This uncertainty is not surprising if we keep in mind that it is the grid of WFs periodically repeated over the whole lattice which represents the energy band and that none of the lattice points has a special meaning. In case of composite bands it is the center of mass of all Wannier functions associated with a given lattice point \mathbf{R} which is unique up to a lattice translation.

Reality of Wannier function. It was found empirically that in systems with real Hamiltonian, e.g., without spin-orbit coupling and external magnetic field, the exponentially localized WFs obtained with the maximum-localization method [13] (discussed in section 4.1) are real, up to a trivial overall phase. It was conjectured [13] that real localized Wannier functions can be constructed in such systems. A simple criterion for existence of real localized WFs is presented in Ref. [12].

Symmetry constraints. In some cases Wannier functions may be uniquely defined by symmetry requirements. We present a simple example here. Let us assume that we have an isolated band and a real Wannier function $w(\mathbf{r})$ of definite parity (e.g., even). We show that such a WF is unique.

$$w(\mathbf{r}) = \frac{V}{(2\pi)^d} \int_{\text{BZ}} \psi_{\mathbf{k}}(\mathbf{r})\tag{30}$$

$$w^*(\mathbf{r}) = w(\mathbf{r}), \quad w(-\mathbf{r}) = w(\mathbf{r}).\tag{31}$$

Let us consider another WF obtained by a gauge transformation

$$\begin{aligned}\tilde{w}(\mathbf{r}) &= \frac{V}{(2\pi)^d} \int_{\text{BZ}} e^{i\phi(\mathbf{k})} \psi_{\mathbf{k}}(\mathbf{r}), \\ &= \frac{V}{(2\pi)^d} \sum_{\mathbf{R}} \left(\int_{\text{BZ}} e^{i\phi(\mathbf{k})} e^{i\mathbf{k}\cdot\mathbf{R}} \right) w(\mathbf{r} - \mathbf{R}).\end{aligned}\quad (32)$$

Requiring reality and evenness of $\tilde{w}(\mathbf{r})$ leads to a constraint on the gauge transformation $e^{i\phi(\mathbf{k})} = e^{-i\phi(\mathbf{k})}$ and thus $\phi(\mathbf{k}) = N\pi$ (with integer N), which means that $w(\mathbf{r})$ is unique up to a trivial change of sign.

While it is possible that reality and symmetry properties define exponentially localized WFs uniquely also for composite bands for cases with simple symmetry, we are not aware of rigorous proofs of such theorems. In general, the exponential localization only fixes the long-range behavior of the WFs, while the short range form depends on the particular computational method. For a discussion of point group aspects of construction of localized orbitals see Ref. [16].

4 Numerical methods for construction of Wannier functions

In the following we will describe two methods that are commonly used in connection with present band structure codes. We will also describe an approach for construction of Wannier functions in case of entangled bands, i.e., a situation where composite bands cannot be isolated from the rest of the spectrum. The first method is the Marzari and Vanderbilt maximally-localized WF construction [13], which consists in finding a gauge in which the second moment of the density distribution is minimized. The second method, proposed by Ku [17], is a realization of the projection construction (21). A common feature of these methods is that they start from a known set of the Bloch eigenstates of the Hamiltonian and search for the unitary transformation (16) according to some criteria. These two approaches are commonly used for construction of first principles tight-binding models that are used in the context of many-electron calculations, in particular LDA+DMFT.

4.1 Maximally localized Wannier functions

The maximal-localization (MALOC) method aims to minimize the spread functional

$$\Omega = \sum_n \langle r^2 \rangle_n - \langle \mathbf{r} \rangle_n^2, \quad (33)$$

where n is the WF index and $\langle \mathbf{r} \rangle_n$ is the position of the center of n th WF given by the \mathbf{r} -matrix element (28), while $\langle r^2 \rangle_n$ is the corresponding r^2 matrix element. The reader is invited to check that Ω depends neither on the choice of the particular lattice site \mathbf{R} associated with WF nor on the choice of the coordinate origin. Marzari and Vanderbilt proceed by splitting Ω into gauge

independent Ω_I and gauge dependent $\tilde{\Omega}$ parts

$$\Omega = \Omega_I + \tilde{\Omega} \quad (34)$$

$$\Omega_I = \sum_n \langle r^2 \rangle_n - \sum_{\mathbf{R}, n, m} |\langle \mathbf{R}, m | \mathbf{r} | 0, n \rangle|^2 \quad (35)$$

$$\tilde{\Omega} = \sum_n \sum_{\mathbf{R}, m \neq 0, m} |\langle \mathbf{R}, m | \mathbf{r} | 0, n \rangle|^2. \quad (36)$$

Expressed in terms of the Berry connection (23), $\tilde{\Omega}$ can be written as a sum of two terms

$$\tilde{\Omega} = \sum_n \frac{V}{(2\pi)^d} \int_{\text{BZ}} d\mathbf{k} |\mathbf{A}_{nn}(\mathbf{k}) - \langle \mathbf{r} \rangle_n|^2 - \sum_{n \neq m} \frac{V}{(2\pi)^d} \int_{\text{BZ}} d\mathbf{k} |\mathbf{A}_{mn}(\mathbf{k})|^2, \quad (37)$$

where $\mathbf{A}_{mn}(\mathbf{k})$ is a d -dimensional vector ($d=1,2,3$) and $|\cdot|^2$ is the corresponding norm. The task of constructing MALOC WFs can be formulated as follows: given some initial gauge we are looking for a gauge transformation which minimizes $\tilde{\Omega}$. Definition (23) of the Berry connection is not useful for numerical calculations and has to be replaced with a discrete approximation, which reduces to (23) and (37) in the continuum limit. Among the different possibilities Marzari and Vanderbilt choose

$$M_{mn}^{(\mathbf{k}, \mathbf{b})} = (u_{m, \mathbf{k}} | u_{n, \mathbf{k} + \mathbf{b}}) \quad (38)$$

and

$$\tilde{\Omega} = \frac{1}{N} \sum_{\mathbf{k}, \mathbf{b}} w_{\mathbf{b}} \sum_n (-\text{Im} \ln M_{nn}^{(\mathbf{k}, \mathbf{b})} - \mathbf{b} \cdot \langle \mathbf{r} \rangle_n)^2 + \frac{1}{N} \sum_{\mathbf{k}, \mathbf{b}} w_{\mathbf{b}} \sum_{n \neq m} |M_{mn}^{(\mathbf{k}, \mathbf{b})}|^2. \quad (39)$$

Here we assume to have a uniform \mathbf{k} -mesh throughout the Brillouin zone. For each \mathbf{k} -point a set of vectors \mathbf{b} defines connections to its neighbors. Moreover, we assume periodicity, which means that a connection is defined between points of the opposite sides of the Brillouin zone. The quadrature of (37) associates a weight $w_{\mathbf{b}}$ with each \mathbf{b} -vector [13]. The same discretization for $\langle \mathbf{r} \rangle_n$ leads to the expression

$$\langle \mathbf{r} \rangle_n = -\frac{1}{N} \sum_{\mathbf{k}, \mathbf{b}} w_{\mathbf{b}} \mathbf{b} \text{Im} \ln M_{nn}^{(\mathbf{k}, \mathbf{b})}. \quad (40)$$

Using the logarithm in the discrete approximation preserves the gauge invariance of $\langle \mathbf{r} \rangle_n$. It also ensures the proper shift of the WF centers by a lattice vector upon the gauge transformation $e^{i\mathbf{k} \cdot \mathbf{R}}$. However, the discretization is not without problems. In particular, for $M_{nn}^{(\mathbf{k}, \mathbf{b})}$ to provide an approximation of the continuum limit, we shall require that $1 - M_{nn}^{(\mathbf{k}, \mathbf{b})} \rightarrow 0$ as $\mathbf{b} \rightarrow 0$. This is typically not the case in numerical calculations, where phases of the eigenvectors are more or less random, no matter how small \mathbf{b} . In other words, the gauge transformation between the computer output and a smooth \mathbf{k} -dependence of $|u_{n, \mathbf{k}}\rangle$, which yields localized WFs, corresponds to an erratic function of \mathbf{k} . It is nearly impossible to find such a transformation by local minimization methods. Another way to see the problem is to realize that $M_{nn}^{(\mathbf{k}, \mathbf{b})}$ is complex and thus its logarithm is a multivalued function. A consistent choice of the branch leading to localized WFs becomes numerically impossible. Using more \mathbf{k} -points makes the problem even worse as it does not remove the random phase, but only makes the phase fluctuations sharper.

The way out consists in preprocessing the Bloch states so that the phase differences between nearby \mathbf{k} -points are small. This can be achieved by defining a set of trial orbitals $|M_n\rangle$, similar to (21). Projecting these onto the Bloch states

$$|\Phi_n(\mathbf{k})\rangle = \sum_m |\psi_{m,\mathbf{k}}\rangle \langle \psi_{m,\mathbf{k}} | M_n \rangle \quad (41)$$

$$S_{mn}(\mathbf{k}) = \langle \Phi_m(\mathbf{k}) | \Phi_n(\mathbf{k}) \rangle \quad (42)$$

and performing symmetric orthogonalization

$$|\tilde{\psi}_{n,\mathbf{k}}\rangle = \sum_m [S^{-1/2}(\mathbf{k})]_{mn} |\psi_{m,\mathbf{k}}\rangle \quad (43)$$

we get the desired result. Now, when constructing $M_{mn}^{(\mathbf{k},\mathbf{b})}$ from $|\tilde{\psi}_{n,\mathbf{k}}\rangle$ the phase of $M_{mn}^{(\mathbf{k},\mathbf{b})}$ is small and we can choose $|\text{Im} \ln M_{nn}^{(\mathbf{k},\mathbf{b})}| < \pi$. The functional $\tilde{\Omega}$ can then be minimized by a sequence of unitary transformations using the method of steepest-descent. For details we refer the reader to Ref. [13].

4.2 Projection method

It is possible to use only the initialization step described in the previous section to construct Wannier functions. This approach was adopted by Ku *et al.* [17] and used subsequently by others [18, 19]. It was successfully applied to investigation of oxides and similar materials for construction of atom-centered WFs. The advantage of this approach is that WFs retain the symmetry of the trial orbitals. The second advantage is that the construction is simple and guaranteed to converge as long as the overlaps with the trial functions are non-zero and localized WFs exist. This makes the method attractive for iterative procedures, where accidental freezing in a wrong local minima might spoil the calculation. This is typically the case when constructing tight-binding models, where we have a good idea of the shape and symmetry of the desired Wannier orbitals and want to get the quantitative information such as hopping parameters. On the other hand, when it comes to molecular orbitals or low-symmetry situations, where constructing good trial orbitals might be complicated, the MALOC method is preferable. This also applies to situations where the positions of Wannier centers are not fixed by symmetry and need to be calculated accurately, e.g., in ferroelectric materials.

4.3 Entangled bands

In some situations the number of constructed WFs is smaller than the number of Bloch bands. Typically this happens when no isolated band or composite bands can be found. In such a case it may also happen that the number of Bloch states per \mathbf{k} -point $N_{\mathbf{k}}$ varies although it must always be at least equal to N , the number of desired WFs. The technique to construct WFs in such situations is known as disentanglement and was introduced by Souza *et al.* [20]. The idea is to find an N -dimensional subspace $\mathcal{S}(\mathbf{k})$ of the $N_{\mathbf{k}}$ -dimensional space spanned by the Bloch states and \mathbf{k} such that the overlap between $\mathcal{S}(\mathbf{k})$ and $\mathcal{S}(\mathbf{k} + \mathbf{b})$ at neighboring \mathbf{k} -points is maximized

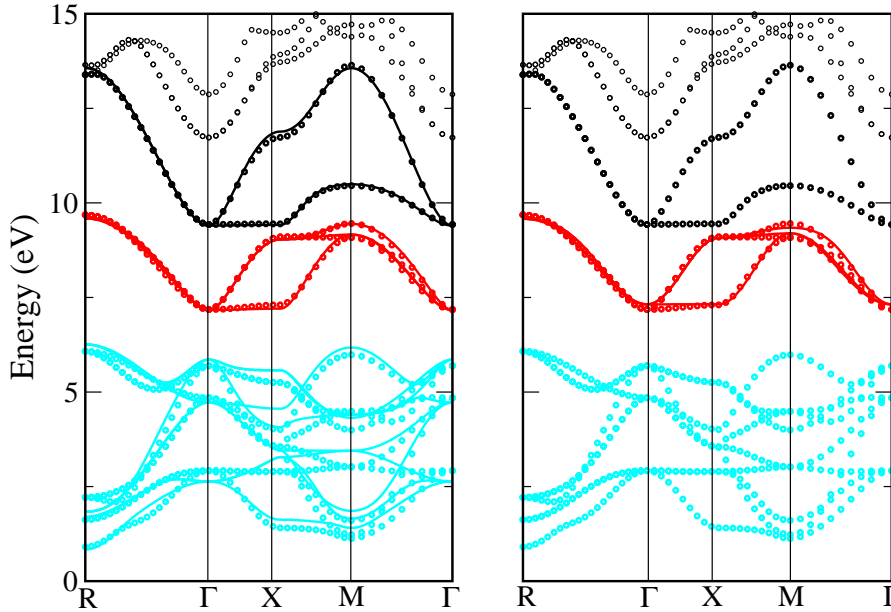


Fig. 5: LDA band structure of SrVO_3 (circles) with dominant contributions marked by color: O - p (blue), V - d - t_{2g} (red), and V - d - e_g (black). Tight-binding bands obtained from WFs for (i) [left panel] all V - d and O - p bands (ii) [right panel] only the t_{2g} band, with spatial cut-offs as described in the text (the Fermi level is at 8.16 eV).

in some sense. Souza *et al.* showed that this is equivalent to minimizing the Ω_I functional. The WF construction then proceeds in two steps. First, iterative minimization of Ω_I with respect to unitary transformations at each \mathbf{k} -point. Second, the above MALOC procedure performed using $\mathcal{S}(\mathbf{k})$'s from the first step. The details and examples of application can be found in [20].

The disentanglement procedure can also be realized with the projection technique [17]. In this case the subspace $\mathcal{S}(\mathbf{k})$ is chosen at each \mathbf{k} -point such that its overlap with a set of trial orbitals is maximal.

We have presented the two currently most popular methods for construction of Wannier functions in periodic solids. Examples of other methods can be found in Refs. [21, 22].

5 Examples of applications

In the following we present several examples that demonstrate applications of Wannier functions. This is a rather small sample of possible applications and focuses on problems connected with the construction of tight-binding models. For other important applications of WFs such as investigation of dielectric polarization the reader is referred to literature [3]. The examples presented below use the MALOC procedure implemented in the wannier90 package [23] starting from the electronic structures obtained with the wien2k code [24] and preprocessed with wien2wannier [25].

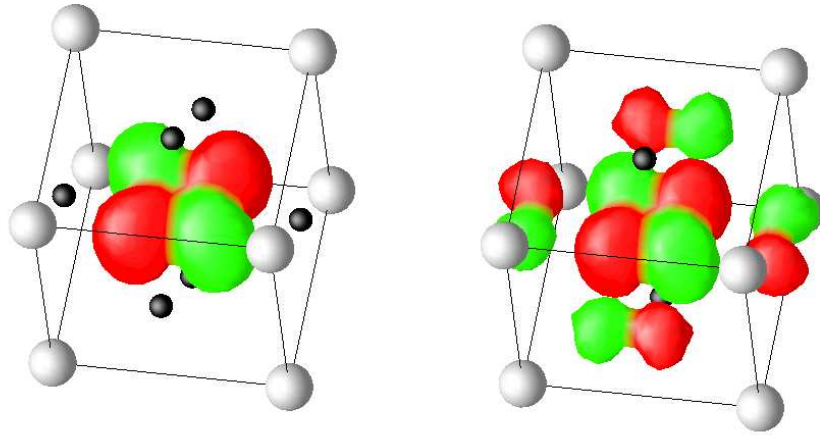


Fig. 6: The xy Wannier orbital plotted as an isosurface of the charge density $|w(\mathbf{r})|^2$ and colored by the sign of $w(\mathbf{r})$. The left panel corresponds to the large energy window (i), the right panel to the small energy window (ii) for the same isovalue.

5.1 Wide versus narrow energy window: SrVO_3

One of the basic parameters defining the WFs is the choice of the Hilbert space they should span. In practice this means the choice of bands or an energy window to be covered. In general, the larger the energy window the more localized WFs can be constructed. Taken to extreme, using the full spectrum of the Hamiltonian (up to infinite energy) we can build a basis out of Dirac delta functions. We will demonstrate the effect of the energy window on the localization of WFs for the example of a transition metal oxide by comparing two possible choices of the energy window (i) $3d+O-2p$ bands, (ii) $3d$ bands only.

SrVO_3 has a rather simple band structure (see Fig. 5) consisting of isolated groups of bands derived from $O-p$, $V-d-t_{2g}$ and $V-d-e_g$ orbitals. For choice (i) we use all $V-d$ and $O-p$ bands, while for (ii) we select only the $V-d$ bands of t_{2g} symmetry. The corresponding xy WF are shown in Fig. 6. The choice of the smaller energy window leads to the xy orbital having substantial weight on its oxygen neighbors. This is easy to understand. In terms of atomic-like V and O orbitals, the $V-t_{2g}$ band is anti-bonding, while the $O-2p$ band is bonding. When we choose to build WFs from the anti-bonding band only, we cannot recover the atomic orbitals. The $O-2p$ atomic character contained in the anti-bonding band cannot disappear. It is preserved in the more extended shape of WFs. Note, that the anti-bonding character is apparent in the WF having a node (changes sign) between the O and V sites. Using a larger energy window (i) we include both bonding and anti-bonding bands. This allows us to construct WFs which closely resemble atomic V and O functions. There is still some $V-xy$ weight on the O sites to ensure orthogonality of the WFs, but it is substantially less than in case (ii) where the O tails of the xy functions are enforced by hybridization.

The spatial extent of the Wannier orbitals is also reflected in the hopping integrals. With the small energy window (ii) the t_{2g} bands are well described when at least nn- and nnn-hoppings

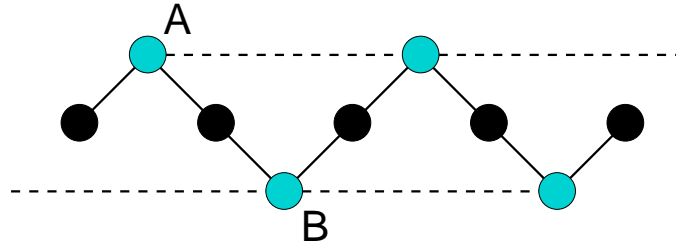


Fig. 7: A simple zig-zag chain with two types of atoms (black and blue). The lines indicate possible hoppings. (including longer-range hoppings does not break the A-B symmetry).

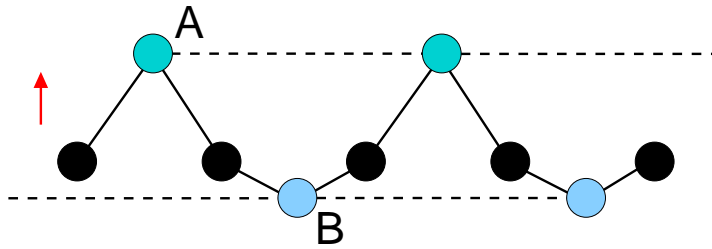


Fig. 8: The zig-zag chain in perpendicular electric field, which breaks the A-B symmetry as indicated by a uniform shift of blue atoms.

are considered (see Fig. 5), which in the $\{xy, yz, zx\}$ basis read

$$t_{100}[\text{meV}] = \begin{pmatrix} -268 & 0 & 0 \\ 0 & -30 & 0 \\ 0 & 0 & -268 \end{pmatrix}, \quad t_{101}[\text{meV}] = \begin{pmatrix} 7 & 10 & 0 \\ 10 & 7 & 0 \\ 0 & 0 & -93 \end{pmatrix}. \quad (44)$$

The longest nnn $t_{2g-t_{2g}}$ hopping corresponds to a length of of 5.4 Å. Using the more localized orbitals (i) we can achieve similar accuracy (see Figure 5) by considering only V-V nn-hopping and V-O nnn-hopping, which translate into a direct spatial cut-off of only 4.3 Å. The obvious price to be paid are larger matrices (14×14 for (i) versus 3×3 in case (ii)).

5.2 Unfolding band structures: LaOFeAs

Wannier functions provide a tool for an unbiased and quantitatively accurate construction of tight binding models. In some situations it is possible that the tight-binding Hamiltonian has a symmetry higher than the original band structure. The recently much studied LaOFeAs is a good example. However, before we discuss this material let us consider the simpler case of a zig-zag chain shown in Fig. 7. Geometrically, the sites A and B are not related by a translation. However, when we consider the symmetry of the graph representing the hoppings (full and dashed lines are shown as an example) only the black and the blue sites are distinct. Another way to see the equivalence between A and B sites is to realize that they are connected by a 'gauge' transformation which changes the sense of 'up' and 'down'. Only when we add a field which breaks the up-down symmetry (e.g. uniform electric field perpendicular to chain) the

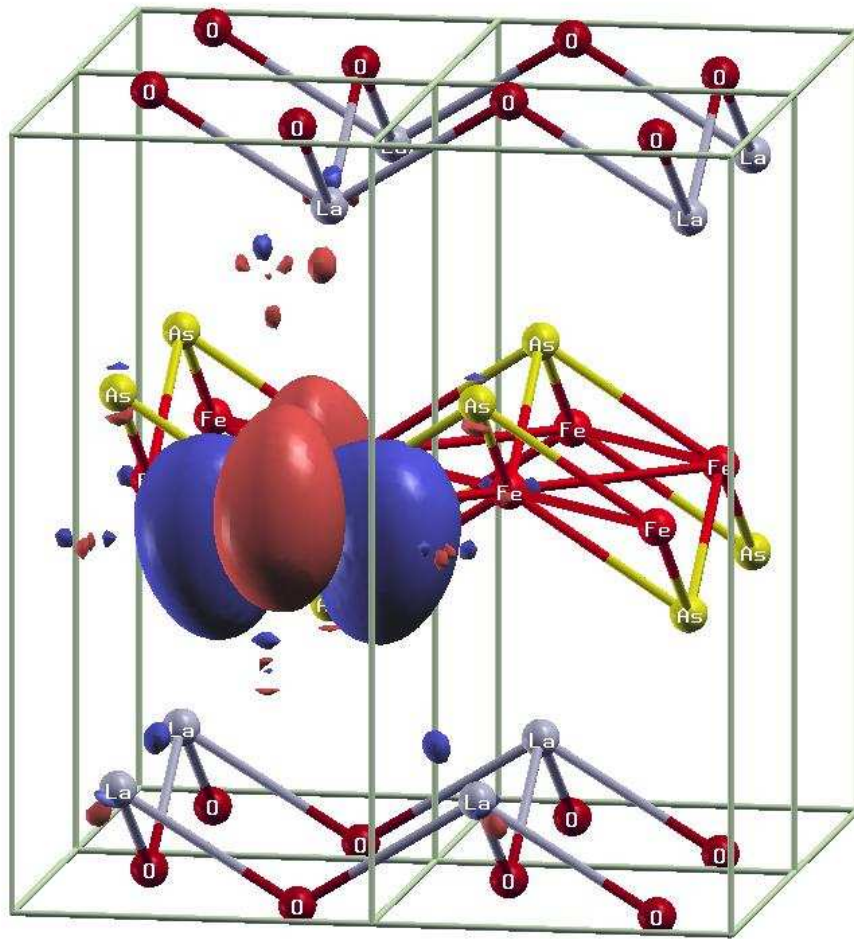


Fig. 9: The Fe x^2-y^2 Wannier orbital for the $d-p$ model of LaOFeAs. The deviation from the atomic x^2-y^2 orbital reflecting the geometry of Fe-As plane is clearly visible.

A-B equivalence is broken (see Fig. 8). The Fe-As plane in LaOFeAs with As atom alternating above and below the Fe plane resembles the zig-zag chain.

As bandstructure codes use the geometrical symmetry of the crystal structure, the electronic structure calculation is performed in a unit cell with two Fe atoms and the corresponding Brillouin zone. Unfolding the bandstructure amounts to doubling the Brillouin zone and halving the number of bands. A straightforward way to unfold the bandstructure is to construct the WFs and the corresponding tight-binding Hamiltonian. After a simple geometrical transformation the intra-cell hoppings can be converted into inter-cell ones, to obtain a tight-binding Hamiltonian for the smaller unit cell with a single Fe atom. The construction of the corresponding unfolded bandstructure is straightforward. The fact that the corresponding crystal structure has a lower symmetry of course cannot disappear. It is reflected in the shape of the WFs (see Fig. 9) and the fact that WFs centered at the two Fe sites in the original unit cell are not connected by a translation, but a translation combined with a 90 deg rotation along the z -axis.

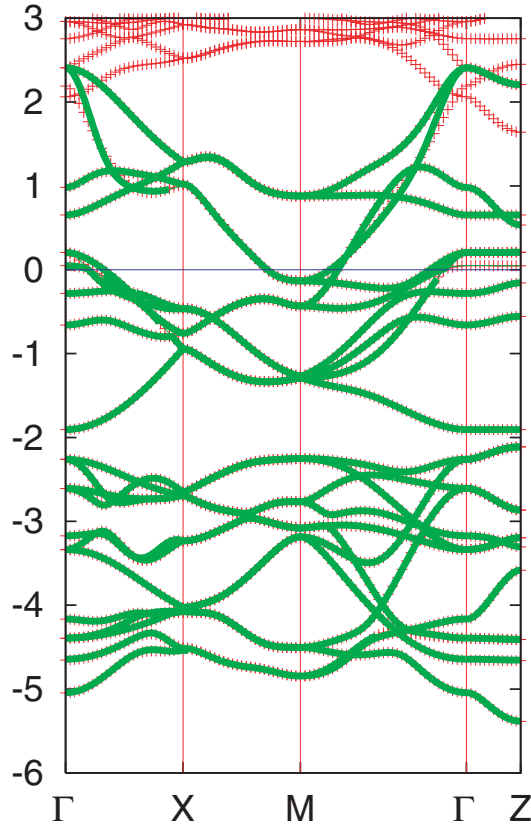


Fig. 10: LDA band structure of LaOFeAs (red) compared with tight-binding bands obtained from WFs (green) for the $d-p$ model. Note the difference above 2 eV.

5.3 Entangled bands

Staying with LaOFeAs, we will demonstrate a typical application of the band disentanglement technique. As before, we are interested in constructing an effective Hamiltonian of LaOFeAs in terms of Fe- d and As- p WFs. However, we see from Fig. 10 that the Fe- d bands overlap with the higher lying bands (above ~ 2 eV). The hybridization between the overlapping bands leads to admixing of Fe- d character to these higher lying bands and to band anti-crossing, which modifies the band topology, e.g., around the Γ point. In order to reproduce the bands around 2 eV at Γ precisely, we would have to either enlarge our model to include other orbitals or work with more extended Fe- d WFs. However, the details of band topology at 2 eV above the Fermi level are of little physical significance. It is preferable to work with an approximate bandstructure, which can be obtained with localized Fe- d WFs in a $d-p$ model. What we want to achieve can be viewed roughly as follows. Imagine we build a WF representation for a much larger energy window than we are interested in. Subsequently we switch off all the hoppings between our target Fe- d and As- p WFs and anything else. The resulting band structure would closely resemble the original one around the chemical potential where the bands are dominated by Fe- d and As- p character. Deviations will appear at energies and k-space regions where the neglected hoppings play a role, e.g. 2 eV at Γ -point. The disentanglement procedure performs essentially the same task, however, without the need to construct the large tight-binding model

first. In fact, it does even better. Using the above picture, the disentanglement procedure does not put all the hoppings between Fe- d and As- p strictly to zero, but minimizes them in such a way that the band dispersion in a specified energy window is reproduced precisely, which in practice means it produces Fe- d and As- p WFs that are a bit more extended than they would have been in the large tight-binding model.

From the discussion it is clear that the disentanglement procedure may find its use even in some cases with isolated (composite) bands when one decides to trade the accuracy away from the region of interest (Fermi level) for having more localized WFs.

5.4 Spin-orbit coupling: Sr_2IrO_4

The electronic states in condensed matter physics are commonly described in the quasi-relativistic approximation, i.e., using two component spinors. In many applications, in particular using lattice models, the spin projection is conserved. Therefore we can work in a basis in which the Bloch eigenstates have only one non-zero spinor component and forget about the spinor structure. The most common situation where non-trivial spinor structure survives are systems with spin-orbit coupling (another less common example are broken-symmetry systems with a non-collinear magnetic order). However, since the construction of WFs amounts to a unitary transformation from the space indexed by (n, \mathbf{k}) , the band indices and \mathbf{k} -vectors, to the space, indexed by (m, \mathbf{R}) , the orbital indices and the lattice vectors, the spinor character of the electronic functions does not change the discussion in the previous sections. It merely changes the definition of the dot product in (23) and (38) from

$$\langle \psi | \psi' \rangle = \int d\mathbf{r} \psi^*(\mathbf{r}) \psi'(\mathbf{r}) \quad (45)$$

to

$$\langle \psi | \psi' \rangle = \sum_{\sigma} \int d\mathbf{r} \psi_{\sigma}^*(\mathbf{r}) \psi'_{\sigma}(\mathbf{r}), \quad (46)$$

where σ indexes the spinor components

$$|\psi\rangle = \begin{pmatrix} \psi_{\uparrow} \\ \psi_{\downarrow} \end{pmatrix}. \quad (47)$$

The transformations (13) or (16) now connect the Bloch spinor functions to Wannier spinor functions. The spinors must be viewed as the basic objects. For example a MALOC procedure applied to the spinor components separately will not lead to a sensible result.

In the following we use Sr_2IrO_4 as an example of a material where spin-orbit coupling substantially modifies the band structure and leads to Wannier orbitals in which both spin projections are mixed. For the sake of simplicity, we have performed the calculations using an idealized double-perovskite crystal structure. The electronic structure (see Fig. 11) can be understood by considering crystal-field splitting, spin-orbit coupling and inter-site hopping between Ir- d orbitals. The crystal-field splitting, being the largest of the three, opens a gap between the t_{2g} and e_g bands, rendering the latter empty, while the former accommodate one hole per Ir atom.

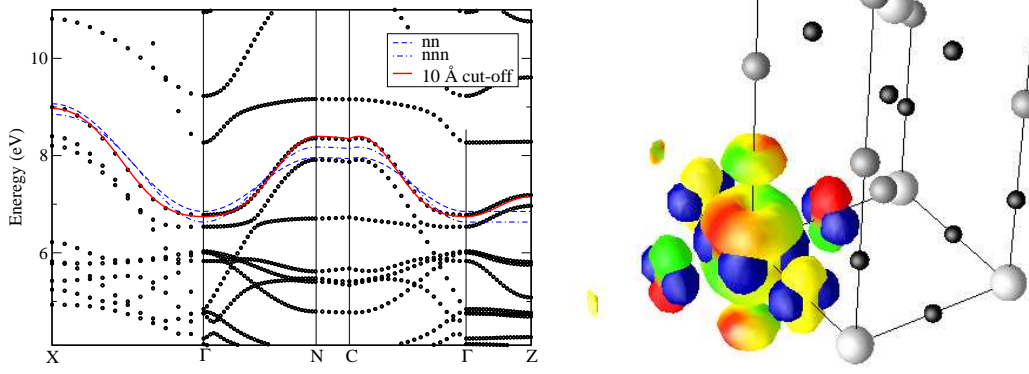


Fig. 11: *Ir* $J = 1/2$ Wannier orbital $w_{+, [0,0,0]}(\mathbf{r})$ visualized as a $|w|^2$ isosurface. The almost real \downarrow -spin component is blue, the \uparrow -spin component is colored with cosine of its phase (red=real positive, green=real negative, yellow=0).

The t_{2g} orbitals may be labeled with a pseudo-spin $I = 1$. The spin-orbit coupling splits the t_{2g} manifold into a quadruplet and doublet with pseudo-spins $J = 3/2$ and $J = 1/2$, respectively. Since the spin-orbit splitting is rather large, the inter-site hopping leads only to moderate mixing of the states with different J . Therefore we may expect the isolated band doublet at the top of the t_{2g} manifold to be predominantly of $J = 1/2$ character. We construct the MALOC WFs for this two-fold degenerate band. Expressed in the local coordinate system pointing towards the O atoms, the $J = 1/2$ spinor functions adopt the form

$$|\phi_+\rangle \sim \begin{pmatrix} -2Y_{21} \\ Y_{2-2} - Y_{22} \end{pmatrix}, \quad |\phi_-\rangle \sim \begin{pmatrix} Y_{2-2} - Y_{22} \\ 2Y_{2-1} \end{pmatrix}. \quad (48)$$

The MALOC WFs are indeed very close to this form as shown in Fig. 11. In particular, the WF, to a very good approximation, consists of a real xy orbital in one spin channel and a complex $(x \pm iy)z$ orbital in the other (\pm) spin channel. Note that the relative phase of the two components is not arbitrary and that the corresponding charge density has approximately cubic symmetry as expected for a $J = 1/2$ orbital. Also, similar to WFs in SrVO_3 the anti-bonding character with respect to oxygen is reflected in the sign-change (red-green) between the lobes on Ir and its O neighbors.

5.5 Wannier functions in dynamical mean-field theory

WFs and the techniques described in these notes are often used to construct Hubbard type Hamiltonians for real materials

$$H = \sum_{\mathbf{k}} \left([h^{dd}(\mathbf{k})]_{\alpha\beta} d_{\mathbf{k}\alpha}^\dagger d_{\mathbf{k}\beta} + [h^{pp}(\mathbf{k})]_{\gamma\delta} p_{\mathbf{k}\gamma}^\dagger p_{\mathbf{k}\delta} + [h^{dp}(\mathbf{k})]_{\alpha\gamma} d_{\mathbf{k}\alpha}^\dagger p_{\mathbf{k}\gamma} + [h^{pd}(\mathbf{k})]_{\gamma\alpha} p_{\mathbf{k}\gamma}^\dagger d_{\mathbf{k}\alpha} \right) + \sum_i U_{\alpha\beta\gamma\delta} d_{i\alpha}^\dagger d_{i\beta}^\dagger d_{i\gamma} d_{i\delta} - H_{\text{dc}}, \quad (49)$$

where we have denoted electrons subject to local interactions as d -type and non-interacting electrons as p -type. The \mathbf{k} -indexed operators are Bloch sums of the corresponding direct space

operators and summation over the Greek indices is assumed. The double-counting term H_{dc} is a one-particle operator discussed below. Obviously, the results of any theory should depend only on physical parameters and not on the coordinate system or basis used to formulate the problem. How can this requirement be reconciled with the non-uniqueness in the definition of the Wannier orbitals?

First, let us emphasize that the first term in (49) does not depend on the choice of the WFs. Its form is the same in any WF basis. While the numerical values of the matrices $h^{dd}(\mathbf{k})$, $h^{pp}(\mathbf{k})$, $h^{pd}(\mathbf{k})$ and $h^{dp}(\mathbf{k})$ are basis dependent, the spectrum of the operator is not. The same is not true for the second term. Here a change of the basis leads to terms, which are not present in (49) such as inter-site interactions. Therefore the model (49) is always basis dependent and we can only require that the dependence is weak and that the choice of a particular basis is physically motivated. We can heuristically reason as follows. The electron-electron scattering exists between any four states (subject to conservation laws). However, to describe the physics on the eV scale, the majority of the interaction terms can be treated at mean-field level. These terms are implicitly included in the one-particle part of (49). Only those terms leading to sizeable dynamical correlation effects have to be treated explicitly as two-particle operators. The choice of the WF basis thus amounts to deciding which part of the electron-electron interaction can be approximated by a mean-field decoupling and which part should be kept explicitly. The double-counting term H_{dc} is used to avoid counting the same interaction twice, once on a mean-field level in the first term and once explicitly. Following this reasoning, it appears physically well justified to work with as localized WF orbitals as possible since this amounts to the best possible separation of intra- and inter-atomic electron-electron interaction. We emphasize that we have just presented a physical picture rather than a mathematical framework. In practice the starting band structures are not obtained from Hartree-Fock theory (i.e., a self-consistent mean-field solution of the electron Hamiltonian), but from density functional theory, which has a similar mathematical structure, but uses essentially an empirical one-particle potential, which is known to yield better results than Hartree-Fock theory. Using the empirical potential makes it impossible to rigorously define what the double-counting term H_{dc} should be. Therefore, we are left with heuristic arguments to define H_{dc} and various forms are being used with varied success for different groups of materials.

In Fig. 12 a flow chart of a typical LDA+DMFT [26] calculation is shown. It consists of two more or less independent steps. First, density-functional theory (usually in LDA or similar approximation) is used to obtain the band structure of a given material. This involves iteration of the charge distribution. Once the LDA calculation is converged, results are postprocessed to obtain the model Hamiltonian (49). This involves the physically motivated selection of the relevant part of the Hilbert space (specified through the energy window or band indices) and the construction of the corresponding Wannier functions. The one-particle part of (49) is obtained by a straightforward unitary transformation between the Bloch and Wannier functions. Determination of the interaction parameters is less straightforward since it has to account for various screening processes, in particular those arising from electronic states excluded from the model. Several approaches are possible: (i) $U_{\alpha\beta\gamma\delta}$ is treated as a free parameter or adjusted to some ex-

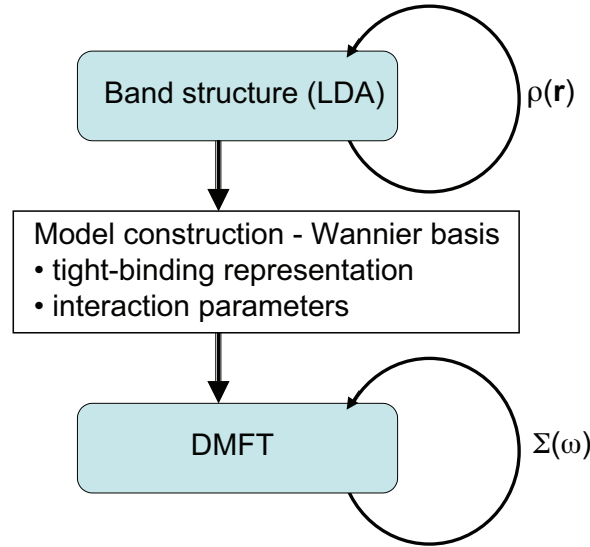


Fig. 12: A flow chart of typical LDA+DMFT calculation. Two separate self-consistency loops over the charge density $\rho(\mathbf{r})$ and self-energy $\Sigma(\omega)$ are involved.

perimental input, or it is calculated using (ii) constrained LDA method [27] or (iii) constrained RPA [28]. The computational approaches (ii) and (iii) are based on the response of the effective LDA band structure to external perturbations. The second step in the LDA+DMFT approach consists in DMFT treatment of model (49), which involves iteration over the one-particle self-energy. Since the dynamical effects in DMFT often lead to some charge redistribution, it may be desirable to have the charge and self-energy self-consistency in the same loop [19].

6 Summary

Wannier functions provide an extremely useful set of orthogonal functions, which can represent isolated single or composite bands exactly or entangled bands with adjustable accuracy or resolution. They are, in a sense, an optimal compromise between localization in direct space and in the energy domain. The long-range behavior of WFs reflects the smoothness of the Bloch functions as functions of quasi-momentum \mathbf{k} . Existence of exponentially localized WFs is a topological property of a given (composite) band. While in commonly studied cases, e.g., time reversal symmetry and no spin-orbit coupling, exponentially localized WFs can be constructed, the bands for which this is not possible are of great interest as they exhibit anomalies such as topologically protected surface currents.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] for review see R.M. Dreizler and E.K.U. Gross, *Density functional theory* (Springer-Verlag, Berlin, 1990)
- [2] for example in Ch. Kittel, *Introduction to Solid State Physics* (Wiley, New York, 1996)
- [3] R.D. King-Smith and D. Vanderbilt, *Phys. Rev. B* **47**, 1651 (1993)
- [4] G.H. Wannier, *Phys. Rev.* **52**, 191 (1937)
- [5] for details see e.g. G. P. Tolstov, *Fourier Series* (Dover Publications, 1962)
- [6] www-math.mit.edu/~stevenj/bump-saddle.pdf
- [7] W. Kohn, *Phys. Rev.* **115**, 809 (1959)
- [8] J. des Cloizeaux, *Phys. Rev.* **135**, A685 (1964)
- [9] J. des Cloizeaux, *Phys. Rev.* **135**, A698 (1964)
- [10] M. Kohmoto, *Ann. Phys.* **160**, 343 (1985)
- [11] G. Nenciu, *Commun. Math. Phys.* **91**, 81 (1983)
- [12] C. Brouder, G. Panati, M. Calandra, C. Mourougane, and N. Marzari, *Phys. Rev. Lett.* **98**, 046402 (2007)
- [13] N. Marzari and D. Vanderbilt, *Phys. Rev. B* **56**, 12847 (1997)
- [14] G. Panati, *Ann. Henri Poincaré* **8**, 995 (2007)
- [15] M.Z. Hasan and C.L. Kane, *Rev. Mod. Phys.* **82**, 3045 (2010)
- [16] J. des Cloizeaux, *Phys. Rev.* **129**, 554 (1963)
- [17] Wei Ku, H. Rosner, W.E. Pickett, and R.T. Scalettar, *Phys. Rev. Lett.* **89**, 167204 (2002)
- [18] V.I. Anisimov *et al.*, *Phys. Rev. B* **71**, 125119 (2005)
- [19] M. Aichhorn *et al.*, *Phys. Rev. B* **80**, 085101 (2009)
- [20] I. Souza, N. Marzari, and D. Vanderbilt, *Phys. Rev. B* **65**, 035109 (2001)
- [21] W. Kohn and J.R. Onffroy, *Phys. Rev. B* **8**, 2485 (1973)
- [22] W. Kohn, *Phys. Rev. B* **10**, 382 (1974)
- [23] A.A. Mostofi, J.R. Yates, Y.-S. Lee, I. Souza, D. Vanderbilt and N. Marzari, *Comput. Phys. Commun.* **178**, 685 (2008)

-
- [24] P. Blaha, K. Schwarz, G.K.H. Madsen, D. Kvasnicka, J. Luitz: *WIEN2K, An Augmented Plane Wave and Local Orbitals Program for Calculating Crystal Properties* (Vienna University of Technology, Austria, 2001), K. Schwarz, P. Sorantin, S.B. Trickey, *Comput. Phys. Commun.* **59**, 399 (1990)
- [25] J. Kuneš, R. Arita, P. Wissgott, A. Toschi, H. Ikeda, and K. Held, *Comput. Phys. Commun.* **181**, 1888 (2010)
- [26] a recent review G. Kotliar *et al.*, *Rev. Mod. Phys.* **78**, 865 (2006)
- [27] O. Gunnarsson, O.K. Andersen, O. Jepsen, and J. Zaanen, *Phys. Rev. B* **39**, 1708 (1989)
V.I. Anisimov and O. Gunnarsson, *Phys. Rev. B* **43**, 7570 (1991)
- [28] F. Aryasetiawan *et al.*, *Phys. Rev. B* **70**, 195104 (2004)

5 Introduction to Dynamical Mean-Field Theory

Marcus Kollar

Center for Electronic Correlations and Magnetism

University of Augsburg

Contents

1	Introduction	2
2	Fermions in infinite dimensions	6
3	Simplifications for many-body theory	9
4	Dynamical mean-field theory	12
5	Summary and outlook	17

1 Introduction

The previous lectures have already discussed several aspects of the physics of correlated electrons in solids, i.e., electrons for which the Coulomb interaction is important and whose behavior cannot be assumed to be independent of one another. Several ingredients are necessary to successfully describe such correlated materials: the Hamiltonian describing the electronic physics must be obtained, both the interaction part and the kinetic part (i.e., the band structure as determined with density functional theory, which also provides a suitable basis). Then the interaction must be treated reliably, and for this dynamical mean-field theory (DMFT) and related theories provide a controlled approach. The spirit and some technical aspects of DMFT were already mentioned in the Lecture of D. Vollhardt. The purpose of the present chapter is to provide a derivation of DMFT (one of many possible derivations, see, e.g., Ref. [1–3]), which becomes exact in the limit of infinite spatial dimensions.

We assume that the band structure and interaction are known, leading to a one- and two-body Hamiltonian of the type

$$H = \sum_{ij\alpha\beta\sigma} t_{ij}^{\alpha\beta} c_{i\alpha\sigma}^+ c_{j\beta\sigma} + \frac{1}{2} \sum_{\substack{ijkl \\ \alpha\beta\gamma\delta\sigma\sigma'}} V_{ijkl}^{\alpha\beta\gamma\delta} c_{i\alpha\sigma}^+ c_{j\beta\sigma'}^+ c_{l\delta\sigma'} c_{k\gamma\sigma}. \quad (1)$$

For the present purpose we assume that this Hamiltonian can be reduced further: we keep only a single band and only the on-site Hubbard interaction $U = V_{iiii}$, leading to the single-band Hubbard model:

$$H = H_0 + H_1, \quad H_1 = U \sum_i n_{i\uparrow} n_{i\downarrow}, \quad (2a)$$

$$H_0 = \sum_{ij\sigma} t_{ij} c_{i\sigma}^+ c_{j\sigma} = \sum_{\mathbf{k}\sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^+ c_{\mathbf{k}\sigma}, \quad (2b)$$

where t_{ij} is the hopping amplitude from site i to j , whose Fourier transform is the dispersion relation $\epsilon_{\mathbf{k}}$.

We begin by reviewing some definitions and basic concepts of many-body physics that are useful for the formulation and application of DMFT. In Sec. 2 we consider the limit of infinite dimensions and analyze what happens to the kinetic Hamiltonian H_0 in this limit; in particular the hopping matrix elements must be scaled correctly with the diverging lattice dimension. In Sec. 3 we discuss what happens to the many-body perturbation series as a consequence of this scaling, i.e., that the self-energy becomes local. Finally, it is shown how this local self-energy can actually be calculated in DMFT (Sec. 4).

Green functions

An important dynamical quantity which measures the equilibrium properties of a correlated electron system is the electronic Green function [4, 5]. In general a Green function G_{AB} is defined as an expectation value of operators A and B taken at different (real or imaginary) times

in a thermal state, i.e., with density matrix $\propto \exp(-\beta(H - \mu N))$ corresponding to the temperature $T = 1/\beta$, or possibly the ground state. Hence it measures the probability amplitude for a propagation of a particle or hole excitation in an equilibrium state if A and B are annihilation and creation operators.

In finite-temperature problems one often uses the imaginary-time-ordered (fermionic) single particle Green function $G_{\alpha\beta}(\tau)$ (we put $A = c_\alpha$, $B = c_\beta^+$):¹

$$G_{\alpha\beta}(\tau) = -\langle T_\tau c_\alpha(\tau) c_\beta^+(0) \rangle = - \begin{cases} \langle c_\alpha(\tau) c_\beta^+(0) \rangle & \tau > 0 \\ -\langle c_\beta^+(0) c_\alpha(\tau) \rangle & \tau \leq 0 \end{cases} \quad (3)$$

$$= -G_{\alpha\beta}(\tau + \beta) \quad \text{for } -\beta < \tau < 0, \quad (4)$$

with imaginary-time Heisenberg operators $A(\tau) = e^{H\tau} A e^{-H\tau}$; note that $A^+(\tau) \neq A(\tau)^+$. Its dependence on time difference only and the anti-periodicity (4) follow from the cyclic properties of the trace. A Fourier transform yields the Matsubara Green function $G_{\alpha\beta}(i\omega_n)$:

$$G_{\alpha\beta}(i\omega_n) = \int_0^\beta d\tau G_{\alpha\beta}(\tau) e^{i\omega_n \tau}, \quad G_{\alpha\beta}(\tau) = T \sum_{n=-\infty}^{+\infty} G_{\alpha\beta}(i\omega_n) e^{-i\omega_n \tau}, \quad (5)$$

with fermionic Matsubara frequencies $i\omega_n = 2\pi T(n + \frac{1}{2})$. It is useful to note the spectral representation

$$G_{\alpha\beta}(i\omega_n) = \int_{-\infty}^{\infty} d\omega \frac{A_{\alpha\beta}(\omega)}{i\omega_n - \omega}, \quad (6)$$

with the spectral function given by its Lehmann representation as (Z : partition function, E_n : eigenvalues of $H - \mu N$)

$$A_{\alpha\beta}(\omega) = \frac{1}{Z} \sum_{n,m} \langle n | c_\beta^+ | m \rangle \langle m | c_\alpha | n \rangle (e^{-\beta E_m} - e^{-\beta E_n}) \delta(\omega - (E_n - E_m)). \quad (7)$$

In particular $A_{\alpha\alpha}(\omega) \geq 0$. Note that in practice the spectral or Green function can be evaluated via the Lehmann representation only for sufficiently small systems, i.e., when the many-body energy eigenvalues and eigenstates can be obtained directly.

From the spectral function other single-particle Green functions can also be obtained, such as the retarded Green function

$$G_{\alpha\beta}^{\text{ret}}(\omega) = \int_{-\infty}^{\infty} d\omega' \frac{A_{\alpha\beta}(\omega')}{\omega + i0^+ - \omega'}, \quad (8)$$

which corresponds to a Green function in the time domain that involves real-time Heisenberg operators. We note that

$$A_{\alpha\beta}(\omega) = -\frac{1}{\pi} \text{Im} G_{\alpha\beta}^{\text{ret}}(\omega), \quad (9)$$

¹Note that the prefactor -1 is omitted from the definition in Ref. [5].

and that the retarded Green function can be obtained from the Matsubara Green function by the analytic continuation from $i\omega_n$ to $\omega + i0^+$. In view of the spectral representations (6) and (8) we will often write $G_{\alpha\beta}(\omega)$ for both the Matsubara or retarded Green function, with the understanding that the argument is either $i\omega_n$ for the former and $\omega + i0^+$ for the latter (and hence is never purely real).

The indices α, β, \dots can represent lattice site or momentum \mathbf{k} , as well as spin index σ (and possibly orbital or band index). The real-space and momentum-space Green functions are related by a Fourier transform. We will work in particular with the local Green function (L : number of lattice sites)

$$G_{ii\sigma}(\omega) = G_{\sigma}(\omega) = \frac{1}{L} \sum_{\mathbf{k}} G_{\mathbf{k}\sigma}(\omega), \quad A_{ii\sigma}(\omega) = A_{\sigma}(\omega) = -\frac{1}{\pi} \text{Im} G_{\sigma}(\omega + i0^+), \quad (10)$$

assuming translational invariance.

Free particles, with Hamiltonian $H - \mu N = \sum_{\mathbf{k}\sigma} (\epsilon_{\mathbf{k}} - \mu) c_{\mathbf{k}\sigma}^+ c_{\mathbf{k}\sigma}$, are characterized by the free Green function $G_{\mathbf{k}\sigma}^{(0)}(\omega)$ and the free density of states $\rho(\epsilon)$,²

$$G_{\mathbf{k}\sigma}^{(0)}(\omega) = \frac{1}{\omega + \mu - \epsilon_{\mathbf{k}}}, \quad \rho(\omega) = A_{\sigma}^{(0)}(\omega) = \frac{1}{L} \sum_{\mathbf{k}} \delta(\omega - \epsilon_{\mathbf{k}}). \quad (11)$$

For interacting systems the self-energy $\Sigma_{\mathbf{k}}(\omega)$ is defined so that it measures the difference between interacting and free Green functions:

$$G_{\mathbf{k}\sigma}(\omega)^{-1} = G_{\mathbf{k}\sigma}^{(0)}(\omega)^{-1} - \Sigma_{\mathbf{k}\sigma}(\omega), \quad G_{\mathbf{k}\sigma}(\omega) = \frac{1}{\omega + \mu - \epsilon_{\mathbf{k}} - \Sigma_{\mathbf{k}\sigma}(\omega)}. \quad (12)$$

For a translationally invariant system the Green function and self-energy are diagonal in momentum space. It can also be useful instead to use a matrix notation in site indices, $G_{ij\sigma}(i\omega_n) = (\mathbf{G})_{ij,\sigma,n}$ etc., for which

$$\mathbf{G}^{-1} = \mathbf{G}^{(0)-1} - \mathbf{\Sigma}, \quad \mathbf{G} = \mathbf{G}^{(0)} + \mathbf{G}^{(0)} \mathbf{\Sigma} \mathbf{G}. \quad (13)$$

Eq. (12) or (13) are referred to as the (lattice) Dyson equation. The Dyson equation (in any basis) can be expressed with Feynman diagrams as

$$\text{====} = \text{——} + \text{——} \circlearrowleft \Sigma \text{====}. \quad (14)$$

We will discuss Feynman diagrams for the self-energy in Sec. 3.

Path-integral formulation

Another useful technique to work with Green functions is the path integral representation [5]. The partition function and the imaginary-time-ordered Green function for the fermionic Hamil-

²In the thermodynamic limit ($L \rightarrow \infty$), the sum over the first Brillouin zone in (11) can be replaced by an integral, see e.g. (28) below.

tonian $H(\{c_\alpha^+\}, \{c_\alpha\})$ can be written in terms of functional integrals over Grassmann variables,

$$Z = \text{Tr} e^{-\beta(H-\mu N)} = \int_{\phi_\alpha(\beta)=-\phi_\alpha(0)} \mathcal{D}(\phi_\alpha^*(\tau), \phi_\alpha(\tau)) \exp(\mathcal{A}), \quad (15)$$

$$G_{\alpha\beta}(\tau) = \frac{1}{Z} \int_{\phi_\alpha(\beta)=-\phi_\alpha(0)} \mathcal{D}(\phi_\alpha^* \phi) \phi_\alpha(\tau) \phi_\beta^*(0) \exp(\mathcal{A}), \quad (16)$$

with the action

$$\mathcal{A} = - \int_0^\beta d\tau \left[\sum_\alpha \phi_\alpha^* (\partial_\tau - \mu) \phi_\alpha + H(\{\phi_\alpha^*\}, \{\phi_\alpha\}) \right]. \quad (17)$$

Note that the Grassmann fields $\phi_\alpha^*(\tau)$ and $\phi_\alpha(\tau)$ are independent (i.e., they are not complex or hermitian conjugates of each other, even though they represent creation and annihilation operators) and antiperiodic boundary conditions are imposed on the latter. Strictly speaking, path-integral expressions such as (16) and (17) are merely shorthand for limits of expressions that are discretized in imaginary time τ . We refer to Ref. [5] for details.

Quasiparticles

Without interactions single-particle excitations can be created and propagated freely. In a large many-body system with interactions, on the other hand, particle or hole excitations will usually be damped and have a finite lifetime. This is encoded in the complex (retarded) self-energy $\Sigma_{\mathbf{k}}(\omega)$, in terms of which the spectral function becomes

$$A_{\mathbf{k}}(\omega) = \frac{1}{\pi} \frac{\text{Im}\Sigma_{\mathbf{k}}(\omega)}{(\omega + \mu - \epsilon_{\mathbf{k}} - \text{Re}\Sigma_{\mathbf{k}}(\omega))^2 + (\text{Im}\Sigma_{\mathbf{k}}(\omega))^2}. \quad (18)$$

This reduces to a δ function only if $\text{Im}\Sigma_{\mathbf{k}}(\omega) \rightarrow 0^-$. On the other hand, if $\text{Im}\Sigma_{\mathbf{k}}(\omega)$ is finite and not too large, the maxima of $A_{\mathbf{k}}(\omega)$ are located approximately at the zeros $\omega = E_{\mathbf{k}}$ of

$$\omega + \mu - \epsilon_{\mathbf{k}} - \text{Re}\Sigma_{\mathbf{k}}(\omega) = 0. \quad (19)$$

In the vicinity of $E_{\mathbf{k}}$ the Green function can then be approximated to lowest order as

$$G_{\mathbf{k}}(\omega) = \frac{Z_{\mathbf{k}}(E_{\mathbf{k}})}{\omega - E_{\mathbf{k}} + i\tau_{\mathbf{k}}(E_{\mathbf{k}})^{-1}}, \quad (20a)$$

$$Z_{\mathbf{k}}(\omega) = [1 - \text{Re}\Sigma_{\mathbf{k}}(\omega)]^{-1}, \quad (20b)$$

$$\tau_{\mathbf{k}}(\omega) = [-Z_{\mathbf{k}}(\omega)\text{Im}\Sigma_{\mathbf{k}}(\omega)]^{-1}, \quad (20c)$$

where $Z_{\mathbf{k}}$ and $\tau_{\mathbf{k}}$ play the role of a quasiparticle weight and lifetime. In analogy to the non-interacting case the maxima $E_{\mathbf{k}}$ of $A_{\mathbf{k}}(\omega)$ yield the *electronic dispersion*, i.e., the relation between crystal momentum and excitation energy, although this maximum may be quite broad.

A reliable quasiparticle picture is guaranteed in a Landau Fermi liquid close to the Fermi surface, i.e., near $\omega = 0$, because then $\text{Re}\Sigma_{\mathbf{k}}(\omega)$ is linear and $\text{Im}\Sigma_{\mathbf{k}}(\omega)$ quadratic in ω for small frequencies. Near $\omega = 0$ this leads to

$$E_{\mathbf{k}} = Z_{\mathbf{k}}(0)(\epsilon_{\mathbf{k}} - \mu + \text{Re}\Sigma_{\mathbf{k}}(0)), \quad (21)$$

i.e., a linear relation between bare and interacting dispersion. However angle-resolved photoemission (ARPES) nowadays provides a means to measure $A_{\mathbf{k}}(\omega)$ (times the Fermi function) even deep below the Fermi energy with high accuracy (see, e.g., Ref. [6]). Therefore it is desirable to understand the origin of resonances given by (19), even if these excitations are not as coherent as low-energy excitations in a Landau Fermi liquid.

Hubbard bands and the Mott transition

Let us consider the atomic limit of the Hubbard model, i.e., no hopping, $t_{ij} = 0$. The Green function then becomes momentum-independent and reads

$$G_{\mathbf{k}\sigma}^{\text{at}}(\omega) = \frac{n_{-\sigma}}{\omega + \mu - U} + \frac{1 - n_{-\sigma}}{\omega + \mu}, \quad (22)$$

which corresponds to a spectral function with two δ peaks separated by an energy U , and for half-filling the system is *insulating*. What happens now if we turn on the hopping t_{ij} ? The δ peaks in the spectral function will broaden so that two subbands develop, the *Hubbard bands*. Note that these subbands are not one-electron bands as in non-interacting systems. For example, the upper Hubbard band describes the spectrum of charge excitations on top of the filled lower Hubbard band. If the hopping is increased further, or the Hubbard interaction U decreased, these Hubbard bands will eventually overlap and the system will become metallic at a critical value U_c on the order of the bandwidth. This correlation-induced *metal-insulator transition* does not break translational invariance and is called the Mott transition, as it was originally put forward by Mott [7].

Starting from the atomic limit, the simplest and rather crude method to capture the Mott phenomenon is the so-called Hubbard-I approximation: one obtains the atomic self-energy from (22) and uses it in the Dyson equation. However, this ad-hoc approximation leads to several undesirable pathologies (see Ref. [8] for a discussion). Starting from the weak-coupling side, a reasonable picture of the Mott transition can be provided by, e.g., the Gutzwiller wave function (see [9] for a review); however, there are also some shortcomings in this and other variational approaches. In fact, one of the successes of DMFT is its description of the Mott transition; in Sec. 4 we will mention some of these results.

2 Fermions in infinite dimensions

Historically, DMFT began with the discovery of simplifications that occur in the limit of infinite spatial dimensions [10], which we will now discuss. First of all, it is of course straightforward

to generalize the three-dimensional simple cubic lattice to the d -dimensional hypercubic lattice. The hypercubic lattice has the unit cell basis vectors

$$\mathbf{e}_1 = (1, 0, 0, \dots, 0), \quad (23)$$

$$\mathbf{e}_2 = (0, 1, 0, \dots, 0), \quad (24)$$

$$\dots \quad (25)$$

$$\mathbf{e}_d = (0, 0, 0, \dots, 1). \quad (26)$$

A nearest-neighbor hopping amplitude t_{ij} and corresponding dispersion then have the form

$$t_{ij} = t(\mathbf{R}_i - \mathbf{R}_j) = \begin{cases} -t & \text{if } \mathbf{R}_i - \mathbf{R}_j = \pm \mathbf{e}_n \\ 0 & \text{otherwise} \end{cases}, \quad \epsilon_{\mathbf{k}} = -2t \sum_{i=1}^d \cos k_i. \quad (27)$$

We now consider the limit $d \rightarrow \infty$ and obtain the limit of the density of states (11), in two ways. The first and rather elegant way proceeds by appealing to the central limit theorem of probability theory [10]. For this, consider the random variables $X_i = \sqrt{2} \cos k_i$, with the independent random variables k_i each distributed uniformly in $[-\pi : \pi]$. X_i has zero mean and unit variance, $\int_{-\pi}^{\pi} \frac{dk_i}{2\pi} X_i^2 = 1$. By the central limit theorem, for $d \rightarrow \infty$ the random variable $X_d = \frac{1}{\sqrt{d}} \sum_{i=1}^d X_i$ converges in law to a normal distributed random variable X with zero mean and unit variance. This means that the distribution function of X_d converges to the normal distribution $f(x) = \exp(-x^2/2)/\sqrt{2\pi}$. Next the density of states $\rho(\epsilon)$ can be regarded as the distribution function of the random variable $\sqrt{2d}t X_d$. We conclude that a finite density of states is obtained for $d \rightarrow \infty$ if the hopping amplitude is scaled proportional to $d^{-1/2}$,

$$\rho(\epsilon) = \int \frac{d^d k}{(2\pi)^d} \delta(\epsilon - \epsilon_{\mathbf{k}}) \quad (28a)$$

$$= \frac{1}{2\pi|t_*|} \exp\left[-\frac{\epsilon^2}{2t_*^2}\right] \quad \text{for } t = \frac{t_*}{\sqrt{2d}}, \quad (28b)$$

with t_* independent of d . Hence nearest-neighbor hopping on the hypercubic lattice, if scaled appropriately, leads to a Gaussian density of states. In Fig. 1, which shows the density of states for hopping on hypercubic lattice for several d , the trend towards the Gaussian density of states for large d can be recognized.

The second way to obtain this density of states uses the Fourier transform of $\rho(\epsilon)$ [12], which factorizes:

$$\Phi(s) = \int_{-\infty}^{\infty} d\epsilon e^{is\epsilon} \rho(\epsilon) = \int \frac{d^d k}{(2\pi)^d} e^{is\epsilon_{\mathbf{k}}} \quad (29)$$

$$= \left[\int_{-\pi}^{\pi} \frac{dk}{2\pi} \exp\left(-\frac{2ist_*}{\sqrt{2d}} \cos k\right) \right]^d = J_0\left(\frac{2t_*}{\sqrt{2d}}\right)^d \quad (30)$$

$$= \left[1 - \frac{t_*^2 s^2}{2d} + O\left(\frac{1}{d^2}\right) \right]^d = \exp\left[-\frac{t_*^2 s^2}{2} + O\left(\frac{1}{d}\right)\right], \quad (31)$$

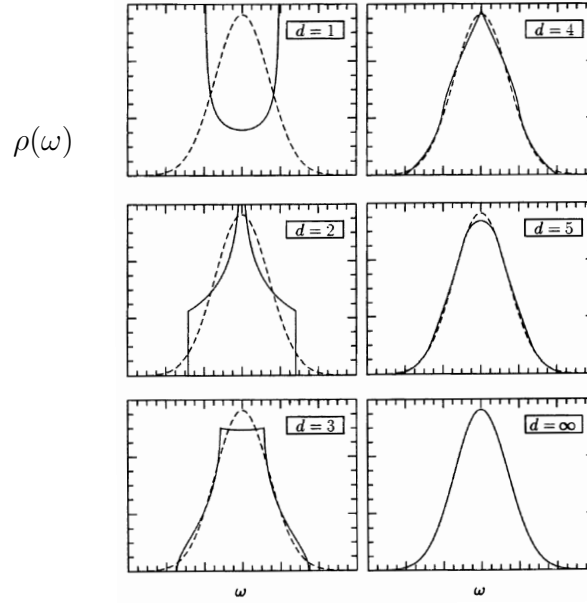


Fig. 1: Density of states for hopping on hypercubic lattice for several d , compared to the Gaussian that is obtained in $d \rightarrow \infty$. From Ref. [2].

where $J_0(z)$ is a Bessel function. The inverse transform is then

$$\rho(\epsilon) = \int_{-\infty}^{\infty} \frac{d\epsilon}{2\pi} e^{-is\epsilon} \Phi(s) = \frac{1}{2\pi|t_*|} \exp\left[-\frac{\epsilon^2}{2t_*^2} + O\left(\frac{1}{d}\right)\right]. \quad (32)$$

In fact, this calculation is similar to the proof ideas behind the central limit theorem, for which Fourier transforms of probability function are also used.

The important conclusion from these considerations is that the nearest-neighbor hopping amplitude must be scaled with $1/\sqrt{d}$ to obtain a meaningful finite limit. This statement can be generalized as follows: each hopping amplitude t_n must be scaled proportional to $1/\sqrt{Z_n}$, where Z_n is the number of sites that are connected by t_n , e.g., $Z_1 = 2d$ for nearest-neighbor hopping and $Z_2 = (2d-1)2d = (Z-1)Z$ for next-nearest-neighbor hopping on the hypercubic lattice, and so on.

Note that the density of states extends up to infinite positive and negative energies ϵ , even after the scaling (28). An infinite bandwidth results also for other generalized lattices, such as the face-centered-hypercubic lattice [11] (which is asymmetric and has one finite band edge) or the hyperdiamond lattice [13] (for which the symmetric density of states vanishes at $\epsilon = 0$). One of the few lattices with finite bandwidth is the Bethe lattice, i.e., an infinite Cayley tree of which each node has Z nearest neighbors. This recursively defined lattice (which is not a crystal lattice) has a semi-elliptic density of states with a finite bandwidth in the limit $Z \rightarrow \infty$ for scaled nearest-neighbor hopping $t = t_*/\sqrt{Z}$,

$$\rho_{\text{Bethe}}(\epsilon) = \begin{cases} \frac{\sqrt{4t_*^2 - \epsilon^2}}{2\pi t_*^2} & \text{for } |\epsilon| \leq 2|t_*| \\ 0 & \text{otherwise} \end{cases}. \quad (33)$$

This result is best obtained with recursive methods and can be generalized to next-nearest-neighbor and longer-range hopping (see Refs. [14, 15] and references therein).

The infinite band edges for nearest-neighbor hopping on regular lattice may seem problematic at first glance. After all we are interested in using the infinite-dimensional limit as an approximation to finite-dimensional lattices, for which finite hopping amplitude always imply finite band edges. The practical solution is to use the simplifications that result from the infinite-dimensional limit (see the following sections) and to plug in the non-interacting density of states of the system of interest everywhere. Some justification for this scheme comes from two observations: (i) In infinite dimensions the dispersion $\epsilon_{\mathbf{k}}$ typically enters only via the density of states (at least into single-particle quantities), so that the detailed lattice structure does not enter. (ii) For any single-band density of states one can always construct a corresponding set of hopping amplitudes (both for the infinite-dimensional hypercubic [16] and Bethe lattice [14]); in particular, densities of states with finite bandwidth are perfectly possible, although they usually require long-range hopping.

3 Simplifications for many-body theory

We now turn to the consequences that the limit $d \rightarrow \infty$ has for many-body theory [12, 17], in particular for the self-energy (defined in 13). This is best discussed in terms of Feynman diagrams for Green functions [4, 5], of which we first review some basics.

A guide to Feynman diagrams

Feynman diagrams for single-particle Green functions (for arbitrary quadratic H_0 and two-particle interaction H_1) are built from the following elements:

$$\text{————} = \text{non-interacting Green function line } \mathbf{G}^{(0)}, \quad (34)$$

$$\text{Y} \text{---} \text{---} \text{---} \text{Y} = \text{interaction vertex}, \quad (35)$$

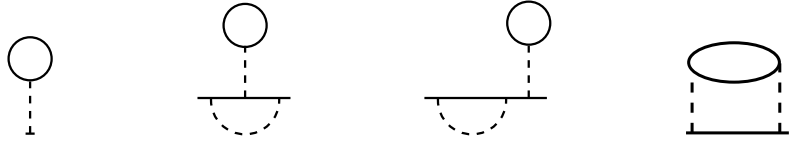
$$\text{====} = \text{full (interacting) Green function line } \mathbf{G}. \quad (36)$$

The perturbation expansion in H_1 then yields the following series of diagrams (unlabeled, and arrows omitted) for the Green function:

$$\text{====} = \text{————} + \text{————} \text{---} \text{O} \text{---} \text{————} + \text{————} \text{---} \text{---} \text{---} \text{---} + \text{————} \text{---} \text{O} \text{---} \text{---} \text{---} \text{---} + \text{————} \text{---} \text{O} \text{---} \text{---} \text{---} \text{---} + \text{————} \text{---} \text{---} \text{---} \text{---} \text{---} + \dots \quad (37)$$

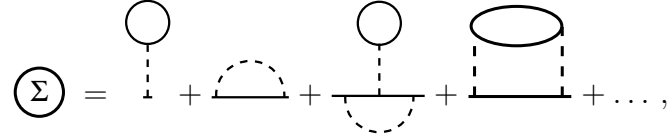
Clearly some parts of the diagram occur repeatedly. Therefore one defines proper self-energy diagrams, which are “one-particle irreducible” (i.e., cannot be cut in two pieces by cutting a

single solid line) and have their external vertices amputated. Here are some examples:



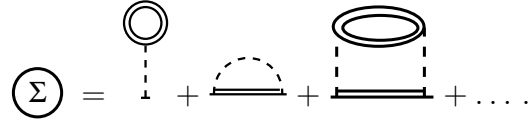
proper
proper
not proper
proper

From these diagrams one builds the self-energy,



$$\textcircled{\Sigma} = \text{---} + \text{---} + \text{---} + \text{---} + \dots, \quad (40)$$

which, when combined with (14), indeed yields (36). So far we have considered an expansion of the form $\Sigma[\mathbf{G}^{(0)}]$,³ i.e., in terms of the free Green function $\mathbf{G}^{(0)}$. These diagrams still contain self-energy insertions in their internal lines. The next step is therefore to construct the *skeleton expansion* which instead uses full (interacting) Green function lines:



$$\textcircled{\Sigma} = \text{---} + \text{---} + \text{---} + \dots. \quad (41)$$

Clearly one must be careful not to include diagrams more than once, especially in higher orders. The skeleton expansion $\Sigma[\mathbf{G}]$ is a useful representation to analyze the self-energy in the limit $d \rightarrow \infty$.

Power counting in $1/d$

We now study first the d dependence of $G_{ij\sigma}(\omega)$ in the limit $d \rightarrow \infty$, for scaled hopping amplitudes,

$$t_{ij} = t_{ij}^* d^{-\frac{1}{2}\|\mathbf{R}_i - \mathbf{R}_j\|}. \quad (42)$$

Here $\|\mathbf{R}_i - \mathbf{R}_j\|$ is the shortest number of lattice steps from \mathbf{R}_i to \mathbf{R}_j on the hypercubic lattice, and hence proportional to the number of sites connected by the hopping amplitude t_{ij} , so that (42) has the correct scaling. By our construction the kinetic energy is finite in the limit $d \rightarrow \infty$, which can be expressed in terms of the Green function,

$$E_{\text{kin},\sigma} = \sum_{ij} t_{ij} \langle c_{i\sigma}^\dagger c_{j\sigma} \rangle = \sum_{ij} t_{ij} \int_{-\infty}^{\infty} \frac{d\omega}{2\pi i} G_{ij\sigma}(\omega) e^{i\omega 0^+} = O(d^0). \quad (43)$$

Here the double sum yields a contribution of order $Ld^{\|\mathbf{R}_i - \mathbf{R}_j\|}$. Hence we can conclude

$$G_{ij\sigma}(\omega) = O(d^{-\frac{1}{2}\|\mathbf{R}_i - \mathbf{R}_j\|}), \quad G_{ii\sigma}(\omega) = O(d^0), \quad (44)$$

i.e., the Green function decays rapidly with distance, which leads to simplifications for the Feynman diagrams.

³This is of course a functional dependence of Σ on $\mathbf{G}^{(0)}$, because the whole matrix $\mathbf{G}^{(0)}(i\omega_n)$ and also the frequency dependence enter into the value of the Feynman diagrams due to summations over internal lines.

Local self-energy

We know analyze the consequences for the self-energy. For this discussion, however, we work with Hugenholtz diagrams instead, which combine direct and exchange diagrams into a box vertex [5]. However, for the Hubbard interaction here are no exchange diagrams anyway. We thus make the replacement

$$i, \sigma \rangle \text{---} \langle i, -\sigma = U n_{i\uparrow} n_{i\downarrow} = \text{[Square with four external lines]} . \quad (45)$$

In terms of these diagrams, the skeleton expansion takes the form

$$\Sigma = \text{[Self-energy diagrams: bubble, two parallel lines, triangle, ...]} , \quad (46)$$

and has the property that by construction any two vertices are joined through Green function lines via at least three independent paths. Namely, suppose there is only one such path; then the diagram is one-particle irreducible, a contradiction. If there are only two paths, then they must run through a diagram part which is a self-energy insertion, which is also a contradiction.

Now consider an arbitrary diagram (in position space, so that the interaction vertices are labeled by lattice site vectors), in which two internal vertices labeled by i and j appear,

$$\text{[Diagram: Two square vertices labeled } i \text{ and } j \text{ inside an oval, connected by three lines]} . \quad (47)$$

Let us hold i fixed for the moment. We now compare the case $j \neq i$ with the case $j = i$. Suppose $j \neq i$. As discussed above, there are three independent paths from the vertex i to the vertex j . The Green function lines on these paths can thus contribute at most $O(d^{-\frac{3}{2}\|\mathbf{R}_i - \mathbf{R}_j\|})$ (for example, less if there is another intermediate site \mathbf{R}_k on a path). The summation over j will then yield an order $O(d^{\|\mathbf{R}_i - \mathbf{R}_j\|})$. As a consequence, any skeleton diagram is suppressed at least by a factor $O(d^{-\frac{1}{2}\|\mathbf{R}_i - \mathbf{R}_j\|})$, for example this one:

$$\text{[Diagram: Two square vertices labeled } i \text{ and } j \text{ connected by three parallel lines]} . \quad (48)$$

By contrast, for $j = i$ the Green functions are of order $O(d^0)$, and there is no summation. We thus conclude that only the case $i = j$ contributes in the limit $d \rightarrow \infty$, i.e., all diagrams in the skeleton expansion $\Sigma[\mathbf{G}]$ have the same lattice site label at all their internal and external vertices. Hence the self-energy is site-diagonal (“local”),

$$\Sigma_{ij\sigma}(\omega) = \delta_{ij} \Sigma_{ii\sigma}(\omega) = \delta_{ij} \Sigma_{\sigma}(\omega) , \quad (49)$$

or, equivalently, momentum-independent in \mathbf{k} space,

$$\Sigma_{\mathbf{k}\sigma}(\omega) = \Sigma_{\sigma}(\omega). \quad (50)$$

Furthermore, *the self-energy* $\Sigma_{\sigma}(\omega)$ *is a functional only of the local Green function* $G_{\sigma}(\omega)$, because all internal vertices in the skeleton expansion have the same site label.

The simple form of the self-energy has some immediate consequences also for the Green function (13), namely

$$G_{\mathbf{k}\sigma}(\omega) = \frac{1}{\omega + \mu - \epsilon_{\mathbf{k}} - \Sigma_{\sigma}(\omega)} = G_{\mathbf{k}\sigma}^{(0)}(\omega - \Sigma_{\sigma}(\omega)), \quad (51)$$

and in particular the local Green function becomes

$$G_{\sigma}(\omega) = \int \frac{d^d k}{(2\pi)^d} \frac{1}{\omega + \mu - \epsilon_{\mathbf{k}} - \Sigma_{\sigma}(\omega)} \quad (52)$$

$$= \int_{-\infty}^{\infty} d\omega' \frac{\rho(\epsilon)}{\omega + \mu - \Sigma_{\sigma}(\omega) - \epsilon}. \quad (53)$$

The last equation thus provides a relation between the local self-energy and the local Green function, and depends only on the dispersion via the free density of states. We will come back to this relation in the next section.

4 Dynamical mean-field theory

In the last section we have seen that the self-energy becomes site-diagonal and momentum-independent in the limit $d \rightarrow \infty$. The last step is now to actually construct the functional $\Sigma_{\sigma}[G_{\sigma}]$ [1–3, 18, 19], which will complete the derivation of the DMFT equations.

Mapping onto effective impurity models

Consider a single-site action, $\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2$, consisting of a quadratic part and an interaction,

$$\mathcal{A}_1 = \int_0^{\beta} d\tau \int_0^{\beta} d\tau' \sum_{\sigma} c_{\sigma}^*(\tau) \mathcal{G}_{\sigma}^{-1}(\tau, \tau') c_{\sigma}(\tau') = \sum_{n,\sigma} c_{\sigma}^*(i\omega_n) \mathcal{G}_{\sigma}(i\omega_n)^{-1} c_{\sigma}(i\omega_n), \quad (54a)$$

$$\mathcal{A}_2 = -U \int_0^{\beta} d\tau c_{\uparrow}^*(\tau) c_{\uparrow}(\tau) c_{\downarrow}^*(\tau) c_{\downarrow}(\tau), \quad (54b)$$

with some as yet unfixed “free” Green function $(\mathcal{G}^{-1})_{\tau,\tau'} = \mathcal{G}_{\sigma}^{-1}(\tau, \tau')$, which also depends only on imaginary-time differences. The goal is now to match this action to that of Hubbard model in infinite dimensions.

Suppose that we calculate the imaginary-time-ordered Green function of the single degree of freedom c from the action (54), and Fourier transform to Matsubara frequencies. This is abbreviated as

$$G_\sigma(i\omega_n) = \langle c_\sigma(i\omega_n) c_\sigma^*(i\omega_n) \rangle_{\mathcal{A}[\mathcal{G}]} . \quad (55)$$

Then define the impurity *impurity self-energy* $\tilde{\Sigma}$ via the *impurity Dyson equation*,

$$\mathbf{G} = [\mathbf{g}^{-1} - \tilde{\Sigma}]^{-1} . \quad (56)$$

Now consider the diagrams in the skeleton expansion of $\tilde{\Sigma}[\mathbf{G}]$,

$$\tilde{\Sigma}[\mathbf{G}] = \begin{array}{c} \text{⊙} \\ \vdots \end{array} + \begin{array}{c} \text{---} \\ \text{---} \end{array} + \begin{array}{c} \text{---} \\ \text{---} \\ \text{---} \end{array} + \dots , \quad (57)$$

in which of course only the single site of (54) occurs. However, since the local Hubbard interaction is the same both for the lattice Hubbard model and the single-site action, this skeleton expansion is exactly the same as that for the Hubbard model (41), i.e.,

$$\tilde{\Sigma}[\mathbf{G}] = \Sigma[\mathbf{G}] . \quad (58)$$

This shows that the desired functional $\Sigma_\sigma[G_\sigma]$ can be obtained by solving the single-site problem (54).

Dynamical mean-field equations

We summarize again the three DMFT equations, which determine three unknowns: the local Green function $G_\sigma(i\omega_n)$, the *dynamical mean field* (or *Weiss field*) $\mathcal{G}_\sigma(i\omega_n)$, and the local self-energy $\Sigma_\sigma(i\omega_n)$:

$$G_\sigma(i\omega_n) = \langle c_\sigma(i\omega_n) c_\sigma^*(i\omega_n) \rangle_{\mathcal{A}[\mathcal{G}]}, \quad (\text{DMFT-1})$$

$$G_\sigma(i\omega_n) = [\mathcal{G}_\sigma(i\omega_n)^{-1} - \Sigma_\sigma(i\omega_n)]^{-1}, \quad (\text{DMFT-2})$$

$$G_\sigma(i\omega_n) = \int d\epsilon \frac{\rho(\epsilon)}{i\omega_n + \mu - \Sigma_\sigma(i\omega_n) - \epsilon}. \quad (\text{DMFT-3})$$

Note that the self-consistency equation (53) provides precisely the needed relation (DMFT-3) to fix the Weiss field \mathcal{G}_σ . After all, it must be ensured that one solves the correct single-site problem, i.e., the one which indeed corresponds to the Hubbard model on a lattice with density of states $\rho(\epsilon)$.

A typical iterative solution then proceeds as follows. Start with some Weiss field \mathcal{G}_σ , obtain G_σ from (DMFT-1), determine Σ_σ from the impurity Dyson equation (DMFT-2), calculate G_σ from self-consistency equation (DMFT-3), obtain \mathcal{G}_σ by using (DMFT-2) again, and repeat until convergence is reached.

Of course the DMFT equations should produce the correct noninteracting and atomic limits. (i) In the noninteracting case we have $U = 0$ and thus $\Sigma_\sigma(i\omega_n) = 0$. Furthermore it follows from (DMFT-3) that then $G_\sigma(i\omega_n) = G_\sigma^{(0)}(i\omega_n)$. Finally (DMFT-2) gives $\mathcal{G}_\sigma(i\omega_n) = G_\sigma(i\omega_n)$, and this agrees with (DMFT-1) for $U = 0$. (ii) On the other hand, in the atomic limit we have $t_{ij} = 0$ and $\epsilon_k = 0$, i.e., $\rho(\epsilon) = \delta(\epsilon)$. From (DMFT-3) we obtain $G_\sigma(i\omega_n) = [i\omega_n + \mu - \Sigma_\sigma(i\omega_n)]^{-1}$, and (DMFT-2) yields $\mathcal{G}_\sigma(i\omega_n)^{-1} = i\omega_n + \mu$, i.e., $\mathcal{G}_\sigma^{-1}(\tau) = -\partial_\tau + \mu$, which agrees with (DMFT-1) for $t_{ij} = 0$.

For general interaction U , the local action (54) clearly represents the most difficult of the DMFT equations. To obtain the impurity Green function from it, a dynamical single-site problem must be solved, usually with numerical methods. For finite temperatures quantum and thermal averages can be stochastically sample with quantum Monte Carlo (QMC) methods, such as the Hirsch-Fye QMC algorithm [19–21, 1] and continuous-time (CT) QMC [22–24]. Methods that also work for zero temperature include exact diagonalization (ED) [25–27], the numerical renormalization group (NRG) [28, 29] and the density-matrix renormalization group (DMRG) [30]. A number of perturbative or semianalytic methods is also available.

To use these “impurity solvers”, the single-site action (54) is not used directly, but rather an impurity problem defined by a Hamiltonian is considered, usually by constructing a single-impurity Anderson model (SIAM):

$$H_{\text{SIAM}} = \sum_{\ell\sigma} \epsilon_\ell a_{\ell\sigma}^\dagger a_{\ell\sigma} + \sum_{\ell\sigma} V_\ell (a_{\ell\sigma}^\dagger c_\sigma + c_\sigma^\dagger a_{\ell\sigma}) + U c_\uparrow^\dagger c_\uparrow c_\downarrow^\dagger c_\downarrow. \quad (59)$$

Here the fermions $a_{\ell\sigma}$ represent a non-interacting bath which hosts the interacting fermion c_σ . This bath can be at once integrated out from the action which represents H_{SIAM} , because this involves only Gaussian integrals. The resulting action is then precisely of the form (54), with

$$G_\sigma^{-1}(i\omega_n) = i\omega_n + \mu - \frac{1}{\pi} \int_{-\infty}^{\infty} d\omega \frac{\Delta(\omega)}{i\omega_n - \omega}, \quad \Delta(\omega) = \pi \sum_{\ell} V_\ell^2 \delta(\omega - \epsilon_\ell), \quad (60)$$

where $\Delta(\omega)$ is called the hybridization function. In the DMFT cycle one must now find the parameters V_ℓ and ϵ_ℓ that allow a self-consistent DMFT solution. Then one has found the appropriate SIAM that represents the Hubbard model in DMFT.

For reference we note that the self-consistency equation (DMFT-3) yields a simple relation for next-neighbor hopping t_* on the Bethe lattice with density of states (33),

$$\mathcal{G}_\sigma(i\omega_n) = i\omega_n + \mu - t_*^2 G(i\omega_n). \quad (61)$$

This relation and generalizations for other types of hopping are discussed in Refs. [1, 16, 14, 15].

Results for the Hubbard model

Some aspects of the spectrum and DMFT phase diagram of the Hubbard model were discussed already in the Lecture of D. Vollhardt. Fig. 2 shows the zero-temperature spectral function for

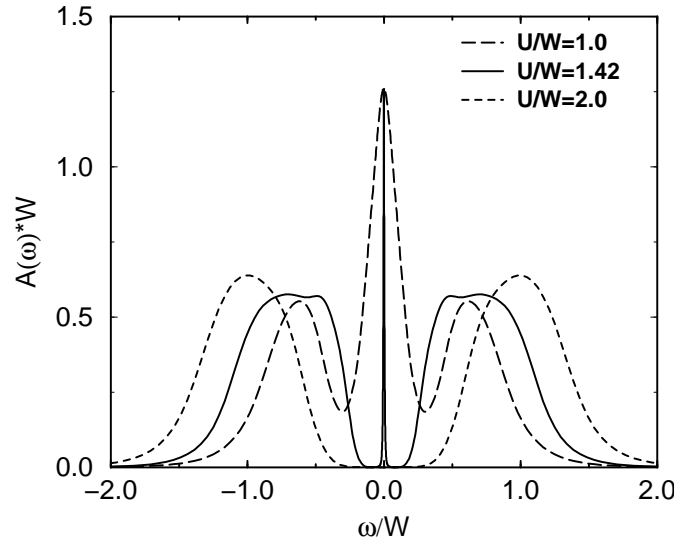


Fig. 2: Zero-temperature spectral function for the homogeneous phase of the Hubbard model on the Bethe lattice with nearest-neighbor hopping and bandwidth $W = 4|t_*|$ at half-filling, evaluated with NRG. From Ref. [28].

the homogeneous phase of the Hubbard model on the Bethe lattice with nearest-neighbor hopping and bandwidth $W = 4|t_*|$ at half-filling, evaluated with NRG. Three values of U are shown, one in the metallic phase (three peaks in the spectral function), one close to the critical value U_c , and one for the insulating phase (with gap in the spectral function). At the Fermi energy the spectral function has the same value for all U in the metallic phase; this is a consequence of Luttinger's theorem [12]. In the metallic phase the weight of the central peak is proportional to the Fermi liquid quasiparticle renormalization factor Z (see (21)), whereas the outer two peaks are the developing Hubbard bands.

Fig. 3 shows the renormalization factor Z obtained with various methods. It starts from $Z = 1$ for the non-interacting case and decreases as U is increased, corresponding to the decreasing width of the central peak in the spectral function and an increasingly flatter dispersion. At U_c , the half-filled system becomes localized and Z vanishes accordingly.

The Falicov-Kimball model, a solvable example

The Falicov-Kimball model is a simplified version of the Hubbard model, in which only one of the two spin species is mobile (reabeled as d_i), while the other (reabeled as f_i) is not. For this model the Green function can be obtained explicitly from the DMFT action [31]. The Hamiltonian reads

$$H = \sum_{ij} t_{ij} d_i^\dagger d_j + E_f \sum_i f_i^\dagger f_i + U \sum_i d_i^\dagger d_i f_i^\dagger f_i, \quad (62)$$

i.e. the d electrons are moving in front of a background of static f electrons, whose configuration is chosen such that it optimizes the free energy. In principle this makes the model quite complicated, as one needs the spectrum of H for all the possible f configurations. In dimensions $d \geq 2$ it is known that at half-filling on a bipartite lattice checkerboard order of the f

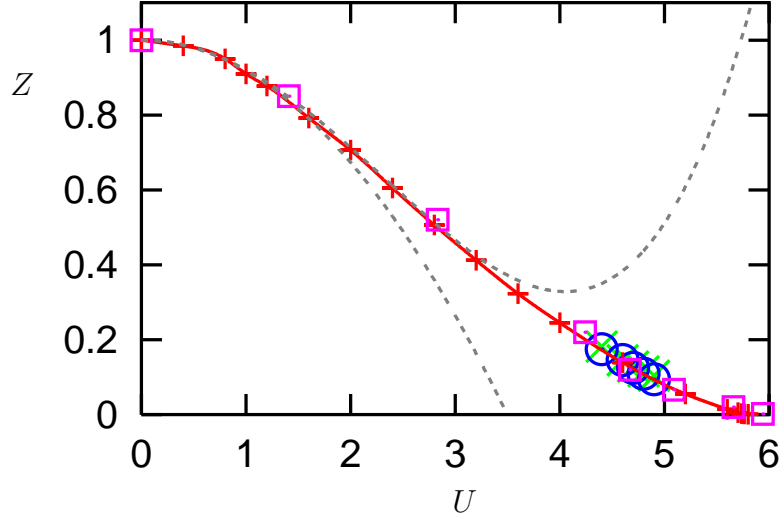


Fig. 3: Quasiparticle weight Z for the half-filled Hubbard model on the Bethe lattice (with $t_* = 1$) in DMFT. Crosses +: NRG; squares: ED; crosses \times and circles: QMC extrapolations; lower dashed line: 2nd order perturbation theory in U , upper dashed line: 4th order perturbation theory in U . From Ref. [16].

electrons appears in the ground state and persists up to a finite critical temperature [32]. Here we consider only the homogeneous phase in DMFT for simplicity.

Since there is no hopping amplitude for the f electrons, the DMFT self-consistency yields at once $\mathcal{G}_f^{-1} = -\partial_\tau + \mu + E_f$, as explained above for the atomic limit. The DMFT action is thus given by

$$\begin{aligned} \mathcal{A} = & \int_0^\beta d\tau \int_0^\beta d\tau' d^*(\tau) \mathcal{G}_d^{-1}(\tau, \tau') d(\tau') \\ & + \int_0^\beta d\tau f^*(\tau) (\partial_\tau - \mu + E_f) f(\tau) - U \int_0^\beta d\tau d^*(\tau) d(\tau) f^*(\tau) f(\tau). \end{aligned} \quad (63)$$

Now the f electrons can be integrated out at each lattice site, i.e., they are in the atomic limit (cf. Sec. 1). This leads to

$$G_d(i\omega_n) = \langle d(i\omega_n) d^*(i\omega_n) \rangle_{\mathcal{A}} = \frac{n_f}{\mathcal{G}_d(i\omega_n)^{-1} - U} + \frac{1 - n_f}{\mathcal{G}_d(i\omega_n)^{-1}}, \quad (64)$$

which must be solved together with the other two DMFT equations

$$G_d(i\omega_n) = \int_{-\infty}^{\infty} \frac{d\epsilon \rho_d(\epsilon)}{i\omega_n + \mu - \Sigma_d(i\omega_n) - \epsilon}, \quad (65)$$

$$G_d(i\omega_n)^{-1} = \mathcal{G}_d(i\omega_n)^{-1} - \Sigma_d(i\omega_n). \quad (66)$$

This set of equations determines the d -electron Green function $G_d(i\omega_n)$ for any density of states $\rho_d(\epsilon)$. Analytically continuation to real frequencies shows at once that the spectra in the homogeneous phase are independent of temperature (but this no longer holds in the checkerboard

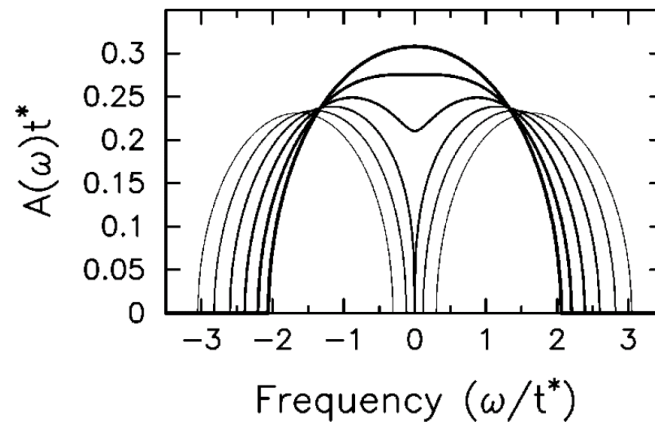


Fig. 4: Spectral function of itinerant d electrons for the Falicov-Kimball model in DMFT for nearest-neighbor hopping on the Bethe lattice, homogeneous phase, $n_d = n_f = \frac{1}{2}$, $U = 0.5, 1.0, \dots 3.0$. From Ref. [31].

phase). Fig. 4 shows the spectral function $A_d(\omega)$ for several U for the Bethe lattice (with nearest-neighbor hopping $t_* = 1$). In particular there is a Mott metal-insulator transition taking place at $U_c = 2$; for larger U , a band gap develops. Nevertheless, the transition is qualitatively different from that in the Hubbard model. For example, for the Falicov-Kimball model it can be shown that from the low-energy form of the self-energy that for $0 < U < U_c$ the metallic state is not a Landau Fermi liquid; as a consequence, the spectral function is not pinned at the Fermi surface.

It is also possible to solve for d self-energy as a functional of the d Green function, i.e., for the skeleton functional⁴ $\Sigma_d[G_d]$ [8]

$$\Sigma_d(i\omega_n) = \frac{U}{2} - \frac{1}{2G_d(i\omega_n)} \pm \sqrt{\left(\frac{U}{2} - \frac{1}{2G_d(i\omega_n)}\right)^2 + \frac{Un_f}{G_d(i\omega_n)}}. \quad (67)$$

Just like any skeleton expansion, this relation holds for any density of state $\rho(\epsilon)$.

5 Summary and outlook

The goal of this lecture was to demonstrate the origin of DMFT, i.e., to show how the infinite-dimensional Hubbard model can be mapped onto a dynamical single-site problem in an effective bath, which has to be determined self-consistently. In the other lectures several further aspects of DMFT will be discussed. For one, some of the numerical approaches that were only mentioned in this lecture will be explained in detail. Second, DMFT will be combined with ab-initio band structure methods to make quantitative predictions about correlated materials. Finally, several extensions of DMFT to correlated clusters (instead of a single correlated site) will be developed, which improve the description for finite-dimensional systems.

⁴Note that in the DMFT solution for the Falicov-Kimball model this functional is in fact a function: $\Sigma_d(i\omega_n)$ depends only on $G_d(i\omega_n)$ at the same frequency. This is certainly not the case for the Hubbard model.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] A. Georges, G. Kotliar, W. Krauth, and M. J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [2] D. Vollhardt, in *Correlated Electron Systems*, edited by V. J. Emery (World Scientific, Singapore, 1993) p. 57
- [3] D. Vollhardt, in *Lectures on the Physics of Strongly Correlated Systems XIV, AIP Conference Proceedings*, vol. 1297, ed. by A. Avella, F. Mancini (American Institute of Physics, Melville, 2010), p. 339; arXiv:1004.5069
- [4] A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, 1971)
- [5] J. W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wesley, 1988)
- [6] X. J. Zhou, T. Cuk, T. Devereaux, N. Nagaosa, and Z.-X. Shen, in: J. R. Schrieffer (ed.), *Handbook of High-Temperature Superconductivity: Theory and Experiment* (Springer, 2007), p. 87; arXiv:cond-mat/0604284
- [7] N. F. Mott, *Proc. Phys. Soc. A* **62**, 416 (1947)
- [8] F. Gebhard, *The Mott Metal-Insulator Transition* (Springer, Berlin, 1997)
- [9] D. Vollhardt, *Rev. Mod. Phys.* **56**, 99 (1984)
- [10] W. Metzner and D. Vollhardt, *Phys. Rev. Lett.* **62**, 324 (1989)
- [11] E. Müller-Hartmann, in: E. Talik and J. Szade (eds.), *Proceedings of the V. Symposium "Physics of Metals"* (Ustroń, 1991)
- [12] E. Müller-Hartmann, *Z. Phys. B* **74**, 507 (1989)
- [13] G. Santoro, M. Airoldi, S. Sorella, and E. Tosatti, *Phys. Rev. B* **47**, 16216 (1993)
- [14] M. Eckstein, M. Kollar, K. Byczuk and D. Vollhardt, *Phys. Rev. B* **71**, 235119 (2005)
- [15] M. Kollar, M. Eckstein, K. Byczuk, N. Blümer, P. van Dongen, M. H. Radke de Cuba, W. Metzner, D. Tanaskovic, V. Dobrosavljevic, G. Kotliar, and D. Vollhardt, *Ann. Phys. (Leipzig)* **14**, 642 (2005)
- [16] N. Blümer, *Metal-Insulator Transition and Optical Conductivity in High Dimensions* (Shaker Verlag, Aachen, 2003)
- [17] E. Müller-Hartmann, *Z. Phys. B* **76**, 211 (1989)
- [18] A. Georges and G. Kotliar, *Phys. Rev. B* **45**, 6479 (1992)
- [19] M. Jarrell, *Phys. Rev. Lett.* **69**, 168 (1992)

-
- [20] M. J. Rozenberg, X. Y. Zhang, and G. Kotliar, *Phys. Rev. Lett.* **69**, 1236 (1992)
- [21] A. Georges and W. Krauth, *Phys. Rev. Lett.* **69**, 1240 (1992)
- [22] A. N. Rubtsov, V. V. Savkin, and A. I. Lichtenstein, *Phys. Rev. B* **72**, 035122 (2005)
- [23] P. Werner, A. Comanac, L. de' Medici, M. Troyer, and A. J. Millis, *Phys. Rev. Lett.* **97**, 076405 (2006)
- [24] K. Haule, *Phys. Rev. B* **75**, 155113 (2007)
- [25] M. Caffarel and W. Krauth, *Phys. Rev. Lett.* **72**, 1545 (1994)
- [26] Q. Si, M. J. Rozenberg, G. Kotliar, and A. E. Ruckenstein, *Phys. Rev. Lett.* **72**, 2761 (1994)
- [27] M. J. Rozenberg, G. Moeller, and G. Kotliar, *Mod. Phys. Lett. B* **8**, 535 (1994)
- [28] R. Bulla, *Phys. Rev. Lett.* **83**, 136 (1999)
- [29] R. Bulla, T. A. Costi, and Th. Pruschke, *Rev. Mod. Phys.* **80**, 395 (2008)
- [30] M. Karski, C. Raas, and G. S. Uhrig, *Phys. Rev. B* **77**, 075116 (2008)
- [31] J. K. Freericks and V. Zlatić, *Rev. Mod. Phys.* **75**, 1333 (2003)
- [32] T. Kennedy and E. H. Lieb, *Physica A* **138** 320 (1986)

6 The LDA+DMFT Approach

Eva Pavarini

Institute for Advanced Simulation

Forschungszentrum Jülich GmbH

Contents

1	The many-body problem	2
2	Low-energy models	7
3	Many-body models from DFT	10
3.1	Towards ab-initio Hamiltonians	10
3.2	Coulomb interaction tensor	12
3.3	Minimal material-specific models	14
4	Methods of solution	17
4.1	LDA+U	17
4.2	LDA+DMFT	21
5	The origin of orbital order	26
6	Conclusions	30
A	Constants and units	32
B	Atomic orbitals	32
B.1	Radial functions	32
B.2	Real harmonics	32
B.3	Slater-Koster integrals	34
B.4	Gaunt coefficients and Coulomb integrals	35

It would indeed be remarkable if Nature fortified herself against further advances in knowledge behind the analytical difficulties of the many-body problem. (Max Born, 1960)

1 The many-body problem

Most of chemistry and solid-state physics is described by the Hamiltonian

$$\hat{H} = -\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i \neq i'} \frac{1}{|\mathbf{r}_i - \mathbf{r}_{i'}|} - \sum_{i,\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} - \sum_\alpha \frac{1}{2M_\alpha} \nabla_\alpha^2 + \frac{1}{2} \sum_{\alpha \neq \alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{R}_\alpha - \mathbf{R}_{\alpha'}|},$$

where $\{\mathbf{r}_i\}$ are the coordinates of the N_e electrons, $\{\mathbf{R}_\alpha\}$ those of the N_n nuclei, Z_α the atomic numbers, and M_α the nuclear masses. The Born-Oppenheimer product Ansatz $\Psi(\{\mathbf{r}_i\}, \{\mathbf{R}_\alpha\}) = \psi(\{\mathbf{r}_i\}; \{\mathbf{R}_\alpha\})\Phi(\{\mathbf{R}_\alpha\})$ simplifies the problem. The Schrödinger equation for the electrons, $H_e\psi = \varepsilon\psi$, with

$$\begin{aligned} \hat{H}_e &= -\frac{1}{2} \sum_i \nabla_i^2 + \frac{1}{2} \sum_{i \neq i'} \frac{1}{|\mathbf{r}_i - \mathbf{r}_{i'}|} - \sum_{i,\alpha} \frac{Z_\alpha}{|\mathbf{r}_i - \mathbf{R}_\alpha|} + \frac{1}{2} \sum_{\alpha \neq \alpha'} \frac{Z_\alpha Z_{\alpha'}}{|\mathbf{R}_\alpha - \mathbf{R}_{\alpha'}|} \\ &= \hat{T}_e + \hat{V}_{ee} + \hat{V}_{en} + \hat{V}_{nn}, \end{aligned} \quad (1)$$

has however a simple solution only in the non-interacting limit ($\hat{V}_{ee} = 0$). In this case, H_e is separable as $\hat{H}_e = \sum_i \hat{h}_e^0(\mathbf{r}_i) + \hat{V}_{nn}$, with

$$\hat{h}_e^0(\mathbf{r}) = -\frac{1}{2} \nabla^2 - \sum_\alpha \frac{Z_\alpha}{|\mathbf{r} - \mathbf{R}_\alpha|} = -\frac{1}{2} \nabla^2 + v_{\text{ext}}(\mathbf{r}).$$

In a crystal the external potential $v_{\text{ext}}(\mathbf{r})$ is periodic, the eigenvectors of $\hat{h}_e^0(\mathbf{r})$ are Bloch functions, $\psi_{n\mathbf{k}\sigma}(\mathbf{r})$. The eigenvalues are the corresponding band energies, $\varepsilon_{n\mathbf{k}}$. Many-electron ($N_e > 1$) states may be obtained by filling energy levels $\varepsilon_{n\mathbf{k}}$ with electrons and anti-symmetrizing the wave-function according to the Pauli principle (Slater determinant). For a half-filled band described by the dispersion relation $\varepsilon_{\mathbf{k}}$, such a Slater determinant has the form

$$\psi(\{\mathbf{r}_i\}; \{\mathbf{R}_\alpha\}) = \frac{1}{\sqrt{N_e!}} \begin{vmatrix} \psi_{\mathbf{k}_1\uparrow}(\mathbf{r}_1) & \psi_{\mathbf{k}_1\uparrow}(\mathbf{r}_2) & \dots & \psi_{\mathbf{k}_1\uparrow}(\mathbf{r}_{N_e}) \\ \psi_{\mathbf{k}_1\downarrow}(\mathbf{r}_1) & \psi_{\mathbf{k}_1\downarrow}(\mathbf{r}_2) & \dots & \psi_{\mathbf{k}_1\downarrow}(\mathbf{r}_{N_e}) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{\mathbf{k}_{\frac{N_e}{2}}\uparrow}(\mathbf{r}_1) & \psi_{\mathbf{k}_{\frac{N_e}{2}}\uparrow}(\mathbf{r}_2) & \dots & \psi_{\mathbf{k}_{\frac{N_e}{2}}\uparrow}(\mathbf{r}_{N_e}) \\ \psi_{\mathbf{k}_{\frac{N_e}{2}}\downarrow}(\mathbf{r}_1) & \psi_{\mathbf{k}_{\frac{N_e}{2}}\downarrow}(\mathbf{r}_2) & \dots & \psi_{\mathbf{k}_{\frac{N_e}{2}}\downarrow}(\mathbf{r}_{N_e}) \end{vmatrix}. \quad (2)$$

Unfortunately, the electron-electron repulsion is strong, and the non-interacting electrons approximation is insufficient to understand real materials. Because \hat{V}_{ee} is not separable, with increasing N_e , finding the solution of the Schrödinger equation $H_e\psi = \varepsilon\psi$ becomes quickly an unfeasible task, even for a single atom.

A big step forward was the development of density-functional theory (DFT) [1, 2], described in detail in the Lecture of Peter Blöchl. DFT is based on the Hohenberg-Kohn theorem, which

establishes the one-to-one correspondence between the ground-state electron density $n(\mathbf{r})$ of an interacting system and the external potential $v_{\text{ext}}(\mathbf{r})$ acting on it. For any material described by the Hamiltonian (1), the ground-state total energy is a functional of the electron density, $E[n]$, which is minimized by the ground-state density. $E[n]$ can be written as

$$E[n] = F[n] + \int d\mathbf{r} v_{\text{ext}}(\mathbf{r})n(\mathbf{r}) + E_{nn} = F[n] + V[n] + E_{nn}.$$

$F[n] = T_e[n] + E_{ee}[n]$, the sum of the kinetic and electron-electron interaction energy, is a (unknown) universal functional (the same for all systems). $V[n]$ is the system-specific potential energy. The shift E_{nn} is the nucleus-nucleus interaction energy. The obstacle is that finding $n(\mathbf{r})$ still requires, in principle, the solution of the many-body problem (1). Kohn and Sham have shown, however, that $n(\mathbf{r})$ can be obtained by solving the Schrödinger equation of a fictitious non-interacting system, whose external potential $v_R(\mathbf{r})$ is chosen such that the ground-state density $n_0(\mathbf{r})$ equals $n(\mathbf{r})$

$$n(\mathbf{r}) = n_0(\mathbf{r}) = \sum_n^{\text{occ}} |\psi_n(\mathbf{r})|^2.$$

To obtain the Hamiltonian $\hat{h}_e^0(\mathbf{r})$ of such an auxiliary problem we rewrite $F[n]$ as

$$F[n] = T_0[n] + E_H[n] + E_{xc}[n] = T_0[n] + \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{xc}[n],$$

where $T_0[n]$ is the kinetic energy of the auxiliary system, $E_H[n]$ the classical electrostatic (or Hartree) energy, and $E_{xc}[n]$ is the small exchange-correlation correction,

$$E_{xc}[n] = E_{ee}[n] - E_H[n] + T_e[n] - T_0[n].$$

By minimizing the total energy with respect to $\{\psi_n\}$, with the constraint $\langle \psi_n | \psi_{n'} \rangle = \delta_{n,n'}$, we find the Kohn-Sham equation

$$\hat{h}_e^0(\mathbf{r}) \psi_n(\mathbf{r}) = \left[-\frac{1}{2} \nabla^2 + v_R(\mathbf{r}) \right] \psi_n(\mathbf{r}) = \varepsilon_n \psi_n(\mathbf{r}). \quad (3)$$

The eigenvalues ε_n are the Lagrange multipliers which enter the minimization through the constraint. The external (or reference) potential is given by

$$v_R(\mathbf{r}) = - \sum_{\alpha} \frac{Z_{\alpha}}{|\mathbf{r} - \mathbf{R}_{\alpha}|} + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{xc}[n]}{\delta n}.$$

The exchange-correlation functional is unknown, and includes a Coulomb ($E_{ee}[n] - E_H[n]$) and a kinetic energy ($T_e[n] - T_0[n]$) term. The latter can be transformed into a correction of the Coulomb term by means of a coupling-constant integration: The interaction V_{ee} is rescaled by a parameter λ (with $0 \leq \lambda \leq 1$), while keeping $n(\mathbf{r})$ fixed; this constraint is fulfilled through a reference potential $v_R^{\lambda}(\mathbf{r})$. Using the Hellmann-Feynman theorem to calculate $\frac{\partial E_{\lambda}}{\partial \lambda}$, where $E_{\lambda} = \langle \Psi_{\lambda} | H_{\lambda} | \Psi_{\lambda} \rangle$ is the ground-state energy at coupling constant λ , and then integrating over λ to obtain $E_1 - E_0$, one may show that

$$E_{xc}[n] = \int d\mathbf{r} \int d\mathbf{r}' \frac{n(\mathbf{r})n(\mathbf{r}')(\bar{g}(\mathbf{r}, \mathbf{r}') - 1)}{|\mathbf{r} - \mathbf{r}'|},$$

where

$$\bar{g}(\mathbf{r}, \mathbf{r}') = \int_0^1 d\lambda g_\lambda(\mathbf{r}, \mathbf{r}').$$

The quantity $n(\mathbf{r}, \mathbf{r}') = \sum_{\sigma, \sigma'} n(\mathbf{r}\sigma, \mathbf{r}'\sigma') = n(\mathbf{r}')n(\mathbf{r})g_\lambda(\mathbf{r}, \mathbf{r}')$ is the joint probability of finding electrons at \mathbf{r} and \mathbf{r}' . The function $g_\lambda(\mathbf{r}, \mathbf{r}')$ is the pair-correlation function. It can be shown that $g_\lambda(\mathbf{r}, \mathbf{r}') - 1$ vanishes in the large $|\mathbf{r} - \mathbf{r}'|$ limit.

In the Hartree-Fock approximation, in which the wavefunction is a Slater determinant, e.g. (2)

$$n(\mathbf{r}\sigma, \mathbf{r}'\sigma') = n(\mathbf{r}\sigma)n(\mathbf{r}'\sigma') - \delta_{\sigma, \sigma'} \left| \sum_i^{N_e/2} \bar{\psi}_{\mathbf{k}_i\sigma}(\mathbf{r})\psi_{\mathbf{k}_i\sigma}(\mathbf{r}') \right|^2, \quad (4)$$

where the last term accounts for the Pauli exclusion principle (*exchange*) and cancels the unphysical interaction of each electron with itself (*self-interaction*) present in the Hartree energy. The following sum rule holds for the pair-correlation function

$$\int d\mathbf{r}' n(\mathbf{r}') (g_\lambda(\mathbf{r}, \mathbf{r}') - 1) = -1. \quad (5)$$

This -1 is, in atomic units (Appendix A), a positive charge $-e$. The exchange-correlation energy $E_{xc}[n]$ may thus be interpreted as the energy gain due to the interaction of each electron with an *exchange-correlation hole* with charge density $n(\mathbf{r}') (g_\lambda(\mathbf{r}, \mathbf{r}') - 1)$ surrounding it. Since the *exchange hole* described by Eq. (4) already satisfy the sum rule (5), the remaining *correlation hole* redistributes the charge density of the hole. In the one-electron case ($N_e = 1$), $E_{xc}[n]$ merely cancels the Hartree self-interaction energy.

The main difficulty of DFT is to find good approximations to $E_{xc}[n]$. The most common is the local-density approximation (LDA), in which $E_{xc}[n]$ is replaced by its expression for a homogeneous interacting electron gas with density equal to the local density $n(\mathbf{r})$

$$E_{xc}[n] = \int d\mathbf{r} \epsilon_{xc}^{\text{LDA}}(n(\mathbf{r}))n(\mathbf{r}). \quad (6)$$

The LDA is particularly justified in systems with slowly varying spatial density $n(\mathbf{r})$. For such materials, we could split space into regions in which the density is basically constant and the system can indeed be described by a homogeneous electron gas; if we add up the contributions of all these regions of space we obtain the integral (6). The spin-polarized extension of the local-density approximation is the local spin-density approximation (LSDA).

The ground-state electron-density $n(\mathbf{r})$ can be obtained by solving (3) self-consistently. Various successful methods have been developed to find the eigenvalues and eigenvectors of (3), for solids and molecules. They are based on atomic-like orbitals (LMTO, NMTO), plane-waves (pseudopotentials), combinations of both (LAPW, PAW), gaussians, or Green functions (KKR) [3]. Through the years, DFT and the LDA have provided insight not only in solid-state physics, but also in chemistry and even in systems of biological interest. For this reason DFT became the *standard model* for electronic-structure calculations [1–3]. Strictly speaking, the Kohn-Sham energies ε_n have no physical meaning except the highest occupied state, which

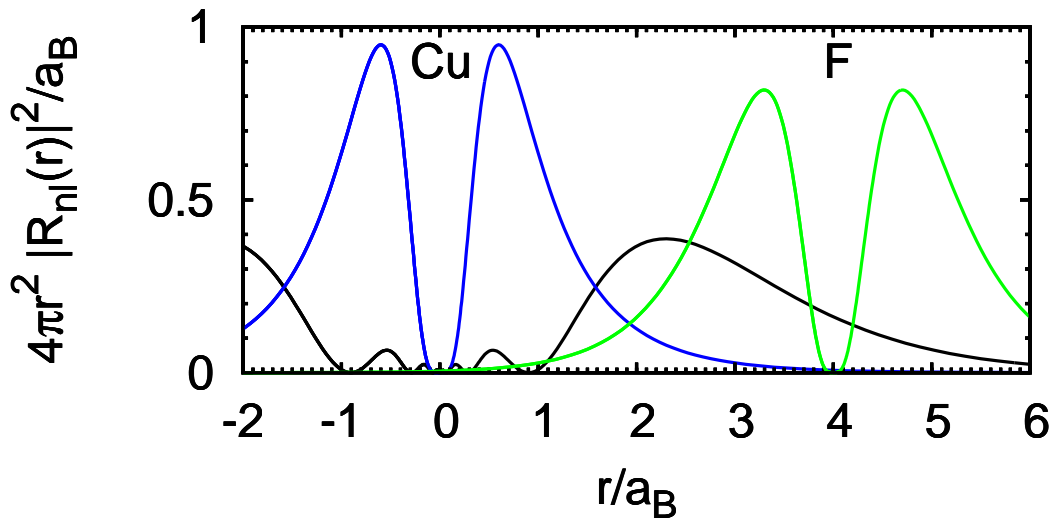


Fig. 1: LDA solution of the Schrödinger equation for a single atom: $4\pi r^2 |R_{nl}(r)|^2/a_B$ as a function of the distance from the nucleus, r/a_B (atomic units, Appendix A). Blue: Cu 3d. Black: Cu 4s. The 2p orbital of a F atom in $r = 4a_B$ is also shown (green). Cu has the electronic configuration $[Ar] 3d^{10}4s^1$ and F the configuration $[He] 2s^22p^5$.

yields the ionization energy, and their identification with one-particle energies is not justified. The Kohn-Sham orbitals $\psi_n(\mathbf{r})$ are just a tool to generate the ground-state density $n(\mathbf{r})$. Nevertheless, in practice Kohn-Sham orbitals turned out to be very useful to explain the properties of solids. Fermi surfaces, chemistry and many features of the electronic structure are qualitatively and often quantitatively well described by DFT in the LDA approximation or its extensions. The energy gap of semiconductors is underestimated, but can be corrected within many-body perturbation theory (GW approximation, discussed in the Lecture of Karsten Held).

LDA fails to capture, however, the essential physics of *strongly-correlated systems*, even at a qualitative level. At the center of this discrepancy are many-body effects between electrons in open d or f shells. Since these electrons are very *localized*, the Coulomb repulsion between them is significant. When Coulomb repulsion is strong, electrons lose their individuality: The dynamics of a single electron depends on the position of all others, the Coulomb repulsion of which it has to avoid (electrons are strongly correlated), and cannot be described by a reference mean-field potential. This happens for example in the case of *Mott insulators*. Because the Kohn-Sham Hamiltonian (3) with the LDA exchange-correlation potential describes independent electrons, many-electron states can be built from the Kohn-Sham orbitals as a single Slater determinant. Thus, a non-magnetic crystal with an odd number of electrons per unit cell has partially filled bands because of spin degeneracy, and therefore is metallic. However, due to Coulomb repulsion, several transition-metal compounds with partially filled d shells are insulating, paramagnetic above the Neel temperature T_N , and sometimes exhibit a large gap. In Fig. 1 the extensions of the atomic radial functions for the outer orbitals, 3d and 4s, of Cu can be compared. While for 3d electrons the radial function decays very rapidly with distance, for

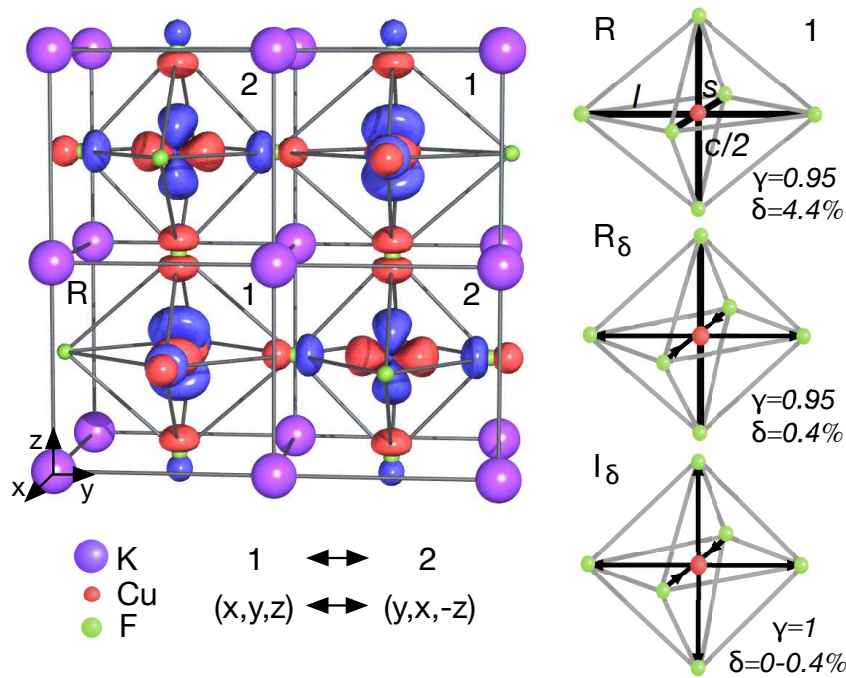


Fig. 2: Crystal structure, distortions and orbital order in KCuF_3 . Cu is at the center of F octahedra enclosed in a K cage. The conventional cell is tetragonal with axes a , b , c . The pseudocubic axes x , y , z pointing towards neighboring Cu are shown in the corner. Short (s) and long (l) CuF bonds alternate between x and y along all pseudocubic axes (co-operative Jahn-Teller distortion). The distortions are measured by $\delta = (l - s)/(l + s)/2$ and $\gamma = c/a\sqrt{2}$. R is the experimental structure ($\gamma = 0.95$, $\delta = 4.4\%$), R_δ ($\gamma = 0.95$) and I_δ ($\gamma = 1$) two ideal structures with reduced distortions. In the I_0 structure the cubic crystal-field at the Cu site splits the $3d$ manifold into a t_{2g} triplet and a e_g doublet. In the R structure, site symmetry is lowered further by the tetragonal compression ($\gamma < 1$) and the Jahn-Teller distortion ($\delta \neq 0$). The figure shows the highest energy d orbital. From Ref. [4].

$4s$ electrons it is still sizable $\sim 2 \text{ \AA}$ away from the nucleus, a typical interatomic distance in a lattice. Thus in a crystal $4s$ electrons are likely to form delocalized states, while $3d$ electrons tend to retain part of their atomic characteristics.

As example we take KCuF_3 . This system has a perovskite structure, shown in Fig. 2, with each Cu surrounded by a F octahedron. The nominal valence for K, Cu and F is K^+ ($4s^0$), F^- ($2p^6$), Cu^{2+} ($3d^9 4s^0$). The cubic crystal field at the Cu site splits the partially filled $3d$ levels into the lower energy t_{2g} ($|xy\rangle, |xz\rangle, |yz\rangle$), and the higher energy e_g ($|x^2 - y^2\rangle, |3z^2 - r^2\rangle$) manifold; the electronic configuration is $t_{2g}^6 e_g^3$. The co-operative Jahn-Teller distortion and the tetragonal compression further reduce the site symmetry of Cu, and the e_g doublet splits into $|3l^2 - 1\rangle$ and $|s^2 - z^2\rangle$. Because long (l) and short (s) CuF bond alternate between x and y along all cubic axes, the highest energy d orbitals, $|s^2 - z^2\rangle$, form the pattern shown in Fig. 2. The LDA band structure of KCuF_3 is shown in Fig. 3. We can identify the bands from their main character as F p -like (filled), Cu t_{2g} -like (filled), Cu e_g -like (occupied by 3 electrons), Cu s - and K s -like (empty). The Fermi level is located in the middle of the e_g -like bands.

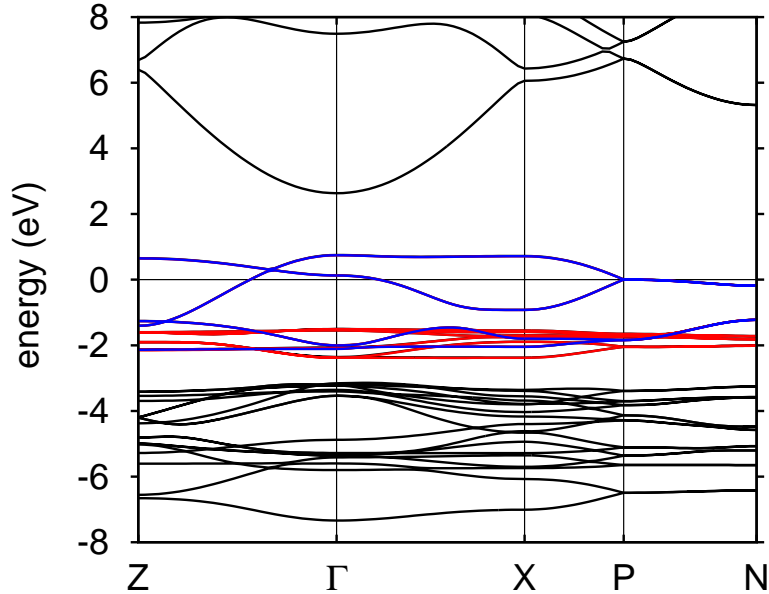


Fig. 3: LDA band structure of KCuF_3 . Blue: Cu e_g -like bands. Red: Cu t_{2g} -like bands. Black: filled F p -like bands and empty bands.

Thus LDA predicts that KCuF_3 is a metal, although it actually is an insulator (paramagnetic down to $T_N = 40$ K). A similar problem occurs in many other transition-metal compounds with partially filled d shells: manganites, vanadates, titanates. This discrepancy cannot be solved by simple improvements of the LDA functional. Coulomb repulsion effects beyond mean field are essential to understand the origin of the insulating state in these materials. Other systems for which similar considerations apply are heavy fermions and Kondo systems (f electrons) or organics (molecular crystals).

2 Low-energy models

Lacking a working *ab-initio* theory, strongly-correlated systems have been studied for a long time through low-energy model Hamiltonians. Within this approach only the states and interactions believed to be most important to describe a given phenomenon are considered. Models can be justified on the ground that at low energy, high-energy degrees of freedom can be, in principle, projected out (*downfolded*), in the spirit of Wilson renormalization group. Their main effect is assumed to be included implicitly in the low-energy model through a renormalization of parameters. In LDA strongly-correlated transition-metal compounds usually have narrow d bands close to the Fermi level (see Fig. 3) and thus the d bands, or a subgroup of those (e_g -bands for KCuF_3) are believed to be the essential degrees of freedom. The minimal model to describe a system with a narrow band at the Fermi level is the Hubbard model

$$\hat{H} = -t \sum_{\sigma \langle ii' \rangle} c_{i\sigma}^\dagger c_{i'\sigma} + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} = \hat{H}_0 + \hat{U}, \quad (7)$$

where $c_{i\sigma}^\dagger$ ($c_{i\sigma}$) creates (destroys) an electron with spin σ at site i , $\hat{n}_{i\sigma} = c_{i\sigma}^\dagger c_{i\sigma}$ gives the i -site occupancy per spin, t is the hopping integral between first neighbors, and U the on-site Coulomb repulsion.

In the non-interacting limit ($U = 0$), the Hamiltonian (7) can be written in diagonal form

$$\hat{H}_0 = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} \hat{n}_{\mathbf{k}\sigma}.$$

The band energy is given by $\varepsilon_{\mathbf{k}} = -t \frac{1}{N} \sum_{\langle ii' \rangle} e^{i\mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_{i'})}$, where \mathbf{R}_i are lattice vectors, and N is the number of sites; the operator $c_{\mathbf{k}\sigma}^\dagger$ is the Fourier transform of $c_{i\sigma}^\dagger$, i.e., $c_{\mathbf{k}\sigma}^\dagger = \frac{1}{\sqrt{N}} \sum_i e^{i\mathbf{k} \cdot \mathbf{R}_i} c_{i\sigma}^\dagger$, and $\hat{n}_{\mathbf{k}\sigma} = c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma}$. At half-filling ($N_e = N$), the ground state is paramagnetic and metallic.

In the atomic limit ($t = 0$), the model (7) describes instead an insulating collection of independent atoms with disordered magnetic moments.

Thus the Hubbard model captures the essence of the paramagnetic metal to paramagnetic insulator (Mott) transition, and can qualitatively explain why systems like KCuF_3 are paramagnetic insulators in a large temperature range. Furthermore, it explains the fact that KCuF_3 and most strongly-correlated transition-metal compounds have an antiferromagnetic ground state. For small t/U , by downfolding doubly occupied states, the Hubbard model (7) can be mapped onto a spin 1/2-antiferromagnetic Heisenberg model

$$\hat{H} \rightarrow J_{\text{AFM}} \frac{1}{2} \sum_{\langle ii' \rangle} \left[\mathbf{S}_i \cdot \mathbf{S}_{i'} - \frac{1}{4} \hat{n}_i \hat{n}_{i'} \right],$$

with coupling $J_{\text{AFM}} = 4t^2/U$. Thus at low temperature, when charge fluctuations play a minor role, a transition to an antiferromagnetic state can take place. In strongly-correlated transition-metal compounds, where the hopping t between correlated d states is mediated by the p orbitals of the atom between two transition metals (e.g., F p states in KCuF_3 , Fig. 2), this many-body exchange mechanism is called *super-exchange*. Because the Hubbard model can be solved exactly only in special cases (e.g., in one dimension), it was for a long time impossible to understand the nature of the Mott transition within this model. Understanding real materials appeared even less likely. Progress came with the development of the dynamical mean-field theory (DMFT) [5]. In DMFT, the Hubbard model, which describes a lattice of correlated sites, is mapped onto an effective Anderson model, which describes a correlated *impurity*

$$\hat{H}_{\text{eff}} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}} \hat{n}_{\mathbf{k}\sigma} + \varepsilon_d \sum_{\sigma} \hat{n}_{d\sigma} + U \hat{n}_{d\uparrow} \hat{n}_{d\downarrow} + \sum_{\mathbf{k}\sigma} (V_{\mathbf{k}d} c_{\mathbf{k}\sigma}^\dagger d_{\sigma} + \bar{V}_{\mathbf{k}d} d_{\sigma}^\dagger c_{\mathbf{k}\sigma}).$$

Here d_{σ}^\dagger (d_{σ}) creates (destroys) an electron at the impurity site, and $\hat{n}_{d\sigma} = d_{\sigma}^\dagger d_{\sigma}$ counts the number of electrons on the impurity; $c_{\mathbf{k}\sigma}^\dagger$ ($c_{\mathbf{k}\sigma}$) creates (destroys) a *bath* electron with energy $\varepsilon_{\mathbf{k}}$, and $V_{\mathbf{k}d}$ is the hybridization between bath and impurity. This auxiliary quantum-impurity model is solved self-consistently. The solution is found when the interacting Green function $G(\omega)$ of the auxiliary model equals the local Green function $G_{ii}(\omega)$ of the Hubbard model (7)

$$G(\omega) = G_{ii}(\omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k} \in \text{BZ}} \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}} - \Sigma(\omega)} = \int d\varepsilon \frac{\rho(\varepsilon)}{\omega + \mu - \varepsilon - \Sigma(\omega)}. \quad (8)$$

Here μ is the chemical potential, the sum is over $N_{\mathbf{k}}$ \mathbf{k} -points of the Brillouin Zone (BZ), $\Sigma(\omega)$ is the self-energy of the quantum impurity model and $\rho(\varepsilon)$ is the density of states. The self-energy $\Sigma(\omega)$ can be obtained from the Dyson equation of the impurity problem

$$\mathcal{G}^{-1}(\omega) = G^{-1}(\omega) + \Sigma(\omega), \quad (9)$$

where $\mathcal{G}(\omega)$ is the non-interacting Green function of the Anderson model (bath Green function). The main approximation in DMFT consists in neglecting spatial fluctuations in the lattice self-energy; this approximation becomes exact in the limit of infinite coordination number [5]. The Anderson model is a full many-body Hamiltonian, known since long in the framework of the Kondo effect [6], but, in contrast to the original Hubbard model, it describes only a single correlated site. It can be solved numerically with different approaches (*quantum-impurity solvers*): the numerical renormalization group [6], various flavors of quantum Monte Carlo (QMC) [7,8], Lanczos [9], or other methods [6, 10]. Some of the most important solvers are presented in the Lectures of Erik Koch, Nils Blümer, and Philipp Werner. If we use QMC, we have to work in imaginary time/frequencies, and replace the frequency ω in (8, 9) with $i\omega_n$, where ω_n are Fermionic Matsubara frequencies, $\omega_n = (2n + 1)\pi k_B T$, and T is the temperature.

The DMFT approach is discussed in detail in the Lecture of Marcus Kollar. We recall here some important conclusions obtained by studying the half-filled Bethe lattice, described for $U = 0$ by a semi-elliptical density of states [10]. In the Fermi-liquid regime (metallic phase, low temperature, $\omega \sim \mu = 0$), the self-energy can be expanded as

$$\Sigma(\omega + i0^+) \sim \frac{U}{2} + (1 - 1/Z)\omega - i\Delta\omega^2 + \dots$$

The effective mass of quasi-particles is $m^* = m/Z$ and their life-time $\propto 1/\Delta$; Z is the quasi-particle weight. In the Mott insulating regime, the self-energy has instead the following low-frequency behavior

$$\Sigma(\omega + i0^+) \sim \frac{U}{2} + \Gamma/\omega - i\pi\Gamma\delta(\omega) + \dots,$$

where Γ can be viewed as an order parameter. Thus the real part of the self-energy diverges at $\omega \sim 0$; the strong ω dependence of $\Sigma(\omega)$ is essential to obtain the Mott metal-insulator transition in the one-band Hubbard model.

The model Hamiltonian approach has proven effective in gaining insight into the behavior of strongly-correlated systems. However, the actual derivation of low-energy models by down-folding the full many-body problem, although formally possible, is in practice unfeasible and would in general lead to complex interactions beyond the Hubbard model [11]. The insight is thus gained at the price of neglecting all interactions that are thought not to have a direct influence on the specific phenomenon, and then relating the few free parameters (here t and U) to experimental data. It is clear that simple models such as the Hubbard model (7), although grasping an essential aspect of Mott physics, are hardly sufficient to describe the complexity of real materials such as KCuF_3 . Thus they have been extended to include many orbitals (e.g., the full d shell), crystal-field splittings (which divides the d shell, e.g., into 3-fold degenerate t_{2g}

and 2-fold degenerate e_g states), multiplets (whose description requires taking into account the full Coulomb interaction tensor and the spin-orbit interaction), lattice distortions (which, in the case of KCuF_3 , split the e_g and the t_{2g} manifold and change the hopping integrals), filled (e.g., $F p$ in KCuF_3) or excited ($\text{Cu } s$, $\text{K } s$, ...) states, and Coulomb repulsion between neighbors. As we will see later in some examples, these *details do matter* when we want to understand real materials; neglecting them easily leads to wrong conclusions. Some of the parameters of such extended Hubbard models can indeed be obtained by fitting to experiments, but with the increasing number of free parameters it becomes impossible to put any theory to a real test.

3 Many-body models from DFT

3.1 Towards ab-initio Hamiltonians

The dream of calculating the parameters of model Hamiltonians *ab-initio* exists since long. We know from the successes of LDA that the Kohn-Sham orbitals obtained within the local-density approximation carry the essential information about the structure and chemistry of a given material. It appears therefore natural to build material-specific many-body models starting from the LDA. This can be achieved by constructing *ab-initio* a basis of localized LDA Wannier functions

$$\psi_{in\sigma}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{-i\mathbf{R}_i \cdot \mathbf{k}} \psi_{n\mathbf{k}\sigma}(\mathbf{r}),$$

and the corresponding many-body Hamiltonian, which is the sum of an LDA term \hat{H}^{LDA} , a Coulomb term \hat{U} , and a double-counting correction \hat{H}_{DC}

$$\hat{H}_e = \hat{H}^{\text{LDA}} + \hat{U} - \hat{H}_{\text{DC}}. \quad (10)$$

The LDA part of the Hamiltonian is given by

$$\hat{H}^{\text{LDA}} = - \sum_{\sigma} \sum_{in, i'n'} t_{n, n'}^{i, i'} c_{in\sigma}^{\dagger} c_{i'n'\sigma}, \quad (11)$$

where $c_{in\sigma}^{\dagger}$ ($c_{in\sigma}$) creates (destroys) an electron with spin σ in orbital n at site i , and

$$t_{n, n'}^{i, i'} = - \int d\mathbf{r} \bar{\psi}_{in\sigma}(\mathbf{r}) \left[-\frac{1}{2} \nabla^2 + v_{\text{R}}(r) \right] \psi_{i'n'\sigma}(\mathbf{r}). \quad (12)$$

The on-site ($i = i'$) terms yield the crystal-field matrix while the $i \neq i'$ contributions are the hopping integrals. The Coulomb interaction \hat{U} is given by

$$\hat{U} = \frac{1}{2} \sum_{ii'jj'} \sum_{\sigma\sigma'} \sum_{nn'pp'} U_{np n'p'}^{ijj'j'} c_{in\sigma}^{\dagger} c_{jp\sigma'}^{\dagger} c_{j'p'\sigma'} c_{i'n'\sigma},$$

with

$$\begin{aligned} U_{np n'p'}^{ijj'j'} &= \langle in\sigma \ jp\sigma' | \hat{U} | i'n'\sigma \ j'p'\sigma' \rangle \\ &= \int d\mathbf{r}_1 \int d\mathbf{r}_2 \bar{\psi}_{in\sigma}(\mathbf{r}_1) \bar{\psi}_{jp\sigma'}(\mathbf{r}_2) \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_{j'p'\sigma'}(\mathbf{r}_2) \psi_{i'n'\sigma}(\mathbf{r}_1). \end{aligned} \quad (13)$$

To avoid double counting, the correction \hat{H}_{DC} should cancel the electron-electron interaction contained in \hat{H}^{LDA} , which in (10) is explicitly described by \hat{U} . However, it is more meaningful to take advantage of the LDA, and subtract from \hat{U} the long-range Hartree and the mean-field exchange-correlation interaction, described by the LDA; from the successes of the LDA we expect that they are well accounted for. Thus the difference $\hat{U} - \hat{H}_{\text{DC}}$ is a short-range many-body correction to the LDA.

The Hamiltonian (10) still describes the complete many-body problem, finding the solution of which remains an impossible task. To make progress, we separate the electrons in two types, the *correlated* or heavy electrons and the *uncorrelated* or light electrons. For the correlated electrons LDA fails qualitatively and $\hat{U} - \hat{H}_{\text{DC}}$ has to be taken into account explicitly; we can assume that $\hat{U} - \hat{H}_{\text{DC}}$ is local (on-site) or almost local (between near neighbors). For the uncorrelated electrons the LDA is overall a good approximation, and we do not consider any correction $\hat{U} - \hat{H}_{\text{DC}}$. By truncating $\hat{U} - \hat{H}_{\text{DC}}$ to the correlated sector we implicitly assume that the effect of light electrons is a mere renormalization or *screening* of Coulomb parameters in the correlated sector. This implies that the Coulomb couplings cannot be calculated directly from (13). The calculation of screened Coulomb integrals remains to date a major challenge, and effects beyond Coulomb parameter renormalization are usually neglected. Since correlated electrons partially retain their atomic character, in the rest of this Lecture we will label them with the quantum numbers $lm\sigma$, as in the atomic limit. To first approximation we can assume that $\hat{U} - \hat{H}_{\text{DC}}$ is local and that correlated electrons belong to a single shell (e.g., d electrons, $l = 2$). Thus we can write the Hamiltonian as the generalized Hubbard model

$$\hat{H}_e = \hat{H}^{\text{LDA}} + \hat{U}^l - \hat{H}_{\text{DC}}^l, \quad (14)$$

where the *screened* Coulomb interaction is

$$\hat{U}^l = \frac{1}{2} \sum_i \sum_{\sigma\sigma'} \sum_{m_\alpha m'_\alpha} \sum_{m_\beta m'_\beta} U_{m_\alpha m_\beta m'_\alpha m'_\beta} c_{im_\alpha\sigma}^\dagger c_{im_\beta\sigma'}^\dagger c_{im'_\beta\sigma'} c_{im'_\alpha\sigma}, \quad (15)$$

and \hat{H}_{DC}^l is, in principle, the mean-field value of \hat{U}^l . More generally, we can include in the Hamiltonian (14) the Coulomb interaction between first neighbors, different shells, etc. Although (14) is simpler than the original model (10), it still describes a full many-body problem; however, as we will see later, such a problem can be solved numerically within the dynamical mean-field approximation.

At the heart of (14) is the assumption that the $\hat{U}^l - \hat{H}_{\text{DC}}^l$ is local, or almost local; thus it is essential to use a localized basis in the correlated sector, a basis in which the separation of heavy and light electrons makes actually sense. Localized Wannier functions can be constructed in different ways. Successful methods are the *ab-initio* downfolding procedure based on the NMTO approach [12] and the maximally-localized Wannier functions algorithm of Marzari and Vanderbilt [13]; a lighter alternative to localized Wannier functions are projected local orbitals [14]. The latter two methods are presented in the Lecture of Jan Kuneš.

3.2 Coulomb interaction tensor

The screened Coulomb interaction, central to building the Hamiltonian (14), has the same form as the bare Coulomb interaction tensor. To identify the different terms in \hat{U}^l , it is useful to derive the bare Coulomb integrals for atomic orbitals $\psi_{nlm}(\mathbf{r}) = R_{nl}(r)Y_m^l(\theta, \phi)$ (see Appendix B). The generalization to Wannier orbitals is straightforward. First, we write the electronic positions in spherical coordinates, $\mathbf{r}_i = r_i(\sin \theta_i \cos \phi_i, \sin \theta_i \sin \phi_i, \cos \theta_i)$, and express the Coulomb interaction as

$$\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} = \sum_{k=0}^{\infty} \frac{r_{<}^k}{r_{>}^{k+1}} \frac{4\pi}{2k+1} \sum_{q=-k}^k Y_q^k(\theta_2, \phi_2) \bar{Y}_q^k(\theta_1, \phi_1), \quad (16)$$

where $r_{<}$ ($r_{>}$) is the smaller (larger) of r_1 and r_2 . By inserting (16) into (13) we obtain

$$U_{m_\alpha m_\beta m'_\alpha m'_\beta} = \sum_{k=0}^{2l} a_k(m_\alpha m'_\alpha, m_\beta m'_\beta) F_k,$$

where a_k are angular integrals

$$a_k(m_\alpha m'_\alpha, m_\beta m'_\beta) = \frac{4\pi}{2k+1} \sum_{q=-k}^k \langle lm_\alpha | Y_q^k | lm'_\alpha \rangle \langle lm_\beta | \bar{Y}_q^k | lm'_\beta \rangle,$$

and F_k radial Slater integrals

$$F_k = \int dr_1 r_1^2 \int dr_2 r_2^2 R_{nl}^2(r_1) \frac{r_{<}^k}{r_{>}^{k+1}} R_{nl}^2(r_2).$$

The most important Coulomb integrals are the two-index terms: the direct ($U_{mm'mm'}$) and exchange ($U_{mm'm'm}$, with $m \neq m'$) integrals, which can be expressed as

$$\begin{aligned} U_{mm'mm'} = U_{m,m'} &= \sum_{k=0}^{2l} a_k(mm, m'm') F_k, \\ U_{mm'm'm} = J_{m,m'} &= \sum_{k=0}^{2l} a_k(mm', m'm) F_k. \end{aligned}$$

It can be shown that $U_{m,m'}$ and $J_{m,m'}$ are positive, and that $U_{m,m'} \geq J_{m,m'}$. If we neglect all terms but the direct and the exchange Coulomb interaction, only density-density terms ($\propto \hat{n}_{im\sigma} \hat{n}_{im'\sigma'}$, with $\hat{n}_{im\sigma} = c_{im\sigma}^\dagger c_{im\sigma}$) remain, and the Coulomb interaction takes a simpler form

$$\hat{U}^l \sim \frac{1}{2} \sum_{i\sigma} \sum_{mm'} U_{m,m'} \hat{n}_{im\sigma} \hat{n}_{im'-\sigma} + \frac{1}{2} \sum_{i\sigma} \sum_{m \neq m'} (U_{m,m'} - J_{m,m'}) \hat{n}_{im\sigma} \hat{n}_{im'\sigma}. \quad (17)$$

The contributions neglected in (17), spin-flip exchange terms ($\propto J_{m,m'}$) and off-diagonal contributions (terms with more than two different orbital indices), are important to get the correct multiplet structure. They are often neglected in DMFT calculations based on QMC solvers because they can generate a strong sign problem.

In electronic-structure calculations real harmonics (Appendix B) rather than spherical harmonics are normally used; therefore here we will give the Coulomb integrals for d electrons in the basis of real harmonics. It is useful to introduce first average Coulomb parameters

$$U_{avg} = \frac{1}{(2l+1)^2} \sum_{m,m'} U_{m,m'} = F_0,$$

$$U_{avg} - J_{avg} = \frac{1}{2l(2l+1)} \sum_{m,m'} (U_{m,m'} - J_{m,m'}).$$

For d -electrons, only F_0 , F_2 and F_4 contribute to the Coulomb integrals, and we can show that $J_{avg} = (F_2 + F_4)/14$. For hydrogen-like $3d$ orbitals, $F_4/F_2 = 15/23$, while for realistic $3d$ orbitals this ratio is slightly smaller. A typical value is $F_4/F_2 \sim 0.625 = 5/8$.

The parameters $U_{m,m'}^l$ may be written as

$U_{m,m'}^l$	$ xy\rangle$	$ yz\rangle$	$ 3z^2 - r^2\rangle$	$ xz\rangle$	$ x^2 - y^2\rangle$
$ xy\rangle$	U_0	$U_0 - 2J_1$	$U_0 - 2J_2$	$U_0 - 2J_1$	$U_0 - 2J_3$
$ yz\rangle$	$U_0 - 2J_1$	U_0	$U_0 - 2J_4$	$U_0 - 2J_1$	$U_0 - 2J_1$
$ 3z^2 - r^2\rangle$	$U_0 - 2J_2$	$U_0 - 2J_4$	U_0	$U_0 - 2J_4$	$U_0 - 2J_2$
$ xz\rangle$	$U_0 - 2J_1$	$U_0 - 2J_1$	$U_0 - 2J_4$	U_0	$U_0 - 2J_1$
$ x^2 - y^2\rangle$	$U_0 - 2J_3$	$U_0 - 2J_1$	$U_0 - 2J_2$	$U_0 - 2J_1$	U_0

where

$$U_0 = U_{avg} + \frac{8}{7}J_{avg} = U_{avg} + \frac{8}{5}\mathcal{J}_{avg}$$

$$J_1 = \frac{3}{49}F_2 + \frac{20}{9} \frac{1}{49}F_4$$

$$J_2 = -2\mathcal{J}_{avg} + 3J_1$$

$$J_3 = 6\mathcal{J}_{avg} - 5J_1$$

$$J_4 = 4\mathcal{J}_{avg} - 3J_1.$$

The parameter \mathcal{J}_{avg} is the actual average of the exchange terms in the basis of real harmonics

$$\mathcal{J}_{avg} = \frac{1}{2l(2l+1)} \sum_{m \neq m'} J_{m,m'} = \frac{5}{7}J_{avg}.$$

For atomic d states, U_{avg} is very large (15 – 20 eV), but screening effects reduce it drastically. The calculation of screening effects is very difficult because it is basically equivalent to finding the solution of the full many-body problem. Approximate schemes are the constrained LDA (cLDA) approach [16] and the constrained RPA (cRPA) method [17]. In cLDA, the screened U is obtained from the second derivative of the total energy as a function of the density; to avoid electron transfer between correlated and uncorrelated sectors, the hopping integrals between heavy and light electrons are cut. In cRPA the polarization (and thus the screened Coulomb interaction) is obtained in the random-phase approximation by downfolding the uncorrelated sector, assuming that the latter is well described by mean field. These approaches are discussed in the Lecture of Ferdi Aryasetiawan.

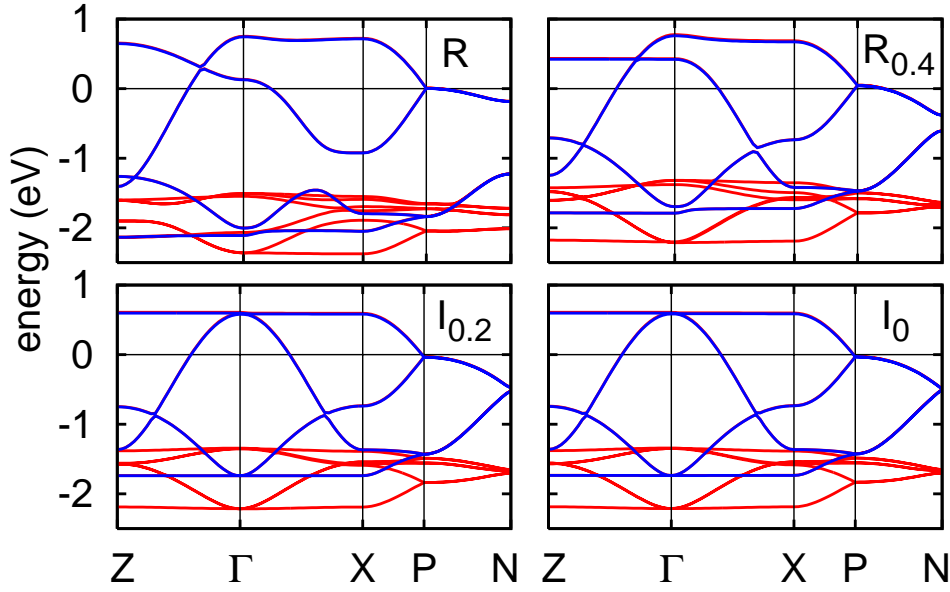


Fig. 4: LDA e_g (blue) and t_{2g} (red) band structure of KCuF_3 for the experimental structure (R) and ideal structures with progressively reduced distortions (see Fig. 2). I_0 : simple cubic. The unit cell contains 2 formula units. From Ref. [4].

3.3 Minimal material-specific models

To understand a given system it is convenient to make the correlated electron sector as small as possible, while still retaining the crucial degrees of freedom. To this end we have to construct *minimal model Hamiltonians*, which are still material specific but have as few degrees of freedom and parameters as possible. Here we will see how this can be achieved through massive downfolding of the LDA Hamiltonian. As example we consider the case of KCuF_3 in tight-binding theory. For simplicity, we neglect the tetragonal and Jahn-Teller distortions. In the cubic structure, the primitive cell contains one formula unit (a single K cube in Fig. 2). The cubic axes are x , y , z , and the lattice constant a . A Cu atom at site \mathbf{R}_i is surrounded by two apical F atoms, F_1 at $\mathbf{R}_i + \frac{1}{2}\mathbf{z}$ and F_2 at $\mathbf{R}_i - \frac{1}{2}\mathbf{z}$, and four planar F atoms, F_3 and F_4 at $\mathbf{R}_i \pm \frac{1}{2}\mathbf{x}$ and F_5 and F_6 at $\mathbf{R}_i \pm \frac{1}{2}\mathbf{y}$. In Fig. 4 one can see the effects of the cubic approximation on the e_g bands: the crystal-field splitting of e_g states is zero, the band width slightly reduced, gaps disappear, and the dispersion relations is sizably modified (e.g., along ΓZ). We take as Wannier basis the atomic $3d$ e_g orbitals for Cu and the $2p$ orbitals for F; we neglect the overlap integrals and all other states. The main contribution to the hopping integrals (12) are the Slater-Koster two-center matrix elements (Appendix B). In the case described, the only relevant Slater-Koster parameter is $V_{pd\sigma}$. The $|3z^2 - r^2\rangle_i$ and $|x^2 - y^2\rangle_i$ states of the Cu at \mathbf{R}_i are coupled via $V_{pd\sigma}$ to $|z\rangle_i$, the p_z orbitals of F_1 and F_2 , to $|x\rangle_i$, the p_x orbitals of F_3 and F_4 and to $|y\rangle_i$, the p_y orbitals of F_5 and F_6 . From the basis $|\alpha\rangle_i$ (where $\alpha = x, y, z, 3z^2 - r^2, x^2 - y^2$), we construct the Bloch states $|\mathbf{k}\alpha\rangle = \frac{1}{\sqrt{N}} \sum_i e^{i\mathbf{k}\cdot\mathbf{R}_i} |\alpha\rangle_i$, and obtain the tight-binding Hamiltonian

H^{TB}	$ \mathbf{k} z\rangle$	$ \mathbf{k} x\rangle$	$ \mathbf{k} y\rangle$	$ \mathbf{k} 3z^2 - r^2\rangle$	$ \mathbf{k} x^2 - y^2\rangle$
$ \mathbf{k} z\rangle$	ε_p	0	0	$-2V_{pd\sigma}s_z$	0
$ \mathbf{k} x\rangle$	0	ε_p	0	$V_{pd\sigma}s_x$	$-\sqrt{3}V_{pd\sigma}s_x$
$ \mathbf{k} y\rangle$	0	0	ε_p	$V_{pd\sigma}s_y$	$\sqrt{3}V_{pd\sigma}s_y$
$ \mathbf{k} 3z^2 - r^2\rangle$	$-2V_{pd\sigma}\bar{s}_z$	$V_{pd\sigma}\bar{s}_x$	$V_{pd\sigma}\bar{s}_y$	ε_d	0
$ \mathbf{k} x^2 - y^2\rangle$	0	$-\sqrt{3}V_{pd\sigma}\bar{s}_x$	$\sqrt{3}V_{pd\sigma}\bar{s}_y$	0	ε_d

(18)

where $s_\alpha = ie^{-ik_\alpha a/2} \sin k_\alpha a/2$, $\alpha = x, y, z$, $\varepsilon_p < \varepsilon_d = \varepsilon_p + \Delta_{pd}$, and $V_{pd\sigma} < 0$. If $|V_{pd\sigma}|/\Delta_{pd}$ is small, the occupied bands are F p -like, while the partially filled bands Cu e_g -like. We now calculate the bands along high-symmetry lines.¹ Along ΓZ , the eigenvalues ε_i ($\varepsilon_i \leq \varepsilon_{i+1}$) of H^{TB} are

$$\begin{aligned} \varepsilon_2 &= \varepsilon_p \\ \varepsilon_3 &= \varepsilon_p \\ \varepsilon_4 &= \varepsilon_d \\ \varepsilon_{1,5} &= \varepsilon_p + \frac{1}{2}\Delta_{pd} \pm \frac{1}{2}\sqrt{\Delta_{pd}^2 + 16V_{pd\sigma}^2|s_z|^2} \end{aligned}$$

where ε_1 is bonding and F z -like, while ε_5 anti-bonding and $3z^2 - r^2$ -like. Along ΓX , we have instead the dispersion relations

$$\begin{aligned} \varepsilon_2 &= \varepsilon_p \\ \varepsilon_3 &= \varepsilon_p \\ \varepsilon_4 &= \varepsilon_d \\ \varepsilon_{1,5} &= \varepsilon_p + \frac{1}{2}\Delta_{pd} \pm \frac{1}{2}\sqrt{\Delta_{pd}^2 + 16V_{pd\sigma}^2|s_x|^2} \end{aligned}$$

where ε_1 is bonding and F x -like, while ε_5 anti-bonding and $x^2 - y^2$ -like.

To obtain the e_g -like bands, instead of diagonalizing H^{TB} as we have done above, we can also use the downfolding procedure, which, for independent electrons, can be done exactly. We divide the orbitals in passive (F p) and active (Cu d), and write the eigenvalues equation as

$$\begin{bmatrix} H_{pp} & H_{pd} \\ H_{dp} & H_{dd} \end{bmatrix} \begin{bmatrix} |\mathbf{k} p\rangle \\ |\mathbf{k} d\rangle \end{bmatrix} = \varepsilon \begin{bmatrix} I_{pp} & 0 \\ 0 & I_{dd} \end{bmatrix} \begin{bmatrix} |\mathbf{k} p\rangle \\ |\mathbf{k} d\rangle \end{bmatrix},$$

where H_{pp} (I_{pp}) is the Hamiltonian (identity matrix) in the p -electron space (3×3), and H_{dd} (I_{dd}) the Hamiltonian (identity matrix) in the d -electron space (2×2). By downfolding to the d sector we obtain the energy-dependent operator H_{dd}^ε , which acts in the d space

$$H_{dd}^\varepsilon = H_{dd} - H_{dp}(H_{pp} - \varepsilon I_{pp})^{-1}H_{pd},$$

and a correspondingly transformed and energy-dependent basis set for the active space, $|\mathbf{k} d\rangle_\varepsilon$. The operator H_{dd}^ε has the same eigenvalues and eigenvectors as the original Hamiltonian. In the case of KCuF_3

H_{dd}^ε	$ \mathbf{k} 3z^2 - r^2\rangle_\varepsilon$	$ \mathbf{k} x^2 - y^2\rangle_\varepsilon$
$ \mathbf{k} 3z^2 - r^2\rangle_\varepsilon$	$\varepsilon'_d - 2t_\varepsilon[\frac{1}{4}(\cos k_x a + \cos k_y a) - \cos k_z a]$	$2t_\varepsilon[\frac{\sqrt{3}}{4}(\cos k_x a - \cos k_y a)]$
$ \mathbf{k} x^2 - y^2\rangle_\varepsilon$	$2t_\varepsilon[\frac{\sqrt{3}}{4}(\cos k_x a - \cos k_y a)]$	$\varepsilon'_d - 2t_\varepsilon[\frac{3}{4}(\cos k_x a + \cos k_y a)]$

(19)

¹Special points: $\Gamma = (0, 0, 0)$, $Z = (0, 0, \pi/a)$, $X = (\pi/a, 0, 0)$, $M = (\pi/a, \pi/a, 0)$, $R = (\pi/a, \pi/a, \pi/a)$.

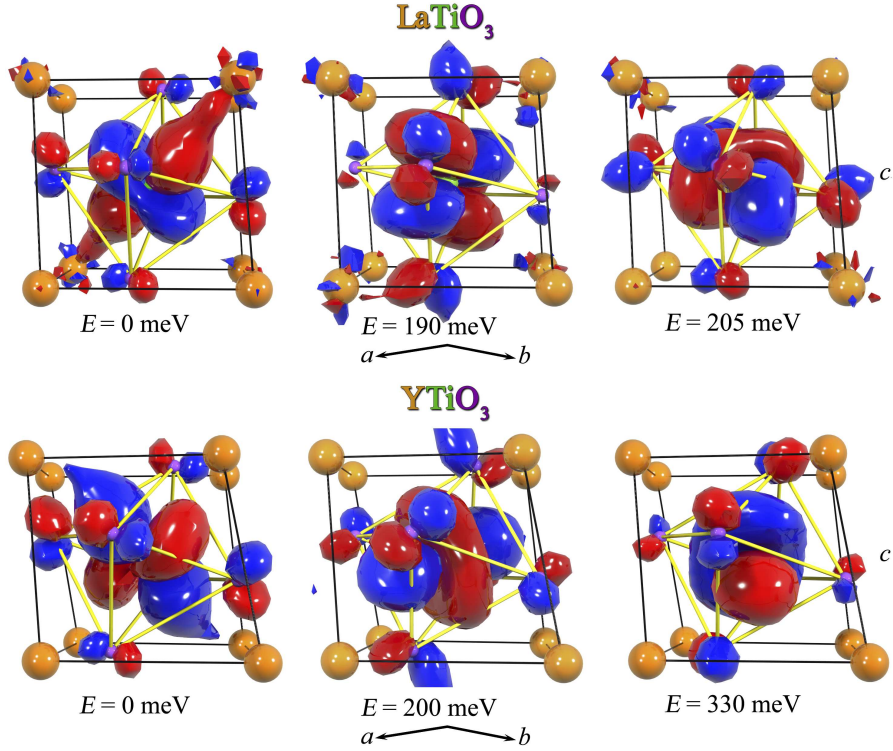


Fig. 5: Crystal-field t_{2g} Wannier orbitals and corresponding crystal-field energies in the $3d^1$ perovskites LaTiO_3 and YTiO_3 . La/Y: orange. Ti: red. O: violet. From Ref. [12].

where $t_\varepsilon = V_{pd\sigma}^2/(\varepsilon - \varepsilon_p)$, $\varepsilon'_d = \varepsilon_d + 3t_\varepsilon$, and $|\mathbf{k} \ 3z^2 - r^2\rangle_\varepsilon$, $|\mathbf{k} \ x^2 - y^2\rangle_\varepsilon$ is the new basis set, which retains however the original ($3z^2 - r^2$ or $x^2 - y^2$) symmetry. The downfolding procedure has *renormalized* the parameters ε_d of the original model (18), but also introduced a new interaction: inter-orbital matrix elements. Furthermore, H_{dd}^ε and the Bloch basis are now energy dependent. Along ΓZ , the eigenvalues of (19) are given implicitly by the equations $\varepsilon = \varepsilon_d + 2t_\varepsilon - 2t_\varepsilon \cos k_z a$ and $\varepsilon = \varepsilon_d$; in second-order perturbation theory $t_\varepsilon \sim t_{\varepsilon_d} = V_{pd\sigma}^2/\Delta_{pd}$, $\varepsilon \sim \varepsilon_d + 2t_{\varepsilon_d} - 2t_{\varepsilon_d} \cos k_z a$, $\varepsilon = \varepsilon_d$. From Hamiltonian (19) it is easy to see that the e_g bands are 2-fold degenerate along ΓR , to find the dispersion along ΓM and $R M$, and obtain the e_g -like bands in Fig. 4. From the Bloch states $|\mathbf{k} \ 3z^2 - r^2\rangle_\varepsilon$ and $|\mathbf{k} \ x^2 - y^2\rangle_\varepsilon$, we may build new Wannier functions. They have $3z^2 - r^2$ or $x^2 - y^2$ symmetry as the original ones but also span, to arbitrary accuracy, the e_g bands. These new Wannier functions are longer range than the original atomic orbitals, since they have p tails on the downfolded neighboring F sites.

The same downfolding procedure can be performed *ab-initio*, e.g., using DFT approaches based on atomic-like orbitals, such as the NMTO method. It leads to *ab-initio* Wannier functions, which carry the information on the lattice structure and the chemistry [12]. The higher energy crystal-field orbital of KCuF_3 (experimental structure), calculated in this way, is shown in Fig. 2. Another example of Wannier functions is shown in Fig. 5. LaTiO_3 and YTiO_3 have a perovskite structure as KCuF_3 , but the O octahedra are tilted and rotated and the La/Y cages are strongly distorted. To obtain the t_{2g} crystal-field Wannier orbitals in Fig. 5 we downfold all other states but the t_{2g} . Although the Ti atoms form an almost cubic lattice, the t_{2g} Wannier functions reflect

the actual distorted structure, as do the corresponding hopping integrals; thus the t_{2g} orbitals are not degenerate, as they would be in a cubic lattice.

Once we have obtained the basis and the hopping integrals, we construct $H = \hat{H}^{\text{LDA}} + \hat{U}^l - \hat{H}_{\text{DC}}^l$. For the e_g doublet we can express \hat{U}^l in a simpler form, because only two-index Coulomb terms are present; furthermore, $U_{m,m'} = U - 2J(1 - \delta_{m,m'})$ and, for $m \neq m'$, $J_{m,m'} = J$, where $U = U_0$ and $J = J_2$. Thus the e_g Hubbard model has the form

$$\begin{aligned}
H = & - \sum_{m,m',i,i',\sigma} t_{mm'}^{i,i'} c_{im\sigma}^\dagger c_{im'\sigma} + U \sum_{i,m} \hat{n}_{im\uparrow} \hat{n}_{im\downarrow} \\
& + \frac{1}{2} \sum_{\substack{i\sigma\sigma' \\ m \neq m'}} (U - 2J - J\delta_{\sigma,\sigma'}) \hat{n}_{im\sigma} \hat{n}_{im'\sigma'} \\
& - J \sum_{i,m \neq m'} \left[c_{im\uparrow}^\dagger c_{im\downarrow}^\dagger c_{im'\uparrow} c_{im'\downarrow} + c_{im\uparrow}^\dagger c_{im\downarrow} c_{im'\downarrow}^\dagger c_{im'\uparrow} \right] - \hat{H}_{\text{DC}}^{e_g}, \quad (20)
\end{aligned}$$

where $m, m' = 3z^2 - r^2, x^2 - y^2$. The last two terms describe the pair-hopping ($U_{mmmm'm'} = J_{m,m'}$ for real harmonics, while for spherical harmonics $U_{mmmm'm'} = 0$) and spin-flip processes. The Hamiltonian has the same form (however with $J = J_1$) for systems with partially filled t_{2g} bands, e.g., LaTiO₃ and YTiO₃.

Massive downfolding to few-band models (for KCuF₃ the e_g bands) necessarily leads to longer range Wannier functions. On the other hand, working with the full Hamiltonian has the drawback that the large number of parameters makes it difficult to gain insight in the problem, losing the advantage of model physics. Another advantage of massive downfolding to a given set of correlated states is that the double-counting correction becomes an energy shift. We can incorporate it into the chemical potential, which is obtained self-consistently from the number of correlated electrons, getting rid of an essentially unknown term.

4 Methods of solution

4.1 LDA+U

The first systematic attempt to construct and solve *ab-initio* many-body Hamiltonians was the LDA+ U method [15]. In this approach the Coulomb interaction is treated in static mean-field theory (in the spirit of Hartree-Fock), and therefore true many-body effects are lost. However, the problems that we have to face in constructing the Hamiltonian (14) are the same, independently of how the Coulomb interaction is then treated. To we explain how LDA+ U works, we assume that Hamiltonian (14) has the simplified form

$$\hat{H}^{\text{LDA}} + \hat{U}^l - \hat{H}_{\text{DC}}^l = \hat{H}^{\text{LDA}} + \frac{1}{2}U \sum_i \sum_{m\sigma \neq m'\sigma'} \hat{n}_{im\sigma} \hat{n}_{im'\sigma'} - \frac{1}{2}U \sum_i \sum_{m\sigma \neq m'\sigma'} \langle \hat{n}_{im\sigma} \rangle \langle \hat{n}_{im'\sigma'} \rangle.$$

We treat the Coulomb interaction in static mean-field at the Hartree level,

$$\hat{n}_{im\sigma} \hat{n}_{im'\sigma'} \rightarrow \langle \hat{n}_{im\sigma} \rangle \hat{n}_{im'\sigma'} + \hat{n}_{im\sigma} \langle \hat{n}_{im'\sigma'} \rangle - \langle \hat{n}_{im\sigma} \rangle \langle \hat{n}_{im'\sigma'} \rangle,$$

and approximate the mean-field energy $\frac{1}{2}U \sum_{m\sigma \neq m'\sigma'} \langle \hat{n}_{im\sigma} \rangle \langle \hat{n}_{im'\sigma'} \rangle$ by the Hartree energy $\frac{1}{2}UN^l N^l$, where $N^l = \sum_{m\sigma} \langle \hat{n}_{im\sigma} \rangle$ is the number of heavy electrons per site. The mean-field Hamiltonian is

$$H = \hat{H}^{\text{LDA}} + \sum_{im\sigma} t_m^\sigma \hat{n}_{im\sigma}, \quad \text{with} \quad t_m^\sigma = U \left(\frac{1}{2} - \langle \hat{n}_{im\sigma} \rangle \right).$$

The levels of the correlated electrons are shifted by $-U/2$ if occupied and by $U/2$ if empty, like in the atomic limit of the half-filled Hubbard model. A total energy functional which shifts the orbital energies in this way is

$$E_{\text{LDA}+U}[n] = E_{\text{LDA}}[n] + \sum_i \left[\frac{1}{2}U \sum_{m\sigma \neq m'\sigma'} \langle \hat{n}_{im\sigma} \rangle \langle \hat{n}_{im'\sigma'} \rangle - E_{\text{DC}} \right],$$

where $E_{\text{DC}} = \frac{1}{2}UN^l(N^l - 1)$ and $E_{\text{LDA}}[n]$ is the total energy obtained using the spin-polarized version of the local-density approximation (LSDA) for the exchange-correlation functional. Indeed

$$\varepsilon_{im\sigma}^{\text{LDA}+U} = \frac{\partial E_{\text{LDA}+U}}{\partial \langle \hat{n}_{im\sigma} \rangle} = \varepsilon_{im\sigma}^{\text{LDA}} + U \left(\frac{1}{2} - \langle \hat{n}_{im\sigma} \rangle \right).$$

More generally, if \hat{U}^l has the form (15), the LDA+ U functional is given by

$$\begin{aligned} E_{\text{LDA}+U}[n] &= E_{\text{LDA}}[n] + \frac{1}{2} \sum_{i\sigma} \sum_{mm'm''m'''} U_{mm'm''m'''} \langle \hat{n}_{imm'm'}^\sigma \rangle \langle \hat{n}_{im''m'''}^{-\sigma} \rangle \\ &+ \frac{1}{2} \sum_{i\sigma} \sum_{mm'm''m'''} [U_{mm'm''m'''} - U_{mm''m''m'}] \langle \hat{n}_{imm'm'}^\sigma \rangle \langle \hat{n}_{im''m'''}^\sigma \rangle - E_{\text{DC}}, \end{aligned}$$

where $\langle \hat{n}_{imm'm'}^\sigma \rangle = \langle c_{im\sigma}^\dagger c_{im'\sigma} \rangle$ is the density matrix, and $\langle \hat{n}_{im\sigma} \rangle = \langle \hat{n}_{imm}^\sigma \rangle$. One of the most common recipe for the double-counting correction is

$$E_{\text{DC}} = \frac{1}{2}U_{\text{avg}}N^l(N^l - 1) - \frac{1}{2}J_{\text{avg}} \sum_{\sigma} N_{\sigma}^l(N_{\sigma}^l - 1). \quad (21)$$

The corresponding one-electron LDA+ U Hamiltonian is

$$\hat{H} = \hat{H}^{\text{LDA}} + \sum_{imm'\sigma} t_{mm'}^\sigma c_{im\sigma}^\dagger c_{im'\sigma}, \quad (22)$$

where²

$$\begin{aligned} t_{mm'}^\sigma &= \sum_{i\sigma} \sum_{m''m'''} U_{mm''m''m'} \langle \hat{n}_{im''m'''}^{-\sigma} \rangle + [U_{mm''m''m'} - U_{mm'm''m'}] \langle \hat{n}_{im''m'''}^\sigma \rangle \\ &- \left[U_{\text{avg}} \left(N^l - \frac{1}{2} \right) - J_{\text{avg}} \left(N_{\sigma}^l - \frac{1}{2} \right) \right] \delta_{m,m'}. \end{aligned}$$

In LDA+ U , differently than in static mean-field for a given Hamiltonian, \hat{H}^{LDA} is obtained self-consistently. The LDA+ U correction in (22) modifies the occupations of the correlated sector

²The LDA+ U correction to the parameters of the LDA Hamiltonian may be obtained from the derivative of $E_{\text{LSDA}+U}[n]$ with respect to the density matrix.

with respect to LDA. If we assume that LDA describes uncorrelated electrons sufficiently well, the readjustments in the uncorrelated sector can be calculated by making the total charge density and the reference potential consistent within the LDA (*charge self-consistency*), however with the constraint given by (22). LDA+ U calculations are usually not performed in a Wannier basis. They are typically based on the identification of an *atomic sphere*, a region of space in which correlated electrons are well described by atomic-like orbitals; the LDA+ U correction is determined through projections onto such atomic orbitals. Thus LDA+ U results are essentially dependent on the choice of the set of correlated electrons and this atomic sphere. However, if the correlated electrons are well localized, they retain indeed to a good extent their atomic character in a solid. Thus the arbitrariness of the choice is less crucial than could be expected. Calculations based on different electronic structure methods confirm this.

LDA+ U describes successfully the magnetic ground-state of Mott insulators. This could appear surprising, because the LDA+ U Hamiltonian (22) describes a non-interacting system, and therefore should in principle have the same defects of the LDA Hamiltonian. However, LDA+ U opens a gap by making long-range order (magnetic, magnetic and orbital, . . .). Indeed, in some cases we obtain a gap already in LSDA, although often such a gap is much smaller than the experimental one. As in LDA, the LDA+ U eigenvalues are real, and quasi-particles have an infinite lifetime. In the Green function language, LDA+ U yields a self-energy which is orbital, spin, and site dependent within the unit cell, but has no ω dependence and no imaginary part.

We present as an example the case of KCuF_3 . Instead of the full LDA+ U calculation, for simplicity we discuss the results for the e_g -band Hubbard model (20) and do not perform any charge self-consistency. We study the ideal cubic structure; the independent-electron part of the Hamiltonian is given to a good approximation by Eq. (19). We consider a unit cell with axes $\mathbf{a} = (-\mathbf{x} + \mathbf{y})$, $\mathbf{b} = (\mathbf{x} + \mathbf{y})$, $\mathbf{c} = 2\mathbf{z}$. This cell contains four octahedra. The four Cu atoms are located at $(0, 0, 0)$, $(a/2, a/2, 0)$, $(0, 0, a/2)$, $(a/2, a/2, a/2)$; we label them as $1_u, 2_u, 1_d, 2_d$. The difference between sites of type 1 and 2 in the experimental structure with the co-operative Jahn-Teller distortion is illustrated in Fig. 2; in the cubic structure they are identical. The metallic paramagnetic LDA band structure for this supercell is shown in Fig. 6. The insulating static mean-field bands are shown on the left. In the basis $|\alpha\sigma\rangle_i, |\beta\sigma\rangle_i$, with $\alpha = 3z^2 - r^2$, and $\beta = x^2 - y^2$, the LDA+ U self-energy for a site i has the following form

$$\Sigma^{i\sigma} = \begin{bmatrix} \Sigma_{\alpha,\alpha}^{i\sigma} & \Sigma_{\alpha,\beta}^{i\sigma} \\ \Sigma_{\beta,\alpha}^{i\sigma} & \Sigma_{\beta,\beta}^{i\sigma} \end{bmatrix}.$$

Setting for simplicity $J = 0$ we find

$$\Sigma^{i\sigma} = U \begin{bmatrix} \frac{1}{2} - \langle \hat{n}_{i\alpha\alpha}^\sigma \rangle & -\langle \hat{n}_{i\beta\alpha}^\sigma \rangle \\ -\langle \hat{n}_{i\alpha\beta}^\sigma \rangle & \frac{1}{2} - \langle \hat{n}_{i\beta\beta}^\sigma \rangle \end{bmatrix},$$

where the density matrix has to be determined self-consistently. The static mean-field bands in the left panel of Fig. 6 correspond approximatively to a state with one hole in $|z^2 - y^2 \uparrow\rangle = \frac{\sqrt{3}}{2}|\alpha \uparrow\rangle + \frac{1}{2}|\beta \uparrow\rangle$ at site 1_u , in $|z^2 - y^2 \downarrow\rangle$ at site 1_d , in $|z^2 - x^2 \uparrow\rangle$ at site 2_u , and in $|z^2 - x^2 \downarrow\rangle$ at site 2_d . Correspondingly, the self-energy matrix may be written as

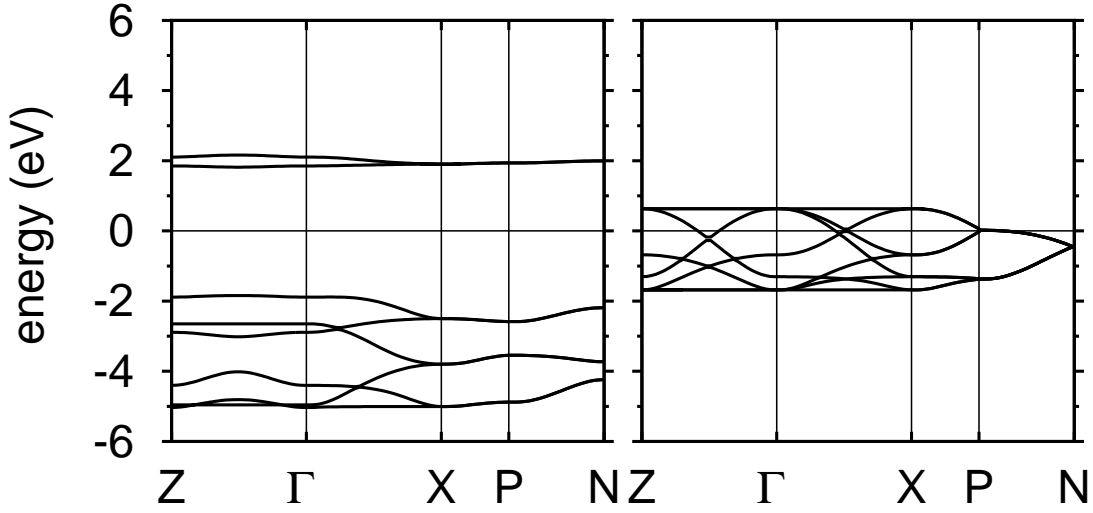


Fig. 6: Right: LDA e_g band structure of cubic $KCuF_3$ calculated using the experimental magnetic unit cell with four formula units. Left: Static mean-field band structure, calculated for the experimental orbital and spin order. Parameters: $U = 7$ eV and $J = 0.9$ eV.

Σ/U	$ \alpha\sigma\rangle_{1u}$	$ \beta\sigma\rangle_{1u}$	$ \alpha\sigma\rangle_{2u}$	$ \beta\sigma\rangle_{2u}$	$ \alpha\sigma\rangle_{1d}$	$ \beta\sigma\rangle_{1d}$	$ \alpha\sigma\rangle_{2d}$	$ \beta\sigma\rangle_{2d}$
$ \alpha\sigma\rangle_{1u}$	$\frac{-2\delta_{\sigma,\downarrow} - \delta_{\sigma,\uparrow}}{4}$	$\frac{\sqrt{3}\delta_{\sigma,\uparrow}}{4}$	0	0	0	0	0	0
$ \beta\sigma\rangle_{1u}$	$\frac{\sqrt{3}\delta_{\sigma,\uparrow}}{4}$	$\frac{-2\delta_{\sigma,\downarrow} + \delta_{\sigma,\uparrow}}{4}$	0	0	0	0	0	0
$ \alpha\sigma\rangle_{2u}$	0	0	$\frac{-2\delta_{\sigma,\downarrow} - \delta_{\sigma,\uparrow}}{4}$	$\frac{-\sqrt{3}\delta_{\sigma,\uparrow}}{4}$	0	0	0	0
$ \beta\sigma\rangle_{2u}$	0	0	$\frac{-\sqrt{3}\delta_{\sigma,\uparrow}}{4}$	$\frac{-2\delta_{\sigma,\downarrow} + \delta_{\sigma,\uparrow}}{4}$	0	0	0	0
$ \alpha\sigma\rangle_{1d}$	0	0	0	0	$\frac{-2\delta_{\sigma,\uparrow} - \delta_{\sigma,\downarrow}}{4}$	$\frac{\sqrt{3}\delta_{\sigma,\downarrow}}{4}$	0	0
$ \beta\sigma\rangle_{1d}$	0	0	0	0	$\frac{\sqrt{3}\delta_{\sigma,\downarrow}}{4}$	$\frac{-2\delta_{\sigma,\uparrow} + \delta_{\sigma,\downarrow}}{4}$	0	0
$ \alpha\sigma\rangle_{2d}$	0	0	0	0	0	0	$\frac{-2\delta_{\sigma,\uparrow} - \delta_{\sigma,\downarrow}}{4}$	$\frac{-\sqrt{3}\delta_{\sigma,\downarrow}}{4}$
$ \beta\sigma\rangle_{2d}$	0	0	0	0	0	0	$\frac{-\sqrt{3}\delta_{\sigma,\downarrow}}{4}$	$\frac{-2\delta_{\sigma,\uparrow} + \delta_{\sigma,\downarrow}}{4}$

The spatial structure of the self-energy, which yields long-range spin and orbital order, is what opens the gap, by resolving spin and orbital degeneracy; the insulating state in LDA+ U has therefore a very different nature than in DMFT.

To understand the nature of the difference, we consider the half-filled one-band Hubbard model (7) for a linear chain ($\varepsilon_k = -2 \cos k_x a$) with lattice constant a . We double the unit cell (lattice constant $b = 2a$) and solve the model with Hartree-Fock, looking for long-range magnetic order. The mean-field Hamiltonian is

$$H^{\text{HF}} = -t \sum_{\langle ii'\rangle\sigma} c_{i\sigma}^\dagger c_{i'\sigma} + U \sum_{i\sigma} \langle \hat{n}_{i\sigma} \rangle \hat{n}_{i-\sigma} - U \sum_i \langle \hat{n}_{i\uparrow} \rangle \langle \hat{n}_{i\downarrow} \rangle,$$

where $\langle \hat{n}_{i\uparrow} \rangle = 1/2 + m(-1)^i$, $\langle \hat{n}_{i\downarrow} \rangle = 1/2 - m(-1)^i$, and m is the magnetic moment per site. We set the Fermi level to zero and assume $m = 1/2$; the self-energy is $\Sigma_{i\sigma} = U \langle \hat{n}_{i-\sigma} \rangle$ and the

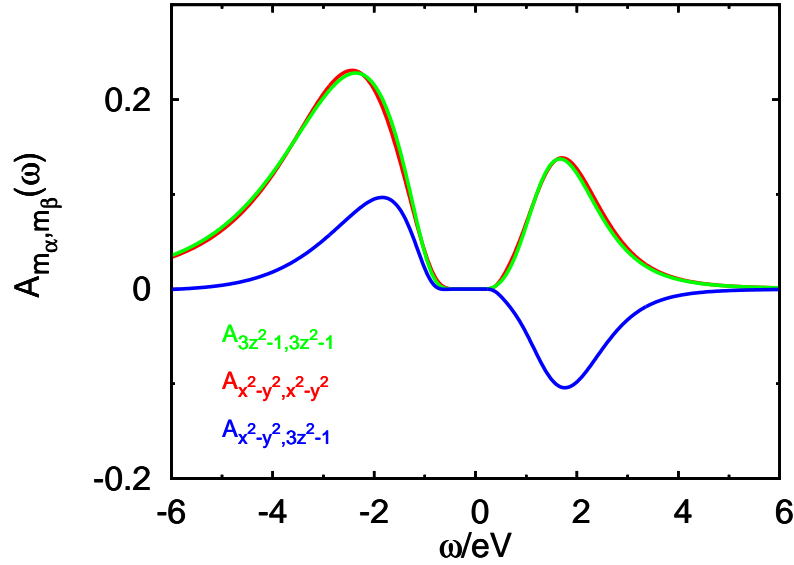


Fig. 7: LDA+DMFT spectral matrix per spin (~ 230 K) in the orbitally-ordered paramagnetic phase of KCuF_3 , cubic structure. Coulomb parameters: $U = 7$ eV, $J = 0.9$ eV. The off-diagonal terms, here large, are zero in the high temperature para-orbital phase.

doubly degenerate eigenvectors are $\varepsilon_{\pm} = \pm\sqrt{\varepsilon_k^2 + U^2/4}$. At half-filling, only ε_- is occupied, and the gap opens at $k = \pi/2a$, the borders of the reduced BZ, $[-\pi/2a, \pi/2a)$. To compare with the DMFT solution of the one band Hubbard model, we unfold the bands back to the BZ of the small unit cell, $[-\pi/a, \pi/a)$. The Hartree-Fock bands can then be written as $\varepsilon_k + \Delta(k)$, where $\Delta(k) = -\varepsilon_k + [\Theta(\frac{\pi}{2a} - k) - \Theta(k - \frac{\pi}{2a})]\sqrt{\varepsilon_k^2 + U^2/4}$, and Θ is the step function. Thus, to obtain the Hartree-Fock bands we would have to correct the band energy ε_k with a function $\Delta(k)$ which has a jump at $k = \pi/2a$, i.e. a strong k -dependence. This should be compared with the divergence $\propto 1/\omega$ found in the real part of the k -independent DMFT self-energy for the paramagnetic half-filled Bethe lattice. Thus, the origin of the gap is completely different in static and dynamical mean-field theory.

LDA+ U has been successfully used to describe the magnetic ground state of many transition-metal compounds; however, since it is based on the static mean-field (Hartree-Fock-like) approximation, it often gives too large band gaps. A related problem is that the tendency to long-range order is overestimated, because fluctuations are neglected.

4.2 LDA+DMFT

The natural extension of LDA+ U to include true correlation effects is LDA+DMFT [18–21]. In this approach we solve the Hubbard model $\hat{H}^{\text{LDA}} + \hat{U}^l - \hat{H}_{\text{DC}}^l$ by means of dynamical mean-field theory. To do this, first we map $\hat{H}^{\text{LDA}} + \hat{U}^l - \hat{H}_{\text{DC}}^l$ onto a multi-orbital Anderson model with the same Coulomb interaction at the impurity site; $\mathcal{G}(\omega)$ is the bath Green-function matrix of such Anderson model. Next, we label with i_c the correlated sites within the unit cell and with $lm\sigma$ the correlated orbitals at sites i_c ; the local lattice Green-function matrix is

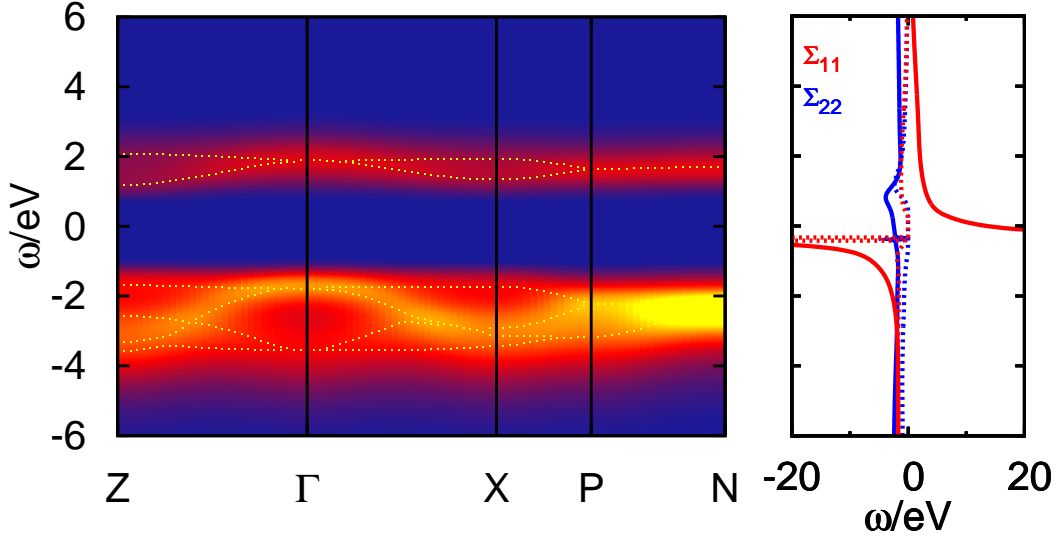


Fig. 8: Left: LDA+DMFT correlated band structure (~ 230 K) in the orbitally-ordered phase of KCuF_3 , cubic structure; the dots are the poles of the Green function. Right: Self-energy matrix in the basis of the natural orbitals. Full lines: real part. Dotted lines: imaginary part. Coulomb parameters: $U = 7$ eV, $J = 0.9$ eV.

$$G_{i_c m \sigma, i'_c m' \sigma}(\omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \left(\left[\omega + \mu I - \hat{H}_{\mathbf{k}}^{\text{LDA}} - \Sigma^l(\omega) + \hat{H}_{\text{DC}}^l \right]^{-1} \right)_{i_c m \sigma, i'_c m' \sigma}.$$

The DMFT self-energy matrix $\Sigma^l(\omega)$ is non-zero in the correlated sector only; furthermore, it is diagonal in the site indices, $\Sigma_{i_c m \sigma, i'_c m' \sigma}^l = \delta_{i_c, i'_c} \Sigma_{m, m'}^{i_c \sigma}$. For two equivalent sites i_c and i'_c , space group symmetries transform $\Sigma_{m, m'}^{i_c \sigma}$ into $\Sigma_{m, m'}^{i'_c \sigma}$. In the paramagnetic case, in which $G_{i_c m \sigma, i'_c m' \sigma} = G_{i_c m - \sigma, i'_c m' - \sigma}$, the additional relation $\Sigma_{m, m'}^{i_c \sigma} = \Sigma_{m, m'}^{i'_c - \sigma} = \Sigma_{m, m'}^{i'_c}$ holds.

Let us consider the example of paramagnetic KCuF_3 . The primitive cell (Fig. 2) contains two formula units (and thus two equivalent Cu sites, labeled as 1 and 2 in Fig. 2). The LDA Hamiltonian which describes the e_g bands is a 4×4 matrix; the transformation between site 1 and 2 is $x, y, z \rightarrow y, x, -z$, and therefore $\Sigma_{\alpha\beta}^1 = -\Sigma_{\alpha\beta}^2$, $\Sigma_{\alpha\alpha}^1 = \Sigma_{\alpha\alpha}^2$, $\Sigma_{\beta\beta}^1 = \Sigma_{\beta\beta}^2$. In matrix form

$$\begin{array}{ccccc} \Sigma^l & |\alpha\sigma\rangle_1 & |\beta\sigma\rangle_1 & |\alpha\sigma\rangle_2 & |\beta\sigma\rangle_2 \\ |\alpha\sigma\rangle_1 & \Sigma_{\alpha\alpha}^1 & \Sigma_{\alpha\beta}^1 & 0 & 0 \\ |\beta\sigma\rangle_1 & \Sigma_{\beta\alpha}^1 & \Sigma_{\beta\beta}^1 & 0 & 0 \\ |\alpha\sigma\rangle_2 & 0 & 0 & \Sigma_{\alpha\alpha}^1 & -\Sigma_{\alpha\beta}^1 \\ |\beta\sigma\rangle_2 & 0 & 0 & -\Sigma_{\beta\alpha}^1 & \Sigma_{\beta\beta}^1 \end{array}.$$

The DMFT self-consistency condition requires that $G_{i_c, i_c}(\omega)$ equals the impurity Green-function matrix of the Anderson model, $G(\omega)$. Thus, in the paramagnetic case

$$G_{m, m'}(\omega) = G_{i_c m \sigma, i_c m' \sigma}(\omega), \quad \mathcal{G}_{m, m'}^{-1}(\omega) = G_{m, m'}^{-1}(\omega) + \Sigma_{m, m'}^{i_c}(\omega).$$

The site i_c is one of the equivalent sites $\{i_c\}$; in the KCuF_3 example, it could be, e.g., a site 1. The solution of the multi-orbital quantum-impurity model requires a method which can deal with realistic Coulomb interactions and Green-function matrices; depending on the problem we want to address, we can choose from quantum Monte Carlo [7, 8], Lanczos [9], the numerical renormalization group, and many more.

The Hirsch-Fye QMC [7] approach, presented in the Lecture of Nils Blümer, is very general; a limitation is that spin-flip and pair-hopping processes yield a strong sign problem, and have to be neglected; in many cases this is, however, a good approximation. In Fig. 7 we show the *paramagnetic* e_g spectral-function matrix of KCuF_3 at ~ 230 K, calculated with a Hirsch-Fye QMC solver. The figure shows that with DMFT we can obtain an insulating orbitally ordered solution even in the absence of long-range magnetic order; this is not the case in static mean-field theory. The off-diagonal elements of the spectral matrix (and correspondingly those of the self-energy matrix) are large and cannot be neglected. In Fig. 8 we show the LDA+DMFT e_g band structure of KCuF_3 , corresponding to the spectral matrix in Fig. 7. We can compare these bands with the static mean-field *antiferromagnetic* band structure in Fig. 6. The LDA+DMFT band gap is significantly smaller. The imaginary part of the self-energy, which is zero in static mean-field theory, makes the Hubbard bands partly incoherent. The real part of the self-energy of the half-filled orbital (Fig. 8), which in static mean-field theory does not depend on ω , diverges at low frequencies, as in the case of the half-filled Bethe lattice.

The continuous-time QMC technique [8], discussed in the Lecture of Philipp Werner, allows us to include in our model spin-flip and pair-hopping terms. As an example we present LDA+DMFT results for Ca_2RuO_4 (Fig. 9), a layered $4d^4$ perovskite which at low energy can be described by a t_{2g} Hubbard model of the form (14); to solve it we use the weak-coupling continuous-time QMC scheme. Fig. 9 shows that neglecting spin-flip and pair-hopping terms changes the degeneracy of multiplets, and leads to an overestimation of the mass renormalization m^*/m [22].

The LDA+DMFT self-consistent procedure works as follows (for QMC solvers $\omega \rightarrow i\omega_n$)

- construct $\hat{H}^{\text{LDA}} + \hat{U}^l - \hat{H}_{\text{DC}}^l$
- calculate the local Green-function matrix for a starting self-energy matrix

$$G_{i_c\sigma, i_c\sigma}(\omega) = \frac{1}{N_{\mathbf{k}}} \sum_{\mathbf{k}} \left([\omega - \hat{H}_{\mathbf{k}}^{\text{LDA}} - \Sigma^l(\omega) + \mu I + \hat{H}_{\text{DC}}^l]^{-1} \right)_{i_c\sigma, i_c\sigma}.$$

For equivalent sites, the LDA Hamiltonian and the self-energy matrix should transform according to space-group symmetries.

- calculate the bath Green-function matrix

$$\mathcal{G}^{-1}(\omega) = G^{-1}(\omega) + \Sigma(\omega), \quad G(\omega) = G_{i_c\sigma, i_c\sigma}(\omega)$$

- obtain the impurity Green-function matrix $G(\omega)$ by solving the quantum-impurity problem defined by $\mathcal{G}(\omega)$ and the Coulomb interaction \hat{U}^l at sites $\{i_c\}$

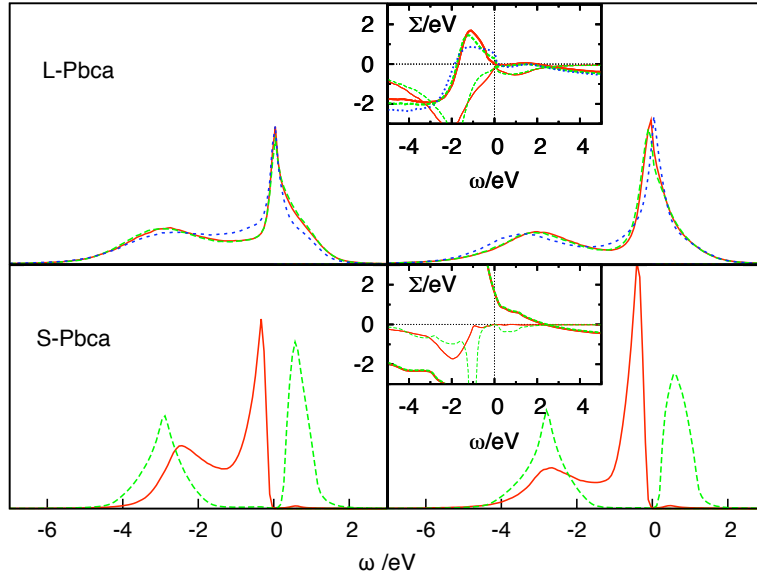


Fig. 9: Ca_2RuO_4 : t_{2g} spectral functions in the high-temperature L - $Pbca$ phase and in the low temperature S - $Pbca$ phase. Left: without spin-flip and pair-hopping terms. Right: with spin-flip and pair-hopping terms. Red: xy . Green and blue: xz and yz . From Ref. [22].

- calculate the self-energy matrix

$$\Sigma(\omega) = \mathcal{G}^{-1}(\omega) - G^{-1}(\omega)$$

- check if self-consistency is reached; if not, recalculate $G_{i_c\sigma, i_c\sigma}(\omega)$ and start over again.

In LDA+DMFT the double-counting correction is the same as in LDA+ U ; in the case of massive downfolding to the correlated-electron sector, as previously discussed, it is incorporated in the chemical potential and does not need to be calculated explicitly.

The extension of LDA+DMFT to the spin-polarized case (spin-dependent self-energy matrix and Green-function matrices) is straightforward. Long-range anti-ferromagnetic order can also be treated, provided that we use the appropriate unit cell. Let us consider the case of $KCuF_3$. Below T_N this system is antiferromagnetic along z and ferromagnetic in the xy plane. To account for this magnetic order, we have to use a unit cell with 4 formula units, the same that we used for the LDA+ U example; the self-energy matrix has therefore the same spatial structure as in LDA+ U

$\Sigma^{l\sigma}$	$ \alpha\sigma\rangle_{1_u}$	$ \beta\sigma\rangle_{1_u}$	$ \alpha\sigma\rangle_{2_u}$	$ \beta\sigma\rangle_{2_u}$	$ \alpha\sigma\rangle_{1_d}$	$ \beta\sigma\rangle_{1_d}$	$ \alpha\sigma\rangle_{2_d}$	$ \beta\sigma\rangle_{2_d}$
$ \alpha\sigma\rangle_{1_u}$	$\Sigma_{\alpha\alpha}^{1\sigma}$	$\Sigma_{\alpha\beta}^{1\sigma}$	0	0	0	0	0	0
$ \beta\sigma\rangle_{1_u}$	$\Sigma_{\beta\alpha}^{1\sigma}$	$\Sigma_{\beta\beta}^{1\sigma}$	0	0	0	0	0	0
$ \alpha\sigma\rangle_{2_u}$	0	0	$\Sigma_{\alpha\alpha}^{1\sigma}$	$-\Sigma_{\alpha\beta}^{1\sigma}$	0	0	0	0
$ \beta\sigma\rangle_{2_u}$	0	0	$-\Sigma_{\beta\alpha}^{1\sigma}$	$\Sigma_{\beta\beta}^{1\sigma}$	0	0	0	0
$ \alpha\sigma\rangle_{1_d}$	0	0	0	0	$\Sigma_{\alpha\alpha}^{1-\sigma}$	$\Sigma_{\alpha\beta}^{1-\sigma}$	0	0
$ \beta\sigma\rangle_{1_d}$	0	0	0	0	$\Sigma_{\beta\alpha}^{1-\sigma}$	$\Sigma_{\beta\beta}^{1-\sigma}$	0	0
$ \alpha\sigma\rangle_{2_d}$	0	0	0	0	0	0	$\Sigma_{\alpha\alpha}^{1-\sigma}$	$-\Sigma_{\alpha\beta}^{1-\sigma}$
$ \beta\sigma\rangle_{2_d}$	0	0	0	0	0	0	$-\Sigma_{\beta\alpha}^{1-\sigma}$	$\Sigma_{\beta\beta}^{1-\sigma}$

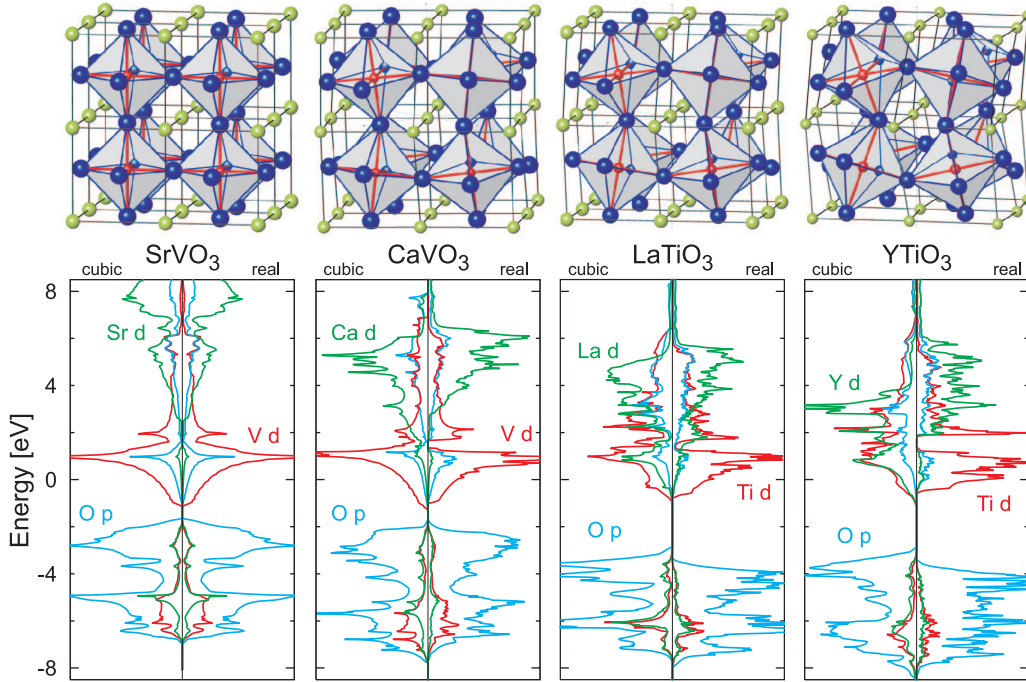


Fig. 10: Evolution of crystal structure and LDA band structure in the series of $3d^1$ perovskites with the $GdFeO_3$ -type distortion. From Ref. [12].

The LDA+DMFT scheme can be easily extended to treat clusters, by using a supercell and treating the supercell as impurity; alternative extensions to account for the \mathbf{k} dependence of the self-energy are the dynamical-cluster approximation (DCA) [23], the dual-fermion approach, or GW+DMFT. Some of these methods will be discussed in the Lectures of Sasha Lichtenstein and Karsten Held. Finally, LDA+DMFT, as LDA+ U , can be also made charge self-consistent. This requires to work, as in LDA+ U , with the full Hamiltonian and to account explicitly for the double-counting correction.

The calculation of the screened Coulomb parameters is a major open problem, in LDA+DMFT as in LDA+ U . In the absence of a definitive method, a useful approach is to analyze trends in similar materials, to single out the effects of chemistry and structural distortions from those of the Coulomb interaction. We adopt this approach in Ref. [24] to study the Mott transition in the series of $3d^1$ ($t_{2g}^1 e_g^0$) perovskites. These materials all have the $GdFeO_3$ -type structure, with distortions (tilting, rotation, deformation of the cation cage) that increase along the series (Fig. 10). By means of massive downfolding based on the NMTO method [12], we obtain the material-specific t_{2g} Wannier basis (crystal-field orbitals) shown in Fig. 5. With increasing distortions, the t_{2g} band-width decreases and the crystal-field splitting increases, reaching ~ 300 meV in $YTiO_3$, still a small fraction of the t_{2g} band width. We find that, despite of its small value, the crystal-field splitting plays a crucial role in helping the metal-insulator transition, by reducing the orbital degeneracy [25] of the many-body states and favoring the formation of an orbitally-ordered state.

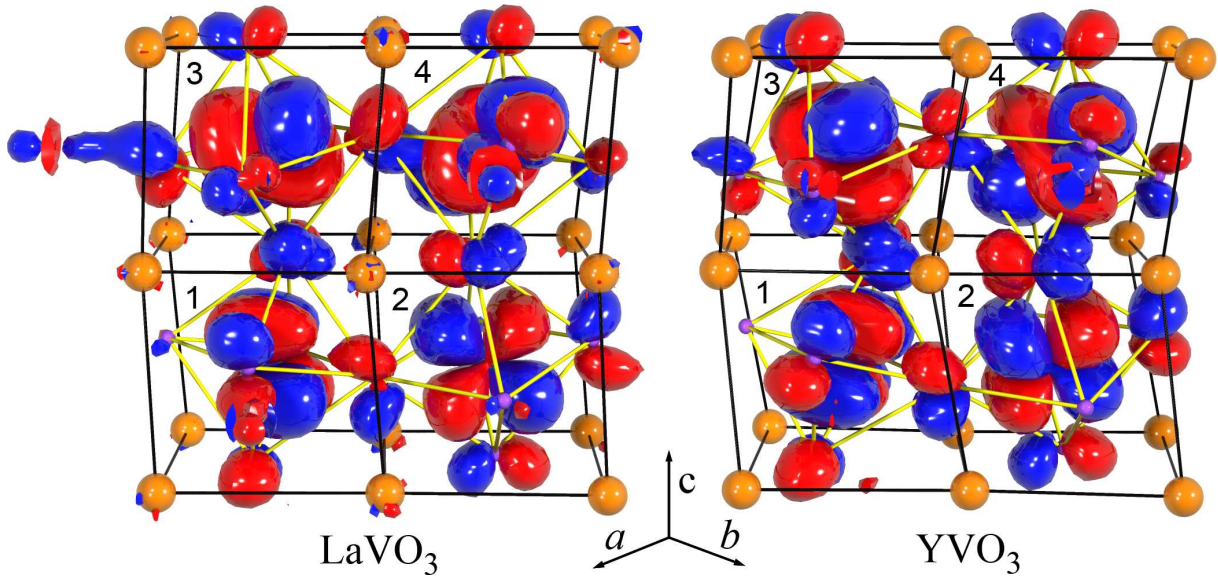


Fig. 11: Orbital-order (empty orbital) in the $3d^2$ perovskites LaVO_3 and YVO_3 . From Ref. [26].

5 The origin of orbital order

In this Section we will show an example of how LDA+DMFT can be used as a tool to understand physical phenomena. Orbital order is believed to play a crucial role in determining the electronic and magnetic properties of many transition-metal compounds. Still, the origin of orbital order in real materials is a subject of hot debate.

The hallmark of orbital order is the co-operative Jahn-Teller distortion. A paradigmatic example is KCuF_3 . The co-operative Jahn-Teller distortion is shown in Fig. 2. This static distortion gives rise to a crystal field, which splits the otherwise degenerate e_g doublet. LDA+DMFT calculations have proven that, due to Coulomb repulsion, even a crystal-field splitting much smaller than the band width can lead to orbital order. The importance of such effect for real materials has been realized first for LaTiO_3 and YTiO_3 [24, 12]. The same effect is at work in a number of other systems with different electronic structure. We discuss few cases. In $3d^2$ vanadates [26] the t_{2g} crystal-field splitting is even smaller than in $3d^1$ perovskites. Still, orbital fluctuations are already strongly suppressed at room temperature, yielding the orbital-order shown in Fig. 11. In Ca_2RuO_4 , due to the layered perovskite structure, the $2/3$ -filled t_{2g} bands split into a wide xy and two narrow xz and yz bands. In the low-temperature phase (S-Pbca) the system is an insulator with a small gap and exhibits xy -orbital order: at each site the xy orbital is filled with two electrons. Above 350 K, in the L-Pbca phase, Ca_2RuO_4 is metallic and no orbital order has been reported. We find (Fig. 9) that the metal-insulator transition is driven by the structural L-Pbca \rightarrow S-Pbca phase transition; furthermore, in the insulating phase the ~ 300 meV crystal-field splitting overcomes the band mismatch and we find xy orbital order [22]. The case of $3d^9$ KCuF_3 and $3d^4$ LaMnO_3 is even more extreme: the e_g crystal-field splitting is $\sim 0.5 - 1$ eV at 300 K; with such a large splitting, orbital fluctuations are suppressed up to melting temperature.

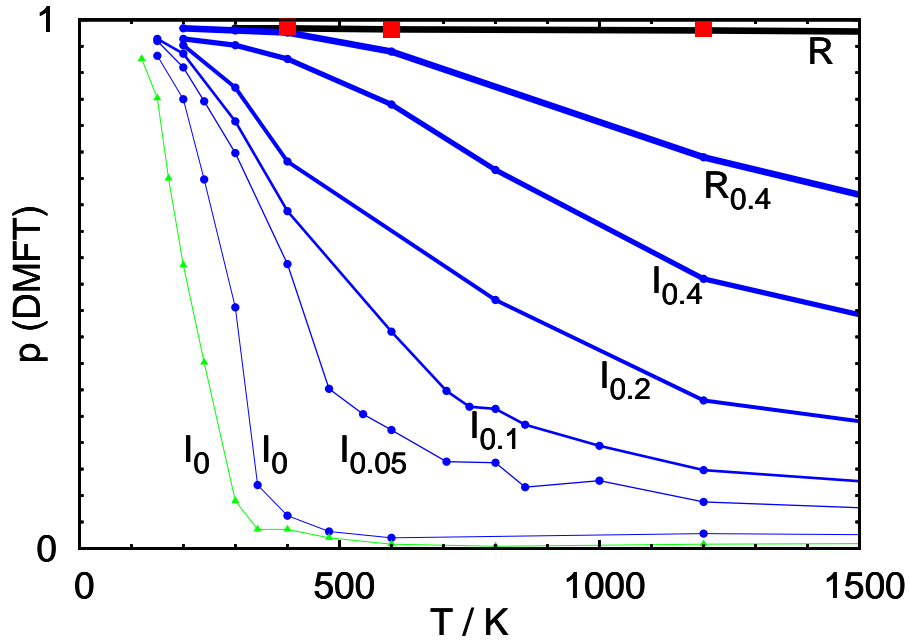


Fig. 12: Orbital order transition in KCuF_3 . Orbital polarization p as a function of temperature calculated in LDA+DMFT. R : $U = 7$ eV, experimental structure. Circles: $U = 7$ eV, idealized structures R_δ and I_δ with decreasing crystal-field. Triangles: $U = 9$ eV, I_0 only. Squares: two-sites CDMFT. From Ref. [4].

Orbital order can already arise, however, in the absence of static distortions. In a seminal work, Kugel and Khomskii [27] showed that in strongly-correlated systems with orbital degrees of freedom (degenerate e_g or t_{2g} levels), many-body effects can give rise to orbital order via a purely electronic mechanism (*spin and orbital super-exchange*). In this picture, the co-operative Jahn-Teller distortion (and thus the crystal-field splitting) is a consequence of orbital order. In the opposite scenario, the co-operative Jahn-Teller distortion is due to the electron-phonon coupling, which removes orbital degeneracy, and orbital order is driven by the static distortion, as discussed above. We analyze these two scenarios for KCuF_3 and LaMnO_3 , the text-book examples [28] of orbitally-ordered systems. LDA+ U total energy calculations show [15, 29] that in these systems the co-operative Jahn-Teller distortion is stabilized by U , a result recently confirmed in LDA+DMFT [30]. This could indicate that super-exchange is the driving mechanism. However, if this is the case it is hard to explain why the magnetic transition temperature, determined by super-exchange, is much lower than the orbital-order transition temperature: $T_N \sim 40$ K for KCuF_3 and $T_N \sim 140$ K for LaMnO_3 , while the co-operative Jahn-Teller distortion persists up to 1000 K or more.

LaMnO_3 and KCuF_3 can both be described by a two-band e_g Hubbard model. In the case of LaMnO_3 we have additionally to take into account the Hund's rule coupling between e_g electrons and t_{2g} spins, $S_{t_{2g}}$. Thus the minimal model to understand orbital order in these two

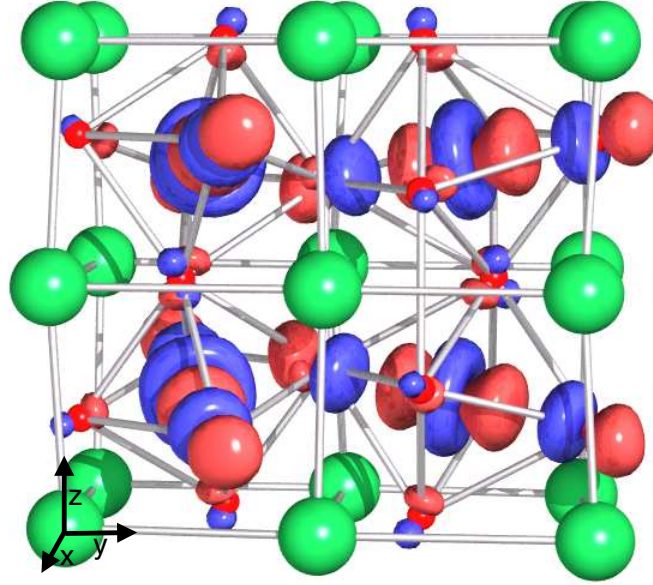


Fig. 13: Orbital order (LDA+DMFT calculations) in the rare-earth perovskite $TbMnO_3$ with the $GdFeO_3$ -type structure. From Ref. [32]. This system has the same structure of $LaMnO_3$.

systems is the Hamiltonian [31]

$$H = - \sum_{im\sigma, i'm'\sigma'} t_{m,m'}^{i,i'} u_{\sigma,\sigma'}^{i,i'} c_{im\sigma}^\dagger c_{i'm'\sigma'} - h \sum_{im} (\hat{n}_{im\uparrow} - \hat{n}_{im\downarrow}) + U \sum_{im} \hat{n}_{im\uparrow} \hat{n}_{im\downarrow} + \frac{1}{2} \sum_{im(\neq m')\sigma\sigma'} (U - 2J - J\delta_{\sigma,\sigma'}) \hat{n}_{im\sigma} \hat{n}_{im'\sigma'}.$$

Here $m, m' = 3z^2 - r^2, x^2 - y^2$. The local magnetic field $h = JS_{t_{2g}}$ describes the Hund's rule coupling to t_{2g} electrons, and $u_{i\sigma, i'\sigma'} = 2/3(1 - \delta_{i,i'})$ accounts for the disorder in orientation of the t_{2g} spins. In the case of $KCuF_3$ $u_{i\sigma, i'\sigma'} = \delta_{\sigma,\sigma'}$ and $h = 0$. For the Coulomb parameters we use the theoretical estimates $J = 0.9$ eV ($KCuF_3$) and $J = 0.75$ eV ($LaMnO_3$) and vary U around 5 eV ($LaMnO_3$) and 7 eV ($KCuF_3$). In the high-spin regime, our results are not very sensitive to h ; we show results for $h \sim 1.3$ eV. We use the massive downfolding technique based on the NMTO method to calculate hopping integrals and crystal-field splittings.

To single out the effects of many-body super-exchange (Kugel-Khomskii mechanism) from the effect of the crystal-field splitting, we perform LDA+DMFT and LDA+CDMFT calculations for a series of hypothetical structures, in which the distortions (and thus the crystal-field splitting) are progressively reduced. In the case of $KCuF_3$, these hypothetical structures are shown in Fig. 2, and the corresponding e_g bands are shown in Fig. 8. For each structure we calculate the order parameter, the orbital polarization p , defined as the difference in the occupations of natural orbitals. In Fig. 12 we show p as a function of temperature. For the experimental structure, $p(T) \sim 1$ till melting temperature; this means that, if the structure stays the same, the system remains orbitally ordered till the crystal melts. The empty orbitals on different sites make the pattern shown in Fig. 2. For the ideal cubic structure I_0 , we find that $p(T) = 0$ at high temperature, but a transition occurs at $T_{KK} \sim 350$ K. This T_{KK} is the critical temperature

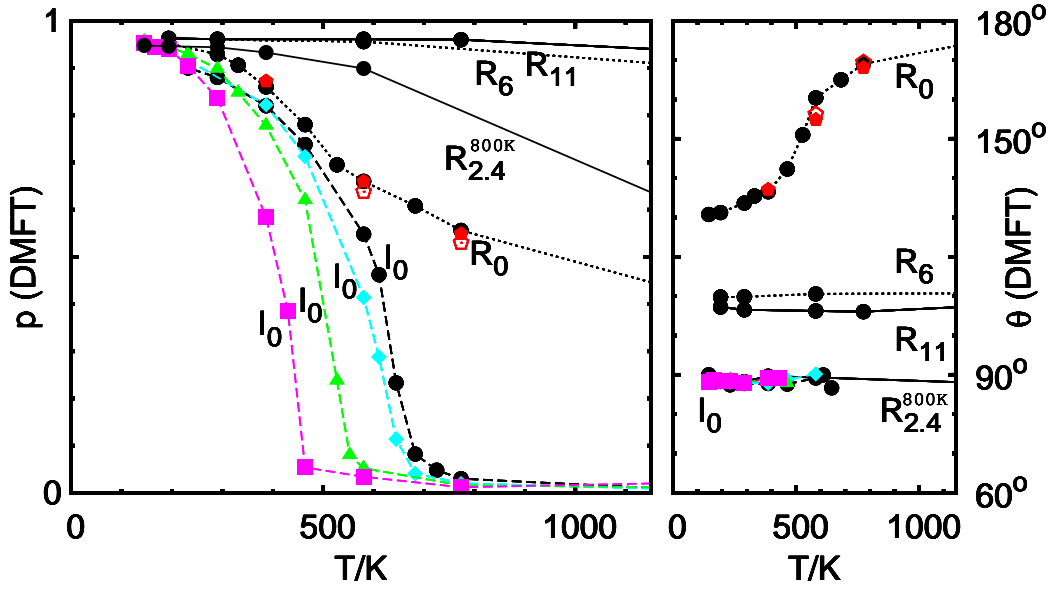


Fig. 14: Orbital order transition in LaMnO_3 . Orbital polarization p (left) and (right) occupied state $|\theta\rangle = \cos \frac{\theta}{2} |3z^2 - r^2\rangle + \sin \frac{\theta}{2} |x^2 - y^2\rangle$ as a function of temperature. Solid line: 300 K experimental structure (R_{11}) and 800 K experimental structure. Dots: orthorhombic structures with half (R_6) or no (R_0) Jahn-Teller distortion. Pentagons: 2 (full) and 4 (empty) sites CDMFT. Dashes: ideal cubic structure (I_0). Circles: $U = 5$ eV. Diamonds: $U = 5.5$ eV. Triangles: $U = 6$ eV. Squares: $U = 7$ eV. Crystal field splittings (meV): 840 (R_{11}), 495 (R_6), 168 ($R_{2.4}^{800\text{K}}$), and 0 (I_0). From Ref. [33].

in the absence of Jahn-Teller effect. Our result shows that around 350 K super-exchange could drive alone the co-operative Jahn-Teller distortion. However, experimentally, the co-operative Jahn-Teller distortion persists up to 800 K or even higher temperature. T_{KK} , although large, is not large enough to explain the presence of a co-operative Jahn-Teller distortion above 350 K; electron-phonon interaction plays a key role. Fig. 12 shows that a ~ 200 meV crystal-field (as in the ideal $R_{0,4}$, which has a Jahn-Teller distortion $\sim 10\%$ of the experimental structure) yields already an almost complete suppression of orbital fluctuations up to at least 1500 K.

In the case of LaMnO_3 we find (see Fig. 14) $T_{\text{KK}} \sim 700$ K. However, besides the co-operative Jahn-Teller distortion and tetragonal compression, LaMnO_3 exhibits a GdFeO_3 -type distortion (Fig. 13), which tends to reduce the e_g band-width [12]. Thus we study in addition an ideal structure R_0 with all distortions except the Jahn-Teller. For such system we cannot obtain T_{KK} from $p(T)$, because, due to the crystal-field splitting ~ 200 meV, Coulomb repulsion strongly suppress orbital fluctuations even at 1500 K. Instead, we study the evolution of the occupied orbital $|\theta\rangle = \cos \frac{\theta}{2} |3z^2 - r^2\rangle + \sin \frac{\theta}{2} |x^2 - y^2\rangle$ with temperature. For the experimental structure (R_{11}) we find $\theta \sim 108^\circ$, in agreement with experiments, and for the I_0 structure we obtain $\theta = 90^\circ$. For the R_0 structure we find two regimes. At high temperature the occupied orbital is the lower energy crystal-field orbital ($\theta = 180^\circ$). At $T_{\text{KK}} \sim 550$ K super-exchange rotates this θ towards 90° , reaching 130° in the zero temperature limit. Such T_{KK} is still very large, but again not sufficient to explain that the Jahn-Teller distortion persist in nanoclusters up to

melting temperature [34]. Thus, as for KCuF_3 , electron-lattice coupling is essential to explain the co-operative Jahn-Teller distortion at high temperature.

Although the co-operative Jahn-Teller distortion persists in domains up to melting temperature, a order-to-disorder (or *orbital melting*) transition has been reported at $T_{\text{OO}} \sim 750$ K [35]. Since $T_{\text{KK}} \sim T_{\text{OO}}$, super-exchange could play a crucial role in such transition. To resolve this issue, we analyze a series of materials for which T_{OO} has been measured: rare-earth manganites with structures similar to LaMnO_3 . For this series, it has been reported that T_{OO} strongly increases with decreasing the rare-earth radius, reaching about 1500 K in TbMnO_3 . Instead, with LDA+DMFT and actually find the opposite trend: T_{KK} is maximum in LaMnO_3 and slightly decreases along the series. Taking the tetragonal crystal-field into account reduces T_{KK} , further increasing the discrepancies with experiment. This proves that, surprisingly, super-exchange effects, although very efficient, in the light of the experimentally reported trends, play a minor role for the orbital order melting observed in rare earth manganites [32].

6 Conclusions

The many-body problem is central to theoretical solid-state physics. Density functional theory is the standard approach for describing the electronic properties of materials. It is a very successful method, which allows us to understand and predict the properties of many systems. However, DFT practice fails completely for *strongly-correlated* materials, in which the movement of one electron depends on the actual, not only the mean position of all other electrons, since it has to avoid their Coulomb repulsion. In this Lecture we have seen a successful scheme to deal with strongly-correlated systems: LDA+DMFT. It is based on the separation of electrons into correlated and uncorrelated. While for uncorrelated electrons we use standard methods based on density-functional theory, for correlated electrons we build material-specific many-body models and solve them with DMFT. Building models requires the construction of localized and material-specific basis sets. To this end, various successful approaches have been devised, such as the *ab-initio* downfolding technique, maximally-localized Wannier functions and projectors. Material-specific many-body models are however complex. Solving them with DMFT requires flexible and efficient quantum-impurity solvers. Examples are Hirsch-Fye QMC, continuous-time QMC, and Lanczos. Short and/or long range spatial correlations can be in principle accounted for within different methods: CDMFT, DCA, dual fermions, GW+DMFT. Thanks to the improvements of impurity solvers and to modern supercomputers, we have been able to include in LDA+DMFT calculations more degrees of freedom, to reach experimental temperatures, to calculate properties beyond the spectral function, and to move towards predictive power. From the many successful applications of LDA+DMFT we have learned that *details do matter*; for example a crystal-field splitting of merely hundred meV, typically neglected in studies based on simple model Hamiltonians, plays a crucial role in stabilizing the orbitally-ordered Mott insulating state [24], and super-exchange, although very efficient, is not the driving mechanism of orbital order in the text-book examples of actual orbitally-ordered materials, LaMnO_3 and KCuF_3 .

Much work lies ahead. A general and effective quantum-impurity solver is not yet available; effective \mathbf{k} -dependent extensions of LDA+DMFT are under development. Central to the success of LDA+DMFT and its extensions is the separation of electrons into light and heavy. This separation, however, is also the main source of trouble. The double-counting correction is essentially unknown. To account for screening effects in a realistic setting is very difficult. Furthermore, by truncating the Coulomb interaction outside the correlated sector, we assume implicitly that the effect of many-body downfolding to the correlated sector is *only* screening; effects which go beyond the mere Coulomb renormalization are usually neglected. Ultimately, these approximations have to be put to a test.

Finally, we have to remember that many-body phenomena are emergent behaviors [36]. Each of such phenomena, although in principle described by the same many-body Hamiltonian (1), the *theory of almost everything* [37], may have a very different nature. To predict new phenomena before they are observed is therefore extremely challenging. After they are discovered, they are often elusive and might remain mysterious for decades, like it is happening for high-temperature superconductivity. In developing *ab-initio* theories for strongly-correlated systems, we have thus always to keep in mind that the crucial aspect to explain a given phenomenon might be hidden in some detail which in more ordinary circumstances would play no role. The challenge is thus to identify *which* details do matter.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

Appendices

A Constants and units

In this Lecture, formulas are written in atomic units. The unit of mass m_0 is the electron mass ($m_0 = m_e$), the unit of charge e_0 is the electron charge ($e_0 = e$), the unit of length a_0 is the Bohr radius ($a_0 = a_B \sim 0.52918 = \text{\AA}$), and the unit of time is $t_0 = 4\pi\epsilon_0\hbar a_0/e^2$. In these units, m_e , a_B , e and $1/4\pi\epsilon_0$ have the numerical value 1; the speed of light is $c = 1/\alpha \sim 137$ in atomic units. The unit of energy is $1Ha = e^2/4\pi\epsilon_0 a_0 \sim 27.211$ eV. These are the natural units for theory. When comparing to experiments, for convenience, we give the energies in eV or meV.

B Atomic orbitals

B.1 Radial functions

The nlm hydrogen-like atomic orbital is given by

$$\psi_{nlm}(\rho, \theta, \phi) = R_{nl}(\rho)Y_l^m(\theta, \phi),$$

where $R_{nl}(\rho)$ is the radial function and $Y_m^l(\theta, \phi)$ a spherical harmonic, $\rho = Zr$ and Z the atomic number. In atomic units, the radial functions are

$$R_{nl}(\rho) = \sqrt{\left(\frac{2Z}{n}\right)^3 \frac{(n-l-1)!}{2n[(n+l)!]^3}} e^{-\rho/n} \left(\frac{2\rho}{n}\right)^l L_{n-l-1}^{2l+1}\left(\frac{2\rho}{n}\right),$$

where L_{n-l-1}^{2l+1} are generalized Laguerre polynomials of degree $n-l-1$.

The radial function for $n = 1, 2, 3$ are

$$\begin{aligned} R_{1s}(\rho) &= 2 Z^{3/2} e^{-\rho} \\ R_{2s}(\rho) &= \frac{1}{2\sqrt{2}} Z^{3/2} (2 - \rho) e^{-\rho/2} \\ R_{2p}(\rho) &= \frac{1}{2\sqrt{6}} Z^{3/2} \rho e^{-\rho/2} \\ R_{3s}(\rho) &= \frac{2}{3\sqrt{3}} Z^{3/2} (1 - 2\rho/3 + 2\rho^2/27) e^{-\rho/3} \\ R_{3p}(\rho) &= \frac{4\sqrt{2}}{9\sqrt{3}} Z^{3/2} \rho(1 - \rho/6) e^{-\rho/3} \\ R_{3d}(\rho) &= \frac{2\sqrt{2}}{81\sqrt{15}} Z^{3/2} \rho^2 e^{-\rho/3} \end{aligned}$$

where we used the standard notation s for $l = 0$, p for $l = 1$ and d for $l = 2$.

B.2 Real harmonics

To study solids, it is usually convenient to work in the basis of real harmonics. The latter are defined in terms of the spherical harmonics as follows:

$$y_{l0} = Y_0^l, \quad y_{lm} = \frac{1}{\sqrt{2}}(Y_{-m}^l + (-1)^m Y_m^l), \quad y_{l-m} = \frac{i}{\sqrt{2}}(Y_{-m}^l - (-1)^m Y_m^l), \quad m > 0.$$

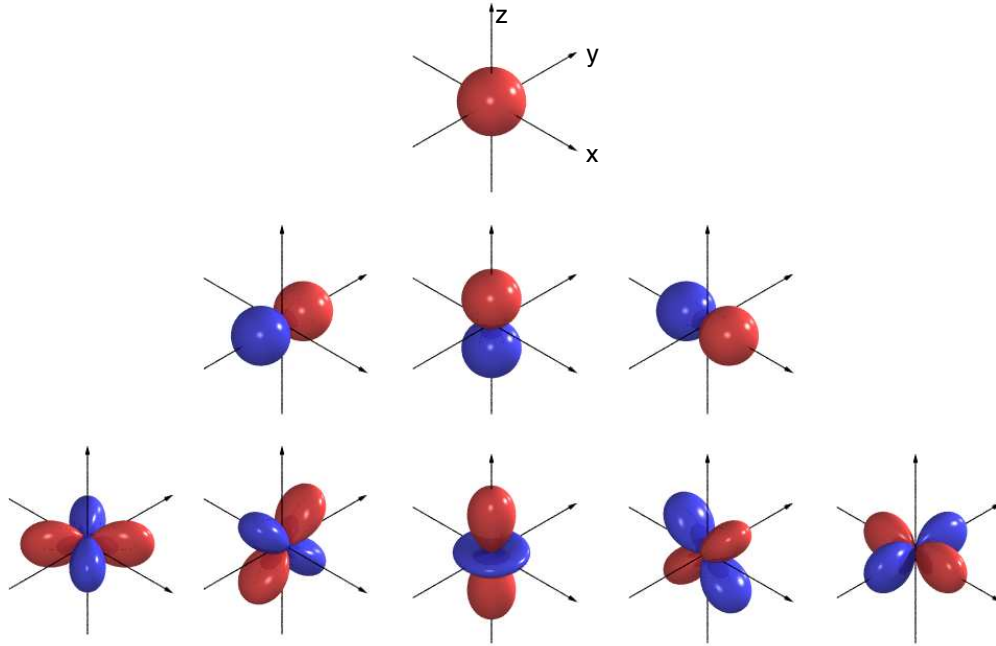


Fig. 15: The s (first row), p_y , p_z , p_x (second row), and d_{xy} , d_{yz} , $d_{3z^2-r^2}$, d_{xz} , $d_{x^2-y^2}$ (last row) real harmonics.

Using the definitions $x = r \sin \theta \cos \phi$, $y = r \sin \theta \sin \phi$, $z = r \cos \theta$, we can express the $l = 0, 1, 2$ real harmonics (Fig. 15) as

$$\begin{aligned}
 s &= y_{00} = Y_0^0 &&= \sqrt{\frac{1}{4\pi}} \\
 p_y &= y_{1-1} = \frac{i}{\sqrt{2}}(Y_1^1 + Y_{-1}^1) &&= \sqrt{\frac{3}{4\pi}} \quad y/r \\
 p_z &= y_{10} = Y_1^0 &&= \sqrt{\frac{3}{4\pi}} \quad z/r \\
 p_x &= y_{11} = \frac{1}{\sqrt{2}}(Y_1^1 - Y_{-1}^1) &&= \sqrt{\frac{3}{4\pi}} \quad x/r \\
 d_{xy} &= y_{2-2} = \frac{i}{\sqrt{2}}(Y_2^2 - Y_{-2}^2) &&= \sqrt{\frac{15}{4\pi}} \quad xy/r^2 \\
 d_{yz} &= y_{2-1} = \frac{i}{\sqrt{2}}(Y_2^2 + Y_{-1}^2) &&= \sqrt{\frac{15}{4\pi}} \quad yz/r^2 \\
 d_{3z^2-r^2} &= y_{20} = Y_2^0 &&= \sqrt{\frac{15}{4\pi}} \frac{1}{2\sqrt{3}} (3z^2 - r^2)/r^2 \\
 d_{xz} &= y_{21} = \frac{1}{\sqrt{2}}(Y_2^2 - Y_{-1}^2) &&= \sqrt{\frac{15}{4\pi}} \quad xz/r^2 \\
 d_{x^2-y^2} &= y_{22} = \frac{1}{\sqrt{2}}(Y_2^2 + Y_{-2}^2) &&= \sqrt{\frac{15}{4\pi}} \frac{1}{2} (x^2 - y^2)/r^2
 \end{aligned}$$

B.3 Slater-Koster integrals

The interatomic Slater-Koster two-center integrals are defined as

$$E_{lm,l'm'} = \int d\mathbf{r} \bar{\psi}_{lm}(\mathbf{r} - \mathbf{d}) V(\mathbf{r} - \mathbf{d}) \psi_{l'm'}(\mathbf{r}).$$

They can be expressed as a function of radial integrals $V_{ll'\alpha}$, which scale with the distance d roughly as $d^{-(l+l'+1)}$ [38], and direction cosines, defined as

$$l = \mathbf{d} \cdot \hat{x}/d, \quad m = \mathbf{d} \cdot \hat{y}/d, \quad n = \mathbf{d} \cdot \hat{z}/d.$$

The Slater-Koster integrals for s , p , and d orbitals [38] are listed below.

$E_{s,s}$	=	$V_{ss\sigma}$		
$E_{s,x}$	=	$lV_{sp\sigma}$		
$E_{x,x}$	=	$l^2V_{pp\sigma}$	$+(1-l^2)V_{pp\pi}$	
$E_{x,y}$	=	$lmV_{pp\sigma}$	$-lmV_{pp\pi}$	
$E_{x,z}$	=	$lnV_{pp\sigma}$	$-lnV_{pp\pi}$	
$E_{s,xy}$	=	$\sqrt{3}lmV_{sd\sigma}$		
E_{s,x^2-y^2}	=	$\frac{1}{2}\sqrt{3}(l^2-m^2)V_{sd\sigma}$		
$E_{s,3z^2-r^2}$	=	$[n^2 - \frac{1}{2}(l^2+m^2)]V_{sd\sigma}$		
$E_{x,xy}$	=	$\sqrt{3}l^2mV_{pd\sigma}$	$+m(1-2l^2)V_{pd\pi}$	
$E_{x,yz}$	=	$\sqrt{3}lmnV_{pd\sigma}$	$-2lmnV_{pd\pi}$	
$E_{x,zx}$	=	$\sqrt{3}l^2nV_{pd\sigma}$	$+n(1-2l^2)V_{pd\pi}$	
E_{x,x^2-y^2}	=	$\frac{\sqrt{3}}{2}l[(l^2-m^2)]V_{pd\sigma}$	$+l(1-l^2+m^2)V_{pd\pi}$	
E_{y,x^2-y^2}	=	$\frac{\sqrt{3}}{2}m[(l^2-m^2)]V_{pd\sigma}$	$-m(1+l^2-m^2)V_{pd\pi}$	
E_{z,x^2-y^2}	=	$\frac{\sqrt{3}}{2}n[(l^2-m^2)]V_{pd\sigma}$	$-n(l^2-m^2)V_{pd\pi}$	
$E_{x,3z^2-r^2}$	=	$l[n^2 - \frac{1}{2}(l^2+m^2)]V_{pd\sigma}$	$-\sqrt{3}ln^2V_{pd\pi}$	
$E_{y,3z^2-r^2}$	=	$m[n^2 - \frac{1}{2}(l^2+m^2)]V_{pd\sigma}$	$-\sqrt{3}mn^2V_{pd\pi}$	
$E_{z,3z^2-r^2}$	=	$n[n^2 - \frac{1}{2}(l^2+m^2)]V_{pd\sigma}$	$+\sqrt{3}n(l^2+m^2)V_{pd\pi}$	
$E_{xy,xy}$	=	$3l^2m^2V_{dd\sigma}$	$+(l^2+m^2-4l^2m^2)V_{dd\pi}$	$+(n^2+l^2m^2)V_{dd\delta}$
$E_{xy,yz}$	=	$3lm^2nV_{dd\sigma}$	$+ln(1-4m^2)V_{dd\pi}$	$+ln(m^2-1)V_{dd\delta}$
$E_{xy,zx}$	=	$3l^2mnV_{dd\sigma}$	$+mn(1-4l^2)V_{dd\pi}$	$+mn(l^2-1)V_{dd\delta}$
E_{xy,x^2-y^2}	=	$\frac{3}{2}lm(l^2-m^2)V_{dd\sigma}$	$2lm(m^2-l^2)V_{dd\pi}$	$\frac{1}{2}lm(l^2-m^2)V_{dd\delta}$
E_{yz,x^2-y^2}	=	$\frac{3}{2}mn(l^2-m^2)V_{dd\sigma}$	$-mn[1+2(l^2-m^2)]V_{dd\pi}$	$+mn[1+\frac{1}{2}(l^2-m^2)]V_{dd\delta}$
E_{zx,x^2-y^2}	=	$\frac{3}{2}nl(l^2-m^2)V_{dd\sigma}$	$+nl[1-2(l^2-m^2)]V_{dd\pi}$	$-nl[1-\frac{1}{2}(l^2-m^2)]V_{dd\delta}$
$E_{xy,3z^2-r^2}$	=	$\sqrt{3}lm[n^2 - \frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$-2\sqrt{3}lmn^2V_{dd\pi}$	$\frac{\sqrt{3}}{2}lm(1+n^2)V_{dd\delta}$
$E_{yz,3z^2-r^2}$	=	$\sqrt{3}mn[n^2 - \frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$+\sqrt{3}mn(l^2+m^2-n^2)V_{dd\pi}$	$-\frac{\sqrt{3}}{2}mn(l^2+m^2)V_{dd\delta}$
$E_{zx,3z^2-r^2}$	=	$\sqrt{3}ln[n^2 - \frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$+\sqrt{3}ln(l^2+m^2-n^2)V_{dd\pi}$	$-\frac{\sqrt{3}}{2}ln(l^2+m^2)V_{dd\delta}$
$E_{x^2-y^2,x^2-y^2}$	=	$\frac{3}{4}(l^2-m^2)^2V_{dd\sigma}$	$+[l^2+m^2-(l^2-m^2)^2]V_{dd\pi}$	$+[n^2+\frac{1}{4}(l^2-m^2)^2]V_{dd\delta}$
$E_{x^2-y^2,3z^2-r^2}$	=	$\frac{\sqrt{3}}{2}(l^2-m^2)[n^2 - \frac{1}{2}(l^2+m^2)]V_{dd\sigma}$	$+\sqrt{3}n^2(m^2-l^2)V_{dd\pi}$	$+\frac{1}{4}\sqrt{3}(1+n^2)(l^2-m^2)V_{dd\delta}$
$E_{3z^2-r^2,3z^2-r^2}$	=	$[n^2 - \frac{1}{2}(l^2+m^2)]^2V_{dd\sigma}$	$+3n^2(l^2+m^2)V_{dd\pi}$	$\frac{3}{4}(l^2+m^2)^2V_{dd\delta}$

B.4 Gaunt coefficients and Coulomb integrals

For the d shell, in the basis of spherical harmonics, the coefficients $G_k(m, m') = \langle lm|Y_q^k|lm'\rangle$ with $k = 2, 4$ are:

$$G_2 = \frac{1}{7\sqrt{4\pi}} \begin{bmatrix} -\sqrt{20} & \sqrt{30} & -\sqrt{20} & 0 & 0 \\ -\sqrt{30} & \sqrt{5} & \sqrt{5} & -\sqrt{30} & 0 \\ -\sqrt{20} & -\sqrt{5} & \sqrt{20} & -\sqrt{5} & -\sqrt{20} \\ 0 & -\sqrt{30} & \sqrt{5} & \sqrt{5} & -\sqrt{30} \\ 0 & 0 & -\sqrt{20} & \sqrt{30} & -\sqrt{20} \end{bmatrix}$$

$$G_4 = \frac{1}{7\sqrt{4\pi}} \begin{bmatrix} 1 & -\sqrt{5} & \sqrt{15} & -\sqrt{35} & \sqrt{70} \\ \sqrt{5} & -4 & \sqrt{30} & -\sqrt{40} & \sqrt{35} \\ \sqrt{15} & -\sqrt{30} & 6 & -\sqrt{30} & \sqrt{15} \\ \sqrt{35} & -\sqrt{40} & \sqrt{30} & -4 & \sqrt{5} \\ \sqrt{70} & -\sqrt{35} & \sqrt{15} & -\sqrt{5} & 1 \end{bmatrix}.$$

The two-index Coulomb integrals can be written as

$$U_{m,m'} = \sum_{k=0}^{2l} a_k(mm, m'm')F_k = \sum_{k=0}^{2l} b_k(m, m')F_k,$$

$$J_{m,m'} = \sum_{k=0}^{2l} a_k(mm', m'm)F_k = \sum_{k=0}^{2l} c_k(m, m')F_k,$$

where

$$a_k(m_\alpha m'_\alpha, m_\beta m'_\beta) = \frac{4\pi}{2k+1} \sum_{q=-k}^k \langle lm_\alpha|Y_q^k|lm'_\alpha\rangle \langle lm_\beta|\bar{Y}_q^k|lm'_\beta\rangle.$$

For $l = 2$, in the basis of spherical harmonics

$$b_0 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad b_2 = \frac{1}{49} \begin{bmatrix} 4 & -2 & -4 & -2 & 4 \\ -2 & 1 & 2 & 1 & -2 \\ -4 & 2 & 4 & 2 & -4 \\ -2 & 1 & 2 & 1 & -2 \\ 4 & -2 & -4 & -2 & 4 \end{bmatrix} \quad b_4 = \frac{1}{49} \frac{1}{9} \begin{bmatrix} 1 & -4 & 6 & -4 & 1 \\ -4 & 16 & -24 & 16 & -4 \\ 6 & -24 & 36 & -24 & 6 \\ -4 & 16 & -24 & 16 & -4 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

$$c_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad c_2 = \frac{1}{49} \begin{bmatrix} 4 & 6 & 4 & 0 & 0 \\ 6 & 1 & 1 & 6 & 0 \\ 4 & 1 & 4 & 1 & 4 \\ 0 & 6 & 1 & 1 & 6 \\ 0 & 0 & 4 & 6 & 4 \end{bmatrix} \quad c_4 = \frac{1}{49} \frac{1}{9} \begin{bmatrix} 1 & 5 & 15 & 35 & 70 \\ 5 & 16 & 30 & 40 & 35 \\ 15 & 30 & 36 & 30 & 15 \\ 35 & 40 & 30 & 16 & 5 \\ 70 & 35 & 15 & 5 & 1 \end{bmatrix}$$

In the basis of real harmonics, we find instead

$$b_0 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \quad b_2 = \frac{1}{49} \begin{bmatrix} 4 & -2 & -4 & -2 & 4 \\ -2 & 4 & 2 & -2 & -2 \\ -4 & 2 & 4 & 2 & -4 \\ -2 & -2 & 2 & 4 & -2 \\ 4 & -2 & -4 & -2 & 4 \end{bmatrix} \quad b_4 = \frac{1}{49 \cdot 9} \begin{bmatrix} 36 & -4 & 6 & -4 & -34 \\ -4 & 36 & -24 & -4 & -4 \\ 6 & -24 & 36 & -24 & 6 \\ -4 & -4 & -24 & 36 & -4 \\ -34 & -4 & 6 & -4 & 36 \end{bmatrix}$$

$$c_0 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad c_2 = \frac{1}{49} \begin{bmatrix} 4 & 3 & 4 & 3 & 0 \\ 3 & 4 & 1 & 3 & 3 \\ 4 & 1 & 4 & 1 & 4 \\ 3 & 3 & 1 & 4 & 3 \\ 0 & 3 & 4 & 3 & 4 \end{bmatrix} \quad c_4 = \frac{1}{49 \cdot 9} \begin{bmatrix} 36 & 20 & 15 & 20 & 35 \\ 20 & 36 & 30 & 20 & 20 \\ 15 & 30 & 36 & 30 & 15 \\ 20 & 20 & 30 & 36 & 20 \\ 35 & 20 & 15 & 20 & 36 \end{bmatrix}$$

For $l = 2$ the transformation matrix M from spherical to real harmonics is

$$M = \frac{1}{\sqrt{2}} \begin{bmatrix} -i & 0 & 0 & 0 & i \\ 0 & i & 0 & i & 0 \\ 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Sometimes the Coulomb integrals are given as linear combination of Racah parameters instead than of Slater integrals F_k . For the d shell, the Racah parameters are $A = F_0 - \frac{49}{441}F_4$, $B = \frac{1}{49}F_2 - \frac{5}{441}F_4$, and $C = \frac{35}{441}F_4$.

References

- [1] W. Kohn, *Nobel Lecture: Electronic structure of matter wave functions and density functionals*, Rev. Mod. Phys. **71**, 1253 (1998)
- [2] U. von Barth, Physica Scripta **109**, 9 (2004)
- [3] R.M. Martin, *Electronic Structure, Basic Theory and Practical Methods* (Cambridge University Press, 2004)
- [4] E. Pavarini, E. Koch and A.I. Lichtenstein, Phys. Rev. Lett. **101**, 266405 (2008)
- [5] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324327 (1989)
A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)
- [6] A.C. Hewson, *The Kondo problem to heavy Fermions* (Cambridge University Press, 1993)
- [7] J.E. Hirsch and R.M. Fye, Phys. Rev. Lett. **56**, 2521 (1986)
- [8] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer and P. Werner, Rev. Mod. Phys. **83**, 349 (2011)
- [9] M. Caffarel, W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)
E. Koch, G. Sangiovanni, and O. Gunnarsson, Phys. Rev. B **78**, 115102 (2008)
- [10] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)
- [11] O. Gunnarsson, Phys. Rev. B **41**, 514 (1990)
- [12] E. Pavarini, A. Yamasaki, J. Nuss and O.K. Andersen, New J. Phys. **7**, 188 (2005)
- [13] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997)
- [14] B. Amadon, F. Lechermann, A. Georges, F. Jollet, T.O. Wehling, and A.I. Lichtenstein, Phys. Rev. B **77**, 205112 (2008)
- [15] V.I. Anisimov, F. Aryasetiawan and A.I. Lichtenstein, J. Phys. Cond. Mat. **9**, 767 (1997)
- [16] O. Gunnarsson, O.K. Andersen, O. Jepsen and J. Zaanen, Phys. Rev. B **37**, 1708 (1989)
- [17] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A.I. Lichtenstein, Phys. Rev. B **70**, 195104 (2004)
- [18] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, and G. Kotliar, J. Phys: Condens. Matter **9**, 7359 (1997)
A.I. Lichtenstein and M.I. Katsnelson, Phys. Rev. B **57**, 6884 (1998)
- [19] A. Georges, in *Lectures on the Physics of Highly Correlated Electron Systems VIII*, American Institute of Physics Conference Proceedings **715**, 3 (2004)

- [20] M. Imada and T. Miyake, *J. Phys. Soc. Jpn.* **79**, 112001 (2010)
- [21] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet and C.A. Marianetti, *Rev. Mod. Phys.* **78**, 865 (2006)
- [22] E. Gorelov, M. Karolak, T. O. Wehling, F. Lechermann, A.I. Lichtenstein and E. Pavarini, *Phys. Rev. Lett.* **104**, 226401 (2010)
- [23] T. Maier, M. Jarrell, T. Pruschke, and M. Hettler, *Rev. Mod. Phys.* **77**, 1027 (2005).
- [24] E. Pavarini, S. Biermann, A. Poteryaev, A.I. Lichtenstein, A. Georges and O. K. Andersen, *Phys. Rev. Lett.* **92**, 176403 (2004)
- [25] O. Gunnarsson, E. Koch and R. Martin, *Phys. Rev. B* **57**, 1146 (1997)
- [26] M. De Raychahudhury, E. Pavarini and O.K. Andersen, *Phys. Rev. Lett.* **99**, 126402 (2007)
- [27] K.I. Kugel and D.I. Khomskii, *Zh. Eksp. Teor. Fiz.* **64**, 1429 (1973) [*Sov. Phys. JETP* **37**, 725 (1973)]
- [28] P. Fazekas, *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, Singapore, 1999)
- [29] W. G. Yin, D. Volja, and W. Ku, *Phys. Rev. Lett.* **96**, 116405 (2006)
- [30] I. Leonov, N. Binggeli, Dm. Korotin, V.I. Anisimov, and D. Vollhardt *Phys. Rev. Lett.* **101**, 096405 (2008); I. Leonov, Dm. Korotin, N. Binggeli, V.I. Anisimov, and D. Vollhardt *Phys. Rev. B* **81**, 075109 (2010)
- [31] K.H. Ahn, and A.J. Millis, *Phys. Rev. B* **61**, 13545 (2000)
- [32] A. Flesch, G. Zhang, E. Koch and E. Pavarini, arXiv:1106.2439 (2011)
- [33] E. Pavarini and E. Koch, *Phys. Rev. Lett.* **104**, 086402 (2010)
- [34] M.C. Sánchez, G. Subías, J. García, and J. Blasco, *Phys. Rev. Lett.* **90**, 045503 (2003)
- [35] J.-S. Zhou and J.B. Goodenough, *Phys. Rev. B* **68**, 144406 (2003) and *Phys. Rev. Lett.* **96**, 247202 (2006)
- [36] P.W. Anderson, *Science* **177**, 393 (1972)
- [37] R. B. Laughlin and D. Pines, *Proceedings of the National Accademy of Science* **97**, 28 (2000)
- [38] W. A. Harrison, *Electronic Structure and The Properties of Solids* (Dover, 1989)

7 The Constrained RPA Method for Calculating the Hubbard U from First-Principles

F. Aryasetiawan,¹ T. Miyake,^{2,3} and R. Sakuma¹

¹ Graduate School of Advanced Integration Science
Chiba University, Japan

² Nanosystem Research Institute, AIST, Japan

³ Japan Science and Technology Agency, CREST

Contents

1	Introduction	2
2	Screening and the random-phase approximation	4
3	Constrained RPA	9
3.1	Theory	9
3.2	Wannier orbitals	12
3.3	cRPA with the maximally localised Wannier function	13
3.4	Example: SrVO ₃	14
3.5	cRPA for entangled bands	16
3.6	Examples: Ni and Ce	19
3.6.1	Nickel	19
3.6.2	Cerium	20
3.7	Further examples	22
A	Basis functions	23
B	Flow chart	24

1 Introduction

Solving the many-electron problem is one of the main goals of condensed matter physics. Were it not for the presence of the Coulomb interaction among the electrons, the many-electron Hamiltonian could be easily solved since it amounts to solving a one-electron problem. The Hamiltonian without the Coulomb interaction is given by

$$H^0 = \sum_{n=1}^N h^0(r_n) = \sum_{n=1}^N \left[-\frac{1}{2} \nabla_n^2 + V_{\text{ext}}(r_n) \right], \quad (1)$$

where we have worked in atomic units ($\hbar = m = e = 1 \rightarrow 1 \text{ a.u.} = 27.2 \text{ eV}$), $r_n = (\mathbf{r}_n, \sigma_n)$ is a combined variable for position and spin, and V_{ext} is an external potential, such as the potential from the nuclear charges. The problem becomes enormously complicated when the Coulomb interaction among the electrons is added to the above Hamiltonian:

$$H = H^0 + V, \quad (2)$$

where

$$V = \frac{1}{2} \sum_{i \neq j} v(r_i - r_j), \quad v(r_i - r_j) = \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (3)$$

Except for small systems, such as atoms and small molecules containing a few tens of electrons, there is little hope of solving the many-electron problem exactly, in particular for excited states, which are our main interest here. Various simplifications and techniques are needed in order to make progress.

In terms of field operators and in the occupation number representation the Hamiltonian takes the form [1]

$$\hat{H} = \int dr \hat{\psi}^\dagger(r) \left[h^0(r) + \frac{1}{2} \hat{V}^H(r) \right] \hat{\psi}(r), \quad (4)$$

where

$$\hat{V}^H(r) = \int dr' v(r - r') \hat{\psi}^\dagger(r') \hat{\psi}(r') = \int dr' v(r - r') \hat{\rho}(r'). \quad (5)$$

We use the convention $\int dr = \sum_\sigma \int d^3r$. Since

$$\hat{\psi}(r) = \sum_n \varphi_n(r) c_n, \quad (6)$$

we obtain

$$\hat{H} = \sum_{ij} c_i^\dagger \left(h_{ij}^0 + \frac{1}{2} \hat{V}_{ij}^H \right) c_j, \quad (7)$$

where

$$\hat{V}_{ij}^H = \sum_{kl} v_{ij,kl} c_k^\dagger c_l, \quad (8)$$

and

$$h_{ij}^0 = \int dr \varphi_i^*(r) h^0(r) \varphi_j(r), \quad (9)$$

$$v_{ij,kl} = \int dr dr' \varphi_i^*(r) \varphi_j(r) v(r-r') \varphi_k^*(r') \varphi_l(r'). \quad (10)$$

The one-particle orbitals $\{\varphi_n\}$ are arbitrary but often chosen to be the Kohn-Sham orbitals. Note that the index n is a combined index of orbital and spin functions:

$$\varphi_n(r) = \varphi_k(\mathbf{r}) \xi(\sigma), \quad n = (k, \xi) \quad (11)$$

and that the definition of the Coulomb matrix is different from the usual convention. The definition in (10) is chosen to conform to the definition of the Hubbard U defined later.

A great simplification to the full many-electron Hamiltonian was introduced by Hubbard when studying the physics of transition metals in the late fifties. He noticed that most of the physics could well be attributed to electrons occupying the partially filled narrow $3d$ bands, which crossed the Fermi level. He then heuristically introduced the following Hamiltonian, now famously known as the Hubbard model [2]:

$$\hat{H} = \sum_{ij \subset 3d} c_i^+ h_{ij}^0 c_j + \frac{1}{2} \sum_{ijkl \subset 3d} U_{ij,kl} c_i^+ c_k^+ c_l c_j. \quad (12)$$

It is the same form as the Hamiltonian in (7) but the orbitals defining the annihilation and creation operators are now confined to the $3d$ orbitals and the Coulomb interaction has been replaced by some effective interaction U . The index i labels the atomic position and the localised $3d$ orbital. He reasoned that the rest of the electrons, that are more extended compared to the localised $3d$ electrons had the role of screening the Coulomb interaction between the $3d$ electrons and therefore the bare Coulomb interaction was reduced to an effective interaction U , the famous Hubbard U , which was assumed to have onsite components only, i.e., the labels i, j, k, l in $U_{ij,kl}$ refer to the same atomic site. Since then this seemingly simple looking model has had an enormous impact in the field of condensed matter physics as witnessed by a huge number of articles on works where the model has been used to study a wide range of problems from magnetism to superconductivity. The Hubbard model is suitable for studying materials with partially filled narrow bands. A large class of materials with this characteristic, often referred to as strongly correlated materials, are hosts to many intriguing physical properties [3] such as the metal-insulator transition and giant magnetoresistance. Small variations in the physical parameters, e.g., pressure or doping, can induce large changes in the physical properties.

Despite its simplicity the Hubbard model has proven to be highly non-trivial to solve, except for a few special cases such as the one-dimensional case. Even in its simplest form with only one orbital per site or one-band model,

$$\hat{H} = t \sum_{\langle i,j \rangle} c_i^+ c_j + U \sum_i n_{i\uparrow} n_{i\downarrow}, \quad (13)$$

no exact solution is known. The notation $\langle ij \rangle$ indicates that the hopping is restricted to the nearest neighbours only. In most cases the Hubbard model is then solved numerically by various methods. Among these we have the Lanczos method (exact diagonalisation), Quantum

Monte Carlo (QMC) method [4], and in recent years the dynamical mean-field theory (DMFT) method [5].

Central to the Hubbard model are the hopping parameters $t_{ij} = h_{ij}^0$ and the effective Coulomb interaction (Hubbard U) $U_{ij,kl}$. For the above simplest model there is actually only one effective parameter, namely, U/t . While it is relatively straightforward to extract the hopping parameters from realistic band structure calculations, it is much more elusive to determine the Hubbard U so that in many cases it is often treated as an adjustable parameter. To understand a generic physical phenomenon, it is quite appropriate to vary U in order to see the effects on the physical properties of interest. However, for a given material under a given condition, there does not seem to be any good reason to vary U since its value ought to be fixed. Different values of U simply correspond to different materials or different conditions. Reliable determination of U is therefore of utmost importance in order to be able to make quantitative predictions and to calculate materials properties from first principles. The present article is focused on the determination of the Hubbard U parameter from realistic first-principles calculations.

The problem of determining the Hubbard U from first principles has been addressed by a number of authors. One of the earliest works is the constrained local density approximation (cLDA) approach [6–8] where the Hubbard U is calculated from the total energy variation with respect to the occupation number of the localised orbital. A further improvement of this scheme was recently proposed [9]. Later, a different approach based on the random-phase approximation (RPA) was introduced [10, 11].

2 Screening and the random-phase approximation

In condensed matter physics, the concept of screening is crucial for understanding many of the physical properties of materials, especially metals. Take for example the famous anomaly associated with the disappearance of the density of states at the Fermi level in metals within the Hartree-Fock approximation, which neglects dynamic screening [12].

When a system of electrons is perturbed by a static external potential, the electrons will rearrange themselves in such a way as to minimize the total energy. If we introduce a positive test charge into the electronic system, the electrons will be attracted to surround the positive charge and in so doing reduce the total energy. The negative potential energy compensates for the increase in the kinetic energy due to the localisation of the electrons around the test charge. As a result of the electron accumulation around the test charge, the effective interaction between the test charge and an electron sufficiently outside the range where the electrons are accumulated becomes much weaker than the bare Coulomb interaction. In other words, the Coulomb interaction is screened. If the test charge is an electron, other electrons will be repelled and a screening hole is created which similarly screens the bare Coulomb interaction. In general, the perturbing field may be time dependent so that screening is a time-dependent or energy-dependent phenomenon. As a consequence, the screened interaction is retarded so that at finite frequencies it may become negative.

Consider applying a time-dependent perturbation $\delta\varphi$ to a system of electrons. The change in the

electron density induced by this perturbation generates in turn a change in the Hartree potential δV_{H} so that the total potential is given by

$$\delta V = \delta\varphi + \delta V_{\text{H}}. \quad (14)$$

The induced Hartree potential δV_{H} screens the applied perturbation $\delta\varphi$ and the ratio between the screened and the applied field is defined to be the inverse dielectric function. To simplify the writing, we use the notation $1 = (r_1, t_1)$ keeping in mind that $r_1 = (\mathbf{r}_1, \sigma_1)$ as previously defined in the Introduction:

$$\begin{aligned} \epsilon^{-1}(1, 2) &= \frac{\delta V(1)}{\delta\varphi(2)} \\ &= \delta(1-2) + \frac{\delta V_{\text{H}}(1)}{\delta\varphi(2)} \\ &= \delta(1-2) + \int d3v(1-3) \frac{\delta\rho(3)}{\delta\varphi(2)} \\ &= \delta(1-2) + \int d3v(1-3)R(3, 2), \end{aligned} \quad (15)$$

or in matrix notation

$$\epsilon^{-1} = 1 + vR. \quad (16)$$

The Coulomb potential $v(1-2)$ is given by

$$v(1-2) = v(|\mathbf{r}_1 - \mathbf{r}_2|)\delta(t_1 - t_2).$$

The δ -function indicates that the Coulomb interaction is instantaneous since we are dealing with non-relativistic cases. We have defined the linear density response function R according to

$$R(1, 2) = \frac{\delta\rho(1)}{\delta\varphi(2)} \quad \text{or} \quad \delta\rho(1) = \int d2R(1, 2)\delta\varphi(2), \quad (17)$$

which describes a change in the electron density $\delta\rho$ induced by an arbitrary time-dependent perturbation $\delta\varphi$ to first order.

The Coulomb interaction $v(1-2)$ may be thought of as the Coulomb potential at point \mathbf{r}_1 arising from a unit point charge at position \mathbf{r}_2 . If we regard this potential as a perturbation, according to (17), the change in the density due to this perturbation is given by,

$$\delta\rho(3, 2) = \int d4R(3, 4)v(4-2). \quad (18)$$

This induced charge generates in turn the Hartree potential

$$\delta V_{\text{H}}(1, 2) = \int d3v(1-3)\delta\rho(3, 2), \quad (19)$$

which screens the Coulomb potential at point \mathbf{r}_1 . The screened Coulomb potential at point \mathbf{r}_1 due to a unit point charge at point \mathbf{r}_2 is therefore given by

$$\begin{aligned}
 W(1, 2) &= v(1 - 2) + \delta V_{\text{H}}(1, 2) \\
 &= v(1 - 2) + \int d3 v(1 - 3)\delta\rho(3, 2) \\
 &= v(1 - 2) + \int d3d4 v(1 - 3)R(3, 4)v(4 - 2) \\
 &= \int d4 \epsilon^{-1}(1, 4)v(4 - 2).
 \end{aligned} \tag{20}$$

The last line is obtained from (15).

To describe formally the screening phenomenon, it is useful to work with the Green function in the interaction representation and employ the Schwinger functional derivative technique [13] as done by Hedin [14]. The Green function in the interaction or Dirac representation is defined as follows:

$$iG(1, 2) = \frac{\langle \Psi_0 | T[\hat{S}\hat{\psi}_D(1)\hat{\psi}_D^\dagger(2)] | \Psi_0 \rangle}{\langle \Psi_0 | \hat{S} | \Psi_0 \rangle} \tag{21}$$

where

$$\hat{S} = T \exp[-i \int d4 \hat{\rho}(4)\varphi(4)]. \tag{22}$$

T is the time-ordering operator that puts the operators chronologically from right to left. The field operators are in the Dirac or interaction representation:

$$\hat{\psi}_D(\mathbf{r}, t) = e^{i\hat{H}t}\hat{\psi}(\mathbf{r})e^{-i\hat{H}t}.$$

\hat{H} is the Hamiltonian of the interacting electron system defined in (4) without the perturbing field φ . The state $|\Psi_0\rangle$ is the ground state of \hat{H} , i.e., it is the same as the Heisenberg ground state in the usual definition of the Green function [1]. As can be easily seen, the above definition of the Green function reduces to the usual definition in terms of the Heisenberg field operators when $\varphi = 0$. One of the merits of the interaction picture is that the field operators do not depend on the perturbing field φ . This property, as will be seen later, is very useful when taking the functional derivative of the Green function with respect to the perturbing field in order to calculate the linear density response function.

The Green function in (21) satisfies the equation of motion

$$\left(i \frac{\partial}{\partial t_1} - h(1)\right) G(1, 2) - \int d3 \Sigma(1, 3)G(3, 2) = \delta(1 - 2), \tag{23}$$

where Σ is the self-energy without the Hartree potential and

$$h = -\frac{1}{2}\nabla^2 + V_{\text{ext}} + V_{\text{H}} + \varphi.$$

Here, V_{ext} and V_{H} are respectively the external field, such as the field from the nuclei, and the Hartree field. By multiplying both sides of (23) from the right by the inverse of the Green function we obtain

$$G^{-1}(1, 2) = \left(i \frac{\partial}{\partial t_1} - h(1) \right) \delta(1 - 2) - \Sigma(1, 2). \quad (24)$$

Since $GG^{-1} = 1$ we also have in matrix notation the identity

$$\frac{\delta G}{\delta \varphi} G^{-1} + G \frac{\delta G^{-1}}{\delta \varphi} = 0 \rightarrow \frac{\delta G}{\delta \varphi} = -G \frac{\delta G^{-1}}{\delta \varphi} G. \quad (25)$$

We are now in the position to derive the equation for the linear response function or the screened Coulomb interaction. Since $\rho(1) = -iG(1, 1^+)$, we find from (17)

$$\begin{aligned} R(1, 2) &= \frac{\delta \rho(1)}{\delta \varphi(2)} \\ &= -i \frac{\delta G(1, 1^+)}{\delta \varphi(2)} \\ &= i \int d3d4 G(1, 3) \frac{\delta G^{-1}(3, 4)}{\delta \varphi(2)} G(4, 1^+), \end{aligned} \quad (26)$$

where we have used the identity in (25). We now use the expression for the inverse of the Green function in (24) to calculate $\delta G^{-1}/\delta \varphi$:

$$\frac{\delta G^{-1}(3, 4)}{\delta \varphi(2)} = - \left[\delta(3 - 2) + \frac{\delta V_{\text{H}}(3)}{\delta \varphi(2)} \right] \delta(3 - 4) - \frac{\delta \Sigma(3, 4)}{\delta \varphi(2)}. \quad (27)$$

The first term on the right-hand side arises from $\delta \varphi(3)/\delta \varphi(2) = \delta(3 - 2)$. At this stage we will only keep the change in the Hartree potential and drop the term $\delta \Sigma/\delta \varphi$. This corresponds to the RPA, which may be regarded as the time-dependent Hartree approximation, since we only consider the change in the Hartree potential upon application of a time-dependent perturbation:

$$\frac{\delta V_{\text{H}}(3)}{\delta \varphi(2)} = \frac{\delta}{\delta \varphi(2)} \int d5 v(3 - 5) \rho(5) = \int d5 v(3 - 5) R(5, 2), \quad (28)$$

Within the RPA we then have

$$\frac{\delta G^{-1}(3, 4)}{\delta \varphi(2)} = - \left[\delta(3 - 2) + \int d5 v(3 - 5) R(5, 2) \right] \delta(3 - 4). \quad (29)$$

Using this in (26) we arrive at

$$\begin{aligned} R(1, 2) &= -i \int d3 G(1, 3) \left[\delta(3 - 2) + \int d5 v(3 - 5) R(5, 2) \right] G(3, 1^+) \\ &= P(1, 2) + \int d3d5 P(1, 3) v(3 - 5) R(5, 2), \end{aligned} \quad (30)$$

where we have defined the polarisation function

$$P(1, 2) = -iG(1, 2)G(2, 1^+). \quad (31)$$

In matrix form

$$R = P + PvR \rightarrow R = [1 - Pv]^{-1} P, \quad (32)$$

which is the well-known RPA equation. We note that, while in the RPA the polarisation function P is approximated by (31), the exact response function R satisfies the same equation with the exact polarization function P . It is straightforward to verify using (16) and (32) that the dielectric function is given by

$$\epsilon = 1 - vP : \quad (33)$$

$$\begin{aligned} \epsilon\epsilon^{-1} &= (1 - vP)(1 + vR) \\ &= 1 + vR - v(P + PvR) \\ &= 1. \end{aligned} \quad (34)$$

Using the convolution theorem, the Fourier transform of P in (31) is given by

$$P(r, r'; \omega) = -i \int \frac{d\omega'}{2\pi} G(r, r'; \omega + \omega') G(r', r; \omega'), \quad (35)$$

where the Fourier transform is defined according to

$$G(\omega) = \int dt e^{i\omega t} G(t), \quad G(t) = \int \frac{d\omega}{2\pi} e^{-i\omega t} G(\omega).$$

Using a non-interacting Green function

$$G^0(r, r'; \omega) = \sum_n^{\text{occ}} \frac{\varphi_n(r)\varphi_n^*(r')}{\omega - \varepsilon_n - i\delta} + \sum_m^{\text{unocc}} \frac{\varphi_m(r)\varphi_m^*(r')}{\omega - \varepsilon_m + i\delta}, \quad (36)$$

where $\{\varphi_n, \varepsilon_n\}$ are usually taken to be the Kohn-Sham orbitals and eigenvalues, the frequency integral can be performed analytically using Cauchy's theorem. Terms involving products of two occupied states or two unoccupied states vanish because the two poles lie on the same plane. Only terms involving the products of occupied and unoccupied states survive. For example, considering only the frequency-dependent parts,

$$-i \int \frac{d\omega}{2\pi} \left(\frac{1}{\omega + \omega' - \varepsilon_n - i\delta} \right)_{\text{occ}} \times \left(\frac{1}{\omega' - \varepsilon_m + i\delta} \right)_{\text{unocc}}$$

can be integrated analytically using the Cauchy theorem by closing the contour along an infinitely large semicircle either in the upper or lower half plane. This yields

$$\frac{-i}{2\pi} (2\pi i \times \text{residue}) = -\frac{1}{\omega + \varepsilon_m - \varepsilon_n - i\delta}.$$

Consequently,

$$\begin{aligned} P(r, r'; \omega) &= - \sum_n^{\text{occ}} \sum_m^{\text{unocc}} \frac{\varphi_n(r)\varphi_n^*(r')\varphi_m(r')\varphi_m^*(r)}{\omega + \varepsilon_m - \varepsilon_n - i\delta} \\ &\quad + \sum_n^{\text{occ}} \sum_m^{\text{unocc}} \frac{\varphi_n(r')\varphi_n^*(r)\varphi_m(r)\varphi_m^*(r')}{\omega - \varepsilon_m + \varepsilon_n + i\delta}, \end{aligned} \quad (37)$$

which can be rewritten more compactly as a sum over occupied and unoccupied pairs of orbitals

$$P(r, r'; \omega) = \sum_{\alpha} \left(\frac{b_{\alpha}(r)b_{\alpha}^*(r')}{\omega - \Delta_{\alpha} + i\delta} - \frac{b_{\alpha}^*(r)b_{\alpha}(r')}{\omega + \Delta_{\alpha} - i\delta} \right), \quad (38)$$

where

$$b_{\alpha} = \varphi_n^* \varphi_m, \quad \Delta_{\alpha} = \varepsilon_m - \varepsilon_n > 0. \quad (39)$$

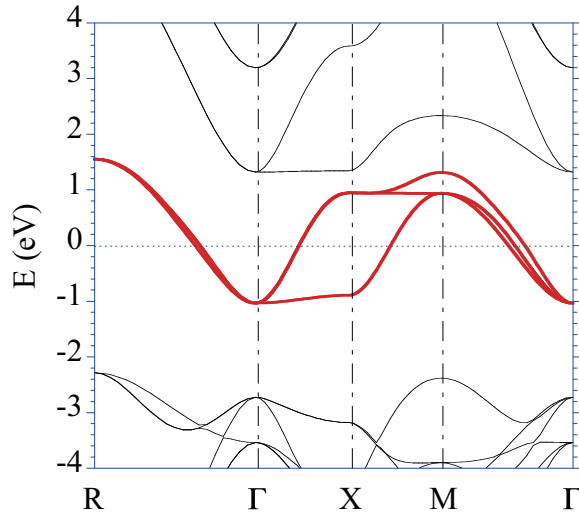


Fig. 1: The band structure of SrVO_3 . The red bands correspond to the vanadium $3d$ orbitals of t_{2g} character. Although the density of states corresponding to the red bands overlap with the rest of the density of states, as can be seen in Fig. 2, the red bands originating from the vanadium $3d$ t_{2g} orbitals are isolated from the rest of the bands.

3 Constrained RPA

3.1 Theory

Let us consider a system with a narrow band, well separated from other bands, crossing the Fermi level. As a concrete example, consider the case of the perovskite SrVO_3 , whose band structure and density of states are shown in Figs. 1 and 2. We first divide the one-particle Hilbert space into two parts, which we call the d and r subspaces. The d subspace is identified with the narrow band, which in the example of SrVO_3 are marked in red.

We may separate the total polarisation of the system into the polarisation within the d subspace, which we shall call P_d , and the rest of the polarisation, which we shall call P_r :

$$P = P_d + P_r. \quad (40)$$

The meaning of P_d and P_r is illustrated in Fig. 3. In the example of SrVO_3 , the red bands in Fig. 1 form our d subspace, which corresponds to the subspace of our model and we wish to determine the Hubbard U or the effective interaction among electrons residing in the red bands. From (20) the fully screened Coulomb interaction is given by

$$W = \epsilon^{-1}v, \quad (41)$$

where from (16) and (33)

$$\epsilon^{-1} = 1 + vR \quad \text{and} \quad \epsilon = 1 - vP. \quad (42)$$

The fully screened interaction can be rewritten as

$$W = [1 - W_r P_d]^{-1} W_r, \quad (43)$$

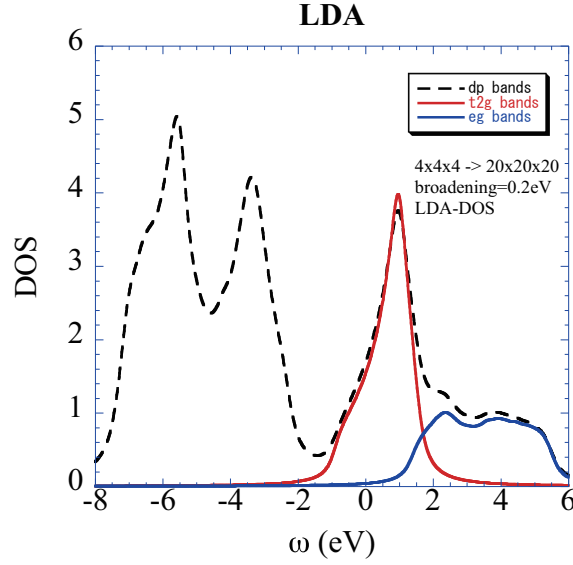


Fig. 2: Density of states of SrVO_3 . The dotted curve is the total density of states. The red and blue curves show the vanadium $3d$ t_{2g} and e_g components, respectively. The density of states between -8 to -2 eV originates from the oxygen $2p$ states.

where

$$W_r = [1 - vP_r]^{-1}v = \epsilon_r^{-1}v. \quad (44)$$

We verify the above identity. From (41) and (42)

$$\begin{aligned} W &= [1 - vP]^{-1}v \\ &= [1 - vP_r - vP_d]^{-1}v \\ &= \{\epsilon_r[1 - \epsilon_r^{-1}vP_d]\}^{-1}v \\ &= [1 - \epsilon_r^{-1}vP_d]^{-1}\epsilon_r^{-1}v \\ &= [1 - W_rP_d]^{-1}W_r. \end{aligned} \quad (45)$$

We observe that the identity in Eq. (43) allows us to interpret W_r as the effective interaction among electrons residing in the d subspace or the Hubbard U [15] because when this effective interaction is screened further in the model by P_d we obtain the fully screened interaction:

$$U(\mathbf{r}, \mathbf{r}'; \omega) = W_r(\mathbf{r}, \mathbf{r}'; \omega). \quad (46)$$

A formal derivation of the Hubbard U from the many-electron Hamiltonian may be found in [16]. The Hubbard U is frequency dependent as a consequence of retarded screening effects.

Eq. (44) is exact, but in practice we approximate $P_r = P - P_d$ within the RPA, which takes the form

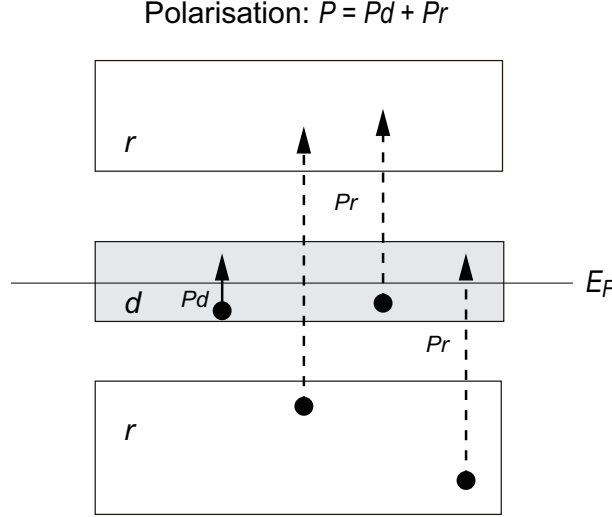


Fig. 3: A schematic picture explaining the meaning of P_d and P_r . While P_d is confined to the transitions within the d subspace, P_r may contain transitions between the d and r subspaces.

$$P(\mathbf{r}, \mathbf{r}'; \omega) = \sum_{kn}^{\text{occ}} \sum_{k'n'}^{\text{unocc}} \left\{ \frac{\psi_{kn}^*(\mathbf{r}) \psi_{k'n'}(\mathbf{r}) \psi_{k'n'}^*(\mathbf{r}') \psi_{kn}(\mathbf{r}')}{\omega - \varepsilon_{k'n'} + \varepsilon_{kn} + i\delta} - \frac{\psi_{kn}(\mathbf{r}) \psi_{k'n'}^*(\mathbf{r}') \psi_{k'n'}(\mathbf{r}) \psi_{kn}^*(\mathbf{r}')}{\omega + \varepsilon_{k'n'} - \varepsilon_{kn} - i\delta} \right\}, \quad (47)$$

where $\{\psi_{kn}, \varepsilon_{kn}\}$ are usually chosen to be the Kohn-Sham eigenfunctions and eigenvalues and $k = (\mathbf{k}, \sigma)$ is a combined index for the \mathbf{k} -vector and the spin σ . For systems without spin-flipping processes, k and k' evidently have the same spin. P_d has exactly the same form as in Eq. (47) but with the bands n and n' restricted to the d subspace. We note that P_r contains not only transitions inside the r subspace but also transitions between the d and r subspaces as illustrated in Fig. 3.

Since P_r does not contain low-energy polarisations that are responsible for metallic screening, U becomes long range. The asymptotic decay of U as a function of distance is expected to behave according to $1/(\alpha r)$ where $\alpha > 1$ rather than exponential, as often assumed. This behaviour is illustrated, e.g., in the case of the BEDT-TTF organic conductors [17].

It may be argued that for narrow-band materials with strong correlations it would not be sufficient to calculate U within the RPA. We would like to point out that from a physical point of view much of the error in the RPA resides in P_d rather than P_r because the former corresponds to the polarisation of the narrow bands, where we expect vertex corrections to the RPA to be large, whereas the latter corresponds to polarisation involving more extended states, for which the RPA is supposed to perform well. Since it is P_r that enters into the calculation of U , we expect that the error in the RPA has much less influence on U than one would anticipate.

In practice, Eq. (44) is solved by introducing a set of basis functions, and the choice of basis functions depends on the band-structure method. For band-structure methods based on pseudopotentials, a plane-wave basis set is a natural choice. For band-structure methods based on

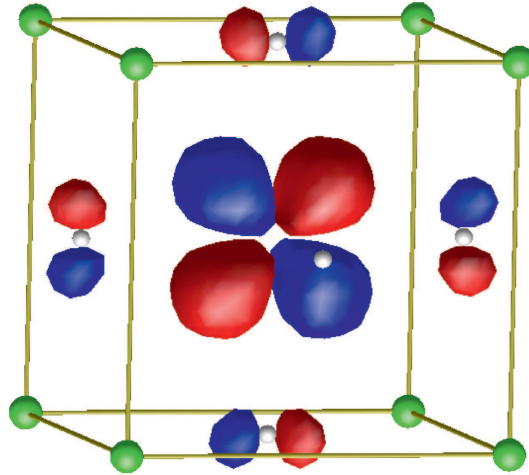


Fig. 4: Contour plot of the maximally localized Wannier function (MLWF) of SrVO_3 . If we take the x axis to be the horizontal direction and the z axis to be the vertical direction, the Wannier function corresponds to xz . The red (blue) represents the positive (negative) contour. The Wannier function is centered at the vanadium site, which is located at the center of the cube. The green spheres at the corners are strontium atoms, and white spheres at the centers of the faces are oxygen atoms. The MLWF is optimized in the t_{2g} model which consists of three t_{2g} -like states. We note that the Wannier function has tails on the oxygen sites.

localised basis functions, such as the linear muffin-tin orbital (LMTO) method, the linearised augmented plane-wave method (LAPW) [18], or the Korringa-Kohn-Rostocker (KKR) method, a product basis set [19, 20] is usually used. This is described in the Appendix.

3.2 Wannier orbitals

After obtaining $U(\mathbf{r}, \mathbf{r}'; \omega)$ the next step is to calculate its matrix elements in some localised orbitals. The Hubbard model in Eq. (12) is defined with respect to a chosen one-particle basis set defining the creation and annihilation operators of the field operators. We must therefore calculate the matrix elements of U in this chosen one-particle basis set. It is of course up to us what basis we choose but an appealing choice is the maximally localised Wannier orbitals which are constructed as follows [21].

The Wannier function with band index n at cell \mathbf{R} is defined by

$$|\varphi_{n\mathbf{R}}\rangle = \frac{V}{(2\pi)^3} \int d^3k e^{-i\mathbf{k}\cdot\mathbf{R}} |\psi_{\mathbf{k}n}^{(w)}\rangle, \quad (48)$$

where $|\psi_{\mathbf{k}n}^{(w)}\rangle$ is the associated Bloch function, which can be expanded as a linear combination of the eigenfunctions of a mean-field Hamiltonian as

$$|\psi_{\mathbf{k}n}^{(w)}\rangle = \sum_m |\psi_{\mathbf{k}m}\rangle \mathcal{U}_{mn}(\mathbf{k}). \quad (49)$$

In practical implementations, the Kohn-Sham wavefunctions are usually used for $|\psi_{\mathbf{k}m}\rangle$. In the maximally localised Wannier function scheme, the coefficients $\mathcal{U}_{mn}(\mathbf{k})$ are determined such

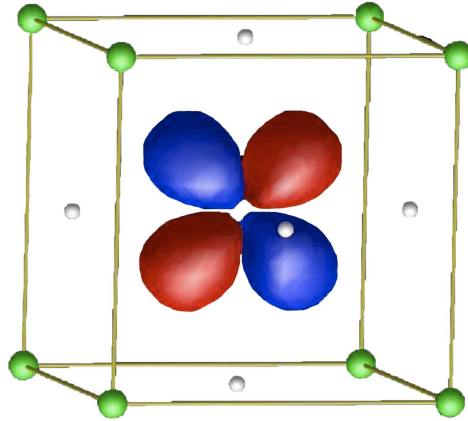


Fig. 5: As in Fig. 4 but the Wannier function is constructed according to the dp model which consists of the vanadium $3d$ bands ($t_{2g}+e_g$) and oxygen $2p$ bands. Compared to the one in Fig. 4 the Wannier function has become much more localised on the vanadium site.

that the quadratic extent of the wavefunctions

$$\Omega = \sum_n (\langle \varphi_{n0} | r^2 | \varphi_{n0} \rangle - |\langle \varphi_{n0} | \mathbf{r} | \varphi_{n0} \rangle|^2) \quad (50)$$

is minimised. When the bands are isolated, the Wannier orbitals are well defined and span the same Hilbert space as that of the isolated bands. However, when the bands are not isolated the Wannier orbitals are not unique. For this case, we introduce an energy window and optimise $\mathcal{U}_{mn}(\mathbf{k})$ with m limited to the states inside the window. The Wannier function is the more localised the larger the energy window, since optimisation is then done in a wider Hilbert space. This is illustrated in Figs. 4 and 5.

3.3 cRPA with the maximally localised Wannier function

Once $\mathcal{U}_{mn}(\mathbf{k})$ is determined on a \mathbf{k} mesh, maximally localised Wannier functions are obtained by Fourier transform as in (48), from which the Hamiltonian corresponding to the d subspace is constructed: $H_{mn}(\mathbf{R}) = \langle \varphi_{m0} | H | \varphi_{n\mathbf{R}} \rangle$. By Fourier transforming $H_{mn}(\mathbf{R})$ back to \mathbf{k} space and diagonalising it, we reproduce the original narrow bands. If the narrow bands forming the d subspace are not completely separated from the rest of the bands, the resulting bands will no longer in general be the same as the original bands. For the case of SrVO_3 we first construct from the red bands in Fig. 1 three Wannier orbitals having strong $3d$ character of t_{2g} symmetry. The next step is to compute the screened Coulomb interaction $W_r(\mathbf{r}, \mathbf{r}'; \omega)$ in the cRPA and take the matrix elements in the maximally localised Wannier basis [22]:

$$W_r(n_1, n_2, n_3, n_4; \mathbf{R}; \omega) \equiv \int \int d^3r d^3r' \varphi_{n_1\mathbf{0}}^*(\mathbf{r}) \varphi_{n_2\mathbf{0}}(\mathbf{r}) W_r(\mathbf{r}, \mathbf{r}'; \omega) \varphi_{n_3\mathbf{R}}^*(\mathbf{r}') \varphi_{n_4\mathbf{R}}(\mathbf{r}'). \quad (51)$$

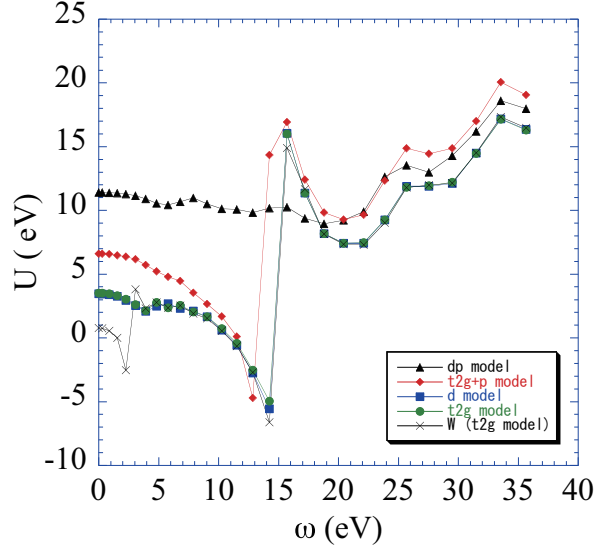


Fig. 6: The real part of the Hubbard U for various d subspaces indicated by the legends in the figure. W is the fully screened interaction. The meaning of the legends is summarised in the table below. For example, the d subspace of the $t_{2g} + p$ model is comprised of the vanadium t_{2g} and oxygen p bands.

model:	t_{2g}	d	$t_{2g} + p$	dp
d subspace:	$V t_{2g}$	$V(t_{2g} + e_g)$	$V t_{2g} + O p$	$V(t_{2g} + e_g) + O p$

The above expression is the most general, but in practice we often restrict ourselves to the on-site values and only consider the direct (charge-charge) and exchange components. The on-site Hubbard U matrix is defined to be

$$U_{nm}(\omega) \equiv \int \int d^3r d^3r' |\varphi_{n0}(\mathbf{r})|^2 W_r(\mathbf{r}, \mathbf{r}'; \omega) |\varphi_{m0}(\mathbf{r}')|^2 \quad (52)$$

and the onsite exchange matrix J .

$$J_{nm}(\omega) \equiv \int \int d^3r d^3r' \varphi_{n0}^*(\mathbf{r}) \varphi_{m0}(\mathbf{r}) W_r(\mathbf{r}, \mathbf{r}'; \omega) \varphi_{n0}(\mathbf{r}') \varphi_{m0}^*(\mathbf{r}'). \quad (53)$$

Note that the definitions of U and J may vary according to convention but in any case the various definitions can be related to (52) and (53). At this point it is worth pointing out that the effective screened interaction $W_r(\mathbf{r}, \mathbf{r}'; \omega)$ calculated using the cRPA method is completely independent of the choice of basis functions. The *matrix elements* are of course dependent on the choice of the orbitals $\{\varphi_{n0}\}$.

3.4 Example: SrVO₃

To illustrate the usefulness of the cRPA method in studying the screening properties of materials, we consider for SrVO₃ the calculations of U for various models. Although we have illustrated the cRPA scheme for a narrow band, the choice of the d subspace is entirely arbitrary and it may not necessarily correspond to a narrow band.

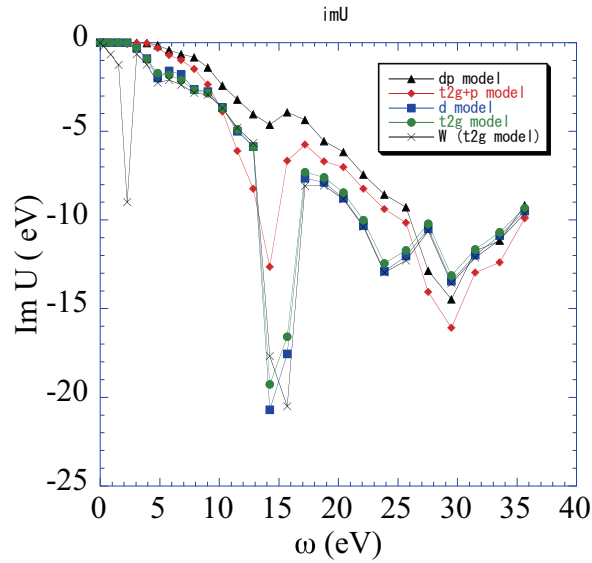


Fig. 7: The imaginary part of the Hubbard U for the models indicated by the legends in the picture. W is the fully screened interaction. For the definitions of the models, see the caption of Fig. 6.

We consider first the fully screened interaction W . Two notable features are clearly discernible in Fig. 7: there is the usual high-energy electron-gas-like plasmon excitation at around 16 eV and more remarkably there is a strong excitation between $2 \sim 3$ eV, which arises mainly from the collective excitation within the partially filled t_{2g} bands. That this is indeed the case can be understood by comparison with $U(\omega)$ for the t_{2g} model in Figs. 7 and 6, where the d subspace is formed by the t_{2g} bands. The structure at $2 \sim 3$ eV is almost absent, due to the elimination of the polarisation within the t_{2g} bands when calculating $U(\omega)$.

Comparison between the t_{2g} - and the d model, where both the t_{2g} and the e_g bands form the d subspace, clearly shows that the corresponding Hubbard U 's are almost the same implying that the $t_{2g} \rightarrow e_g$ screening channel is essentially ineffective as can be seen in Fig. 6.

When the t_{2g} model is enlarged to the $t_{2g} + p$ model, where the d subspace is formed by the vanadium t_{2g} and the oxygen $2p$ bands, the corresponding static Hubbard U is almost doubled from 3.5 eV to 6.5 eV as shown in Fig. 6, demonstrating the important role of the oxygen p electrons in screening the Coulomb interaction. When calculating U in the $t_{2g} + p$ model, the $O_p \rightarrow t_{2g}$ screening channel is left out, and it is the elimination of this screening channel that is responsible for the large increase in the low-energy U . Furthermore, since the amount of screening channels is reduced compared with that of the t_{2g} only model, the onset of the plasmon excitation at around 16 eV is lowered by about 2 eV. In other words, the effective number of electron participating in the formation of plasmon excitations is reduced. It is noteworthy that the onset of the plasmon excitation in the t_{2g} and d models is unchanged, indicating that the plasmon excitation is not coupled to the polarisation within the d bands.

In the most expanded dp model, where the d subspace consists of the vanadium t_{2g} and e_g and the oxygen $2p$ bands, the value of the static U is approximately doubled further to 11 eV as can

be seen in Fig. 6, due to the elimination of the $O_p \rightarrow e_g$ screening, giving further evidence for the importance of the oxygen p electrons in the screening process. It can also be seen that the plasmon onset almost disappears, showing that the plasmon excitation mainly couples to the oxygen p electrons.

3.5 cRPA for entangled bands

Although the cRPA method is rather general, a serious technical problem arises when the narrow band is entangled with other bands, i.e., the narrow band is not completely isolated from the rest of the bands, a situation which occurs in many materials. For example, in $3d$ transition metals, the $3d$ bands mix with the $4s$ and $4p$ bands as illustrated in Fig. 8 and similarly the $4f$ bands of the $4f$ metals hybridise with the more extended $6s$ band. For such cases, it is not obvious anymore how to determine P_r in order to calculate U using the cRPA method.

A number of procedures have been proposed to handle the problem of determining U for entangled bands. One proposal is to choose a set of band indices and define the corresponding bands as the one-particle bands in the Hubbard model. Another proposal is to introduce an energy window and define the one-particle bands to be those that have energies within the energy window. Yet another proposal is to have a combination of energy window and band indices. These procedures, however, suffer from a number of difficulties. When choosing band indices it is likely that some of the states will have a character very different from that of the intended model. For example, in the case of the $3d$ transition metals, choosing five "3d" bands will include at some k-points states which contains a considerable $4s$ component but little $3d$ character. Moreover, the chosen bands will not in general form smoothly connected bands. A similar problem is encountered when choosing an energy window. A hybrid construction using band indices and energy window [23] removes part of the problem but it is somewhat arbitrary. Another procedure is, as we will discuss in detail later, to project the polarisation to the orbitals of interest, e.g., $3d$ orbitals, but this procedure has been found to yield an unphysical result of negative static U .

To overcome the problem with entangled bands we propose the following procedure. We first construct a set of localised Wannier orbitals from a given set of bands defined within a certain energy window by following the post-processing procedure of Souza, Marzari and Vanderbilt [21] or other methods, such as the preprocessing scheme proposed by Andersen *et al.* within the Nth-order muffin-tin orbital (NMTO) method [24]. We then choose this set of Wannier orbitals as the d subspace and use them as a basis for diagonalising the one-particle Hamiltonian, which is usually the Kohn-Sham Hamiltonian in the local density approximation (LDA) or generalised gradient approximation (GGA). The so obtained set of bands, which equivalently define the d subspace, may be slightly different from the original bands defined within the chosen energy window. It is therefore important to confirm that the dispersions near the Fermi level well reproduces the original Kohn-Sham bands. From these bands we calculate the polarisation \tilde{P}_d ,

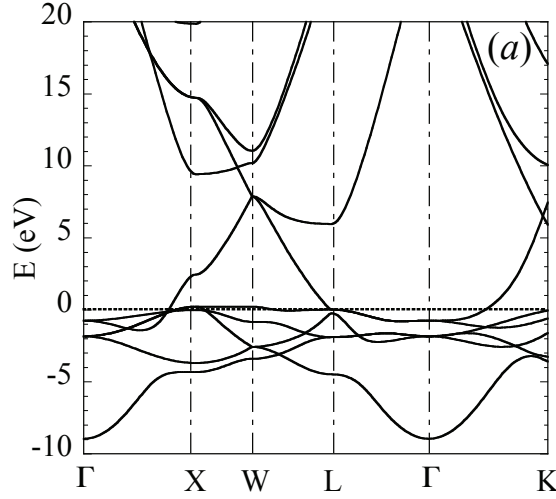


Fig. 8: The band structure of paramagnetic nickel taken from [25]. The 3d band is approximately between -4 eV and slightly above the Fermi level (zero energy).

$$\tilde{P}_d(\mathbf{r}, \mathbf{r}'; \omega) = \sum_i^{\text{occ}} \sum_j^{\text{unocc}} \left[\frac{\tilde{\psi}_i^*(\mathbf{r})\tilde{\psi}_j(\mathbf{r})\tilde{\psi}_j^*(\mathbf{r}')\tilde{\psi}_i(\mathbf{r}')}{\omega - \tilde{\varepsilon}_j + \tilde{\varepsilon}_i + i\delta} - \frac{\tilde{\psi}_i(\mathbf{r})\tilde{\psi}_j^*(\mathbf{r})\tilde{\psi}_j(\mathbf{r}')\tilde{\psi}_i^*(\mathbf{r}')}{\omega + \tilde{\varepsilon}_j - \tilde{\varepsilon}_i - i\delta} \right], \quad (54)$$

where $\{\tilde{\psi}_i\}$, $\{\tilde{\varepsilon}_i\}$ ($i = 1, \dots, N_d$) are the wavefunctions and eigenvalues obtained from diagonalising the one-particle Hamiltonian in the Wannier basis.

It would seem sensible to define the rest of the polarisation as $P_r = P - \tilde{P}_d$, where P is the full polarisation calculated using the *original* (Kohn-Sham) wavefunctions and eigenvalues $\{\psi_i\}$, $\{\varepsilon_i\}$ ($i = 1, \dots, N$), and calculate W_r according to Eq. (44). We have found, however, that this procedure is numerically very unstable, resulting in some cases in unphysical negative static U and large oscillations at low energy. The reason is that \tilde{P}_d does not completely encompass the low-energy excitations so that low energy screening channels associated with the d - d transitions are not completely excluded from P_r . Due to the singular nature of the expression in Eq. (44) these remaining low-energy excitations can cause large fluctuations in W_r .

Another way of calculating P_r is to project the wavefunctions to the d space,

$$|\bar{\psi}_i\rangle = \hat{\mathcal{P}}|\psi_i\rangle, \quad (55)$$

where the projection operator $\hat{\mathcal{P}}$ is defined as

$$\hat{\mathcal{P}} = \sum_{j=1}^{N_d} |\tilde{\psi}_j\rangle\langle\tilde{\psi}_j|. \quad (56)$$

The effective d polarisation may be expressed as

$$\bar{P}_d(\mathbf{r}, \mathbf{r}'; \omega) = \sum_i^{\text{occ}} \sum_j^{\text{unocc}} \left[\frac{\bar{\psi}_i^*(\mathbf{r})\bar{\psi}_j(\mathbf{r})\bar{\psi}_j^*(\mathbf{r}')\bar{\psi}_i(\mathbf{r}')}{\omega - \varepsilon_j + \varepsilon_i + i\delta} - \frac{\bar{\psi}_i(\mathbf{r})\bar{\psi}_j^*(\mathbf{r})\bar{\psi}_j(\mathbf{r}')\bar{\psi}_i^*(\mathbf{r}')}{\omega + \varepsilon_j - \varepsilon_i - i\delta} \right], \quad (57)$$

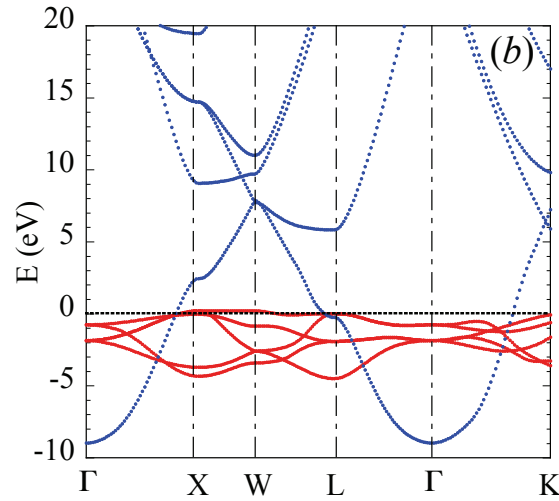


Fig. 9: The disentangled band structure of paramagnetic nickel obtained by diagonalising the Hamiltonian in (59) taken from [25]. The red bands correspond to maximally localised Wannier orbitals of 3d character which form the d subspace. The blue curves correspond to the r subspace.

and $P_r = P - \bar{P}_d$ can be used to calculate W_r . We found that this procedure does not work either and is again unstable for a similar reason as the one described above. Moreover, $\bar{\psi}_i$'s are not orthogonal with each other, and transitions between the states do not correspond to single particle-hole excitations.

Based on these observations we propose the following procedure [25]. We define the r subspace by

$$|\phi_i\rangle = (1 - \hat{\mathcal{P}})|\psi_i\rangle \quad (58)$$

which is orthogonal to the d subspace constructed from the Wannier orbitals. In practice it is convenient to orthonormalise $\{\phi_i\}$ and prepare $N - N_d$ basis functions. By diagonalising the Hamiltonian in this subspace a new set of wavefunctions $\{\tilde{\phi}_i\}$ and eigenvalues $\{\tilde{\epsilon}_i\}$ ($i = 1, \dots, N - N_d$) is obtained. Since the subspaces formed by $\{\phi_i\}$ and $\{\tilde{\psi}_j\}$ are orthogonal, the set of $(N - N_d)$ bands $\{\tilde{\epsilon}_i\}$ are completely disentangled from those of the d space $\{\tilde{\epsilon}_j\}$, and they are slightly different from the original band structure $\{\epsilon_i\}$. Numerical tests show that the disentangled band structure is close to the original one as may be seen in the example of nickel in Figs. 8 and 9. The form of the Hamiltonian is illustrated below

$$H = \begin{bmatrix} H_{dd} & 0 \\ 0 & H_{rr} \end{bmatrix}, \quad (59)$$

where H_{dd} is the Hamiltonian matrix taken in the d subspace $\{\tilde{\psi}_j\}$ and H_{rr} is taken in the subspace of $\{\phi_i\}$. In other words, the coupling between the d and r subspaces is set to zero.

The Hubbard U is then calculated according to Eq. (44) with $P_r = \tilde{P} - \tilde{P}_d$, where \tilde{P} is the full polarisation calculated for the *disentangled* band structure. We note that the screening processes between the d and r subspaces are not neglected but included in P_r , although the d - r coupling is

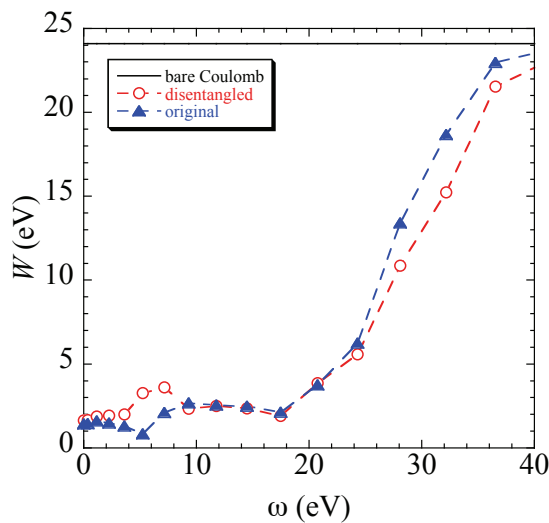


Fig. 10: Comparison between the fully screened interactions W of nickel for the normal case with the original band structure (triangle, blue) and for the case where the band structure is disentangled (circle, red).

cut off in the construction of the wavefunctions and eigenvalues. In the Appendix, the flowchart for the calculation of the Hubbard U is shown.

3.6 Examples: Ni and Ce

3.6.1 Nickel

As applications of the cRPA method for entangled bands, we have calculated the Hubbard U for the $3d$ transition metal series. In Fig. 10 we compare for the case of nickel the fully screened interaction W calculated using the disentangled $3d$ bands with W calculated using the original band structure. The agreement between the two are quite satisfactory for our purpose. Most of the error arising from the disentanglement originates from regions in k -space where the $3d$ bands and the $4s$ - $4p$ bands hybridise and repel each other. After the disentanglement, these bands cross rather than repel each other, as can be seen by comparing Figs. 8 and 9.

The resulting frequency-dependent Hubbard U and the exchange J for nickel are shown in Figs. 11 and 12, respectively. The complete removal of low-energy excitations within the d subspace when calculating P_r ensures that U has little structure at low energy within the band width of the d subspace. The only remaining low-energy transitions come mainly from the $4s$ band. The increase in U at around 20 eV is due to the coupling to plasmon excitations, which in the case of transition metals form a rather broad excitation.

J has a relatively weak dependence on energy, and its static value is approximately given by the unscreened value although some screening effects reducing the unscreened value from 0.8 eV to 0.7 eV at zero frequency can be observed. This is in agreement with the usual practice of taking the atomic J value, which corresponds approximately to the unscreened value, implicitly assuming that screening effects are small for the Coulomb potential arising from the exchange

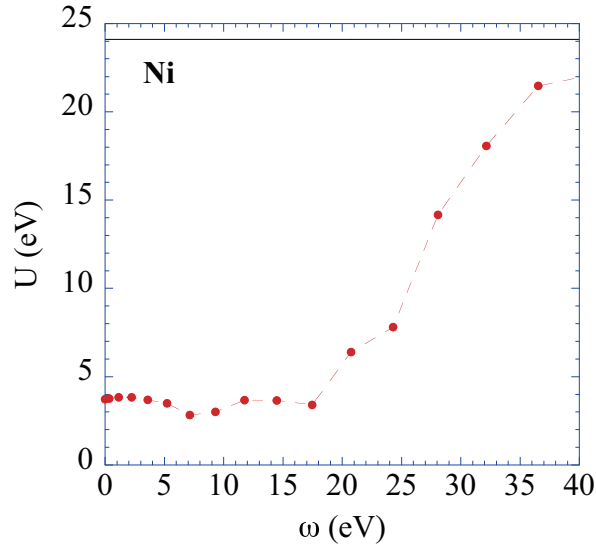


Fig. 11: The Hubbard U of paramagnetic nickel as defined in (52) obtained using the *cRPA* method for disentangled bands as explained in the article. The value is averaged over the diagonal elements of the $3d$ orbitals.

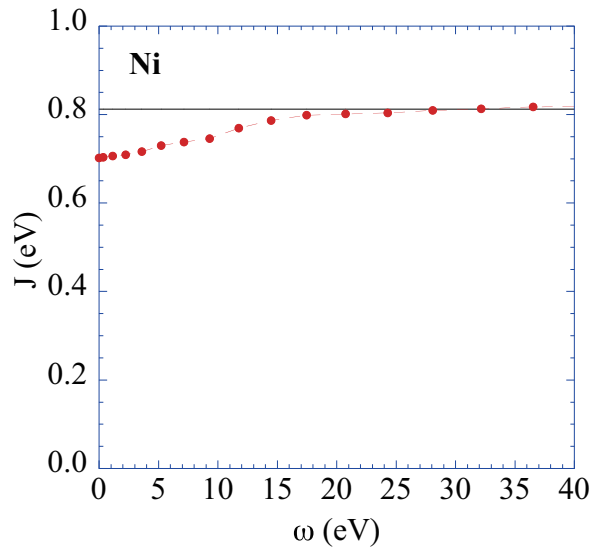


Fig. 12: The exchange integral J as defined in (53) averaged over the $3d$ orbitals.

charge distribution with no $l = 0$ component. In Fig. 13 the static U values for the $3d$ transition metals series are shown (red circles).

3.6.2 Cerium

As a further application, we have calculated the Hubbard U of the isostructural α and γ fcc ceriums, where the former has a smaller unit volume than the latter. As in the case of transition metals, the narrow $4f$ bands for which U is to be calculated, are entangled with the $5d$ and $6s$ bands. The Hubbard U as a function of frequency is surprisingly rich in structure with no less than five prominent peaks with smaller additional features in $\text{Im}(U)$ inducing the Kramers-

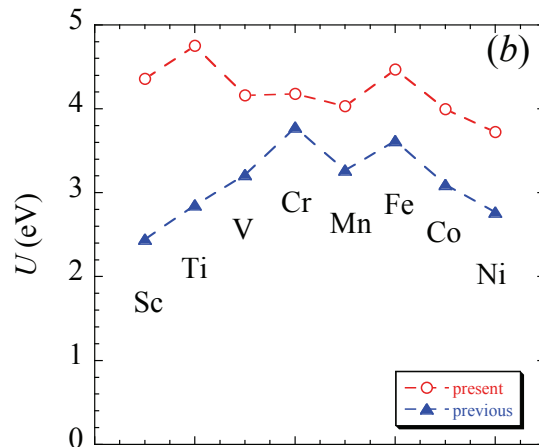


Fig. 13: The static values of the Hubbard U for the 3d transition metal series taken from [25]. The results using the disentanglement method are compared with previous results, where the d subspace was defined according to a combination of band indices and energy window [23]. The significant difference between the two sets of results indicates that the Hubbard U can be sensitive to the choice of the d subspace when the bands are entangled.

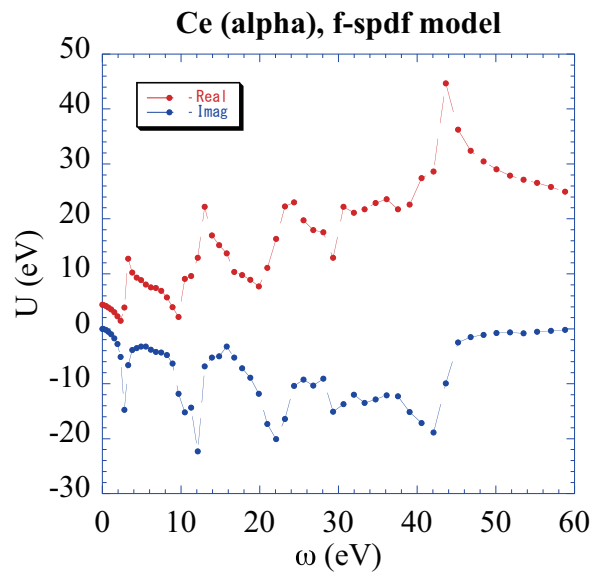


Fig. 14: The real and imaginary parts of the Hubbard U of cerium. The d subspace is taken to be the $4f$ bands extracted using the maximally localised Wannier function method. The calculation is done using the cRPA method for entangled bands as described in the text.

Kronig structures in $\text{Re}(U)$. Unlike the usual case where there is only one prominent plasmon excitation, there appears to be several high-energy sub-plasmon excitations. The structure at low energy around 4 eV indicates that there is a large screening contribution arising from the polarisation between the d and r subspaces at low energy. It suggests that model calculations with a static U may not be sufficient for describing the electronic spectra of cerium even at low energies. Fortunately, very recently a new method to solve the impurity problem within

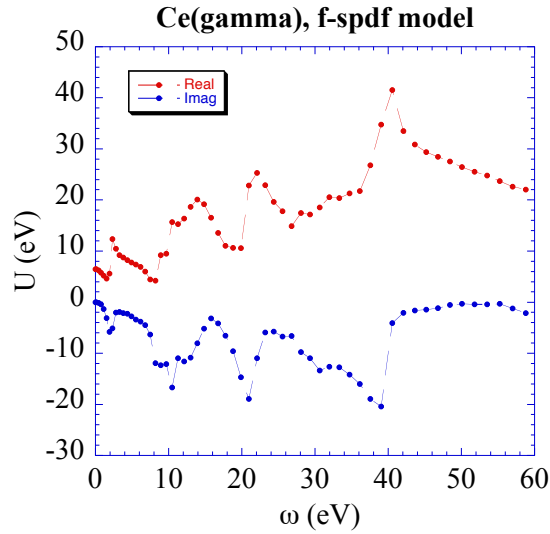


Fig. 15: The same as Fig. 14 but for the γ phase.

the dynamical mean-field theory (DMFT) with a frequency-dependent U has been developed [26, 27].

The subtle difference between the α and γ phases is revealed in U . The peak in $\text{Im}(U)$ at low energy is larger for the α than the γ phase. This is consistent with the fact that the γ phase has a larger unit cell volume so that the $4f$ bands are narrower than in the α phase. This means that $P_r = P - P_d$ for the γ phase contains less low energy transitions between the d and r subspaces because there is less hybridisation between the $4f$ states and other states, compared with the α phase.

3.7 Further examples

The cRPA method has by now been applied to a wide range of materials from simple ones like $3d$ transition metals to complex ones such as the BEDT-TTF organic conductors [17], alkali cluster-loaded soladites [28] and the parent compounds of the recently discovered superconducting iron-based pnictides [29, 30]. In the latter reference, the U and J for a series of pnictides have been systematically calculated. It appears from this study that FeSe is more correlated compared to the other pnictides.

Other applications include calculations of the Hubbard U of MnO as a function of pressure [31]. Recently, the frequency-dependent Hubbard U of the parent cuprate superconductor La_2CuO_4 was calculated [32]. With the development of a new method it is now possible to solve the impurity problem within the DMFT method with a frequency-dependent Hubbard U [26, 27]. This method was very recently applied to study the electronic structure of BaFe_2As_2 , one of the parent compounds of the iron-based superconducting pnictides [33].

Appendices

A Basis functions

The Bloch wavefunctions $\psi_{\mathbf{k}n}(\mathbf{r})$ are expanded in terms of the LAPW basis, i.e., in the interstitial region

$$\psi_{\mathbf{k}n}(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \sum_{\mathbf{G}} c_{\mathbf{k}n,\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} \quad (60)$$

with the unit-cell volume Ω and inside the muffin-tin sphere of atom a

$$\psi_{\mathbf{k}n}(\mathbf{r}) = \sum_{lm} [A_{alm,n}(\mathbf{k})u_{al}(r) + B_{alm,n}(\mathbf{k})\dot{u}_{al}(r)] Y_{lm}(\hat{\mathbf{r}}), \quad (61)$$

where \mathbf{r} is measured from the sphere centre. The coefficients $c_{\mathbf{k}n,\mathbf{G}}$, $A_{alm,n}(\mathbf{k})$, and $B_{alm,n}(\mathbf{k})$ are determined such that the wave functions and their radial derivatives are continuous at the muffin-tin sphere boundaries. The radial functions $u_{al}(r)$ and $\dot{u}_{al}(r)$ are the solution of the radial Schrödinger equation or the scalar-relativistic Dirac equation [34] and its energy derivative, respectively. Evidently for spin-polarised or relativistic systems all the above quantities depend on the spin variable.

The basis functions needed to calculate the response functions and the screened interaction or the Hubbard U are constructed as follows [19, 20]. From (38) it is clear that the space spanned by the polarisation function P is formed by products of orbitals. In terms of the LAPW basis within the muffin-tin spheres these products are

$$\{u_{al}u_{al'}\}, \{u_{al}\dot{u}_{al'}\}, \{\dot{u}_{al}\dot{u}_{al'}\} \otimes Y_{lm}Y_{l'm'}, \quad (62)$$

which form a complete basis for P and R inside the muffin-tin spheres. That the basis is also complete for R may be seen by expanding R in (32) in terms of P :

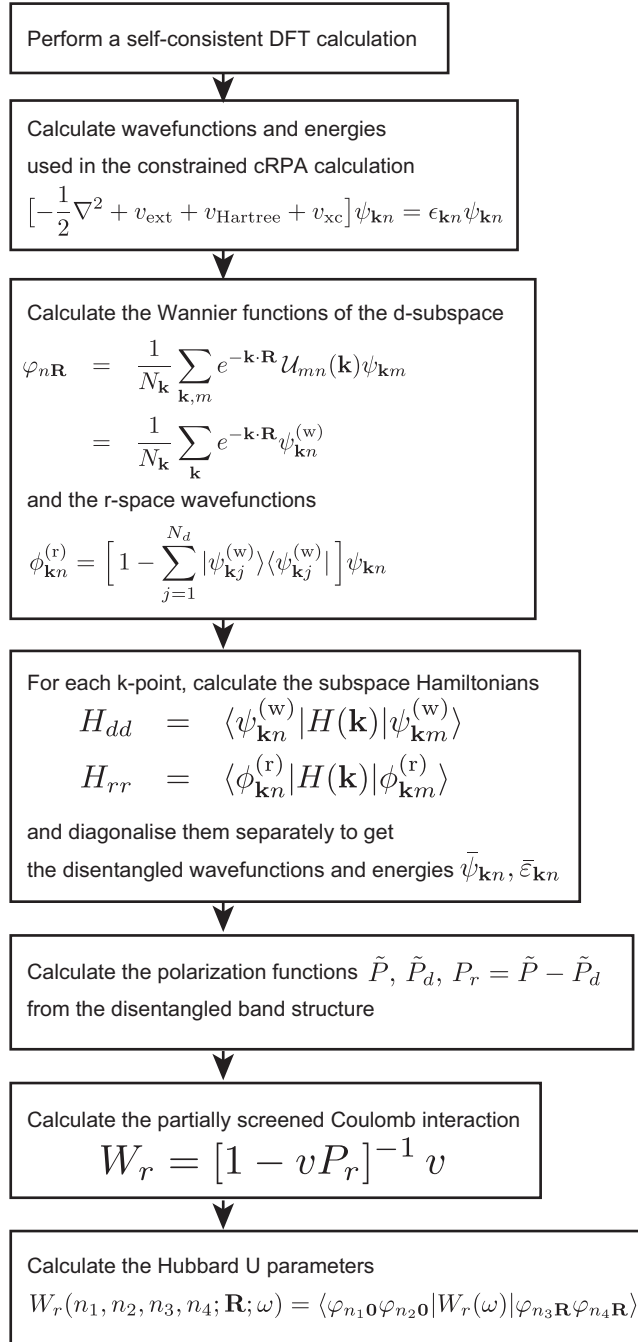
$$R = P + PvP + PvPvP + \dots \quad (63)$$

Since the left and right of R are both projected onto P , the space spanned by P and R are the same. It turns out that the products in (62) can be linearly dependent since they are not orthogonal. To remove this linear dependency and to construct the optimal basis set we follow the procedure in [19, 20]. Calling the orbital products $\{b_\alpha\}$ as in (39) we calculate the overlap matrix

$$O_{\alpha\beta} = \langle b_\alpha | b_\beta \rangle, \quad (64)$$

and diagonalise it. Linear dependency is indicated by zero or very small eigenvalues. The eigenvectors of O form an orthogonal basis and by discarding those eigenvectors with eigenvalues lower than a certain tolerance, set according to desired accuracy, we obtain an optimal basis for the muffin-tin region. The basis for the interstitial part is naturally given by the plane waves, which already form a product basis since any product of two plane waves yields another plane wave.

B Flow chart



References

- [1] A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems* (Dover, New York 2003)
- [2] J. Hubbard, Proc. Roy. Soc. A 276 238 (1963)
- [3] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)
- [4] W.M.C. Foulkes, L. Mitas, R.J. Needs and G. Rajagopal, Rev. Mod. Phys. **73**, 33–83 (2001)
- [5] A. Georges *et al.*, Rev. Mod. Phys. **68**, 13 (1996)
G. Kotliar and D. Vollhardt, *Physics Today*, 53 (March 2004)
- [6] O. Gunnarsson, O.K. Andersen, O. Jepsen, and J. Zaanen, Phys. Rev. B **39**, 1708 (1989)
O. Gunnarsson, Phys. Rev. B **41**, 514 (1990)
V.I. Anisimov and O. Gunnarsson, Phys. Rev. B **43**, 7570 (1991)
- [7] M.S. Hybertsen, M. Schlüter and N.E. Christensen, Phys. Rev. B **39**, 9028 (1989)
- [8] A.K. McMahan, R.M. Martin and S. Satpathy, Phys. Rev. B **38**, 6650 (1988)
- [9] M. Cococcioni and S. de Gironcoli, Phys. Rev. B **71**, 035105 (2005)
- [10] M. Springer and F. Aryasetiawan, Phys. Rev. B **57**, 4364 (1998)
- [11] T. Kotani, J. Phys: Condens. Matter **12**, 2413 (2000)
- [12] N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (Brookes Cole, 1976)
- [13] J. Schwinger, *Proc. Natl. Acad. Sci., USA*, **37**, 452 (1951)
- [14] L. Hedin, Phys. Rev. **139**, A796 (1965); L. Hedin and S. Lundqvist, *Solid State Physics* Vol. 23, eds. H. Ehrenreich, F. Seitz, and D. Turnbull (Academic, New York, 1969)
- [15] F. Aryasetiawan *et al*, Phys. Rev. B **70**, 195104 (2004)
- [16] F. Aryasetiawan, J. Tomczak, T. Miyake, and R. Sakuma, Phys. Rev. Lett. **102**, 176402 (2009)
- [17] K. Nakamura, Y. Yoshimoto, T. Kosugi, R. Arita, and M. Imada, J. Phys. Soc. Jap. **78**, 083710 (2009)
- [18] O.K. Andersen, Phys. Rev. B **12**, 3060 (1975)
- [19] F. Aryasetiawan and O. Gunnarsson, Phys. Rev. B **49**, 16214 (1994)
- [20] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998)

- [21] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997)
I. Souza, N. Marzari and D. Vanderbilt, Phys. Rev. B **65**, 035109 (2001)
- [22] T. Miyake and F. Aryasetiawan, Phys. Rev. B **77**, 085122 (2008)
- [23] F. Aryasetiawan, K. Karlsson, O. Jepsen, and U. Schönberger, Phys. Rev. B **74**, 125106 (2006)
- [24] O.K. Andersen and T. Saha-Dasgupta, Phys. Rev. B **62**, R16219 (2000)
F. Lechermann *et al.*, Phys. Rev. B **74**, 125120 (2006)
- [25] T. Miyake, F. Aryasetiawan, and M. Imada, Phys. Rev. B **80**, 155134 (2009)
- [26] P. Werner and A.J. Millis, Phys. Rev. Lett. **99**, 146404 (2007)
- [27] P. Werner and A.J. Millis, Phys. Rev. Lett. **104**, 146401 (2010)
- [28] K. Nakamura, T. Koretsune, and R. Arita, Phys. Rev. B **80**, 174420 (2009)
- [29] K. Nakamura, R. Arita and M. Imada, J. Phys. Soc. Jap. **77**, 093711 (2008)
- [30] T. Miyake, K. Nakamura, R. Arita, and M. Imada, J. Phys. Soc. Jap. **79**, 044705 (2010)
- [31] J. Tomczak, T. Miyake, and F. Aryasetiawan, Phys. Rev. B **81**, 115116 (2010)
- [32] A.V. Kozhevnikov, A.G. Eguiluz, and T.C. Schultess,
<http://www.computer.org/portal/web/sc10>
- [33] P. Werner, M. Casula, T. Miyake, F. Aryasetiawan, A.J. Millis, and S. Biermann,
<http://arxiv.org/abs/1107.3128>
- [34] D.D. Koelling and B.N. Harmon, J. Phys. C **10**, 3107 (1977)

8 The Lanczos Method

Erik Koch

Computational Materials Science

German Research School for Simulation Sciences

Contents

1	Lanczos Method	3
1.1	Krylov space	5
1.2	Spectral functions	8
2	Application to the Hubbard model	11
2.1	Representation of basis and Hamiltonian	13
2.2	Green functions	14
2.3	Parallelization strategies	16
3	Application to DMFT	18
3.1	Cluster methods	21
3.2	Anderson impurity model	21
3.3	Hybridization sum-rules	23
3.4	Symmetries	26
4	Conclusions	29

The Lanczos iteration [1] was conceived as a method for tridiagonalizing Hermitian matrices. Like the related Arnoldi method [2] for non-Hermitian matrices, it initially received widespread attention. Iterative approaches were, however, soon eclipsed by direct methods (Householder transformations and QR factorization), that are better suited for solving the eigenvalue problem for general matrices. Actually, the Lanczos method is particularly suited for the determination of extreme eigenvalues and v -vectors. Therefore, it was rediscovered in the 1970s [3], when computers had become sufficiently powerful to treat matrices large enough for the Lanczos algorithm to outperform general methods, nicely illustrating the Fundamental Law of Computer Science: the faster the computer, the greater the importance of the speed of algorithms [4]. By now iterative methods are an integral part of the numerical linear algebra curriculum [4–6].

For finding eigenvalues of a matrix H of dimension N , the Lanczos method requires the evaluation of matrix-vector products $H \cdot v$ as the only problem-specific step. This matrix-vector product can be calculated particularly efficiently when the matrix H is sparse, i.e., when the number of non-zero matrix elements per row does not scale with the matrix dimension. Storing such a matrix takes only $\mathcal{O}(N)$ memory and $H \cdot v$ can be evaluated in $\mathcal{O}(N)$ time. Calculating the extremal eigenvalues requires $\mathcal{O}(1)$ iterations, i.e., overall $\mathcal{O}(N)$ time. For comparison, a direct diagonalization takes $\mathcal{O}(N^2)$ for storing the matrix and $\mathcal{O}(N^3)$ time to diagonalize. Besides their favorable scaling for sparse matrix problems, iterative methods have the advantage that they systematically approach the desired result. Typically the iteration converges geometrically and can be stopped as soon as the desired accuracy is reached. In contrast, direct methods appear to make no progress towards the solution until all $\mathcal{O}(N^3)$ operations are completed and the full result is obtained.

Since the Lanczos method is particularly suited for dealing with large sparse Hamiltonians, it is the method of choice for systems with short-range interactions. For band-structure calculations in a linear combination of atomic orbitals (LCAO) or tight-binding (TB) basis, it is known as the recursion method [7]. The basic idea here is to switch from the Bloch picture of a perfectly periodic solid to a local picture, replacing the solution of the Schrödinger equation in terms of Bloch waves by the calculation of the local density of states. The crucial technical point is to calculate the density of states not via a spectral representation (in terms of Bloch waves), but by repeated application of the Hamiltonian H to a localized single-electron state. With each application of H the electron explores more and more sites. Thus, if the hopping matrix elements beyond a certain distance are zero, such calculations can be performed without having to restrict the system to finite size.

For many-body models like quantum-spin- or Hubbard-models [8] this is unfortunately not possible. They have to be defined on a finite cluster, giving rise to a finite-dimensional Hamiltonian matrix. Since the size of the Hilbert space grows exponentially with system-size, actual calculations are restricted by the available computer memory. In a typical simulation, first the ground-state is calculated by a Lanczos iteration. Building on this, spectral functions are calculated in a similar way as in the recursion method. The great advantage of this approach is that it gives the dynamical properties of the ground state ($T=0$) directly on the real axis. The price is the restriction to (small) finite-size systems.

1 Lanczos Method

We can find the ground-state $|\Psi_0\rangle$ and its energy E_0 for a Hamiltonian H from the variational principle. The wavefunction-functional

$$E[\Psi] = \frac{\langle \Psi | H | \Psi \rangle}{\langle \Psi | \Psi \rangle} \quad (1)$$

is minimized for $\Psi = \Psi_0$, with $E[\Psi_0] = E_0$. The functional gradient

$$\frac{\delta E[\Psi]}{\delta \langle \Psi |} = \frac{H|\Psi\rangle - E[\Psi]|\Psi\rangle}{\langle \Psi | \Psi \rangle} = |\Psi_a\rangle \quad (2)$$

gives the direction of steepest-ascent of the functional from the point $|\Psi\rangle$. Moving in the opposite direction will thus result in a wavefunction with lower energy expectation value: $E[\Psi - \alpha\Psi_a] < E[\Psi]$ for small, positive α .

To find the optimum value of α , we minimize $E[\Psi - \alpha\Psi_a]$. For this, it is convenient to introduce an orthogonal basis in the space spanned by the two vectors $|\Psi\rangle$ and $|\Psi_a\rangle$. From (2) we see that $\text{span}(|\Psi\rangle, |\Psi_a\rangle) = \text{span}(|\Psi\rangle, H|\Psi\rangle)$. As first basis vector, we normalize $|\Psi\rangle$

$$|v_0\rangle = |\Psi\rangle / \sqrt{\langle \Psi | \Psi \rangle},$$

for the second vector we orthogonalize $H|v_0\rangle$ to $|v_0\rangle$

$$|\tilde{v}_1\rangle = H|v_0\rangle - |v_0\rangle\langle v_0 | H | v_0 \rangle \quad (3)$$

and normalize to obtain $|v_1\rangle$. With $a_n = \langle v_n | H | v_n \rangle$ and $b_1^2 = \langle \tilde{v}_1 | \tilde{v}_1 \rangle$ we thus have

$$H|v_0\rangle = b_1|v_1\rangle + a_0|v_0\rangle \quad (4)$$

from which we see that $\langle v_1 | H | v_0 \rangle = b_1$.

We can then write any normalized wavefunction in $\text{span}(|\Psi\rangle, H|\Psi\rangle) = \text{span}(|v_0\rangle, |v_1\rangle)$ as

$$|v\rangle = \cos(\theta)|v_0\rangle + \sin(\theta)|v_1\rangle. \quad (5)$$

Minimizing the expectation value

$$\langle v | H | v \rangle = a_0 \cos^2(\theta) + 2b_1 \sin(\theta) \cos(\theta) + a_1 \sin^2(\theta), \quad (6)$$

with respect to θ , we obtain, dividing by $\cos^2(\theta)$, the quadratic equation

$$b_1 \tan^2(\theta) + (a_0 - a_1) \tan(\theta) - b_1 = 0. \quad (7)$$

Solving for θ we find the lowest-energy state on the subspace spanned by $|v_0\rangle$ and $H|v_0\rangle$. Alternatively, we can diagonalize the Hamiltonian matrix on the two-dimensional subspace, which in the basis $|v_0\rangle, |v_1\rangle$ is given by

$$H_{\text{span}(|\Psi\rangle, H|\Psi\rangle)} = \begin{pmatrix} a_0 & b_1 \\ b_1 & a_1 \end{pmatrix}. \quad (8)$$

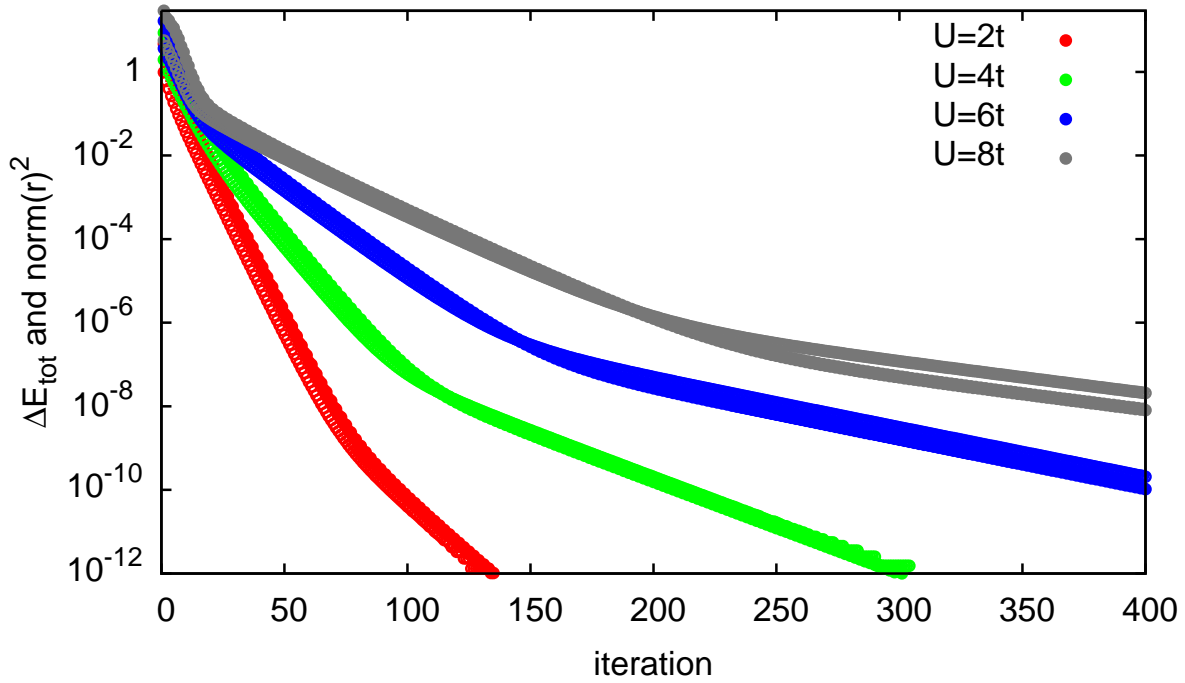


Fig. 1: Convergence of the residual (filled circles) and the lowest eigenvalue (open circles) for a steepest-descent minimization of a Hubbard-chain of 10 sites at half-filling, starting from a random initial vector.

Naturally, we can use the variational state of lowest energy

$$|\Psi^{(2)}\rangle = \cos(\theta_{\min})|v_0\rangle + \sin(\theta_{\min})|v_1\rangle \quad (9)$$

as the starting point for another steepest-descent minimization. Doing this repeatedly, we obtain a series of vectors with decreasing energy expectation value, which rapidly converge to a minimum. For a generic functional, this would usually be a local, not the global minimum, which makes the optimization of high-dimensional functions a hard problem. The energy functional (1), however, has only minima for the ground-states, all other stationary points are saddle points. We can thus expect rapid convergence to the ground state, examples given in figure 1, except in the case where the the gradient (2) vanishes, i.e., if $|\Psi\rangle$ happens to be an eigenfunction of H .

For checking convergence of this steepest-descent method, introduced by Kantorovich [9] and, independently, by Hestenes and Karush [10], we can monitor the change in the energy expectation value or determine when the residual

$$r[\Psi] = \|(H - E[\Psi])|\Psi\rangle\|^2 = \langle\Psi|H^2|\Psi\rangle - E[\Psi]^2, \quad (10)$$

which measures the quality of the eigenstate, becomes sufficiently small.

1.1 Krylov space

If we apply the method of steepest-descent L times, starting from a vector $|v_0\rangle$, the resulting vector will lie in $\mathcal{K}^L(|v_0\rangle) = \text{span}(|v_0\rangle, H|v_0\rangle, H^2|v_0\rangle, \dots, H^L|v_0\rangle)$, the $L + 1$ -dimensional *Krylov space* [11] of H over $|v_0\rangle$. Instead of repeatedly minimizing the energy in two-dimensional subspaces, we could directly find the state of lowest energy in $\mathcal{K}^L(|v_0\rangle)$. Having more degrees of freedom for the minimization will lead to even faster convergence.

To implement this idea, we construct an orthonormal basis $|v_n\rangle$ of the Krylov space. We start with the normalized vector $|v_0\rangle$. The second basis vector $|v_1\rangle$ is constructed as in the steepest-descent method (3):

$$b_1|v_1\rangle = |\tilde{v}_1\rangle = H|v_0\rangle - a_0|v_0\rangle. \quad (11)$$

The next basis vector is likewise constructed as $H|v_n\rangle$ orthogonalized to all previous vectors, and normalized

$$b_2|v_2\rangle = |\tilde{v}_2\rangle = H|v_1\rangle - \sum_{i=0}^1 |v_i\rangle\langle v_i|H|v_1\rangle = H|v_1\rangle - a_1|v_1\rangle - b_1|v_0\rangle. \quad (12)$$

where $a_n = \langle v_n|H|v_n\rangle$ and $b_n^2 = \langle \tilde{v}_n|\tilde{v}_n\rangle$. The fourth basis vector is

$$b_3|v_3\rangle = |\tilde{v}_3\rangle = H|v_2\rangle - \sum_{i=0}^2 |v_i\rangle\langle v_i|H|v_2\rangle = H|v_2\rangle - a_2|v_2\rangle - b_2|v_1\rangle. \quad (13)$$

Here the last term in the orthogonalization vanishes, because (11) together with the orthogonality of the basis vectors for $n = 0 \dots 2$ implies $\langle v_2|H|v_0\rangle = 0$.

The construction of the further basis vectors follows the same scheme

$$b_{n+1}|v_{n+1}\rangle = |\tilde{v}_{n+1}\rangle = H|v_n\rangle - \sum_{i=0}^n |v_i\rangle\langle v_i|H|v_n\rangle = H|v_n\rangle - a_n|v_n\rangle - b_n|v_{n-1}\rangle$$

with $a_n = \langle v_n|H|v_n\rangle$ and $b_n^2 = \langle \tilde{v}_n|\tilde{v}_n\rangle$. Rearranging shows that H is tridiagonalized

$$H|v_n\rangle = b_n|v_{n-1}\rangle + a_n|v_n\rangle + b_{n+1}|v_{n+1}\rangle$$

which in turn implies that $H|v_i\rangle$ is orthogonal to all basis states, except $|v_i\rangle$ and $|v_{i\pm 1}\rangle$. This tridiagonalization of H is the essence of the *Lanczos method* [1].

After L steps the Hamiltonian on the $L + 1$ -dimensional Krylov space is given by

$$H_{\mathcal{K}^L(|v_0\rangle)} = \begin{pmatrix} a_0 & b_1 & 0 & 0 & \dots & 0 & 0 \\ b_1 & a_1 & b_2 & 0 & \dots & 0 & 0 \\ 0 & b_2 & a_2 & b_3 & & 0 & 0 \\ 0 & 0 & b_3 & a_3 & & 0 & 0 \\ & \vdots & & & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & & a_{L-1} & b_L \\ 0 & 0 & 0 & 0 & \dots & b_L & a_L \end{pmatrix} \quad (14)$$

v=init	
b0=norm2(v)	not part of tridiagonal matrix
scal(1/b0,v)	v= v ₀ ⟩
w=0	
w=w+H*v	w= H v ₀ ⟩
a[0]=dot(v,w)	
axpy(-a[0],v,w)	w= ṽ ₁ ⟩ = H v ₀ ⟩ - a ₀ v ₀ ⟩
b[1]=norm2(w)	
for n=1,2,...	
if abs(b[n])<eps then exit	invariant subspace
scal(1/b[n],w)	w= v _n ⟩
scal(-b[n],v)	v= -b _n v _{n-1} ⟩
swap(v,w)	
w=w+H*v	w= H v _n ⟩ - b _n v _{n-1} ⟩
a[n]=dot(v,w)	a[n] = ⟨v _n H v _n ⟩ - b _n ⟨v _n v _{n-1} ⟩
axpy(-a[n],v,w)	w= ṽ _{n+1} ⟩
b[n+1]=norm2(w)	
diag(a[0]..a[n], b[1]..b[n])	getting a _{n+1} needs another H v⟩
if converged then exit	
end	

Table 1: *The implementation of the Lanczos iteration requires only two N -dimensional vectors for tridiagonalizing H and thus for calculating the ground-state energy. Constructing the Lanczos-approximation of the ground-state vector requires a second iteration and one additional N -dimensional vector. The by far most expensive operation is the matrix-vector product.*

If we do not normalize the basis vectors, we obtain an iteration of the form

$$|\Phi_{n+1}\rangle = H|\Phi_n\rangle - \frac{\langle\Phi_n|H|\Phi_n\rangle}{\langle\Phi_n|\Phi_n\rangle}|\Phi_n\rangle - \frac{\langle\Phi_n|\Phi_n\rangle}{\langle\Phi_{n-1}|\Phi_{n-1}\rangle}|\Phi_{n-1}\rangle \quad (15)$$

where $|\Phi_n\rangle = \prod_{i=1}^n b_i |v_n\rangle$ in terms of which we have

$$a_n = \frac{\langle\Phi_n|H|\Phi_n\rangle}{\langle\Phi_n|\Phi_n\rangle}, \quad b_n^2 = \frac{\langle\Phi_n|\Phi_n\rangle}{\langle\Phi_{n-1}|\Phi_{n-1}\rangle}. \quad (16)$$

In this unnormalized basis the Hamiltonian appears non-Hermitian

$$H|\Phi_n\rangle = b_n^2|\Phi_{n-1}\rangle + a_n|\Phi_n\rangle + |\Phi_{n+1}\rangle, \quad (17)$$

but it actually is

$$\langle\Phi_{n+1}|H|\Phi_n\rangle = \langle\Phi_{n+1}|\Phi_{n+1}\rangle = b_{n+1}^2\langle\Phi_n|\Phi_n\rangle = \langle\Phi_n|H|\Phi_{n+1}\rangle. \quad (18)$$

The numerical implementation only requires keeping two N -dimensional vectors in memory. It is shown in table 1.

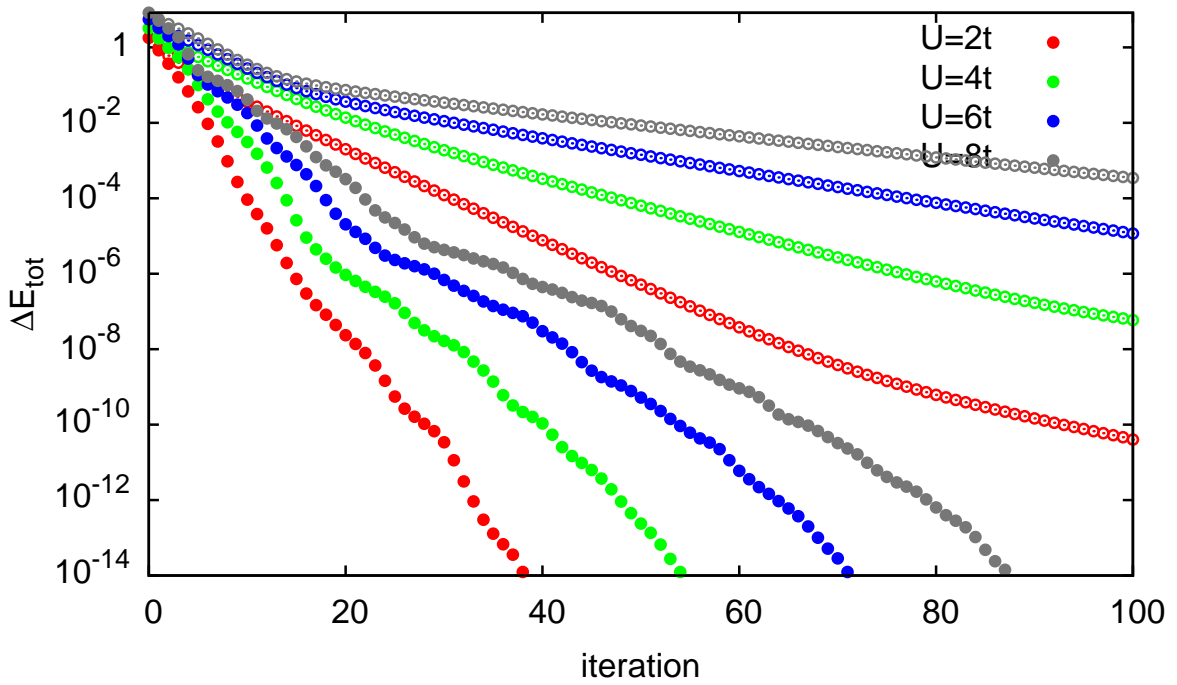


Fig. 2: *Covergence of the lowest eigenvalue for a Lanczos iteration (full circles) compared to steepest-descent minimization (open circles) of a 10-site Hubbard-chain at half-filling, starting from a random initial vector. Due to the additional variational degrees of freedom, Lanczos converges significantly faster. Overall, convergence for the half-filled system gets harder for larger U , as the distance to the lowest excited states is reduced ($\sim t^2/U$) and the spectrum widens ($\sim U$). In all cases, convergence is reached after less than $L \approx 100$ Lanczos iterations, to be compared to the dimension $N=63\,504$ of the Hilbert space.*

Diagonalizing (14), after a few tens to hundred iterations, the lowest eigenvalue of the tridiagonal representation of H on the Krylov space gives an excellent approximation to the ground-state energy of H in the full Hilbert space (Fig. 2). A formal estimate of the convergence was given by Kaniel and Paige [5]. For an $N+1$ -dimensional, symmetric matrix H with eigenvalues E_n , the lowest eigenvalue \check{E}_0 of the tridiagonal representation of H on the $(L+1)$ -dimensional Krylov space over $|v_0\rangle$ fulfills

$$\frac{\check{E}_0 - E_0}{E_N - E_0} \leq \left(\frac{\tan(\arccos(\langle \check{\Psi}_0 | \Psi_0 \rangle))}{T_L \left(1 + 2 \frac{E_1 - E_0}{E_N - E_1} \right)} \right)^2 \quad (19)$$

where $T_L(x)$ is the Chebyshev polynomial of order L and $\langle \check{\Psi}_0 | \Psi_0 \rangle$ the overlap of the Lanczos approximation to the ground-state $\check{\Psi}_0$ with the ground-state of H . Thus, if the initial state $|v_0\rangle$ is not orthogonal to the non-degenerate ground-state, convergence is exponential with a rate roughly increasing with the square root of the gap to the first excited measured in units of the width of the spectrum.

The approximate ground-state vector is given by the linear combination

$$|\check{\Psi}_0\rangle = \sum_{n=0}^L \check{\psi}_{0,n} |v_n\rangle, \quad (20)$$

where $\check{\psi}_0$ is the ground-state vector of the $L + 1$ -dimensional tridiagonal matrix (14). Instead of storing all $L + 1$ basis vectors $|v_n\rangle$, we can restart the Lanczos iteration from the same $|v_0\rangle$, accumulating the sum (20) iteration by iteration. This only requires keeping one additional N -dimensional vector in memory.

So far we have tacitly assumed that the Krylov vectors $H^n|v_0\rangle$ are linearly independent. If not, there will be a vector $H|\check{v}_m\rangle$ that vanishes when orthogonalized to the previous states, i.e., $b_n = 0$. This means that the Krylov space $\text{span}(|v_0\rangle, |v_1\rangle, \dots, |v_m\rangle)$ is invariant under H , i.e., we have found an exact eigenspace of H . For a large matrix H it is quite unlikely to be that lucky. Still, as the Lanczos iteration approaches the ground-state, we encounter a similar situation: Close to an eigenstate, the functional (1) becomes almost stationary, i.e., the coefficients b_n almost vanish. Normalization of the very short vector $|\check{v}_n\rangle$ then amplifies numerical noise in the small vector. This makes the numerical $|v_n\rangle$, which in theory should automatically be orthogonal to all $|v_m\rangle$ with $m < n - 2$, actually have finite overlaps with these vectors. This loss of orthogonality manifests itself in the appearance of multiple copies of eigenvectors (ghost states) which are unrelated to the actual multiplicities of the eigenvalues. This is the problem, which makes the Lanczos method unpractical for tridiagonalizing dense matrices. For the ground-state the variational principle prevents severe problem from the loss of orthogonality. An example of the appearance of ghost states is shown in figure 3.

If we want to reliably obtain excited states, we need to explicitly orthogonalize to the previous basis states. This leads to the Lanczos method with complete reorthogonalization [5]. A similar orthogonalization is performed in the Arnoldi method [2], which, however, is devised for unsymmetric matrices.

1.2 Spectral functions

Given the orthogonality problems of the Lanczos method, it appears hopeless to use it to obtain matrix elements of the resolvent, as they contain information about the full spectrum $H|\Psi_n\rangle = E_n|\Psi_n\rangle$. Still we are tempted to approximate the Lehmann representation

$$G_c(z) = \left\langle \Psi_c \left| \frac{1}{z - H} \right| \Psi_c \right\rangle = \sum_{n=0}^N \frac{\langle \Psi_c | \Psi_n \rangle \langle \Psi_n | \Psi_c \rangle}{z - E_n} \quad (21)$$

in terms of the eigenstates on the Krylov space $\mathcal{K}^L(|\Psi_c\rangle)$

$$\check{G}_c(z) = \left\langle \Psi_c \left| \frac{1}{z - \check{H}_c} \right| \Psi_c \right\rangle = \sum_{n=0}^L \frac{\langle \Psi_c | \check{\Psi}_n \rangle \langle \check{\Psi}_n | \Psi_c \rangle}{z - \check{E}_n}. \quad (22)$$

This is straightforward to calculate: We run L Lanczos iterations, starting from the (normalized) vector $|\Psi_c\rangle$, to create the tridiagonal \check{H}_c . The matrix element of the resolvent is the top left

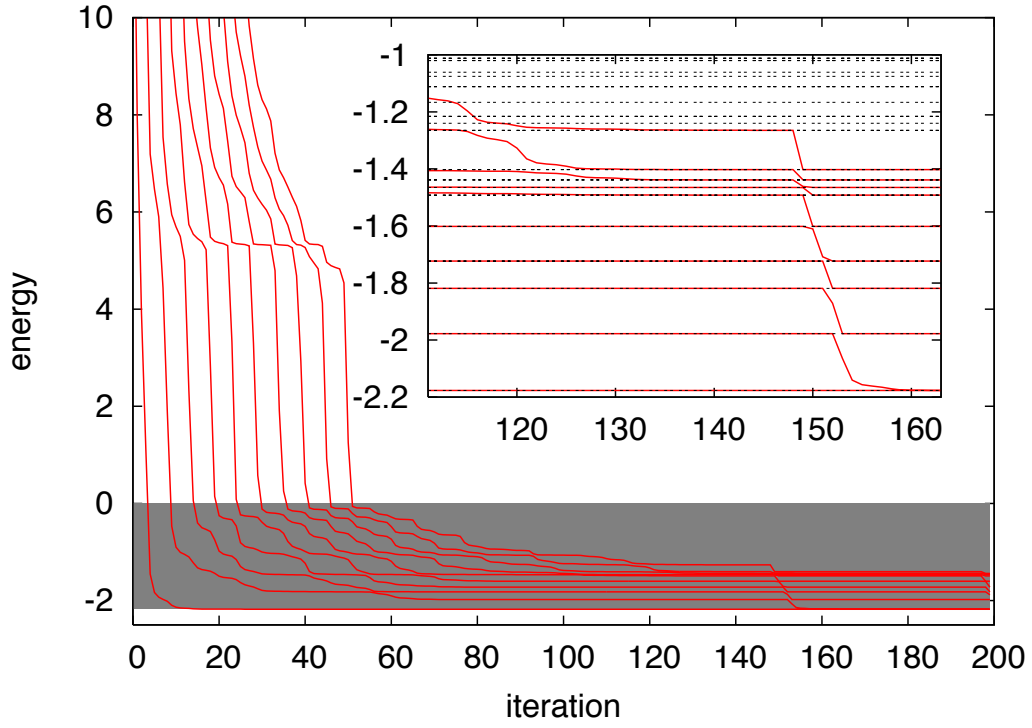


Fig. 3: Appearance of ghost states in an overconverged Lanczos iteration. The ground-state energy for a half-filled 8-site Hubbard-chain with $U = 10t$ is converged to numerical accuracy (10^{-16}) after about 85 iterations. Forcing the Lanczos iteration to continue, we see that at first also the higher excited states converge to the exact eigenvalues (dashed lines). But, as shown in the inset, they eventually start collapsing to the ground state. The appearance of these ghost states is due to the orthogonality problem introduced by small normalization parameters b_n , when the iteration is very close to a stationary point.

matrix element of the inverse of

$$z - \check{H}_c = \begin{pmatrix} z - a_0 & -b_1 & 0 & 0 & \cdots & 0 & 0 \\ -b_1 & z - a_1 & -b_2 & 0 & \cdots & 0 & 0 \\ 0 & -b_2 & z - a_2 & -b_3 & \cdots & 0 & 0 \\ 0 & 0 & -b_3 & z - a_3 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & z - a_{L-1} & -b_L \\ 0 & 0 & 0 & 0 & \cdots & -b_L & z - a_L \end{pmatrix}. \quad (23)$$

This is easily determined, partitioning the matrix as indicated

$$z - \check{H}_c = \begin{pmatrix} z - a_0 & B^{(1)T} \\ B^{(1)} & z - \check{H}_c^{(1)} \end{pmatrix} \quad (24)$$

and inverting the block-matrix, giving

$$[(z - \check{H}_c)^{-1}]_{00} = \left(z - a_0 - B^{(1)T} (z - \check{H}_c^{(1)})^{-1} B^{(1)} \right)^{-1} = \left(z - a_0 - b_1^2 [(z - \check{H}_c^{(1)})^{-1}]_{00} \right)^{-1}.$$

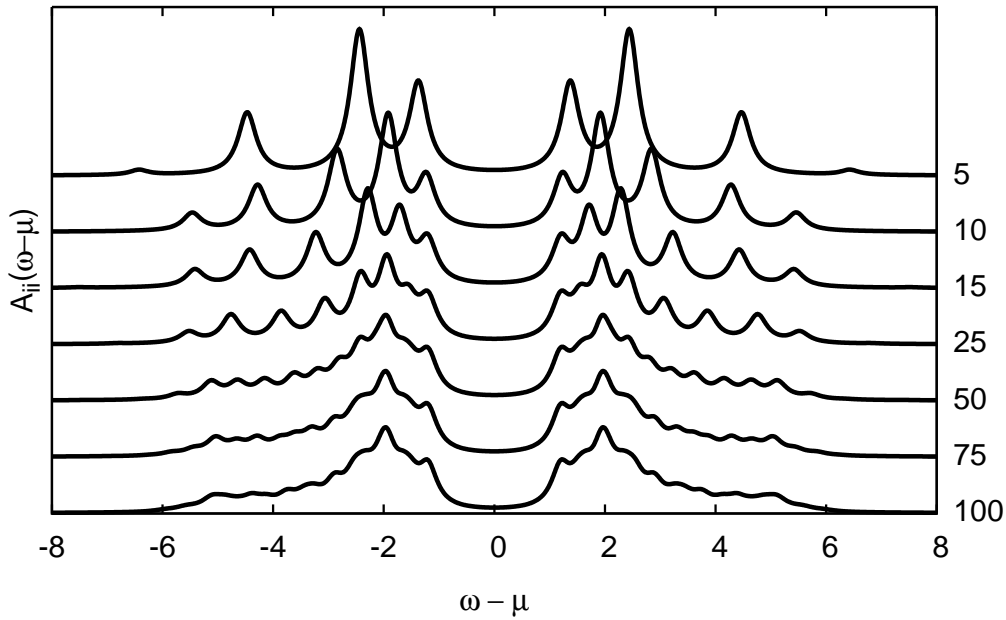


Fig. 4: Convergence of the spectral function with increasing number of Lanczos steps, $L=5, 10, 15, 25, 50, 75,$ and 100 , for a 14-site Hubbard chain with $U = 5t$ at half filling. With increasing L , more and more moments of the photoemission and inverse photoemission part of the spectrum are reproduced correctly.

Repeating inversion by partitioning for the submatrices $\check{H}^{(n)}$ we obtain the continued fraction

$$\check{G}_c(z) = [(z - \check{H}_c)^{-1}]_{00} = \frac{1}{z - a_0 - \frac{b_1^2}{z - a_1 - \frac{b_2^2}{z - a_2 - \dots}}}, \quad (25)$$

which terminates with $-b_L^2/(z - a_L)$. The spectral representation (22) is obtained by diagonalizing the Lanczos matrix \check{H}_c giving us the $L + 1$ eigenvalues \check{E}_n and eigenvectors $\check{\psi}_n$. Since

$$|\check{\Psi}_n\rangle = \sum_{l=0}^L \check{\psi}_{n,l} |v_l\rangle \quad (26)$$

the matrix elements are given by $\langle \check{\Psi}_n | \Psi_c \rangle = \check{\psi}_{n,0}$. Thus

$$\check{G}_c(z) = \sum_{n=0}^L \frac{|\check{\psi}_{n,0}|^2}{z - \check{E}_n} \quad (27)$$

The spectral function

$$\check{A}(\omega \pm i\eta) = \mp \frac{1}{\pi} \Im \check{G}(\omega \pm i\eta) \quad (28)$$

obtained this way converges very quickly. An example is shown in figure 4.

To understand how the $L + 1$ eigenstates of \check{H} can represent the full spectrum so well, we consider the moments of the spectral function

$$\int_{-\infty}^{\infty} d\omega \omega^m \check{A}(\omega) = \sum_{n=0}^L |\check{\psi}_{n,0}|^2 \check{E}_n^m = \sum_{n=0}^L \langle \check{\Psi}_c | \check{\Psi}_n \rangle \langle \check{\Psi}_n | \check{\Psi}_c \rangle \check{E}_n^m = \langle \check{\Psi}_c | \check{H}^m | \check{\Psi}_c \rangle \quad (29)$$

Since \check{H} is the projection of H onto the Krylov space $\mathcal{K}^L(|\check{\Psi}_c\rangle)$, we have $\check{H}^m |\check{\Psi}_c\rangle = H^m |\check{\Psi}_c\rangle$ for $m \leq L$. Thus the Lanczos representation $\check{A}(z)$ correctly reproduces the first $2L + 1$ moments of the spectral function $A(z)$. A further Lanczos step adds one new level to the continued fraction (25), leaving all previous terms unchanged. $b_m^2 = 0$ then implies that the continued fraction terminates, and all moments are given correctly. A near vanishing $b_m^2 \approx 0$, which gives rise to the loss of orthogonality of the Lanczos vectors, for the spectral function merely means that further terms in the continued fraction hardly contribute any more.

So far we have only considered diagonal elements of the resolvent. Off-diagonal matrix elements

$$G_{c_1, c_2}(z) = \left\langle \check{\Psi}_{c_2} \left| \frac{1}{z - H} \right| \check{\Psi}_{c_1} \right\rangle \quad (30)$$

are easily obtained by considering the diagonal elements for the linear combinations

$$\left\langle \check{\Psi}_{c_1} \pm \check{\Psi}_{c_2} \left| \frac{1}{z - H} \right| \check{\Psi}_{c_1} \pm \check{\Psi}_{c_2} \right\rangle = G_{c_1, c_1}(z) \pm G_{c_1, c_2}(z) \pm G_{c_2, c_1}(z) + G_{c_2, c_2}(z). \quad (31)$$

2 Application to the Hubbard model

The Hubbard model

$$H = -t \sum_{\langle i, j \rangle \sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum n_{i\uparrow} n_{i\downarrow} \quad (32)$$

describes the fundamental dichotomy between itinerancy and locality for correlated electrons on a lattice: the hopping tends to delocalize electrons and is diagonal in k -space. This makes it possible to solve the band-structure problem for the infinite solid. In k -space the single electron Hamiltonian is block-diagonal. For the one-band Hubbard model each block is just the band energy ε_k . In general, each block defines the band structure problem for one k -point. Including electron-electron repulsion destroys this symmetry. The two-body Coulomb term is diagonal in real space, while in k -space it is dense

$$H = \sum_{k\sigma} \varepsilon_k c_{k\sigma}^\dagger c_{k\sigma} + \frac{U}{M} \sum_{k, k', q} c_{k\uparrow}^\dagger c_{k-q, \uparrow} c_{k'\downarrow}^\dagger c_{k'+q, \downarrow}. \quad (33)$$

This has two important consequences:

1. Since we know no general approach to transform the full Hamiltonian into finite-dimensional blocks, we have to restrict ourselves to finite-dimensional systems. For a cluster

M	N_{\uparrow}	N_{\downarrow}	dim of Hilbert space	memory
2	1	1	4	
4	2	2	36	
6	3	3	400	
8	4	4	4 900	
10	5	5	63 504	
12	6	6	853 776	6 MB
14	7	7	11 778 624	89 MB
16	8	8	165 636 900	1263 MB
18	9	9	2 363 904 400	17 GB
20	10	10	34 134 779 536	254 GB
22	11	11	497 634 306 624	3708 GB
24	12	12	7 312 459 672 336	53 TB
20	1	1	400	
20	2	2	36 100	
20	3	3	1 299 600	9 MB
20	4	4	23 474 025	179 MB
20	5	5	240 374 016	1833 MB
20	6	6	1 502 337 600	11 GB
20	7	7	6 009 350 400	44 GB
20	8	8	15 868 440 900	118 GB
20	9	9	28 210 561 600	210 GB
20	10	10	34 134 779 536	254 GB

Table 2: Dimension of Hilbert space $\dim(\mathcal{H})$ and computer memory required for storing a single many-body wave-function for Hubbard models with M orbitals and $N_{\uparrow} + N_{\downarrow}$ electrons. The first group of numbers gives the dimensions for half-filling, where the Hilbert space is largest. The second group shows how the dimension grows with the filling (dimensions are symmetric about half-filling). Note that the fourth column resembles a semi-logarithmic plot of $\dim(\mathcal{H})$ as a function of system size or filling.

of M sites with N_{\uparrow} electrons with spin up and N_{\downarrow} with spin down, the dimension of the Hilbert space is

$$\dim(\mathcal{H}) = \dim(\mathcal{H}_{\uparrow}) \times \dim(\mathcal{H}_{\downarrow}) = \binom{M}{N_{\uparrow}} \times \binom{M}{N_{\downarrow}}. \quad (34)$$

Examples for the single-band Hubbard model, illustrating the enormous growth of the Hilbert space are given in table 2. Actual calculations are therefore limited to quite small systems.

2. For a tight-binding system where hopping matrix-elements are restricted to close neighbors the many-body Hamiltonian is a sparse matrix when expressed in a real-space basis of localized orbitals (32). The basis states are then configurations $|\{n_{i\sigma}\}\rangle = \prod (c_{i\sigma}^{\dagger})^{n_{i\sigma}} |0\rangle$, characterized by their occupation numbers $\{n_{i\sigma}\}$.

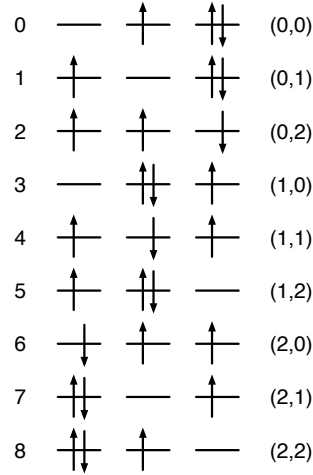


Fig. 5: Basis configurations for a three site system with two up and one down spin electron. The left label denotes the index of the configuration. Equivalently, a state is also unambiguously pointed to by a tuple of up- and down-configuration index.

2.1 Representation of basis and Hamiltonian

Since the many-body basis states

$$|\{n_{i\sigma}\}\rangle = \prod_{i\sigma} \left(c_{i\sigma}^\dagger \right)^{n_{i\sigma}} |0\rangle \quad (35)$$

can be represented by Fermionic occupation numbers, it is natural to encode them in a string of bits. For a Hamiltonian like (32) that conserves spin, we can write

$$|\{n_{i\sigma}\}\rangle = \prod_{i=0}^{L-1} \left(c_{i\downarrow}^\dagger \right)^{n_{i\downarrow}} \left(c_{i\uparrow}^\dagger \right)^{n_{i\uparrow}} |0\rangle \quad (36)$$

with $\sum n_{i\sigma} = N_\sigma$, and encode each spin-component as the integer $m_\sigma = \sum n_{i\sigma} 2^i$. Enumerating all basis states with N_σ electrons on L sites is then as simple as looping over all integers from 0 to $2^L - 1$ and storing each integer m_σ with N_σ bits set. For $N_\uparrow = 2$ and $N_\downarrow = 1$ electrons on $L = 3$ sites we obtain

m_\uparrow	bits	state	i_\uparrow	m_\downarrow	bits	state	i_\downarrow
0	000			0	000		
1	001			1	001	$c_{0\downarrow}^\dagger 0\rangle$	0
2	010			2	010	$c_{1\downarrow}^\dagger 0\rangle$	1
3	011	$c_{0\uparrow}^\dagger c_{1\uparrow}^\dagger 0\rangle$	0	3	011		
4	100			4	100	$c_{2\downarrow}^\dagger 0\rangle$	2
5	101	$c_{0\uparrow}^\dagger c_{2\uparrow}^\dagger 0\rangle$	1	5	101		
6	110	$c_{1\uparrow}^\dagger c_{2\uparrow}^\dagger 0\rangle$	2	6	110		
7	111			7	111		

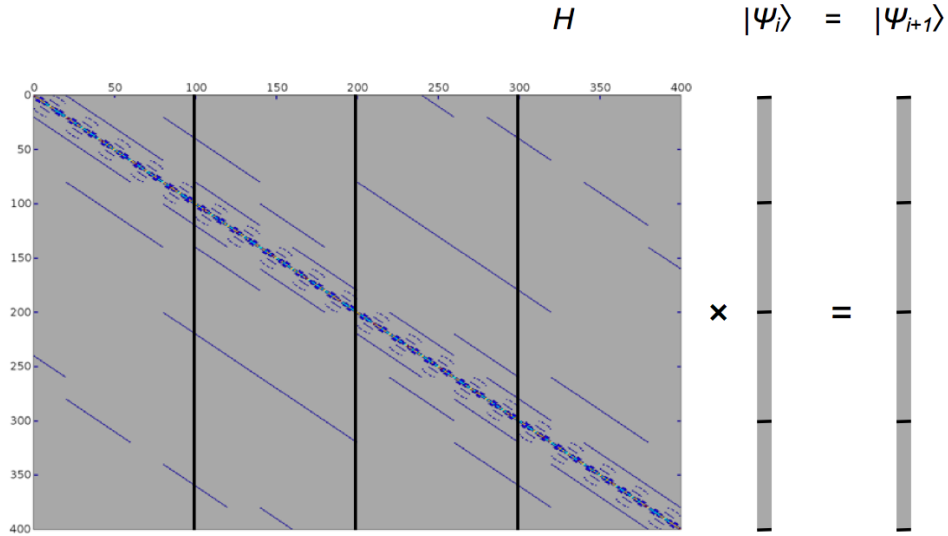


Fig. 6: Matrix-vector product for the Lanczos iteration showing (in blue) the non-zero elements of the Hamiltonian matrix for a six-site Hubbard model at half-filling. Vector Ψ_i is only read, Ψ_{i+1} can be written sequentially: $\Psi_{i+1,n} = \sum_m H_{n,m} \Psi_{i,m}$. Access to elements of Ψ_i is highly non-local.

We number the basis states in the order they are found $i_\sigma = 0 \dots \dim(\mathcal{H}_\uparrow)$. A full basis state (36) is then indexed by the $\mathbf{i} = \mathbf{i}_\downarrow + \dim(\mathcal{H}_\downarrow) \cdot \mathbf{i}_\uparrow$. This corresponds to writing the basis as the tensor product of the up- and down-states. Alternatively we could use $\mathbf{i} = \mathbf{i}_\uparrow + \dim(\mathcal{H}_\uparrow) \cdot \mathbf{i}_\downarrow$. The corresponding configuration is given by the integers $\mathbf{m}_\uparrow[\mathbf{i}_\uparrow]$ and $\mathbf{m}_\downarrow[\mathbf{i}_\downarrow]$

For calculating the matrix elements it is convenient to store the two lookup tables for converting between the integer encoding the basis state \mathbf{m}_σ and its index i_σ in the basis. Since $\dim(\mathcal{H}_\sigma)$ is normally (i.e, close to half-filling) much smaller than $\dim(\mathcal{H})$, this does not use much memory.

The hopping term connects basis state that differ only in two occupation numbers of the same spin, e.g., $n_{i\sigma}$ and $n_{j\sigma}$. The matrix element is $\mp t_{ij}$, where the sign depends on the number $N_{i,j}$ of electrons of spin σ between site i and site j : $-(-1)^{N_{i,j}} t_{i,j}$. As an example we give the matrix of the hopping between the basis states for $N_\uparrow = 2$ electrons on a linear cluster with $L = 3$ sites, nearest neighbor hopping t and with periodic boundary conditions:

$$T_\uparrow = \begin{pmatrix} 0 & -t & +t \\ -t & 0 & -t \\ +t & -t & 0 \end{pmatrix}. \quad (37)$$

The full hopping matrix is then given by the tensor product of T_\uparrow and T_\downarrow . The matrix above looks fairly dense, for larger system the T quickly becomes very sparse, as shown in figure 6.

2.2 Green functions

In a basis of spin-orbitals α and β , the elements of the Green matrix are given by

$$\begin{aligned}
 G_{\alpha\beta}(\omega) &= \left\langle \Psi_0 \left| c_\alpha^\dagger \frac{1}{\omega + (H - E_0 - i\eta)} c_\beta \right| \Psi_0 \right\rangle + \left\langle \Psi_0 \left| c_\alpha \frac{1}{\omega - (H - E_0 - i\eta)} c_\beta^\dagger \right| \Psi_0 \right\rangle \\
 &= \sum_n \frac{\langle \Psi_0 | c_\alpha^\dagger | \Psi_n^{(N-1)} \rangle \langle \Psi_n^{(N-1)} | c_\beta | \Psi_0 \rangle}{\omega + \left(E_n^{(N-1)} - E_0^{(N)} \right) - i\eta} + \sum_n \frac{\langle \Psi_0 | c_\alpha | \Psi_n^{(N+1)} \rangle \langle \Psi_n^{(N+1)} | c_\beta^\dagger | \Psi_0 \rangle}{\omega - \left(E_n^{(N+1)} - E_0^{(N)} \right) + i\eta},
 \end{aligned} \tag{38}$$

where the sums are over the eigenstates of the Hilbert space with one electron less (first term) and one additional electron (second term). Diagonal elements are calculated in Lanczos as described in section 1.2: To find $G_{\alpha\alpha}(\omega)$, we need the ground state vector $|\Psi_0\rangle$ and two additional Lanczos runs, giving the two terms in (38). For the first term, we start the Lanczos iteration from the normalized vector $|\Psi_c^<\rangle = c_\alpha |\Psi_0\rangle / \sqrt{n_\alpha}$, where $n_\alpha = \langle \Psi_0 | c_\alpha^\dagger c_\alpha | \Psi_0 \rangle$. Likewise, for the second term, we start from $|\Psi_c^>\rangle = c_\alpha^\dagger |\Psi_0\rangle / \sqrt{1 - n_\alpha}$. The Green function is then given, in terms of the Lanczos coefficients, by

$$\check{G}_{\alpha\alpha}(\omega) = \frac{n_\alpha}{\omega - E_0 - i\eta + a_0^< - \frac{b_1^{<2}}{\omega - E_0 - i\eta + a_1^< - \dots}} + \frac{1 - n_\alpha}{\omega + E_0 + i\eta - a_0^> - \frac{b_1^{>2}}{\omega + E_0 + i\eta - a_1^> - \dots}}. \tag{39}$$

If the ground state is degenerate, e.g., for $N_\uparrow \neq N_\downarrow$, where $E_0(N_\uparrow, N_\downarrow) = E_0(N_\downarrow, N_\uparrow)$, we average the Green functions calculated from the different ground-states. This is the $T \rightarrow 0$ limit of the finite-temperature Green function

$$G_{\alpha\alpha}(\omega) = \frac{1}{Z} \sum_m e^{-\beta E_m^{(N)}} G_{\alpha\alpha}^{(m)}(\omega), \tag{40}$$

where $Z = \sum_n e^{-\beta E_n^{(N)}}$ is the partition function and $G_{\alpha\alpha}^{(m)}(\omega)$ has the same form as (38), only with Ψ_0 replaced by Ψ_m . For finite, but sufficiently low temperatures, the Boltzmann factor is negligibly small, except for the lowest few states. If we calculate those, taking care of the orthogonality problem (ghost states), we can easily obtain the finite-temperature Green function. A more elaborate method is given in [12].

Off-diagonal elements of the Green matrix are calculated from diagonal elements of linear combinations of spin-orbitals, e.g., $(c_\alpha^\dagger \pm c_\beta^\dagger) |\Psi_0\rangle$, as described in section 1.2.

2.3 Parallelization strategies

Because of the enormous size of the many-body Hilbert space, see table 2, Lanczos calculations are limited by the available memory. On shared-memory systems the most time consuming operation of the Lanczos iteration, the multiplication of the Hamiltonian-matrix with a many-body vector, can be parallelized very easily, when it is written such that the elements of the resulting vector are calculated independently: As illustrated in figure 6, different threads can

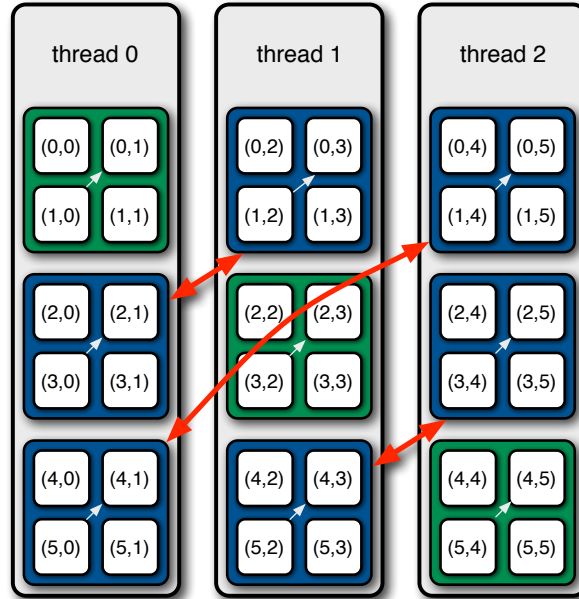


Fig. 7: Transpose operation that makes memory access thread-local when calculating the operation of the Hamiltonian on the state-vector. The communication (red arrows) is realized by a call to `MPI_Alltoall`, which is very efficiently implemented on Blue Gene/L. The small grey arrows indicate the local operations needed to complete the matrix-transpose.

work on different chunks of $|\Psi_{i+1}\rangle$. The off-diagonal elements of the kinetic energy part of (32) lead to non-local memory access, but the elements of $|\Psi_i\rangle$ as well as the matrix elements are only read, so that there is no need for locking. An OpenMP parallelization thus needs only a single pragma. Parallelizing also the scalar products in a similar way, we obtain almost ideal speedup on an IBM p690 frame of JUMP in Jülich. The implementation is however limited to a single node, i.e., about 120 GBytes. To use significantly more memory we need to find an efficient distributed-memory implementation.

A naive approach on distributed memory systems uses MPI2 one-sided communication to emulate the shared-memory approach by direct remote memory access. This leads, however, to a severe speed-down, i.e., the more processors we use, the longer we have to wait for the result.

An efficient distributed-memory implementation [14] is instead based on the fact that hopping does not change spin. Hopping of the up-electron mixes only different up-hopping configurations, while the down-electron configuration remains unchanged. If we group all up configurations for a fixed down configuration together in a single thread, this hopping can be carried out locally. Figure 5 illustrates this: for a fixed index i_\downarrow , all i_\uparrow configurations are stored in adjacent memory locations and can be stored in a thread. We see, that this basis can be naturally indexed by a tuple $(i_\downarrow, i_\uparrow)$ (right labels in figure 5) instead of the global index (left labels). We can therefore equivalently regard the vectors as matrices $v(i_\downarrow, i_\uparrow)$ with indices i_\downarrow and i_\uparrow . Now it is easy to see that a matrix transpose reshuffles the data elements such that the down configurations are sequentially in memory and local to the thread.

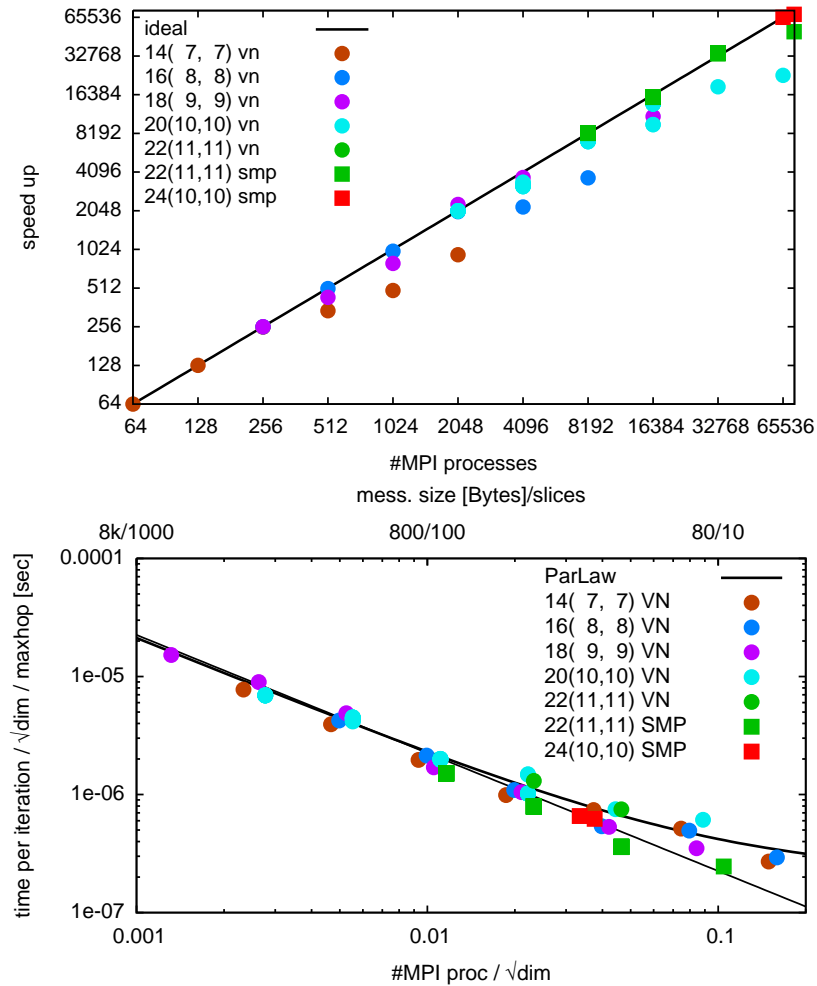


Fig. 8: Timings of the parallel implementation of the Lanczos algorithm for the Hubbard model on the Jülich IBM BlueGene/P. Sizes of the state vectors of the half-filled systems are given in table 2. For the 24 site system with 10+10 electrons, $\text{dim}(\mathcal{H}) = 3\,846\,525\,097\,536$, the state vector takes about 28 TBytes. The simulation of such a system requires the entire machine, using only one processor per node (SMP mode), to make most efficient use of the available memory. For smaller systems we can use all four processors per node (VN mode). Despite massive communication in each iteration, the code shows excellent speed up. Only when the message size per process become too small, performance degrades because of network latency. This is shown in the lower plot. Properly scaling the execution times we obtain a universal scaling (ParLaw) for system sizes ranging over more than five and process counts ranging over three orders of magnitude.

We implement an efficient matrix transpose using `MPI_Alltoall`. This routine expects that the data packages which will be sent to a given process to be stored contiguously in memory. This does not apply to our case, since we would like to store the spin-down electron configurations sequentially in memory. Thus, the matrix is stored column wise. For `MPI_Alltoall` to work properly, we would have to bring the data elements into row-major order. This could be done by performing a local matrix transpose. The involved matrices are, however, in general rectangular, leading to expensive local-copy and reordering operations. We can avoid this by calling `MPI_Alltoall` for each column separately (red arrows in figure 7). After this, only

a local strided transposition has to be performed (small white arrows) to obtain the fully transposed matrix or Lanczos vector. The implementation described so far uses `MPI_Alltoall` which assumes that the matrix to be transposed is a square matrix and that the dimension $dim_{\uparrow} = dim_{\downarrow}$ is divisible by the number of MPI processes. To overcome these restrictions we have generalized the algorithm to `MPI_Alltoallv`. This is the implementation that is used in practice. The speed-up shown at the top of figure 8 shows that our parallelization based on collective communication is indeed very efficient. Even for a system of 24 sites with 10 electrons of either spin, where a single many-body vector takes about 28 TB of memory, our implementation works very efficiently despite the fact that in each Lanczos iteration 28 TB of data have to be moved across the entire machine twice.

The lower plot in figure 8 shows that the execution times for runs of systems varying by more than five orders of magnitude in size (of the Hilbert space) for processor counts varying over three orders of magnitude fall on a universal curve, which is only determined by the bandwidth and the latency of the network. This suggests that the implementation should scale to even larger systems than the present Jülich BlueGene/P with almost 300 000 CPUs and an aggregate memory of 144 terabytes.

3 Application to DMFT

Using the Lanczos method as a solver for DMFT gives results at zero temperature and directly on the real axis. An important limitation is, however, the need to approximate the bath Green's function

$$\mathcal{G}^{-1}(\omega) = \omega + \mu - \int_{-\infty}^{\infty} d\omega' \frac{\Delta(\omega')}{\omega - \omega'} \quad (41)$$

by a discretized version, e.g., of the form

$$\mathcal{G}_{\text{And}}^{-1}(\omega) = \omega + \mu - \sum_{l=1}^{N_b} \frac{V_l^2}{\omega - \varepsilon_l}, \quad (42)$$

corresponding to an Anderson impurity model with a finite number of sites

$$H_{\text{And}} = \varepsilon_0 \sum_{\sigma} n_{\sigma} + U n_{\uparrow} n_{\downarrow} + \sum_{\sigma} \sum_{l=1}^{N_b} \left(\varepsilon_l n_{l\sigma} + V_l \left(a_{l\sigma}^{\dagger} c_{\sigma} + c_{\sigma}^{\dagger} a_{l\sigma} \right) \right), \quad (43)$$

where c_{σ}^{\dagger} and $a_{l\sigma}^{\dagger}$ create an electron of spin σ on the impurity or bath-site l , respectively, $n_{\sigma} = c_{\sigma}^{\dagger} c_{\sigma}$ and $n_{l\sigma} = a_{l\sigma}^{\dagger} a_{l\sigma}$. Writing the non-interacting part of H_{And} as a matrix

$$H_{\text{And}}^0 = \begin{pmatrix} 0 & V_1 & V_2 & V_3 & \cdots \\ V_1 & \varepsilon_1 & 0 & 0 & \\ V_2 & 0 & \varepsilon_2 & 0 & \\ V_3 & 0 & 0 & \varepsilon_3 & \\ \vdots & & & & \ddots \end{pmatrix} \quad (44)$$

we see that (42) is easily recovered from inversion by partitioning.

Since practical calculations are limited by the rapidly increasing size of the Hilbert space to small numbers of bath-sites N_b , it is crucial for the reliability of the calculations to find a good representation $\mathcal{G}_{\text{And}}^{-1}$ for the bath Green's function. The most common approach is to use a least squares fit [15]: Because of the spectral poles on the real axis, such a fit in practice is done on the imaginary axis, where the Green's functions are smooth and the optimization of the distance function is not easily trapped in local minima. One then minimizes a function of the form

$$\chi^2(\{V_l, \varepsilon_l\}) = \sum_{n=0}^{n_{\max}} |\mathcal{G}^{-1}(i\omega_n) - \mathcal{G}_{\text{And}}^{-1}(i\omega_n)|^2 \quad (45)$$

on a set of Matsubara frequencies, corresponding to some fictitious temperature. The choice of this temperature and of n_{\max} essentially determines the relative weighing of high versus low frequency features in the fit. If low $i\omega_n$ are weighted too little, the fit easily becomes underdetermined, since for large imaginary frequencies the hybridization function contains only little information about the system (which is the reason why the analytic continuation back to the real axis is so difficult). To emphasize different frequency ranges, it is possible to introduce frequency-dependent weight functions in (45).

Instead of fitting, we could use a moment expansion of the Weiss function $W(\omega) = \int d\omega' \frac{\Delta(\omega')}{\omega - \omega'}$ similar to that discussed in section 1.2. Such an approach [16] has been used for the Bethe lattice with infinite coordination, where the self-consistency condition simplifies to $W(\omega) = t^2 G_{\text{imp}}(\omega)$: As Lanczos gives a continued-fraction representation for the photoemission and inverse-photoemission part separately, the hybridization function is written as

$$W^<(\omega) + W^>(\omega) = t^2 G^<(\omega) + t^2 G^>(\omega) = \frac{t^2 b_0^{<2}}{\omega + a_0^< - \frac{b_1^{<2}}{\omega + a_1^< - \dots}} + \frac{t^2 b_0^{>2}}{\omega - a_0^> - \frac{b_1^{>2}}{\omega - a_1^> - \dots}} \quad (46)$$

Truncating the continued fractions at $N_b^<$ and $N_b^>$, this corresponds to the impurity model with

$$H_{\text{And}}^0 = \begin{pmatrix} 0 & t^2 b_0^< & & \dots & t^2 b_0^> \\ t^2 b_0^< & -a_0^< & b_1^< & & \\ & b_1^< & -a_1^< & b_2^< & \\ & & b_2^< & -a_2^< & \ddots \\ \vdots & & & \ddots & \ddots \\ t^2 b_0^> & & & & a_0^> & b_1^> \\ & & & & b_1^> & a_1^> & b_2^> \\ & & & & & b_2^> & a_2^> & \ddots \\ & & & & & & & \ddots \end{pmatrix}, \quad (47)$$

where the bath forms two chains, coupled to the impurity. Diagonalizing the bath, it is easily brought to the form (43). This bath-parametrization works very well for systems with large gap. An example is shown in figure 9. Since the approach uses moment expansions for the two parts

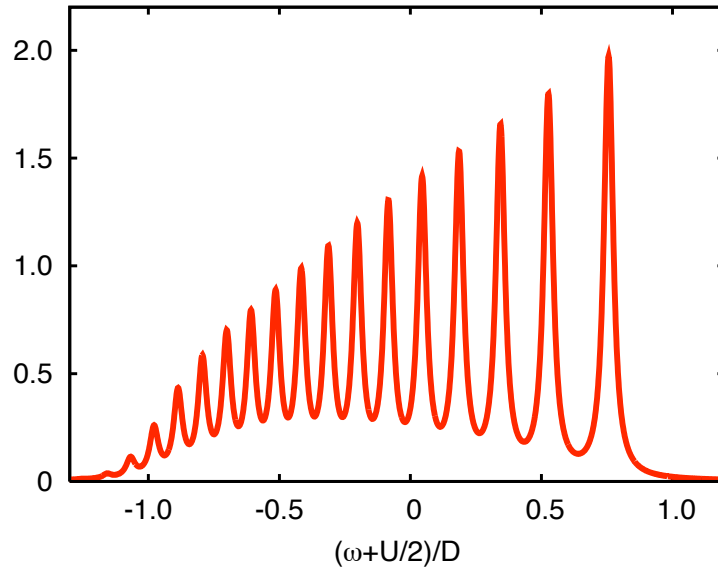


Fig. 9: Spectral function of the lower Hubbard band of a Hubbard model on the infinite Bethe lattice with half-bandwidth D and $U = 8D$ in antiferromagnetic DMFT. The bath was obtained from the continued-fraction expansion of the impurity Green's functions, $N_b = 24$ [18].

$W^<$ and $W^>$ of the Weiss function separately, it does not converge as quickly as a moment expansion for the full hybridization function would. This makes itself particularly felt, when the gap is small or the system is even metallic. To improve the description of the hybridization function close to the Fermi level, it has been proposed to introduce an additional bath site with fixed energy close to μ [17].

3.1 Cluster methods

For cluster versions of DMFT we can use exact sum-rules and symmetries to find the structure of the bath. Our discussion will closely follow [19]. To fix the notation we briefly sketch the self-consistency loop for cellular DMFT (CDMFT) and the dynamical cluster approximation (DCA) using Lanczos as impurity solver. Let N_c be the number of cluster-sites, N_b the number of bath-sites. For simplicity we suppress spin-indices.

Given an $N_c \times N_c$ bath Green matrix \mathcal{G}^{-1} ,

1. fit parameters of an Anderson model with N_b bath-sites

$$\mathcal{G}_{\text{And}}^{-1}(\omega) = \omega + \mu - \mathbf{H}_c - \mathbf{\Gamma} [\omega - \mathbf{E}]^{-1} \mathbf{\Gamma}^\dagger \quad (48)$$

to \mathcal{G}^{-1} , where $\mathbf{\Gamma}$ is the $N_c \times N_b$ -dimensional hybridization matrix, and \mathbf{E} the $N_b \times N_b$ -dimensional bath-matrix. \mathbf{H}_c is specified below,

2. solve the $(N_c + N_b)$ -site Anderson model H_{And} (specified below) to obtain the $N_c \times N_c$ cluster Green matrix \mathbf{G}_c ,

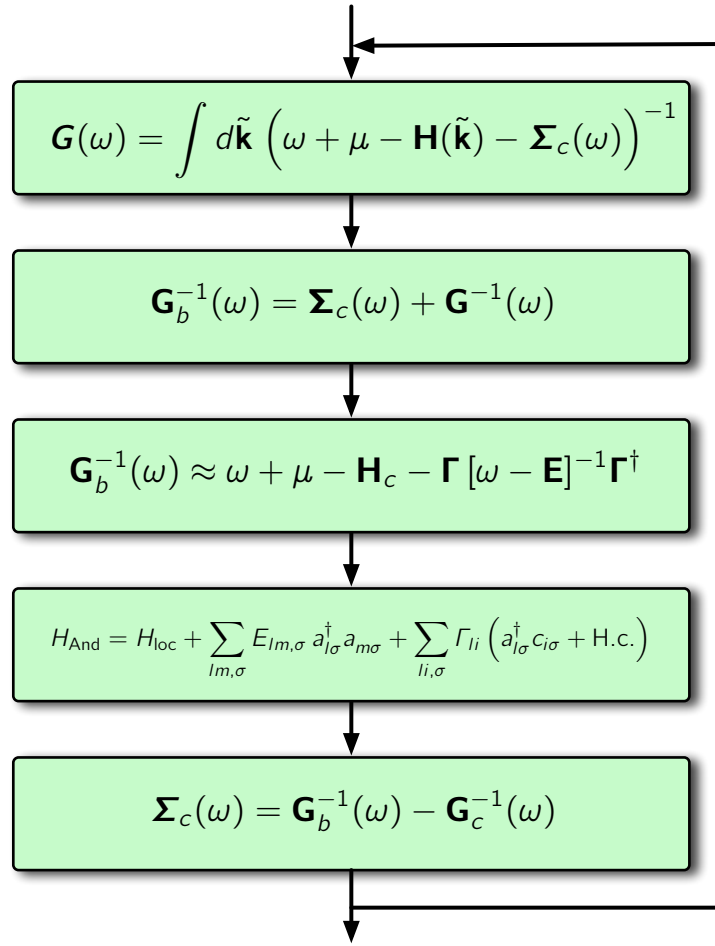


Fig. 10: Self-consistency loop for (cluster) DMFT.

3. get the cluster self-energy matrix

$$\boldsymbol{\Sigma}_c(\omega) = \mathcal{G}^{-1}(\omega) - \mathbf{G}_c^{-1}(\omega), \quad (49)$$

4. calculate the local Green matrix for the cluster by integrating over the reduced Brillouin-zone of the cluster

$$\mathbf{G}(\omega) = \int d\tilde{\mathbf{k}} \left(\omega + \mu - \mathbf{H}(\tilde{\mathbf{k}}) - \boldsymbol{\Sigma}_c(\omega) \right)^{-1}, \quad (50)$$

where $\mathbf{H}(\tilde{\mathbf{k}})$ is the single-electron part of the of the Hubbard Hamiltonian (32) in the reduced Brillouin-zone of the cluster,

5. determine the new bath Green matrix (self-consistency condition)

$$\mathcal{G}^{-1}(\omega) = \boldsymbol{\Sigma}_c(\omega) + \mathbf{G}^{-1}(\omega). \quad (51)$$

These steps are iterated to self-consistency.

3.2 Anderson impurity model

The Anderson model to be solved in step 2 is given by

$$H_{\text{And}} = H_{\text{clu}} + \sum_{lm,\sigma} E_{lm,\sigma} a_{l\sigma}^\dagger a_{m\sigma} + \sum_{li,\sigma} \Gamma_{li} \left(a_{l\sigma}^\dagger c_{i\sigma} + \text{H.c.} \right) \quad (52)$$

where the operator $a_{l\sigma}^\dagger$ creates an electron of spin σ on bath-site l . The cluster Hamiltonian H_{clu} is obtained from the lattice Hamiltonian by transforming to the reciprocal space of the super-lattice of the clusters, and projecting to the cluster. Writing the single-electron part of $H(\tilde{\mathbf{k}})$ as the matrix $\mathbf{H}(\tilde{\mathbf{k}})$, the single-electron part of H_{clu} is given by

$$\mathbf{H}_c = \int d\tilde{\mathbf{k}} \mathbf{H}(\tilde{\mathbf{k}}). \quad (53)$$

The (local) interaction terms are simply those of the lattice model, restricted to the cluster. The Hamiltonian $H(\tilde{\mathbf{k}})$ in the reciprocal space of the super-lattice $\{\tilde{\mathbf{r}}\}$ of clusters can be obtained by changing to the basis of operators

$$\tilde{c}_{\mathbf{R}_i\sigma}^{\text{CDMFT}}(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} e^{-i\tilde{\mathbf{k}}\tilde{\mathbf{r}}} c_{\tilde{\mathbf{r}}+\mathbf{R}_i,\sigma}. \quad (54)$$

The resulting quantum cluster approximation is CDMFT. Alternatively, we can start from the operators in the reciprocal space of the *lattice* to obtain

$$\tilde{c}_{\mathbf{R}_i\sigma}^{\text{DCA}}(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} e^{-i\tilde{\mathbf{k}}(\tilde{\mathbf{r}}+\mathbf{R}_i)} c_{\tilde{\mathbf{r}}+\mathbf{R}_i,\sigma}. \quad (55)$$

Now we obtain the DCA. The choice of the operators in the two approaches differs just by local phase factors. In CDMFT this Kohn-gauge [20] is chosen such that phases appear only in matrix elements involving different clusters. Thus all matrix elements on the cluster are the same as in the original Hamiltonian. The price for retaining the original matrix elements on the cluster is a breaking of the translation-symmetry of the original lattice. DCA opts instead to retain this symmetry by distributing the phase-change uniformly over the cluster-sites. The price for retaining translation-invariance is that the matrix elements in the cluster Hamiltonian differ from those in the original Hamiltonian (coarse graining). In both cases, CDMFT and DCA, the eigenvalues of $\mathbf{H}(\tilde{\mathbf{k}})$ are identical to the eigenvalues of the non-interacting part of H . Obviously, we can construct other cluster extensions to DMFT by different choices of the Kohn-gauge $\varphi(\tilde{\mathbf{k}}; \mathbf{R}_i)$ on the cluster

$$\tilde{c}_{\mathbf{R}_i\sigma}^\varphi(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{r}}} e^{-i(\tilde{\mathbf{k}}\tilde{\mathbf{r}}+\varphi(\tilde{\mathbf{k}};\mathbf{R}_i))} c_{\tilde{\mathbf{r}}+\mathbf{R}_i,\sigma}. \quad (56)$$

3.3 Hybridization sum-rules

While the most general parametrization for the bath is given by expression (48), we can always diagonalize the hopping matrix \mathbf{E} among the bath-sites to obtain

$$\mathcal{G}_{\text{And}}^{-1}(\{\varepsilon_l, \mathbf{V}_l\}; \omega) = \omega + \mu - \mathbf{H}_c - \sum_l \frac{\mathbf{V}_l \mathbf{V}_l^\dagger}{\omega - \varepsilon_l}. \quad (57)$$

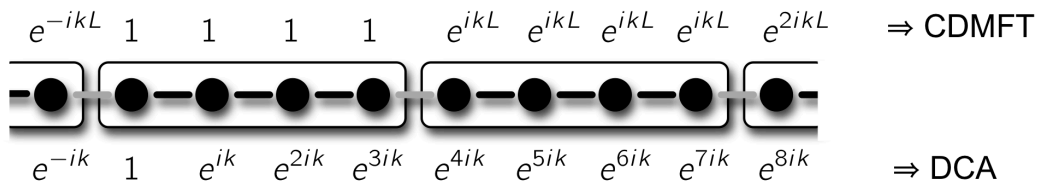


Fig. 11: Phase choice on the cluster that leads to CDMFT or DCA.

The hybridization matrix is then given by the tensor product of the vectors \mathbf{V}_l , where

$$V_{l,i} = \sum_m \Gamma_{i,m} \phi_{l,m} \quad (58)$$

and ϕ_l are the eigenvectors of \mathbf{E} with eigenvalues ε_l .

To obtain sum-rules for the hybridizations, we write the inverse of the bath Green matrix as

$$\mathcal{G}^{-1}(\omega) = \Sigma_c(\omega) + \left(\int d\tilde{\mathbf{k}} \left(\omega + \mu - \mathbf{H}(\tilde{\mathbf{k}}) - \Sigma_c(\omega) \right)^{-1} \right)^{-1}.$$

Considering the limit $\omega \rightarrow \infty$, expanding to order $1/\omega^2$, using (53), and comparing to (57) we find

$$\sum_l \mathbf{V}_l \mathbf{V}_l^\dagger = \int d\tilde{\mathbf{k}} \mathbf{H}^2(\tilde{\mathbf{k}}) - \left(\int d\tilde{\mathbf{k}} \mathbf{H}(\tilde{\mathbf{k}}) \right)^2. \quad (59)$$

To illustrate this hybridization sum-rule we consider a representative set of examples.

Single site

We consider a d -dimensional lattice with hoppings t_n to the z_n n^{th} -nearest neighbors. For $N_c = 1$ we have $\mathbf{H}(\mathbf{k}) = \varepsilon_{\mathbf{k}}$. Thus we find for the hybridizations

$$\sum_l V_l^2 = \frac{1}{(2\pi)^d} \int_{-\pi}^{\pi} d^d \mathbf{k} \varepsilon_{\mathbf{k}}^2 = \sum_n z_n t_n^2, \quad (60)$$

where the integral is just the second moment of the density of states, so that the last equation follows as in the recursion method [7]. For a Bethe lattice of connectivity z with hopping matrix element t/\sqrt{z} the sum-rule reduces to $\sum_l V_l^2 = t^2$.

CDMFT

We start by considering a linear chain with nearest neighbor hopping t and a three-site cluster $N_c = 3$. In the CDMFT gauge we have

$$\mathbf{H}(\tilde{k}) = -t \begin{pmatrix} 0 & 1 & e^{-3i\tilde{k}} \\ 1 & 0 & 1 \\ e^{3i\tilde{k}} & 1 & 0 \end{pmatrix} \quad (61)$$

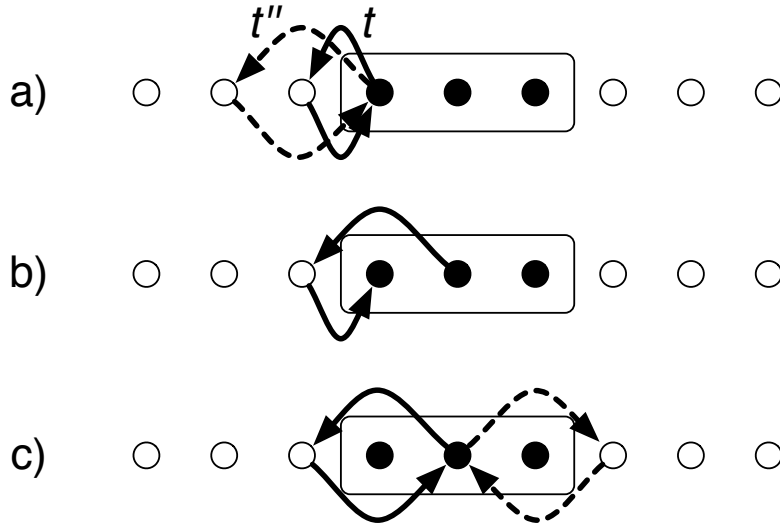


Fig. 12: CDMFT sum-rules for a one-dimensional 3-site cluster with nearest and next-nearest neighbor hoppings t and t'' , respectively: a) $\sum_l |V_{l,1}|^2 = t^2 + t''^2$, b) $\sum_l \bar{V}_{l,1} V_{l,2} = t t''$, and c) $\sum_l |V_{l,2}|^2 = 2t''^2$. The hybridizations are given by the two-step hopping processes that are lost when the cluster is cut out of the original lattice.

so that \mathbf{H}_c is the original single-electron Hamiltonian restricted to the cluster:

$$\mathbf{H}_c = \frac{3}{2\pi} \int_{-\pi/3}^{\pi/3} d\tilde{k} \mathbf{H}(\tilde{k}) = -t \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (62)$$

The sum-rule (59) then is

$$\left(\sum_l V_{l,i} \bar{V}_{l,j} \right) = \begin{pmatrix} t^2 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & t^2 \end{pmatrix}, \quad (63)$$

i.e., only the sites on the surface of the cluster couple to the bath. If we allow also second nearest neighbor hopping with matrix element t'' , we find

$$\left(\sum_l \bar{V}_{l,\mu} V_{l,\nu} \right) = \begin{pmatrix} t^2 + t''^2 & t t'' & 0 \\ t t'' & 2t''^2 & t t'' \\ 0 & t t'' & t^2 + t''^2 \end{pmatrix}. \quad (64)$$

The general CDMFT hybridization sum-rule (59) can be easily visualized: The integral over the Brillouin-zone of the cluster projects the single-electron part of the full Hamiltonian to the cluster (see eqn. (53)). The matrix elements of \mathbf{H}_c^2 are thus the two-step hoppings that are possible on the cluster. Likewise the integral over the Hamiltonian squared gives the second moments, only that here the intermediate site is not restricted to the cluster. Thus the sum-rule matrix is given by the second-order paths between cluster-sites that proceed via a site outside the cluster. This is illustrated in figure 12. As a special case, for a single site we recover the second equality in (60).

The vanishing of a matrix element in the sum-rule merely implies that the corresponding matrix element of the bath Green matrix decays faster than $1/\omega$ for large ω . For a diagonal element, however, all terms in $\sum_l V_{l,i} \bar{V}_{l,i}$ are positive. Thus a vanishing sum means that all terms must be zero. Hence the sum-rule implies that cluster-sites that are so far in the interior that they cannot be reached by hopping from outside the cluster do not couple to bath and that all matrix elements of the bath Green function involving such a site i are given by $\mathcal{G}_{ij}^{-1}(\omega) = \omega + \mu - (\mathbf{H}_c)_{ij}$ for all ω . In that sense the bath hybridizes only to the surface of the cluster and we see that the hybridization-strength to these sites does not decrease for increasing cluster size N_c .

DCA

We start again by considering the 3-site cluster. In the DCA gauge we write

$$\mathbf{H}(\tilde{k}) = -t \begin{pmatrix} 0 & e^{i\tilde{k}} & e^{-i\tilde{k}} \\ e^{-i\tilde{k}} & 0 & e^{i\tilde{k}} \\ e^{i\tilde{k}} & e^{-i\tilde{k}} & 0 \end{pmatrix}. \quad (65)$$

Now \mathbf{H}_c has translation symmetry, but the hopping is rescaled by $\sin(\pi/N_c)/(\pi/N_c)$

$$\mathbf{H}_c = \frac{3}{2\pi} \int_{-\pi/3}^{\pi/3} d\tilde{k} \mathbf{H}(\tilde{k}) = -\frac{3\sqrt{3}}{2\pi} t \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \quad (66)$$

Since all matrices in (59) are periodic, it is convenient to transform to k -space. With $V_{l,K} = \sum_i V_{l,i} e^{iKr_i} / \sqrt{N_c}$ and the coarse-graining factor $\tau = 3\sqrt{3}/2\pi$ we find

$$\begin{aligned} \sum_l |V_{l,K=0}|^2 &= (2 + \tau - 4\tau^2) t^2 \\ \sum_l |V_{l,K=\pm 2\pi/3}|^2 &= (2 - \tau/2 - \tau^2) t^2. \end{aligned}$$

The hybridization sum-rule (59) is then, likewise, diagonal in the cluster-momenta \mathbf{K}

$$\sum_l |V_{l,\mathbf{K}}|^2 = \int d\tilde{\mathbf{k}} \varepsilon_{\mathbf{K}+\tilde{\mathbf{k}}}^2 - \left(\int d\tilde{\mathbf{k}} \varepsilon_{\mathbf{K}+\tilde{\mathbf{k}}} \right)^2, \quad (67)$$

while all terms $V_{l,\mathbf{K}} \bar{V}_{l,\mathbf{K}'}$ mixing different cluster momenta vanish. As a special case, for a single site the above sum-rule is just the first equality in (60). Expanding $\varepsilon_{\mathbf{K}+\mathbf{k}}$ around \mathbf{K} , we find that for a d -dimensional system $\sum_l |V_{l,\mathbf{K}}|^2$ decreases with cluster size as $1/N_c^{2/d}$, while all cluster-sites couple with the same strength to the bath.

Discussion

Besides providing exact relations for the bath-parametrization, in particular which sites need not be coupled to a bath, the sum-rules contain important information about the scaling of cluster methods with cluster size: In CDMFT individual hybridizations are independent of cluster size,

while for DCA they decrease with cluster size as $N_c^{-2/d}$. Interestingly this means that for a d -dimensional system in CDMFT the overall coupling to the bath scales as $N_c^{(d-1)/d}$, while in DCA it scales as $N_c^{(d-2)/d}$. For non-local properties a DCA calculation is therefore expected to converge faster with cluster size. For a calculation where we represent the bath by discrete degrees of freedom this decrease in hybridization strength does, however, not help very much as we still need bath-sites to fit the hybridizations, even if they are small. With increasing DCA cluster size we thus have to parametrize N_c baths, one for each \mathbf{K} . In CDMFT the situation is more fortunate, as the sum-rules imply that many hybridizations vanish and we only need to parametrize the coupling of surface-sites to the bath.

The lack of translational invariance in CDMFT has, however, two important practical implications. First, the full Green matrix has to be calculated, instead of just its diagonal. Second, when calculating local quantities, like the density per site, in CDMFT we have a choice of inequivalent sites, or we could consider the average over all sites. In a gapped system the best choice is the innermost site. In such a situation it might, however, be better to do a straight Lanczos calculation with $N_c + N_b$ cluster sites, instead of using N_b bath sites.

3.4 Symmetries

In the absence of spontaneous symmetry breaking the symmetries of the cluster (point-symmetries in CDMFT and additionally translation symmetry in DCA) are reflected in the Green matrix. In a symmetry-broken state with long-range order, like an antiferromagnet or a charge-density wave, the symmetry of the Green matrix is accordingly lowered. To exploit the symmetry we introduce vectors on the cluster that transform according to its irreducible representations. We write these vectors as $\mathbf{w}_{I,\nu}$ where I is the irreducible representation and $\nu = 1 \dots N_I$ counts linearly independent vectors transforming according to I . On an N_c -site cluster we can choose N_c such vectors that are orthonormal. Defining the matrix $\mathbf{W} = (\mathbf{w}_{I,\nu})$ of these vectors, we can block-diagonalize the bath Green matrix: $\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W}$ has blocks of dimension N_I corresponding to the different irreducible representations I . Since $\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W}$ is block diagonal for all ω , it follows from equation (57), that \mathbf{W} must also block-diagonalize the individual hybridization matrices $\mathbf{V}_l \mathbf{V}_l^\dagger$. Therefore the hybridization vectors must transform according to an irreducible representation: They can be written as $\mathbf{V}_l = \sum_\nu V_{l;I,\nu} \mathbf{w}_{I,\nu}$ for some irreducible representation I . If the \mathbf{V}_l also had components $\mathbf{w}_{J,\nu}$ of a different irreducible representation $J \neq I$ this would produce a hybridization matrix that could not be block-diagonalized. This can only happen for bath-sites with identical energy ε_l (accidental degeneracy): Assume \mathbf{V}_l and $\mathbf{V}_{l'}$ are the hybridizations for two bath-sites with $\varepsilon_l = \varepsilon_{l'}$. Then we can form arbitrary linear combinations of the hybridization matrices and hence of the hybridization vectors. For all these linear combinations the sum of the hybridization matrices must be block diagonal, and hence we can choose the hybridization vectors such that they transform according to irreducible representations.

We thus find that the bath-sites can be arranged into sets corresponding to the different irreducible representations. For fitting a block of the symmetrized bath Green matrix we need then only consider bath-sites of the respective irreducible representation. If the block is one-

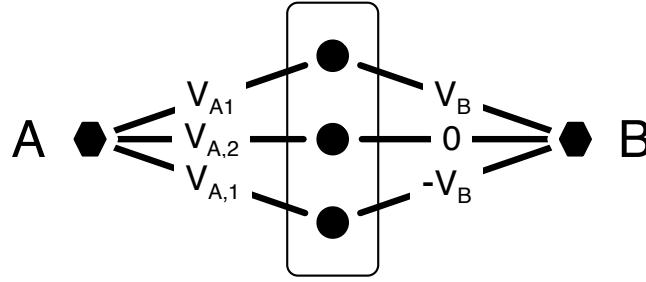


Fig. 13: Hybridization of bath-sites of symmetry A and B to a 3-site cluster. As defined in table 3, A is the unit representation, so a bath-site of type A has the same hybridization V to all cluster-sites that are equivalent by symmetry. B is the antisymmetric representation, so the hybridization of a bath-site of type B to cluster-sites that are related by mirror symmetry have the opposite sign. Consequently the hybridization to the central site of a linear cluster with an odd number of sites vanishes in the B representation.

dimensional we can choose the corresponding hybridizations real.

Sectors corresponding to different irreducible representations are only coupled through the Hubbard interaction U when solving the Anderson impurity model. Note that the coupling to bath-sites corresponding to an irreducible representation other than the unit representation lowers the symmetry of the impurity Hamiltonian with respect to that of the Green matrix.

CDMFT

As an example we consider a linear cluster of 3 sites as shown in figure 13. The symmetry is C_2 (see table 3). Transforming to the basis vectors $\mathbf{w}_{A,1} = (|1\rangle + |3\rangle)/\sqrt{2}$ and $\mathbf{w}_{A,2} = |2\rangle$ of symmetry A (see table 3) and $\mathbf{w}_B = (|1\rangle - |3\rangle)/\sqrt{2}$, we find the transformed bath Green matrix

$$\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W} = \begin{pmatrix} \mathcal{G}_{11}^{-1} + \mathcal{G}_{13}^{-1} & \sqrt{2} \mathcal{G}_{12}^{-1} & 0 \\ \sqrt{2} \mathcal{G}_{21}^{-1} & \mathcal{G}_{22}^{-1} & 0 \\ 0 & 0 & \mathcal{G}_{11}^{-1} - \mathcal{G}_{13}^{-1} \end{pmatrix}.$$

A bath-site of irreducible representation A contributes to the first block and has the same hybridization $V_{A,1}$ to the outer cluster-sites plus an independent hybridization parameter $V_{A,2}$ to the central site. A bath-site of irreducible representation B contributes to the second block. For such a bath-site the hybridization to cluster-sites that are related by mirror symmetry have opposite signs. Consequently, the hybridization to the central site vanishes.

The situation is slightly more complicated when the symmetry group has irreducible representations of dimension higher than one. The simplest example is the 2×2 cluster with C_{4v} symmetry. With $\mathbf{w}_{A_1} = (|1\rangle + |2\rangle + |3\rangle + |4\rangle)/2$, $\mathbf{w}_{B_2} = (|1\rangle - |2\rangle + |3\rangle - |4\rangle)/2$, and the pair $\mathbf{w}_{E,1} = (|1\rangle - |2\rangle - |3\rangle + |4\rangle)/2$, $\mathbf{w}_{E,2} = (|1\rangle + |2\rangle - |3\rangle - |4\rangle)/2$ we find that $\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W}$ is

C_2	E	σ_v		C_{2v}	E	C_2	σ_v	σ'_v
A	1	1		A_1	1	1	1	1
B	1	-1		A_2	1	1	-1	-1
				B_1	1	-1	1	-1
				B_2	1	-1	-1	1

C_{3v}	E	$2C_3$	$3\sigma_v$		C_{4v}	E	$2C_4$	C_4^2	$2\sigma_v$	$2\sigma_d$
A_1	1	1	1		A_1	1	1	1	1	1
A_2	1	1	-1		A_2	1	1	1	-1	-1
E	2	-1	0		B_1	1	-1	1	1	-1
					B_2	1	-1	1	-1	1
					E	2	0	-2	0	0

Table 3: Character tables of the point groups C_{1v} , C_{2v} , C_{3v} , and C_{4v} .

diagonal with diagonal elements

$$\begin{aligned}
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{11} &= \mathcal{G}_{11}^{-1} + 2\mathcal{G}_{12}^{-1} + \mathcal{G}_{13}^{-1} \\
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{22} &= \mathcal{G}_{11}^{-1} - 2\mathcal{G}_{12}^{-1} + \mathcal{G}_{13}^{-1} \\
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{33} &= \mathcal{G}_{11}^{-1} - \mathcal{G}_{13}^{-1} \\
(\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W})_{44} &= \mathcal{G}_{11}^{-1} - \mathcal{G}_{13}^{-1}
\end{aligned}$$

A bath-site of symmetry A_1 has the same hybridization to all cluster-sites while for a bath-site of symmetry B_2 the hybridizations have alternating signs: $\mathbf{V}_l^\dagger = \bar{V}_l (1, -1, 1, -1)$. To realize the two-dimensional representation E we need two bath-sites l_1 and l_2 with degenerate energies $\varepsilon_{l_1} = \varepsilon_{l_2} = \varepsilon_l$ and hybridizations: $\mathbf{V}_{l_1}^\dagger = \bar{V}_l (1, -1, -1, 1)$ and $\mathbf{V}_{l_2}^\dagger = \bar{V}_l (1, 1, -1, -1)$. This is illustrated in figure 14.

DCA

As an example for DCA we consider a 3-site cluster with periodic boundary conditions. The symmetry group is C_{3v} (translations and inversion). Hence we introduce the basis vector $\mathbf{w}_{A_1} = (|1\rangle + |2\rangle + |3\rangle)/\sqrt{3}$, corresponding to $k = 0$, while the vectors formed by $\sin(2\pi/3)$ and $\cos(2\pi/3)$ give the E representation: $\mathbf{w}_{E,1} = (|1\rangle - |2\rangle)/\sqrt{2}$ and $\mathbf{w}_{E,2} = (|1\rangle + |2\rangle - 2|3\rangle)/\sqrt{6}$.

$$\mathbf{W}^\dagger \mathcal{G}^{-1} \mathbf{W} = \begin{pmatrix} \mathcal{G}_{11}^{-1} + 2\mathcal{G}_{12}^{-1} & 0 & 0 \\ 0 & \mathcal{G}_{11}^{-1} - \mathcal{G}_{12}^{-1} & 0 \\ 0 & 0 & \mathcal{G}_{11}^{-1} - \mathcal{G}_{12}^{-1} \end{pmatrix}.$$

In general bath-sites corresponding to the gamma point have the same hybridization to all cluster-sites, while those corresponding to $k = \pi$ have alternating hybridizations. For all other k -points we need two degenerate bath-sites, with hybridizations $V_{l_1, \mu} = V_l \sin(k\mu)$ and $V_{l_2, \mu} = V_l \cos(k\mu)$ to cluster-site μ .

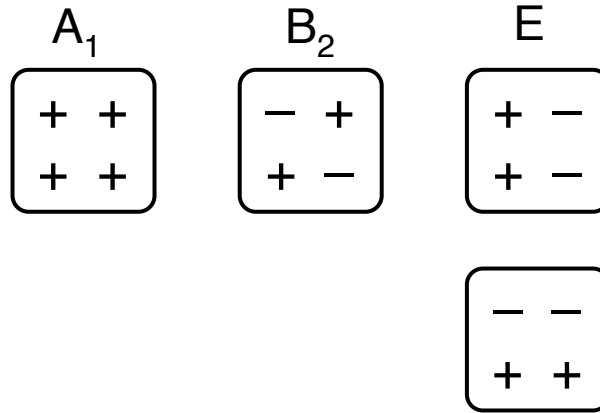


Fig. 14: Hybridization of bath-sites of symmetry A_1 , B_2 , and E to a 2×2 cluster. For a given irreducible representation the absolute value of the hybridization to all cluster-sites is the same, while the signs are indicated in the figure. Non-trivial hybridizations corresponding to irreducible representations A_2 or B_1 only appear for larger clusters.

4 Conclusions

We have seen that the Lanczos method is unbelievably efficient for calculating ground-state and dynamical response functions of many-body Hamiltonians. The determination of the ground-state takes only about $\mathcal{O}(\dim(\mathcal{H}))$ in time and memory. The iteration already converges after about a hundred steps, even for Hilbert spaces with dimensions in the trillions. This astounding convergence is based on the idea of steepest-descent to the ground state, which the Lanczos method even improves upon. In addition, we can very efficiently calculate Green functions. Here the rapid convergence is due to the fact that the Lanczos iteration reproduces more and more moments of the spectral function.

The great advantages of the Lanczos approach to strongly correlated systems is that it provides us with expressions for the Green function on the entire complex plane, i.e., in particular for real frequencies. Calculations are directly for the ground-state, i.e., $T=0$, but can easily be extended to finite, but low, temperatures.

The greatest disadvantage is the need to store full many-body vectors. Calculations are therefore restricted by the available memory to relatively small systems. To minimize the effects of finite system size it is therefore crucial to (i) efficiently use the vast distributed memories of current massively parallel machines and to (ii) find bath-parametrizations that minimize the effect of truncating it to finite size.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] C. Lanczos: *An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators*, J. Res. Nat. Bur. Stand. **49**, 255 (1950)
- [2] W.E. Arnoldi, Quarterly of Applied Mathematics **9**, 17 (1951)
- [3] C.C. Paige: *The Computation of Eigenvalues and Eigenvectors of Very Large Sparse Matrices* (PhD thesis, London University, 1971)
- [4] L.N. Trefethen and D. Bau III: *Numerical Linear Algebra* (Society for Industrial and Applied Mathematics, Philadelphia, 1997)
- [5] G.H. Golub and C.F. van Loan: *Matrix Computations* (Johns Hopkins University Press, 1996)
- [6] J.K. Cullum and R.A. Willboughby: *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. 1: Theory, Vol. 2: Programs* (Birkhäuser, Boston, 1985)
- [7] R. Haydock: *The Recursive Solution of the Schrödinger Equation*, in H. Ehrenreich, F. Seitz, and D. Turnbull (eds.): *Solid State Physics* **35**, 216 (Academic Press, 1980)
- [8] E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994)
- [9] L.V. Kantorovich, Uspekhi Mat. Nauk, **3**, 89 (1948)
- [10] M.R. Hestenes and W. Karush, J. Res. Nat. Bureau Standards **47**, 45 (1951)
- [11] A.N. Krylov, Izv. Akad. Nauk SSSR, Otd. Mat. Estest. **7**, 491 (1931)
- [12] J. Jaklič and P. Prelovšek, Adv. Phys. **49**, 1 (2000)
M. Aichhorn et al., Phys. Rev. B **67**, 161103 (2003)
- [13] A. Dolfen: *Massively parallel exact diagonalization of strongly correlated systems* (Diploma Thesis, RWTH Aachen University, 2006)
- [14] A. Dolfen, T.L. Luo, and E. Koch, Advances in Parallel Computing **15**, 601 (2008)
- [15] M. Caffarel and W. Krauth, Phys. Rev. Lett. **72**, 1545 (1994)
- [16] Q. Si, M.J. Rozenberg, G. Kotliar, and A.E. Ruckenstein, Phys. Rev. Lett. **72**, 2761 (1994)
- [17] M.J. Rozenberg, G. Moeller, and G. Kotliar, Mod. Phys. Lett. B **8**, 535 (1994)
- [18] G. Sangiovanni, A. Toschi, E. Koch, *et al.*, Phys. Rev. B **73**, 205121 (2006)
- [19] E. Koch, G. Sangiovanni, and O. Gunnarsson, Phys. Rev. B **78**, 115102 (2008)
- [20] W. Kohn, Phys. Rev. **133**, A171 (1964)

9 Hirsch-Fye Quantum Monte Carlo Method for Dynamical Mean-Field Theory

Nils Blümer

Institut für Physik

Johannes Gutenberg-Universität Mainz

Contents

1	Introduction	2
1.1	Dynamical mean-field theory	3
2	Hirsch-Fye QMC solution of the single-impurity Anderson model	5
2.1	Wick's theorem for the discretized impurity problem	5
2.2	Monte Carlo importance sampling	8
2.3	Single-spin flip implementation of the HF-QMC method	10
2.4	Choice of simulation parameters: discretization and number of sweeps	12
3	Achieving DMFT self-consistency	13
3.1	Fourier transformation and splining strategies	14
3.2	Search for solutions, hysteresis, overrelaxation	19
4	Computing observables	21
4.1	Observables: mass renormalization, energetics, and spectra	21
4.2	Estimation of errors and extrapolation $\Delta\tau \rightarrow 0$	24
5	Conclusion and outlook	26
A	Multi-band Hubbard models	28
B	Maximum entropy method	29

1 Introduction

A conventional starting point for the study of strongly correlated electron systems is the Hubbard model, which in its single-band version reads

$$\hat{H} = -t \sum_{\langle ij \rangle, \sigma} (\hat{c}_{i\sigma}^\dagger \hat{c}_{j\sigma} + \text{h.c.}) + U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow} \quad (1)$$

where $\hat{c}_{i\sigma}$ is the fermion annihilation operator at site i with spin $\sigma \in \{\uparrow, \downarrow\}$, $\hat{n}_{i\sigma} = \hat{c}_{i\sigma}^\dagger \hat{c}_{i\sigma}$ is the density of fermions with spin σ , t parameterizes the hopping amplitude, and U is the on-site interaction. Unfortunately, numerical methods for its direct solution are either restricted to one-dimensional cases or suffer in general from severe finite-size errors and/or sign problems. Insight into the physics of higher-dimensional systems, thus, requires the use of additional approximations. The dynamical mean-field theory (DMFT) neglects inter-site correlations by assuming a momentum-independent self-energy; it becomes exact in the limit of infinite coordination number. The DMFT maps the lattice problem onto a single-impurity Anderson model (SIAM), supplemented by a self-consistency condition [1]. Its enormous success within the last 20 years would not have been possible without the availability of controlled numerical solvers for (multi-orbital) SIAMs, in particular of the auxiliary-field Hirsch-Fye quantum Monte Carlo (HF-QMC) algorithm [2].

These lecture notes are aimed at a pedagogical introduction to the HF-QMC based method of solving the DMFT self-consistency problem, i.e., of computing electronic properties of Hubbard-type models (possibly derived *ab initio* for a specific material from density functional theory or using the GW method, see subsequent lecture by K. Held) at the DMFT level. As will become clear in the following, the usage of the HF-QMC impurity solver within the DMFT self-consistency cycle has several important implications that would not arise in HF-QMC solutions of some fixed SIAM. For example, statistical Monte Carlo errors might prevent the detection of metastable DMFT fixed points and can, in general, lead to systematic biases in observables (in all stochastic methods including continuous-time QMC); the formulation in imaginary time (in all QMC methods and some IPT variants) implies that (i) analytic continuation is necessary for obtaining spectral information and (ii) that Fourier transforms have to be used within the DMFT cycle which are especially problematic in the case of a uniform time grid (as in HF-QMC); finally, the discretization error of estimated observables in (conventional) HF-QMC is impacted in a quite irregular way due to the self-consistency (cf. Fig. 17) and makes the location of phase boundaries difficult. Thus, a large part of these lectures will go beyond the HF-QMC algorithm itself; we will try to discuss all issues that a researcher should understand (or at least be aware of) when evaluating or producing DMFT data on the basis of HF-QMC. On the other hand, we wish to avoid unnecessary complexity in the presentation, in particular in the formalism. Therefore, we will mainly restrict the treatment to the single-band case (1) which is representative also of multi-band models with (spin) density-density type interactions for which HF-QMC is most competitive and mainly used, but will also point out generalizations where appropriate.

In the remainder of this introduction, we will briefly recapitulate the DMFT and establish the

notation for the following. We will then discuss the HF-QMC algorithm for computing (an estimate of) the Green function of a given impurity problem (defined by the bath Green function \mathcal{G}), as established by Hirsch and Fye, in Sec. 2. In Sec. 3 we will outline all the steps necessary for achieving a self-consistent DMFT solution on the basis of HF-QMC; thus sections 2 and 3 together contain the essential steps for obtaining Green functions and self-energies for Hubbard-type models at the DMFT level (but with a systematic Trotter error). The computation of observables (such as the energy) and of spectra will follow in Sec. 4; here, we will also fully quantify errors and introduce extrapolation techniques for eliminating the Trotter error. Finally, we conclude and give an outlook in Sec. 5.

1.1 Dynamical mean-field theory

As discussed in the preceding lectures by Vollhardt and Kollar, the absence of momentum dependence in the self-energy greatly simplifies the treatment of the Hubbard model [3, 4]. It allows, in fact, to single out one of the lattice sites and replace the influence of its neighbors by the interaction with a single, frequency-dependent bath, i.e., map the Hubbard model onto a single impurity Anderson model (SIAM) in the limit of large coordination number. In order to restore the periodicity of the original lattice, this medium has to be determined self-consistently [1, 5–7]. Written in terms of fermionic Matsubara frequencies¹ $\omega_n = (2n + 1)\pi T$, self-energy $\Sigma_{\sigma n} \equiv \Sigma_{\sigma}(i\omega_n)$, and Green function $G_{\sigma n} \equiv G_{\sigma}(i\omega_n)$ as well as its Fourier transform $G_{\sigma}(\tau)$ (cf. subsection 3.1) the resulting coupled equations read

$$G_{\sigma n} = \int_{-\infty}^{\infty} d\varepsilon \frac{\rho(\varepsilon)}{i\omega_n + \mu - \Sigma_{\sigma n} - \varepsilon} \quad (2)$$

$$G_{\sigma}(\tau) = -\langle T_{\tau} \psi_{\sigma}(\tau) \psi_{\sigma}^*(0) \rangle_{\mathcal{A}}, \quad (3)$$

in the homogeneous phase. Here, T_{τ} is the time ordering operator; properties of the lattice only enter via the density of states (DOS) (for dispersion $\epsilon_{\mathbf{k}}$ and volume V_B of the Brillouin zone)

$$\rho(\epsilon) = \frac{1}{V_B} \int d\mathbf{k} \delta(\epsilon - \epsilon_{\mathbf{k}}) \quad (4)$$

of the noninteracting electrons. The thermal average $\langle \hat{C} \rangle_{\mathcal{A}}$ of some observable \hat{C} is defined as a functional integral over Grassmann variables² ψ, ψ^* (with differentials denoted by \mathcal{D})

$$\langle \hat{C} \rangle_{\mathcal{A}} = \frac{1}{\mathcal{Z}} \int \mathcal{D}[\psi] \mathcal{D}[\psi^*] C[\psi, \psi^*] e^{\mathcal{A}[\psi, \psi^*, \mathcal{G}]}, \quad (5)$$

using the partition function

$$\mathcal{Z} = \int \mathcal{D}[\psi] \mathcal{D}[\psi^*] e^{\mathcal{A}[\psi, \psi^*, \mathcal{G}]}, \quad (6)$$

¹Here, we choose an imaginary-time formulation in anticipation of its use in the context of the imaginary-time quantum Monte Carlo algorithm to be discussed in Sec. 2.

²Grassmann variables are anticommuting numbers for which integration and differentiation rules are defined [8, 9]. Strictly speaking, ψ and ψ^* are (independent) Grassmann fields, i.e., infinite-dimensional vectors with the discrete index σ and the continuous index/argument τ (or t in a real-time formulation). Thus, $\psi_{\uparrow}(\tau_i)$, $\psi_{\downarrow}(\tau_i)$, $\psi_{\uparrow}^*(\tau_i)$, and $\psi_{\downarrow}^*(\tau_j)$ are all linearly independent Grassmann variables for $\tau_i \neq \tau_j$.

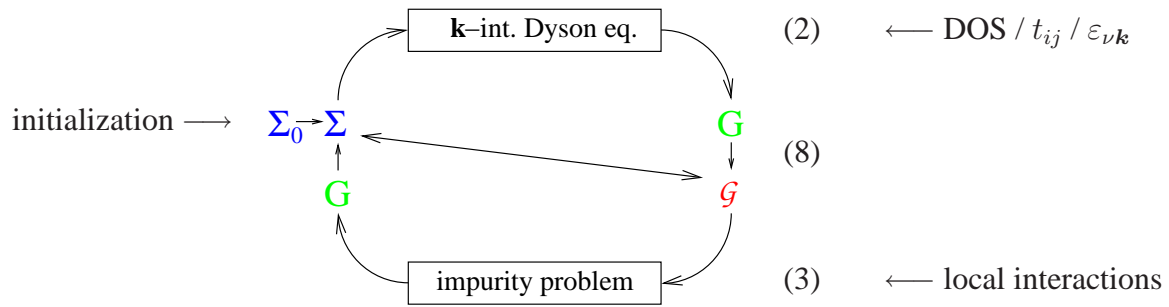


Fig. 1: DMFT self-consistency cycle in conventional form: starting, e.g., with an initial guess for self-energy Σ , the \mathbf{k} integrated Dyson equation (2) yields the lattice Green function G . Both Σ and G are used to compute the bath Green function \mathcal{G} via (8) which defines the impurity problem (3). Its solution using QMC (or IPT, NCA, ED, NRG, etc.) provides a new estimate for G . The cycle is then closed by application of (8) to the new G and the old Σ .

and the single-site action

$$\mathcal{A}[\psi, \psi^*, \mathcal{G}] = \int_0^\beta \int_0^\beta d\tau d\tau' \sum_\sigma \psi_\sigma^*(\tau) \mathcal{G}_\sigma^{-1}(\tau, \tau') \psi_\sigma(\tau') - U \int_0^\beta d\tau \psi_\uparrow^*(\tau) \psi_\uparrow(\tau) \psi_\downarrow^*(\tau) \psi_\downarrow(\tau). \quad (7)$$

Here, $\beta = 1/k_B T$ is the inverse temperature (in the following, we set $k_B \equiv 1$) and

$$\mathcal{G}_{\sigma n}^{-1} = G_{\sigma n}^{-1} + \Sigma_{\sigma n} \quad (8)$$

is the effective local propagator.³ Functions related by a Fourier transformation (here from imaginary time τ to fermionic Matsubara frequencies ω_n or vice versa) are denoted by the same symbol, but can be distinguished by their indices.

The solution of the DMFT problem by iteration is illustrated in Fig. 1. Here, the solution of the \mathbf{k} -integrated Dyson equation (2) is straightforward; it can be performed analytically for the semi-elliptic Bethe DOS commonly used for model studies in the literature. In contrast, the solution of the impurity problem (3) is highly nontrivial.⁴ Most numerical methods developed for the treatment of SIAMs with fixed bath could be adapted to the DMFT problem, e.g., solutions based on exact diagonalization (ED) [1, 10], the non-crossing approximation (NCA) [11–14], the fluctuation-exchange approximation (FLEX) [15–17], the numerical renormalization group (NRG) [18–21], density-matrix renormalization group (DMRG) [22] and quantum Monte Carlo (QMC) algorithms. We will focus on the Hirsch-Fye QMC method in this lecture; recently developed diagrammatic continuous-time QMC algorithms will be covered by Philipp Werner.

³ $\mathcal{G}_{\sigma n}$ may be regarded as a Weiss field in a (frequency-dependent) generalization of the usual static mean field. In contrast with spin models, for which the Weiss field replaces all (NN) interactions, the bath propagator replaces the hybridization of one site with the rest of the lattice.

⁴An exception is the application of (2) and (3) to a Lorentzian DOS $\rho(\epsilon) = t/(\pi(\epsilon^2 + t^2))$ which can be realized on lattices with long range hopping [6]. For this DOS (which is clearly pathological due to its infinite variance), the Weiss function is independent of U ; furthermore (3) is solvable by Bethe ansatz in this case so that many properties can be obtained analytically.

2 Hirsch-Fye QMC solution of the single-impurity Anderson model

In this section, we will discuss the (original)⁵ auxiliary-field quantum Monte Carlo (QMC) algorithm for solving the quantum impurity problem (3). It was originally formulated by Hirsch and Fye for treating a small number of magnetic impurities in metals [2] and later applied to arbitrary hybridization functions, i.e., in the form required for the solution of the DMFT problem [5, 24–26]. Here, we will concentrate on the solution of the impurity problem and postpone aspects specific to the DMFT context to Sec. 3. For simplicity, we also specialize on the single-band homogeneous case (1); generalizations for the multi-band case (and implications for the minus-sign problem) will be pointed out where appropriate, see also App. A. A very general formulation of the HF-QMC method in the multi-band case, including the cases of interorbital hybridization and complex interactions, can be found in [27].

The functional integral equation for the Green function has the structure

$$G_\sigma(\tau_1 - \tau_2) = -\frac{1}{\mathcal{Z}} \int \mathcal{D}[\psi] \mathcal{D}[\psi^*] \psi_\sigma(\tau_1) \psi_\sigma^*(\tau_2) \exp \left[\mathcal{A}_0 - U \int_0^\beta d\tau \psi_\uparrow^* \psi_\uparrow \psi_\downarrow^* \psi_\downarrow \right], \quad (9)$$

where \mathcal{A}_0 denotes the hybridization part, i.e., the first term in (7). For $U = 0$, only quadratic terms in operators or Grassmann variables occur; then, this expression can be solved using Wick's theorem. The idea of the Hirsch-Fye QMC method is to transform the interaction term (quartic in operators or Grassmann variables) to a quadratic term and to solve the resulting problem, again, by Wicks theorem. Such a Hubbard Stratonovich transformation requires, however, a decoupling of interacting and noninteracting terms which only commute in the limit of infinite temperatures. Therefore, HF-QMC has to include some kind of numerical high-temperature expansion by Trotter decomposition. A high-level overview of the full scheme is shown in Fig. 2, which illustrates (in step iii) that the interaction between electrons with different spin (and/or different orbitals in the multi-band case) is replaced by the interaction of each electron with a binary auxiliary field. For each of the configurations the path integral in (9) evaluates to a term of the structure $M^{-1} \det\{M\}$ where the Matrix M depends onto the field configuration $\{s\}$. This analytic part will be discussed in the following subsection 2.1. We will then introduce the concept of Monte Carlo importance sampling in a general way in subsection 2.2 before applying it in the HF-QMC context in subsection 2.3.

2.1 Wick's theorem for the discretized impurity problem

The difficulty in solving the functional integral equation (3) arises from the noncommutativity of the kinetic term and the interaction term in the single-site action (7). These terms can be

⁵Here and in the following, we discuss the original QMC version using a discretization of the imaginary time and based on Trotter decomposition and Hubbard-Stratonovich transformation. It should not be confused with the recent auxiliary-field formulation of the weak-coupling continuous-time QMC method often referred to as CT-AUX [23].

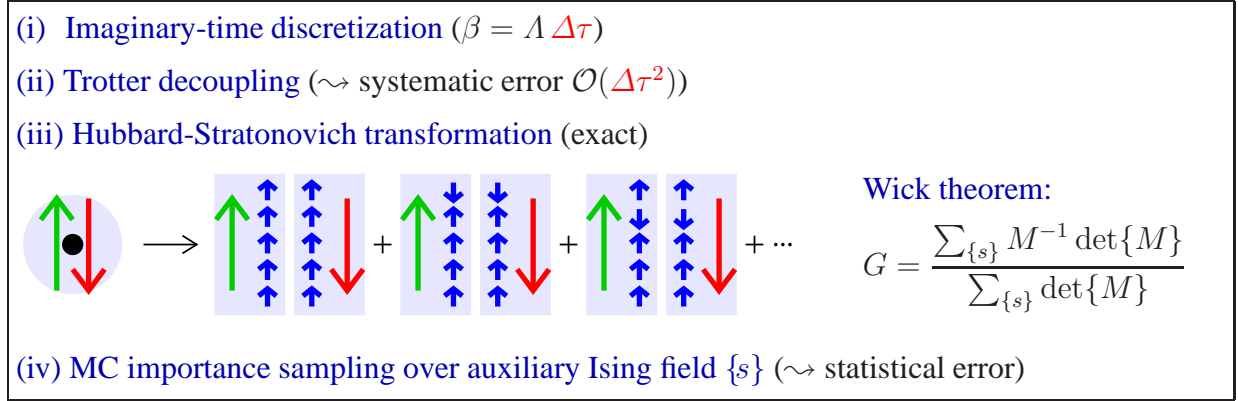


Fig. 2: High-level overview of the HF-QMC approach.

separated by use of the Trotter-Suzuki formula [28, 29] for operators \hat{A} and \hat{B} :

$$e^{-\beta(\hat{A}+\hat{B})} = (e^{-\Delta\tau\hat{A}} e^{-\Delta\tau\hat{B}})^\Lambda + \mathcal{O}(\Delta\tau), \quad (10)$$

where $\Delta\tau = \beta/\Lambda$ and Λ is the number of (imaginary) time slices.⁶ Rewriting the action (7) in discretized form

$$\begin{aligned} \mathcal{A}_\Lambda[\psi, \psi^*, \mathcal{G}, U] &= (\Delta\tau)^2 \sum_{\sigma} \sum_{l, l'=0}^{\Lambda-1} \psi_{\sigma l}^* (\mathcal{G}_{\sigma}^{-1})_{ll'} \psi_{\sigma l'} \\ &\quad - \Delta\tau U \sum_{l=0}^{\Lambda-1} \psi_{\uparrow l}^* \psi_{\uparrow l} \psi_{\downarrow l}^* \psi_{\downarrow l}, \end{aligned} \quad (11)$$

where the matrix \mathcal{G}_{σ} consists of elements $\mathcal{G}_{\sigma ll'} \equiv \mathcal{G}_{\sigma}(l\Delta\tau - l'\Delta\tau)$ and $\psi_{\sigma l} \equiv \psi_{\sigma}(l\Delta\tau)$, we apply (10) and obtain to lowest order

$$\begin{aligned} \exp(\mathcal{A}_\Lambda[\psi, \psi^*, \mathcal{G}, U]) &= \prod_{l=0}^{\Lambda-1} \left[\exp\left((\Delta\tau)^2 \sum_{\sigma} \sum_{l'=0}^{\Lambda-1} \psi_{\sigma l}^* (\mathcal{G}_{\sigma}^{-1})_{ll'} \psi_{\sigma l'} \right) \right. \\ &\quad \left. \times \exp(-\Delta\tau U \psi_{\uparrow l}^* \psi_{\uparrow l} \psi_{\downarrow l}^* \psi_{\downarrow l}) \right]. \end{aligned} \quad (12)$$

Shifting the chemical potential by $U/2$, the four-fermion term can be rewritten as a square of the magnetization (in terms of operators: $(\hat{n}_{\uparrow} - \hat{n}_{\downarrow})^2 = \hat{n}_{\uparrow}^2 + \hat{n}_{\downarrow}^2 - 2\hat{n}_{\uparrow}\hat{n}_{\downarrow} = \hat{n}_{\uparrow} + \hat{n}_{\downarrow} - 2\hat{n}_{\uparrow}\hat{n}_{\downarrow}$) which makes it suitable for the following discrete Hubbard-Stratonovich transformation [30] (the correctness of which is checked easily by inserting the four possible combinations $[(0, 0), (0, 1), (1, 0), (1, 1)]$ of eigen values of the associated density operators):

$$\exp\left(\frac{\Delta\tau U}{2}(\psi_{\uparrow l}^* \psi_{\uparrow l} - \psi_{\downarrow l}^* \psi_{\downarrow l})^2\right) = \frac{1}{2} \sum_{s_l = \pm 1} \exp(\lambda s_l (\psi_{\uparrow l}^* \psi_{\uparrow l} - \psi_{\downarrow l}^* \psi_{\downarrow l})), \quad (13)$$

⁶Since $\beta = 1/k_B T$ is the inverse temperature, small $\Delta\tau$ on each “time slice” corresponds to a higher temperature, for which the operators effectively decouple. Thus, we may view the Trotter approach as a numerical extension of a high-temperature expansion to lower temperatures.

with $\cosh \lambda = \exp(\Delta\tau U/2)$. Here, the interaction between electrons is replaced by the interaction with an auxiliary binary field $\{s\}$ with components $s_l = \pm 1$ for $0 \leq l \leq \Lambda$. Acting like a local, but time-dependent magnetic field, $\{s\}$ can be regarded as an ensemble of Ising spins (as depicted in Fig. 2).

Applying the Trotter formula again, these transformations yield an expression for the functional integral

$$G_{\sigma l_1 l_2} = \frac{1}{\mathcal{Z}} \sum_{\{s\}} \int \mathcal{D}[\psi] \mathcal{D}[\psi^*] \psi_{\sigma l_1}^* \psi_{\sigma l_2} \exp \left(\sum_{\sigma, l, l'} \psi_{\sigma l}^* M_{\sigma l l'}^{s_l} \psi_{\sigma l'} \right), \quad (14)$$

with⁷

$$M_{\sigma l l'}^{s_l} = (\Delta\tau)^2 (\mathcal{G}_{\sigma}^{-1})_{ll'} - \lambda \sigma \delta_{ll'} s_l, \quad (15)$$

where in (14) the sum is taken over all configurations of the Ising spin field, and each term of the sum involves independent fermions only. Now Wick's theorem [32] (or, equivalently, the rules for Gaussian Grassmann integrals) can be applied to get the solution⁸

$$G_{\sigma l l'} = \frac{1}{\mathcal{Z}} \sum_{\{s\}} (M_{\sigma}^{\{s\}})^{-1}_{ll'} \det M_{\uparrow}^{\{s\}} \det M_{\downarrow}^{\{s\}}, \quad (16)$$

where $M_{\sigma}^{\{s\}}$ is the matrix with elements $M_{\sigma l l'}^{s_l}$, and the partition function has the value

$$\mathcal{Z} = \sum_{\{s\}} \det M_{\uparrow}^{\{s\}} \det M_{\downarrow}^{\{s\}}. \quad (17)$$

Equations (16) and (17) already constitute the solution of the impurity problem, as the physical problem has been reduced to algebra which can, in principle, be evaluated exactly for an arbitrary number Λ of time slices. In fact, such so-called full summation can be useful in some contexts (even though the number of terms is exponential). However, the independent computation of the matrix inverses and determinants for each auxiliary field (operations scaling with $\mathcal{O}(\Lambda^3)$ and worse than $\mathcal{O}(\Lambda^4)$, respectively) would be highly inefficient.

Fast update scheme

At least for the case of full summation (i.e., when all the terms are explicitly calculated) one may arrange the sum over the auxiliary spins in (16) so that only one auxiliary spin s_m is flipped between subsequent configuration (Gray code). For the matrices $M_{\sigma}^{\{s\}}$ (with components l_1, l_2) this means [33]:

$$\begin{aligned} M_{\sigma} \xrightarrow{s_m \rightarrow -s_m} M_{\sigma}' &= M_{\sigma} + \Delta^{\sigma m} \\ &= (1 + \Delta^{\sigma m} (M_{\sigma})^{-1}) M_{\sigma} \end{aligned} \quad (18)$$

$$\text{with} \quad \Delta_{ll'}^{\sigma m} = \delta_{ll'} \delta_{lm} 2\Delta\tau \lambda \sigma s_l. \quad (19)$$

⁷A more precise form including subleading corrections is $M_{\sigma l l'}^{s_l} = (\Delta\tau)^2 (\mathcal{G}_{\sigma}^{-1})_{ll'} e^{\lambda \sigma s_l} + \delta_{ll'} (1 - e^{\lambda \sigma s_l})$ [1, 31, 27].

⁸The form shown in Fig. 2 is obtained when G_{\uparrow} and G_{\downarrow} and, correspondingly, M_{\uparrow} and M_{\downarrow} are each arranged as the block-diagonal super-matrices G and M , respectively.

The ratio of the determinants of new and old matrix can be easily determined using the inverse of the old matrix:⁹

$$\begin{aligned} R^{\sigma m} &:= \frac{\det(\mathbf{M}_{\sigma}')}{\det(\mathbf{M}_{\sigma})} = \det(\mathbf{1} + \Delta^{\sigma m}(\mathbf{M}_{\sigma})^{-1}) \\ &= 1 + 2\Delta\tau \lambda \sigma s_m (M_{\sigma})_{mm}^{-1}. \end{aligned} \quad (20)$$

The inversion of \mathbf{M} is also elementary, one obtains:

$$(\mathbf{M}_{\sigma}')^{-1} = (\mathbf{M}_{\sigma})^{-1} + \frac{1}{R^{\sigma m}} (\mathbf{M}_{\sigma})^{-1} \Delta^{\sigma m} (\mathbf{M}_{\sigma})^{-1}. \quad (21)$$

This reduces the effort for the recalculation of a term of (16) after a spin flip to $\mathcal{O}(\Lambda^2)$.

Only for $\Lambda \lesssim 30$ can all terms be summed up exactly. Computations at larger Λ are made possible by Monte Carlo importance sampling which reduces the number of terms that have to be calculated explicitly from 2^{Λ} to order $\mathcal{O}(\Lambda)$.

2.2 Monte Carlo importance sampling

Monte Carlo (MC) procedures in general are stochastic methods for estimating large sums (or high-dimensional integrals) by picking out a comparatively small number of terms (or evaluating the integrand only for a relatively small number of points). Let us assume we want to compute the average $X := \frac{1}{M} \sum_{l=1}^M x_l$, where l is an index (e.g., an Ising configuration $l \equiv \{s\}$) and x some observable with the (true) variance $v_x = \frac{1}{M} \sum_{l=1}^M (x_l - X)^2$. In a simple MC approach, one may select a subset of $N \ll M$ indices independently with a uniform random distribution $P(l_j) = \text{const.}$ (for $1 \leq j \leq N$),

$$X_{\text{MC}} = \frac{1}{N} \sum_{j=1}^N x_{l_j}, \quad (22)$$

$$\Delta X_{\text{MC}} := \langle (X_{\text{MC}} - X)^2 \rangle = \frac{v_x}{N} \approx \frac{1}{N(N-1)} \sum_{j=1}^N (x_{l_j} - X_{\text{MC}})^2. \quad (23)$$

Here, the averages are taken over all realizations of the random experiment (each consisting of a selection of N indices). In the limit of $N \rightarrow \infty$, the distribution of X_{MC} becomes Gaussian according to the central limit theorem. Only in this limit is the estimate of v_x from the QMC data reliable. An application for a continuous set is illustrated in Fig. 3.

Smaller errors and faster convergence to a Gaussian distribution for the estimate may be obtained by importance sampling. Here, the function x_l is split up,

$$x_l = p_l o_l; \quad p_l \geq 0; \quad \sum_{l=1}^M p_l = c, \quad (24)$$

⁹Since $\Delta^{\sigma m}$ has only one non-zero element, it is clear that only the row m of $\Delta^{\sigma m}(\mathbf{M}_{\sigma})^{-1}$ will contain non-zero elements. The determinant of $1 + \Delta^{\sigma m}(\mathbf{M}_{\sigma})^{-1}$ is then equal to the product of its diagonal elements.

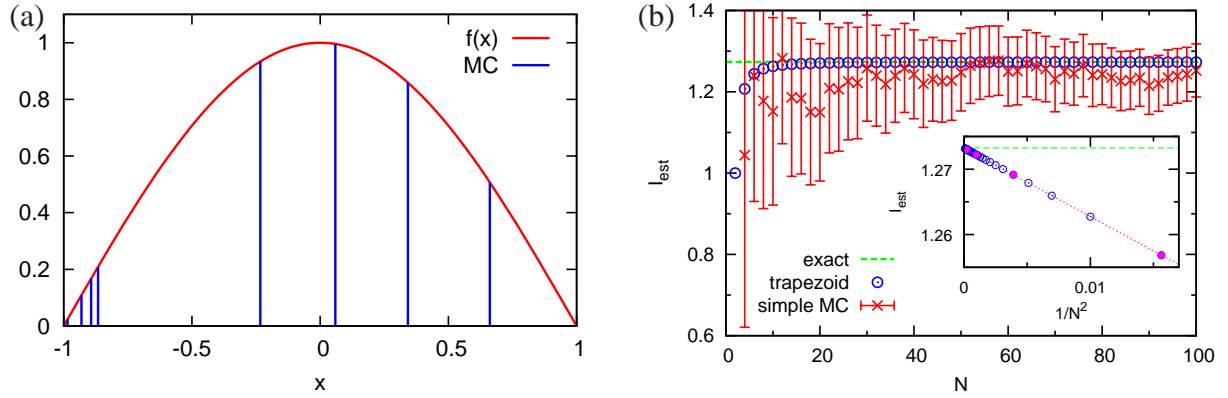


Fig. 3: Illustration of simple MC: (a) instead of using deterministic quadrature schemes such as the trapezoid rule, integrals over some function $f(x)$ can be evaluated by averaging over function values $f(x_i)$ of stochastically generated support points x_i (and multiplying by the integration volume). (b) The resulting statistical error decreases as $1/\sqrt{N}$ with the number of function evaluations (crosses and error bars). The deterministic trapezoid discretization scheme (circles) converges with an error proportional to $N^{-2/d}$, i.e., is superior in dimensions $d < 4$; furthermore, an extrapolation of the systematic error is straightforward (inset).

where we may regard p_l as a (unnormalized) probability distribution for the indices and o_l as a remaining observable. If both the normalization c is known (i.e., the sum over the weights p_l can be performed exactly) and the corresponding probability distribution can be realized (by drawing indices l with probability $P(l) = p_l/c$), we obtain

$$X_{\text{imp}}^{\text{MC}} = \frac{c}{N} \sum_{j=1}^N o_{l_j} \quad \text{and} \quad \Delta X_{\text{imp}}^{\text{MC}} = c \sqrt{\frac{v_o}{N}}. \quad (25)$$

Thus, the error can be reduced ($v_o < c^2 v_x$), when the problem is partially solvable, i.e., the sum over p_l with $p_l \approx x_l$ can be computed.¹⁰ Since this is not possible in general, one usually has to treat the normalization c as an unknown and realize the probability distribution $P(l_j) = p_{l_j}/c$ in a stochastic Markov process: Starting with some initial configuration l_1 , a chain of configurations is built up where in each step only a small subset of configurations l' is accessible in a “transition” from configuration l . Provided that the transition rules satisfy the detailed balance principle,

$$p_l \mathcal{P}(l \rightarrow l') = p_{l'} \mathcal{P}(l' \rightarrow l), \quad (26)$$

and the process is ergodic (i.e., all configurations can be reached from some starting configuration), the distribution of configurations of the chain approaches the target distribution in the limit of infinite chain length, as illustrated in Fig. 4. Since the normalization remains unknown, importance sampling by a Markov process can only yield ratios of different observables evaluated on the same chain of configurations. Another consequence of using a Markov process is that initial configurations have to be excluded from averages since the true associated probabilities might be vanishingly small. They would otherwise be overrepresented in any run of

¹⁰The possible reduction of the variance is limited when x_l is of varying sign. This “minus-sign problem” seriously restricts the applicability of Monte Carlo methods for finite-dimensional fermion problems.

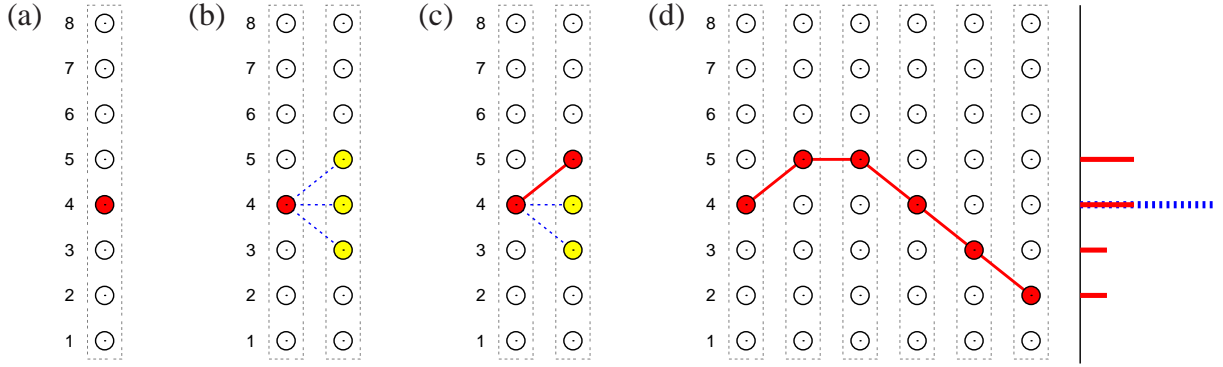


Fig. 4: Illustration of importance sampling Monte Carlo, here for 8 states: (a) the simulation is initialized by randomly choosing a starting configuration, here state 4. (b) Transitions to the neighborhood of the old state are proposed, here for states $s_{new} = s_{old} \pm 1$. (c) A move (here +1) is accepted. (d) The history of the simulation (here with 5 attempted updates) consists both of accepted and declined updates. The normalized histogram of the visited configurations (red lines) approaches the target probability distribution (only) in the limit of infinite simulation length. For finite runs, states in the vicinity of the starting configuration may be overrepresented; this “warm-up” part should be discarded.

finite length. Consequently, we will later distinguish “warm-up sweeps” from “measurement sweeps”.

For the computation of errors, one has to take into account the finite autocorrelation induced by the Markov process, i.e., correlation between subsequent measurements. This correlation may be characterized by the autocorrelation time¹¹ $\kappa_o \geq 1$ which effectively reduces the number of independent samples, so that $\Delta X = c\sqrt{v_o\kappa_o/N}$. The numerical effort necessary to reach some target statistical accuracy ΔX nevertheless increases only as $(1/\Delta X)^2$.

2.3 Single-spin flip implementation of the HF-QMC method

Returning to the evaluation of the Green function using (16) and (17), the obvious choice is to sample configurations $\{s\}$ according to the (unnormalized) probability

$$P(\{s\}) = \left| \det \mathbf{M}_{\uparrow}^{\{s\}} \det \mathbf{M}_{\downarrow}^{\{s\}} \right|. \quad (27)$$

The Green function can then be calculated as an average $\langle \dots \rangle_s$ over these configurations:

$$G_{\sigma ll'} = \frac{1}{\tilde{\mathcal{Z}}} \left\langle (\mathbf{M}_{\sigma}^{\{s\}})^{-1}_{ll'} \text{sign} \left(\det \mathbf{M}_{\uparrow}^{\{s\}} \det \mathbf{M}_{\downarrow}^{\{s\}} \right) \right\rangle_s, \quad (28)$$

$$\tilde{\mathcal{Z}} = \left\langle \text{sign} \left(\det \mathbf{M}_{\uparrow}^{\{s\}} \det \mathbf{M}_{\downarrow}^{\{s\}} \right) \right\rangle_s. \quad (29)$$

Here, $\tilde{\mathcal{Z}}$ deviates from the full partition function by an unknown prefactor which cancels in (28). The same is true for other expectation values of the form (5). The inability to compute

¹¹For a set $\{o_1, o_2, \dots, o_N\}$ of measurements, the autocorrelation function (for the observable o) is $c_l^o = \langle (o_k - \langle o \rangle)(o_{k+l} - \langle o \rangle) \rangle_k$. An associated autocorrelation time may then be defined as $\kappa_o = c_0^o + 2 \sum_{l=1}^{N_0} c_l^o$, where the cutoff N_0 is determined by $c_l^o > 0$ for $l \leq N_0$ and $c_{N_0+1}^o < 0$.

the partition function itself (and, consequently, the free energy and entropy) is a consequence of the importance sampling and is thus a general characteristic of QMC methods which has severe consequences for the study of phase transitions.

If the sign in (28) does not depend on $\{s\}$, i.e. if the so-called sign problem is absent (as in HF-QMC for density-type interactions), the expressions can be simplified:

$$G_{\sigma ll'} = \frac{1}{\tilde{Z}} \left\langle \left(M_{\sigma}^{\{s\}} \right)_{ll'}^{-1} \right\rangle_s, \quad \tilde{Z} = \langle 1 \rangle_s. \quad (30)$$

A practical HF-QMC calculation proceeds as follows: (i) A starting configuration $\{s\}$ is chosen, e.g., taken from a uniform random distribution or from a previous run with similar parameters. (ii) The initial Green function matrix M^{-1} is computed for this configuration from scratch. Note that the associated determinant is not needed. (iii) A number N_{warm} of warm-up sweeps (see below) is performed in order to generate a Markov chain; intermediate Green functions M^{-1} are discarded. (iv) A (larger) number N_{meas} of measurement sweeps is performed, accumulating the sum of all occurring matrices M^{-1} (and possibly of other observables). (v) The estimate of the Green function and of observables is computed by dividing each of the sums by the number of terms, i.e., by the number of attempted configuration updates.

Each of the sweeps mentioned above consists of a loop in which Λ spins are selected for a possible update (either in sequence or randomly). Within the loop, the ratio of determinants for each candidate spin flip is evaluated using (20) and accepted or rejected according to a detailed balance principle. Often, the acceptance probability is chosen according to the Metropolis transition rule [34]

$$\mathcal{P}(\{s\} \rightarrow \{s'\}) = \min\{1, P(\{s'\})/P(\{s\})\}; \quad (31)$$

an alternative is the symmetric heat-bath algorithm.

Roughly, the number of sweeps has to be held constant (rather than the number of attempted single-spin flips) for constant statistical error when the number of spins Λ is varied. Since the systematic error depends on $\Delta\tau$, the necessary number of time slices is proportional to the inverse temperature, $\Lambda \propto \beta$, so that the total numerical cost for given accuracy is proportional to β^3 .

Memory issues and parallelization

The numerically costly part of HF-QMC consists of the update of the Green function matrix (and the associated determinant), specifically of the addition of a dyadic (outer) product of two vectors (dimension Λ) onto a $\Lambda \times \Lambda$ matrix (for each spin and orbital). This BLAS-1 operation (DGER) breaks down to elementary multiply-add pairs where, in general, all operands of subsequent operations are different and have to be loaded from memory. Thus, an efficient execution depends on high memory bandwidth; in practice, this requires that the matrices can be held in the processor cache (typically L2). Fortunately, usual cache sizes (of a few megabytes) suffice for many applications; our code can use fine-grained OpenMP parallelization in order to extend these limits (by spreading the matrices across multiple processor cores and caches). Another strategy for optimizing memory issues is the use of delayed updates [35, 36].

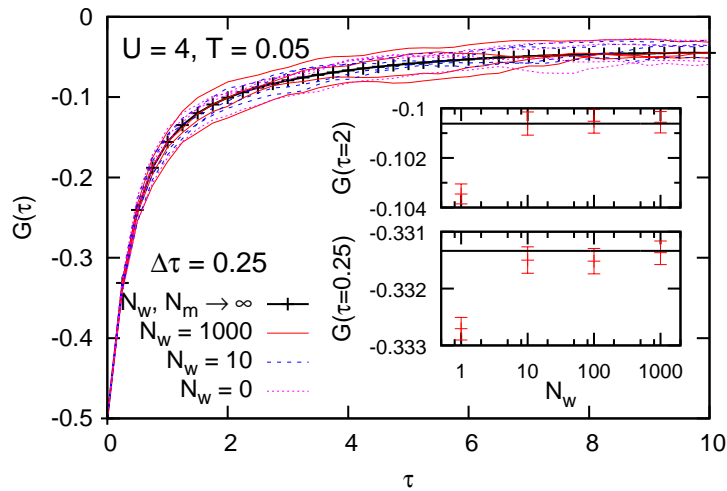


Fig. 5: *Impact of proper warm-up: Green functions estimated from short HF-QMC runs with varying number N_w of warm-up sweeps (see text).*

We also use MPI for the parallel solution of identical impurity problems, with near-perfect scaling up to some 20–30 MPI tasks (and reasonable scaling beyond 100 tasks). These limits could easily be extended by at least a factor of 10 by amortizing the warm-up of the auxiliary field over several (typically 20) DMFT iterations. It should be noted that the averaging over multiple independent solutions also stabilizes the procedure.

2.4 Choice of simulation parameters: discretization and number of sweeps

The important practical question how the simulation parameters should be chosen is most easily answered for the number of measurement sweeps, at least on the SIAM level (not taking the DMFT self-consistency into account): since the error scales as N_{meas} for each observable (including the Green function), it can be determined from the desired precision.

A much harder question is the appropriate value of N_{warm} . A too small value should be avoided at all costs since it will lead to a systematic bias which is near-impossible to detect in the result data of a single run.¹² On the other hand, a too large value wastes resources. We typically use $N_{\text{warm}} \geq 1000$; in high-precision runs, we devote 10% of the sweeps of each serial runs to equilibration (possibly an overkill which, however, costs only 5% in statistical precision). In order to quantify the impact of the warm-up at least for one test case, we have performed a large number of independent simulations at fixed $\Delta\tau$ and for fixed hybridization function \mathcal{G} with N_{warm} ranging from 1 to 1000. In order to see the effect, these runs have to be short; we have chosen $N_{\text{meas}} = 100$. This, on the other hand, gives rise to very large fluctuations in the measured Green functions as shown in the main panel in Fig. 5. At this level, a possible effect of N_{warm} is hidden in the noise. Averaging over a large number (here 1000) of realizations, however, allows

¹²However, one might keep track of the relative probabilities of the visited configurations (or rather of their logarithm) and derive a cutoff from the number of sweeps needed for reaching a typical probability level for the first time.

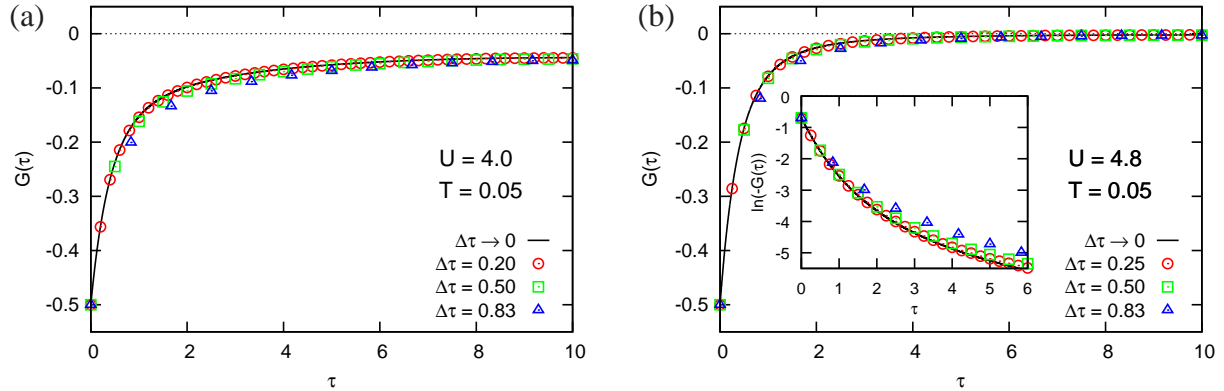


Fig. 6: Impact of the Trotter discretization on HF-QMC estimates of the Green function at $n = 1$ and $T = 0.05$ for $U = 4.0$ (a) and $U = 4.8$ (b), respectively. Only results for $\tau \leq \beta/2$ are shown; the rest follows from particle-hole symmetry $G(\tau) = G(\beta - \tau)$. For each case, $\Delta\tau$ -independent hybridization functions (corresponding to exact DMFT solutions obtained from multigrid HF-QMC) have been used.

to characterize the probability distributions associated with the selected 100 sweeps (offset by 0, 10, 100, or 1000 warm-up sweeps from the random initialization of the Markov chain). The surprising result, shown in the inset for the Green function at two representative grid points, is that 10 warm-up sweeps suffice for removing any detectable equilibration bias. One may suspect that the equilibration times are much longer for insulating or ordered phases; otherwise, significant savings in computer time would appear possible. Fig. 5 also illustrates that imprecise HF-QMC estimates of Green functions can violate causality, i.e., be non-monotonous.

For a sufficiently large number of sweeps, the error of the resulting Green functions is typically dominated by the systematic Trotter error. Such systematic shifts, consistent with a quadratic dependence on $\Delta\tau$, are seen in Fig. 6 for hybridization functions corresponding to metallic and insulating DMFT solutions, respectively. At the same time, the grids vary for each discretization which (at least at first sight) seems to limit the low- τ resolution for large values of $\Delta\tau$. The inset in Fig. 6b shows that the relative Trotter error increases for $\tau \rightarrow \beta/2$ in the insulating case. Still, the qualitative shape of the Green functions remains unchanged even for extremely large values of $\Delta\tau$ in both cases and it is hard to derive an intrinsic limit for reasonable values of $\Delta\tau$ on this level.

3 Achieving DMFT self-consistency

In the main part of this section, we will discuss two aspects of the quantum Monte Carlo (QMC) algorithm for the DMFT self-consistency problem: the discrete Fourier transformations between imaginary time and Matsubara frequencies and the search strategy for solutions of the DMFT equations. Both aspects are particularly relevant, e.g., for the detection of the Mott transition and the associated coexistence region.

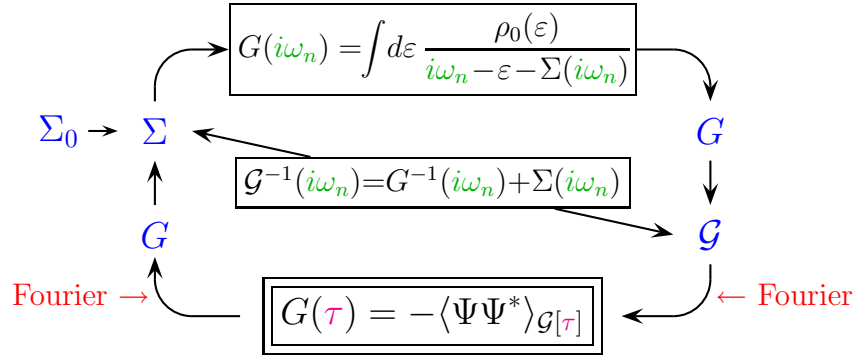


Fig. 7: Schematic DMFT self-consistency scheme showing the two Fourier transformations needed per iteration.

3.1 Fourier transformation and splining strategies

A QMC simulation within the DMFT framework consists of a simultaneous solution of two principal equations: the lattice Dyson equation (2) and the defining expression for the impurity Green function (3). The IPT (iterative perturbation theory [6, 37]) and QMC solutions of the impurity problem are formulated in imaginary time, i.e., the bath Green function \mathcal{G} is needed as function of τ and the result is expressed as $G(\tau)$. In contrast, the Dyson equation is formulated (and is local) in the frequency domain, here for Matsubara frequencies $i\omega_n = i(2n + 1)\pi T$. Therefore, two Fourier transformations (from frequency to imaginary time and vice versa) per self-consistency cycle are necessary (as shown in Fig. 7), which for G read

$$G(i\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} G(\tau) \quad (32)$$

$$G(\tau) = \frac{1}{\beta} \sum_{n=-\infty}^{\infty} e^{-i\omega_n \tau} G(i\omega_n). \quad (33)$$

Note that (33) implies antiperiodicity of $G(\tau)$ for translations β (since $e^{i\omega_n \beta} = -1$) and allows for a discontinuity of $G(\tau)$ (at $\tau = 0$) since the number of terms is infinite. The spectral representation of G

$$G(\omega) = \int_{-\infty}^{\infty} d\omega' \frac{A(\omega')}{\omega - \omega'}; \quad A(\omega) = -\frac{1}{\pi} \text{Im} G(\omega + i0^+)$$

implies a decay of $G(i\omega_n)$ as $1/i\omega_n$ for $|n| \rightarrow \infty$. Furthermore, $G(i\omega_n)$ is purely imaginary when $G(\tau) = G(\beta - \tau)$ as in the case of half band filling on a bipartite lattice (with symmetric DOS).

Discretization problem, naive Fourier transformation

Numerically, however, the integral in (32) needs to be discretized and the Matsubara sum in (33) has to be truncated. Since the numerical effort in QMC scales with the number $\Lambda = \beta/\Delta\tau$ of

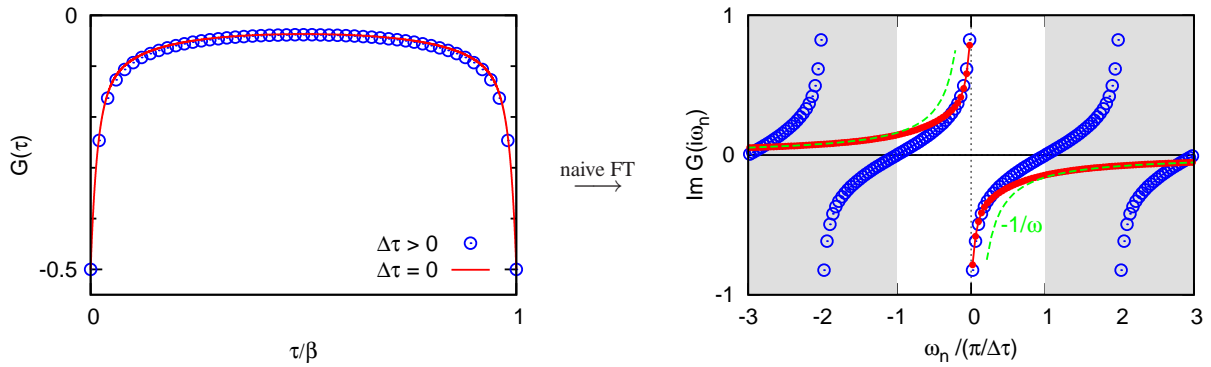


Fig. 8: Illustration of the Fourier transformation (FT) from imaginary time to Matsubara frequencies, based on HF-QMC data for $U = 4$ and $T = 0.04$. The naive discrete FT (open circles) shows oscillatory behavior with poles or zeros at multiples of the Nyquist frequency. Advanced spline-based methods (filled circles and solid line), in contrast, recover the correct large-frequency asymptotics (dashed line).

discretized time slices at least¹³ as Λ^3 , this method is presently restricted to $\Lambda \lesssim 400$. Typically, 200 time slices and less are used. A naive discrete version of the Fourier transform,

$$\tilde{G}(i\omega_n) = \Delta\tau \left(\frac{G(0) - G(\beta)}{2} + \sum_{l=1}^{\Lambda-1} e^{i\omega_n \tau_l} G(\tau_l) \right); \quad \tau_l := l\Delta\tau \quad (34)$$

$$\tilde{G}(\tau_l) = \frac{1}{\beta} \sum_{n=-\Lambda/2}^{\Lambda/2-1} e^{-i\omega_n \tau_l} G(i\omega_n), \quad (35)$$

fails for such a coarse grid. The problems are that the Green function $\tilde{G}(\tau)$ estimated from a finite Matsubara sum cannot be discontinuous at $\tau = 0$ (as required analytically for $G(\tau)$ and also for $\mathcal{G}(\tau)$) while the discrete estimate $\tilde{G}(i\omega_n)$ oscillates with periodicity $2\pi i\Lambda/\beta$ instead of decaying for large frequencies. This implies a large error of $G(i\omega_n)$ when $|\omega_n|$ approaches or exceeds the Nyquist frequency $\pi\Lambda/\beta$ as illustrated in Fig. 8. Both (related) effects would make the evaluation of the corresponding self-consistency equations pointless. In particular, in the naive scheme, the self-energy diverges near the Nyquist frequency. Finally, the sum in (35) is numerically somewhat unstable since $\tilde{G}(\tau)$ also oscillates between the grid points τ_l .

Splining method

Fortunately, there is physical information left that has not been used in the naive scheme: $G(\tau)$ is known to be a smooth function in the interval $[0, \beta)$. In fact, it follows from (52) and (53) that $G(\tau)$ and all even-order derivatives are positive definite and, consequently, reach their maxima at the edges $\tau = 0$ and $\tau = \beta$. This knowledge of “smoothness” can be exploited in an interpolation of the QMC result $\{G(\tau_l)\}_{l=0}^{\Lambda}$ by a cubic spline.¹⁴ The resulting functions may

¹³In practice, the scaling is even worse on systems with a hierarchy of memory systems (i.e. caches and main memory) of increasing capacity and decreasing speed.

¹⁴A spline of degree n is a function piecewise defined by polynomials of degree n with a globally continuous $(n-1)^{\text{th}}$ derivative. For a natural cubic spline, the curvature vanishes at the end points. The coefficients of the

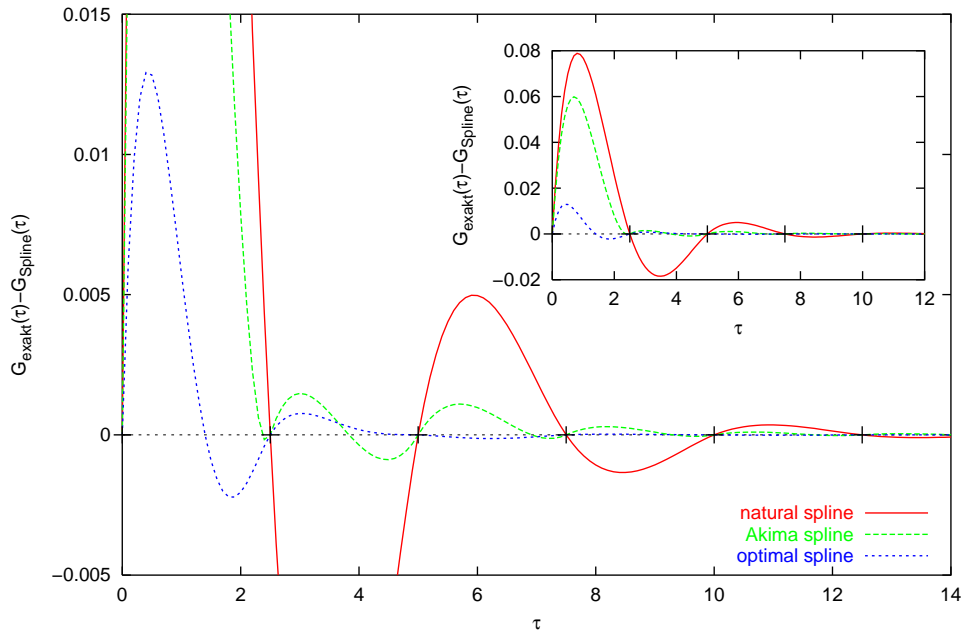


Fig. 9: Difference between exact noninteracting Green function (for semi-elliptic DOS and $\beta = 100$) and cubic spline approximations to the discretized function for $\Lambda = 40$. The strong oscillations observed for the natural spline and for the Akima spline are due to their unphysically vanishing second derivatives (at $\tau = 0$). An optimal choice of the boundary condition reduces the discrepancies by an order of magnitude and doubles the oscillation frequency.

then either be used for oversampling, i.e., for generating $G(\tau)$ on a finite grid or for piecewise direct analytic Fourier transforms. In both cases, $G(i\omega_n)$ can be calculated for a much larger frequency range than before.

The simplest practical approach in this context consists of a direct interpolation of the discrete QMC data by a cubic spline (with a continuous second derivative) as implemented by Krauth [1]. While this step suffices for closing the self-consistency equations without any further adjustments (like Ulmke's smoothing trick, see below), it leads to nonanalytic behavior of the self-energy near and beyond the Nyquist frequency [38]. This problem can be traced back to the fact that the natural spline chosen by Krauth is inadequate for this problem: By definition, the second derivative of a natural spline vanishes at its boundaries. However, all even derivatives of the true Green function are maximal at the edges of the interval $[0, \beta]$. The resulting misfit leads to unphysical ringing as illustrated in Fig. 9. Here, the noninteracting Green function for a semi-elliptic DOS is chosen as an example since it can be computed with arbitrary precision. Furthermore, moderate interactions do not lead to qualitative changes in the Green function so that the example is representative. It is clearly seen that a fit of a discrete set of data points (here for $\Lambda = 40$ and $\beta = 100$, i.e., a large discretization $\Delta\tau = 2.5$) using a natural cubic spline with continuous second derivative leads to a large error oscillating with the Nyquist frequency. While the initial misfit at $\tau = 0$ is similarly large for a natural Akima spline, the oscillation decays significantly faster due to jumps in the second derivative for this

spline fitting a given data set can be easily computed by solving a linear equation.

fit. The discrepancies can be reduced by about an order of magnitude and a smooth curve can be obtained by choosing an appropriate boundary condition (short-dashed line) which has here been obtained by a numerical minimization procedure for the variation of the second derivatives (see [38]). It is possible to avoid the minimization by computing the correct second derivative $G^{(2)}(0)$ analytically (for arbitrary interaction U):

$$\left. \frac{d^2 G(\tau)}{d\tau^2} \right|_{\tau=0+} = -\frac{1}{2} \left(1 + \frac{U^2}{4} \right); \quad (36)$$

in any case, however, a cubic spline interpolation of the full Green function suffers from the problem that derivatives of fourth and higher order vanish on segments of the splines while they can be large for the true Green function, in particular at $\tau = 0$.¹⁵

Splicing the difference Green function using a model self-energy

A related method for improving on the natural spline scheme is to interpolate not $\{G(\tau_l)\}_{l=0}^A$, but only the difference with respect to some reference Green function obtained from, e.g., plain second order perturbation theory or IPT. This approach, discussed in a general context by Deisz [39], is implemented in Jarrell's code [5]. In addition, the high-frequency part may be directly stabilized by supplementing the QMC estimates with IPT (using low-pass/high-pass filters).

In the following, we will describe a stable and accurate method which achieves excellent high-frequency behavior without requiring any filters [38, 40, 41]. It is based on the following exact high-frequency expansion for the self-energy [42]:¹⁶

$$\Sigma_\sigma(\omega) = U \left(\langle \hat{n}_{-\sigma} \rangle - \frac{1}{2} \right) + U^2 \frac{\langle \hat{n}_{-\sigma} \rangle (1 - \langle \hat{n}_{-\sigma} \rangle)}{\omega} + \mathcal{O}(\omega^{-2}). \quad (37)$$

One possible choice of a model self-energy with this asymptotic behavior (including the ω^{-1} term) which is nonsingular (and purely imaginary) on the imaginary axis is given by the following two-pole approximation [38, 40, 41].

$$\Sigma_{\text{model},\sigma}(\omega) = U \left(\langle \hat{n}_{-\sigma} \rangle - \frac{1}{2} \right) + \frac{1}{2} U^2 \langle \hat{n}_{-\sigma} \rangle (1 - \langle \hat{n}_{-\sigma} \rangle) \left(\frac{1}{\omega + \omega_0} + \frac{1}{\omega - \omega_0} \right). \quad (38)$$

While the quality of the low-frequency part of this fit could be tuned by adjusting the parameter ω_0 , it does not depend very sensitively on it as long as it is not much larger than the bandwidth or the Nyquist frequency. By evaluating Σ_{model} on the imaginary axis, the corresponding Green function G_{model} can be computed for an arbitrary number of Matsubara frequencies. Consequently, the Fourier transformation to imaginary time is unproblematic (when the “free” term $1/(i\omega_n)$ is taken care of analytically). Thus, the difference of the Green functions can be accurately evaluated at all time slices. Since the second derivative of the transformed model Green function at $\tau = 0$ exactly reproduces that of the true Green function, the difference is well

¹⁵This also implies that the boundary value (36) may not be the best choice for the whole interval $\tau \in [0, \Delta\tau]$.

¹⁶The multi-band generalization [40] contains additional pair occupancy terms which have to be computed numerically even at fixed band filling.

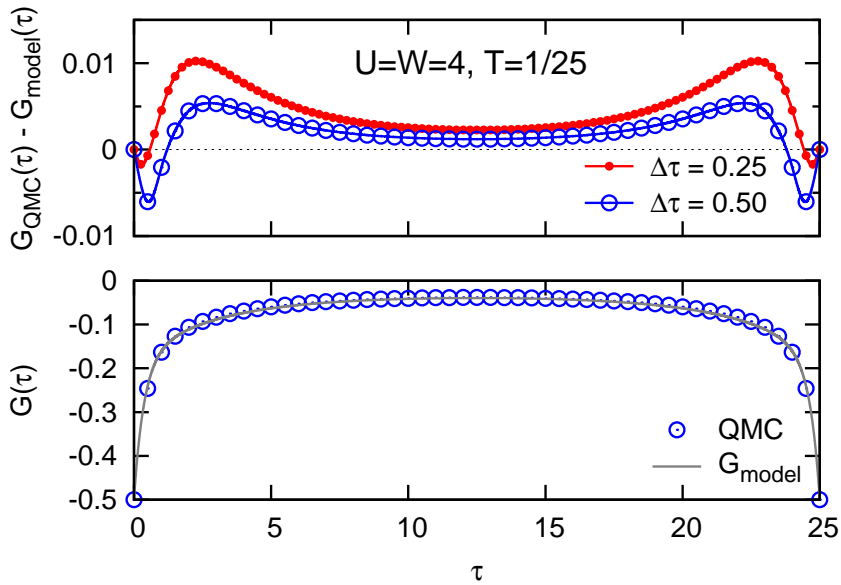


Fig. 10: Example for model based Green function splining method: discrete Green functions obtained from QMC are complemented by (continuous) reference Green functions (lower panel); the difference is generally small and has vanishing second derivatives at the boundaries, i.e. is well approximated by a natural cubic spline (upper panel).

represented by a natural spline. The Green function is obtained as a function of Matsubara frequencies by Fourier transforming the oversampled spline and adding the Matsubara-frequency model Green function to the result.

We note that the concept of the method described here of supplementing QMC data with high-frequency information derived from a weak-coupling approach is similar to Jarrell’s QMC implementation [5]. The latter is, however, less stable since its Fourier transformation relies on a numerical IPT calculation. At least for a symmetric DOS at half filling, our method is unconditionally stable for arbitrarily large frequencies without the need for bandpass filters. This is illustrated in Fig. 11: The raw HF-QMC data (dashed lines) is indistinguishable from the (numerically) exact result (solid line) already for $|\omega_n| \gtrsim 5$ and converges quickly in the low-frequency region for $\Delta\tau \rightarrow 0$. In contrast, continuous-time methods achieve results without systematic bias, however with significant noise at large frequencies, especially in the hybridization variant.

Since each of the approaches discussed so far generates a number of Matsubara frequencies which is much greater than the number of time slices, the inverse Fourier transform (then for \mathcal{G}) is unproblematic: the rounding-off near $\tau = 0$ is already small for the first grid point $\tau = \Delta\tau$ while at the end points $\tau = 0$ and $\tau = \beta$, the numerical Fourier transforms can be shifted by the value $1/2$ to exactly cancel the rounding effect.

Ulmke’s smoothing trick

A completely different approach was taken by Ulmke [44]. Here, the number of Matsubara frequencies is chosen equal to the number of time slices. In order to enforce the correct analytical

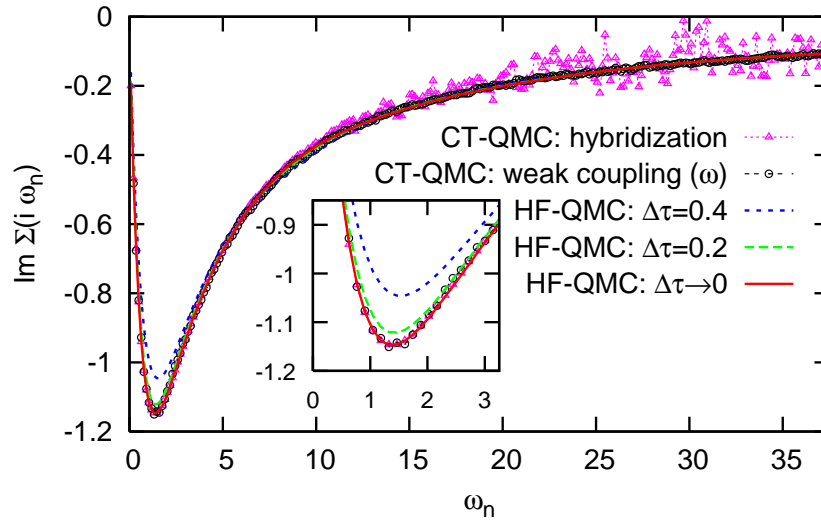


Fig. 11: (Color online) Imaginary part of self-energy on the imaginary axis for $T = 1/45$ as estimated by CT-QMC [43] using the hybridization expansion (triangles) or the weak-coupling expansion (circles) and by Hirsch-Fye QMC using splines for difference Green functions (dashed lines for $\Delta\tau = 0.4$, $\Delta\tau = 0.2$, solid line for $\Delta\tau \rightarrow 0$).

behavior of the Fourier transforms, a “smoothing” trick is employed: the naive Fourier transform is combined with an approximate correcting transformation which approaches the identity for $\Delta\tau \rightarrow 0$. While this method recovers the high-frequency behavior of $G(i\omega_n)$ and $\Sigma(i\omega_n)$ for $|\omega| \leq \pi/\Delta\tau$ about at the level of Krauth’s code, it (unnecessarily) introduces errors at small Matsubara frequencies in insulating phases which may prevent the detection of coexistence regions. This artifact can be cured by introducing a frequency cutoff in the transformation [38]; however, the resulting method is inferior to advanced splining schemes and also less elegant.

3.2 Search for solutions, hysteresis, overrelaxation

An important point not discussed so far is the initialization of the self-consistency cycle in a HF-QMC run for some given set of physical parameters (e.g. U, T, μ) and discretization $\Delta\tau$. Since in most cases the results and also self-energy and Green function will continuously depend on the parameters, one will almost initialize the simulation with a self-energy previously obtained for a related parameter set. One possibility is a sequence of runs for increasing or decreasing values of a physical parameters such as the interaction U . In the vicinity of first-order phase transitions, which show up in the DMFT due to its mean-field character as coexistence regions, such hysteresis runs may be used for determining the phase boundaries. In principle, one could also explore hysteresis in the temperature domain; this is, however, difficult for frequency-based state variables (i.e. $\Sigma(i\omega_n)$) since the Matsubara frequencies depend on the temperature.

In principle much can be gained from hysteresis runs at constant physical parameters, in which the Trotter discretization $\Delta\tau$ is initially chosen quite large; due to the scaling of the effort with $\Delta\tau^{-3}$ convergency can then be obtained rather cheaply. Subsequent runs at intermediate values of $\Delta\tau$, each initialized from the previous run, will then require only few iterations (which are

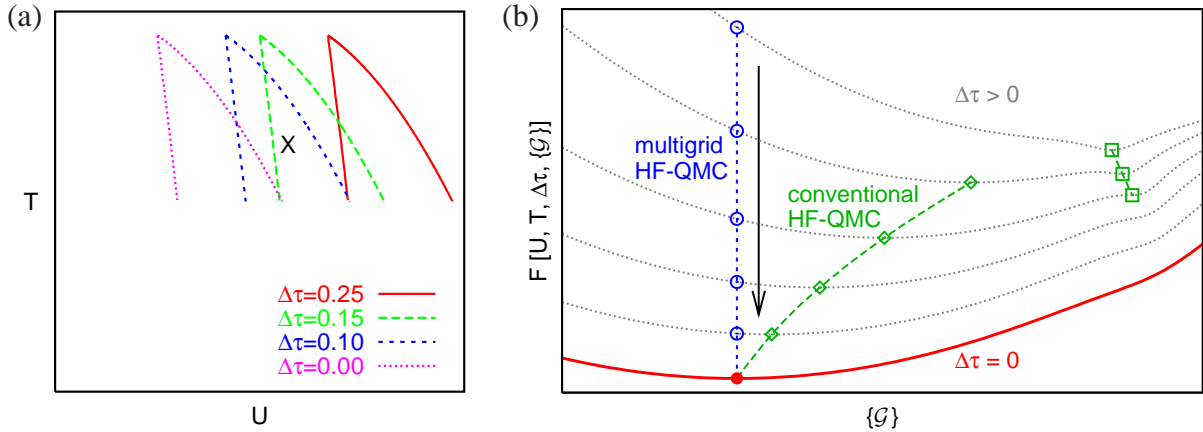


Fig. 12: (a) Scenario for the effect of a $\Delta\tau \rightarrow 0$ extrapolation on a metal-insulator coexistence region, namely a shift to smaller U at roughly constant shape. (b) Schematic view of Ginzburg-Landau free energy in multidimensional space of hybridization functions $\{G\}$. The fixed points of conventional HF-QMC (diamonds, squares) are adiabatically connected to the exact fixed point (full circle) only for small $\Delta\tau$. In contrast, the multigrid method [45] solves all impurity problems at the $\Delta\tau = 0$ fixed point (circles).

still relatively cheap). For a sufficiently large number of steps, about 5 iterations will then suffice at the most costly target discretization. This strategy is, however, not without danger, especially near phase transitions: as the location of coexistence regions will depend on $\Delta\tau$ and may, e.g., shift towards smaller interactions with decreasing $\Delta\tau$ as illustrated in Fig. 12a, a phase that is stable for some physical parameter set (cross) at low values of $\Delta\tau$ may not be stable at high $\Delta\tau$ (and vice versa), so it cannot be obtained in such $\Delta\tau$ hysteresis run.

A more abstract view on this problem is given in Fig. 12b: each DMFT fixed point can be understood as the minimum of a generalized Ginzburg-Landau free energy functional in the (high-dimensional) space of hybridization functions (full line). In the presence of Trotter errors, this functional will in general shift and also change its shape, in particular the number of relative minima. Such phase transition as a function of $\Delta\tau$ clearly limit the possibilities for hysteresis schemes as discussed above and also for $\Delta\tau$ extrapolation to be discussed in the Sec. 4.

We should also stress that the convergence of a given iteration scheme (such as shown in Fig. 1) to its fixed point is far from trivial; e.g., a reversed execution of the same scheme would have the same fixed point, but the roles of stable and unstable fixed points would be reversed. [38] However, the rate of convergence for a stable iteration scheme can be improved by overrelaxation where the new solution is replaced by a linear combination with the old solution: $f_{\text{new}} \rightarrow \alpha f_{\text{new}} + (1 - \alpha) f_{\text{old}}$ with $\alpha > 0$ (while $0 < \alpha < 1$ in an under-relaxation scheme could be used for making a previously unstable scheme stable). While such strategies can save computer time they are to be used with care close to phase transitions.

4 Computing observables

So far, we have collected all essential concepts for achieving self-consistent DMFT solutions on the basis of the HF-QMC impurity solver; these solutions are characterized by the resulting self-energies $\Sigma(i\omega_n)$ or, equivalently, by the lattice and bath Green functions $G(i\omega_n)$ and $\mathcal{G}(i\omega_n)$, i.e., high-dimensional vectors in Matsubara space, or their Fourier transforms. Each of them depends on the physical parameters, on the Trotter discretization $\Delta\tau$, and, within coexistence regions, on the initialization.

We will now discuss the derivation of other properties from HF-QMC calculations, characterize the errors of such estimates and show how the systematic Trotter bias can be removed (or at least drastically reduced) by extrapolation.

4.1 Observables: mass renormalization, energetics, and spectra

The characterization of correlation properties and associated phase transitions is greatly simplified if one can find a scalar variable which plays (more or less) the role of an order parameter. In the cases of magnetic or orbital order, an order parameter is trivially obtained from the spin, orbital and/or sublattice resolved densities. The situation is more complicated at metal-insulator transitions, at least at finite temperatures. However, the quasiparticle weight Z to be discussed below often appears as a useful characteristic. In addition, we will explain how to compute the energetics (i.e., kinetic and interaction energy), which are obviously relevant for all thermodynamic properties, and illustrate the derivation of spectra.

Quasiparticle weight

One quantity closely related to the self-energy (which quantifies interaction effects) is the quasiparticle weight also known as mass renormalization factor Z ,¹⁷ which is defined in terms of the real-frequency self-energy,

$$Z = \frac{m}{m^*} = \frac{1}{1 - \left. \frac{\partial}{\partial \omega} \text{Re} \Sigma(\omega) \right|_{\omega=0}}. \quad (39)$$

In the context of QMC simulations, one usually approximates this quantity in the discrete form

$$Z \approx \frac{1}{1 - \frac{\text{Im} \Sigma(i\omega_1)}{\pi T}}. \quad (40)$$

Due to the rules for complex derivatives (i.e., the Cauchy/Riemann equation $\partial \text{Re} f(z) / \partial \text{Re} z = \partial \text{Im} f(z) / \partial \text{Im} z$), both definitions agree in the limit $T \rightarrow 0$ as long as the Luttinger theorem is fulfilled, i.e., for $\text{Im} \Sigma(\omega = 0) = 0$. For $U \rightarrow 0$, no mass renormalization takes place so that $Z = 1$. With increasing U , Z decreases until the quasiparticle peak (in the spectrum) vanishes (cf. Fig. 14). Extending the Fermi liquid picture, one usually associates the disappearance of the quasiparticle peak with $Z \approx 0$. Z in the definition (39), however, is not positive near a

¹⁷This should not be confused with the coordination number also denoted by Z .

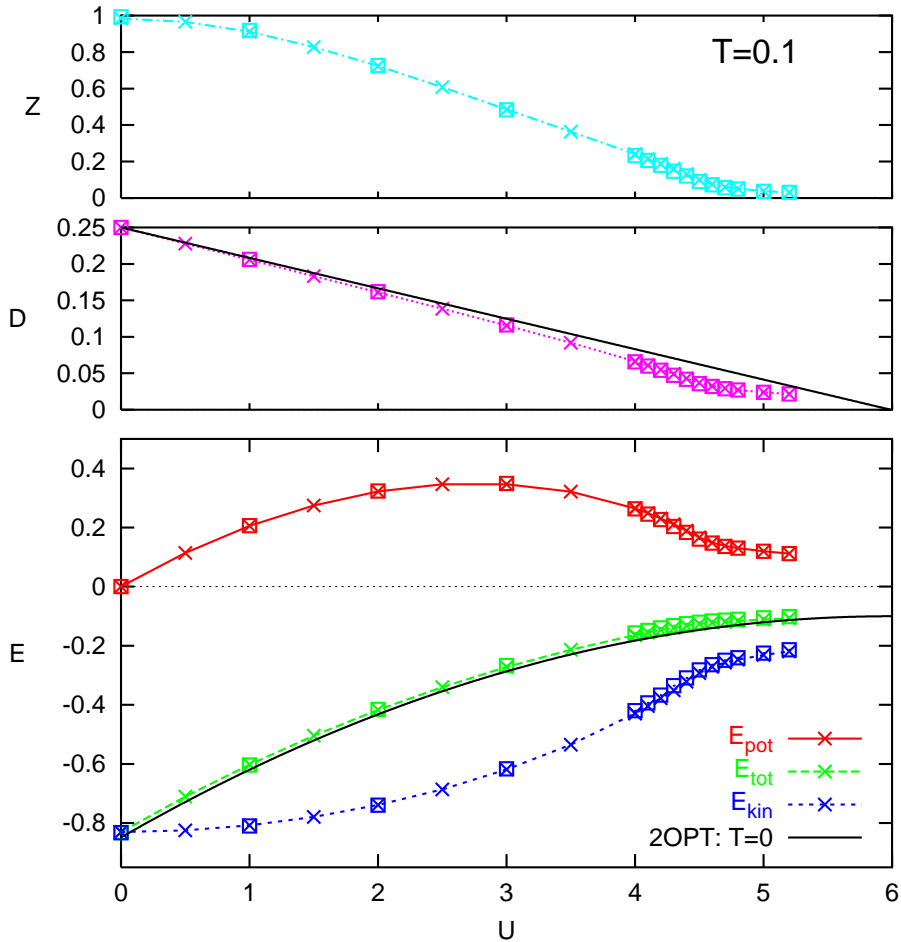


Fig. 13: Quasiparticle weight Z (discrete estimate), double occupancy D , and energy contributions for the 1-band Hubbard model on the Bethe lattice (bandwidth $W = 4$) at $T = 0.1$ in the paramagnetic phase. Crosses (connected with lines) denote QMC results for $\Delta\tau = 0.2$, squares are for $\Delta\tau = 0.125$. For comparison, results of second order perturbation theory (2OPT) are shown for the total energy and the double occupancy for $T = 0$ (solid black lines).

metal-insulator transition so that its interpretation as a quasiparticle weight breaks down. In contrast, the discrete version (40) always leads to positive Z and may therefore appear more physical. In any case, Z loses its theoretical foundation outside the Fermi liquid phase where it remains only a heuristic indicator of a metal-insulator transition. In the uppermost part of Fig. 13, Z is shown for the relatively high temperature $T = 0.1$. A rapid change of slopes indicates a transition or crossover near $U \approx 4.7$.

Energy

Within the DMFT, the energy per lattice site is given as [1, 8]

$$E = \lim_{\eta \rightarrow 0^+} T \sum_{n,\sigma} \int_{-\infty}^{\infty} d\epsilon \frac{e^{i\omega_n \eta} \epsilon \rho(\epsilon)}{i\omega_n - \epsilon - \Sigma_\sigma(i\omega_n)} + \frac{1}{2} T \sum_{n,\sigma} \Sigma_\sigma(i\omega_n) G_\sigma(i\omega_n). \quad (41)$$

Note the convergence factor $e^{i\omega_n\eta}$ which is essential in order to get the correct result. Obviously, such a term is difficult to handle numerically; in practice, η may be replaced, e.g., by the time discretization parameter $\Delta\tau$, which also determines the cutoff frequency in the infinite sum. This approximation can be avoided by evaluating the noninteracting part separately. For the kinetic energy, this implies

$$E_{\text{kin}} = \lim_{\eta \rightarrow 0^+} 2T \sum_{n=-\infty}^{\infty} e^{i\omega_n\eta} \int_{-\infty}^{\infty} d\epsilon \epsilon \rho(\epsilon) \frac{1}{i\omega_n - \epsilon + \mu - \Sigma(i\omega_n)} \quad (42)$$

$$= 2 \int_{-\infty}^{\infty} d\epsilon \frac{\epsilon \rho(\epsilon)}{e^{\beta(\epsilon-\mu)} + 1} + 2T \sum_{n=-\infty}^{\infty} \int_{-\infty}^{\infty} d\epsilon \epsilon \rho(\epsilon) (G_{\epsilon}(i\omega_n) - G_{\epsilon}^0(i\omega_n)) \quad (43)$$

$$\approx 2 \int_{-\infty}^{\infty} d\epsilon \frac{\epsilon \rho(\epsilon)}{e^{\beta(\epsilon-\mu)} + 1} + 2T \sum_{n=-L/2+1}^{L/2} \int_{-\infty}^{\infty} d\epsilon \epsilon \rho(\epsilon) (G_{\epsilon}(i\omega_n) - G_{\epsilon}^0(i\omega_n)) , \quad (44)$$

where we have assumed the paramagnetic case. Here, the interacting and noninteracting “momentum-dependent” Green functions read

$$G_{\epsilon}(i\omega_n) = \frac{1}{i\omega_n - \epsilon + \mu - \Sigma(i\omega_n)} ; \quad G_{\epsilon}^0(i\omega_n) = \frac{1}{i\omega_n - \epsilon + \mu} . \quad (45)$$

In (44), the terms in the Matsubara sum fall off at least as $1/\omega^2$, which makes it well-defined also without convergence factor. At the same time, the truncation error is reduced significantly. The complementary ingredient to the energy is the double occupancy D with $E = E_{\text{kin}} + UD$. In the context of QMC calculations, this observable is best calculated directly from Wick’s theorem (i.e. as $\langle n_{i\uparrow} n_{i\downarrow} \rangle$ or the corresponding expression in Grassmann variables) when sampling over the auxiliary field. The overall behavior of D and E in the Hubbard model can be read off (for $T = 0.1$) from the middle and lower parts of Fig. 13, respectively. For small U , the kinetic energy increases quadratically while D and, consequently, E_{pot} and E increase linearly. The potential energy reaches a maximum below $U = 3$. A region of strong curvature of D , E_{pot} , and E_{kin} near $U = 4.6$ gives a rough indication of the metal-insulator crossover. The total energy E , however, hardly shows any anomalies at this scale. Note also that the solutions for D and E are quite close to the results of plain zero-temperature second-order perturbation theory. The agreement actually becomes better for low-temperature QMC data, extrapolated to $T \rightarrow 0$ (not shown). The offset of the curves for E gives (for not too large U) an indication of the specific heat $c_V = dE/dT$ which is linear within the Fermi-liquid phase and is, in general, best evaluated by fitting the temperature dependence of E .¹⁸

Susceptibilities

The direct evaluation of the compressibility (as of most other susceptibilities) is numerically costly since it requires the QMC computation of 2-particle vertex functions. The formalism is omitted here; it can be found, e.g., in [1, 46].

¹⁸In Fermi liquid phases, the linear coefficient γ of the specific heat may also be obtained via the quasiparticle weight Z (extrapolated to $T = 0$).

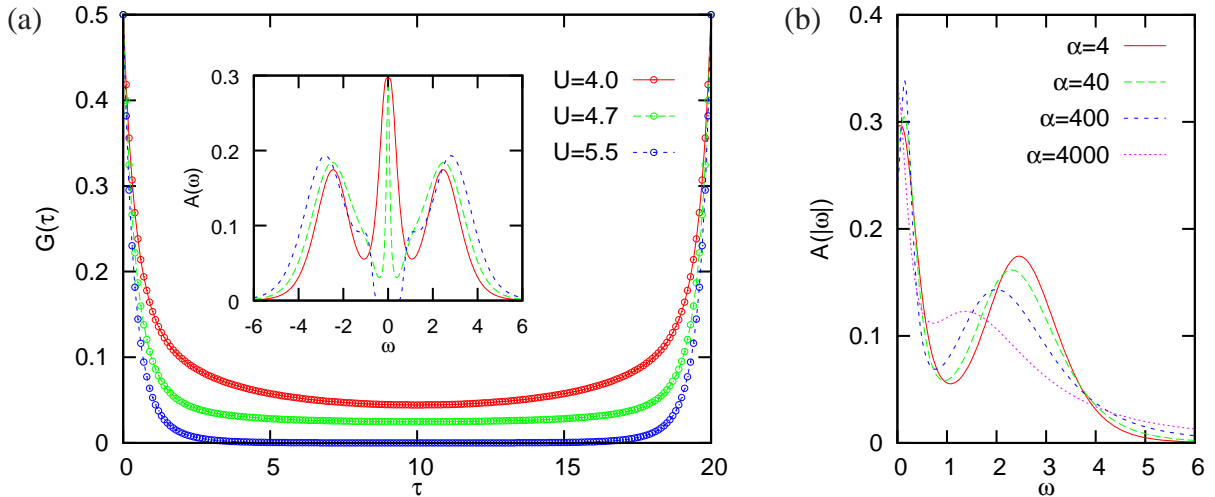


Fig. 14: Example for analytic continuation of Green functions using the maximum entropy method (MEM). (a) While the Green functions (Bethe lattice, $T = 0.05$) are featureless at all interactions and differ mainly by their value and curvature near $\tau = \beta/2$, the spectra derived using MEM (inset) show clearly the narrowing and disappearance of the quasiparticle peak with increasing interaction U . (b) Impact of the weighing factor α on the MEM result for $U = 4.0$.

Spectra

Spectra and optical conductivity data also shed light on systems near an MIT and are essential input for quantitative comparisons with experiments. Unfortunately, an analytic continuation is needed for imaginary-time based algorithms (such as the HF-QMC method) which is inherently unstable. In App. B, we discuss the maximum entropy method (MEM) which regularizes the procedure by a constraint on the smoothness of the spectrum (quantified by the entropy function). As seen in Fig. 14 the essential correlation physics is much more transparent on the level of the spectra than on that of the imaginary-time Green function. Thus, even significant efforts in MEM related questions appear well spent.

4.2 Estimation of errors and extrapolation $\Delta\tau \rightarrow 0$

As we have already mentioned, any “raw” estimate of an observable measured at some iteration using HF-QMC contains various sources of errors: (i) the statistical MC error, (ii) the convergence error associated with incomplete convergence of the self-consistency cycle, and (iii) the discretization error stemming both from the Trotter decoupling and the approximate Fourier transformations discussed in Sec. 2 and Sec. 3, respectively.

The first two errors can be treated at equal footing when we compute the observables at each iteration and extract averages from their “traces” (i.e. their functional dependence as a function of iteration number) as depicted in Fig. 15a for the double occupancy D . We see that the measurements fluctuate significantly for each value of $\Delta\tau$; the amplitude of such fluctuations can be controlled by the number of sweeps. Also apparent is a gradient in the initial iterations for the largest discretization shown, $\Delta\tau = 0.20$. It is clear that such lead-in data should not

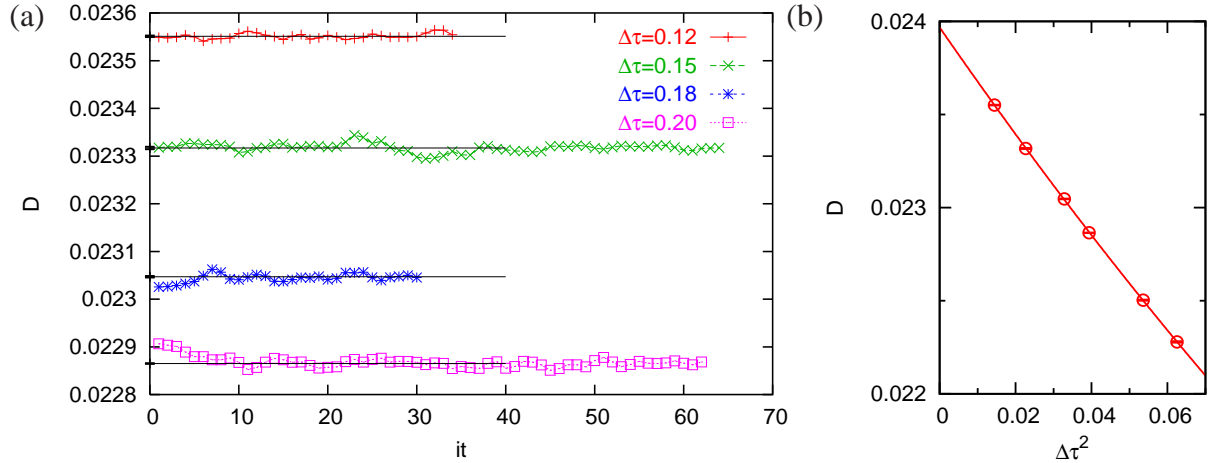


Fig. 15: Steps in the HF-QMC based computation of the double occupancy at $U = 5$, $T = 0.04$: (a) estimates can be calculated as averages with statistical error bars from the analysis of time series (traces), taking autocorrelation into account, for each value of $\Delta\tau$. (b) Numerically exact results are obtained in a second step using extrapolation by least-squares fits.

be included in averages. Moreover, all curves show significant autocorrelation which has to be taken into account for error analysis. Taking these issues into account, standard analysis techniques for time series yield raw HF-QMC results for each value of the discretization plus an error bar which takes statistical and convergence error into account. Such data is shown in Fig. 15b as a function of the squared discretization. Evidently, the discretization dependence is very regular; a straightforward extrapolation using standard least-square fit methods (here with the 3 free parameters corresponding to the orders $\Delta\tau^0$, $\Delta\tau^2$, and $\Delta\tau^4$) essentially eliminates this systematic error, i.e. reduces the total error by two orders of magnitude.¹⁹

Due to the cubic scaling of the effort with the number of time slices, the total cost of achieving extrapolated results for some grid range is dominated by the smallest discretization. Thus the possibility of extrapolation comes (with the use also of coarser grids) at no significant cost; in contrast, DMFT convergence can be accelerated by using the $\Delta\tau$ hysteresis technique outlined above. Taking all of this into account, HF-QMC with extrapolation can be competitive with the recently developed continuous-time QMC solvers, as demonstrated in Fig. 16: at fixed total computer time, this method achieves the highest precision [47] in energy estimates for a test case established in [43].

Note that the use of insufficiently converged solutions is potentially a very significant source of errors. It is important to realize that in principle measurements have to be performed exactly at the solution of the self-consistency equations, i.e., for the exact bath Green function. Averages over measurements performed for different impurity models corresponding to approximate solutions do not necessarily converge to the exact answer in the limit of an infinite number of models (i.e., iterations) and measurements.²⁰ Still, the most important practical point when

¹⁹The reader may note a similarity to the MC example shown in Fig. 3.

²⁰Trivially, a measurement of the free energy F itself (using a suitable impurity solver) is a good example. Since F is minimal for the true solution, all measurements taken for approximate solutions will be too large. The correct answer can, therefore, not be approached by averaging over many measurements, but only by reducing the

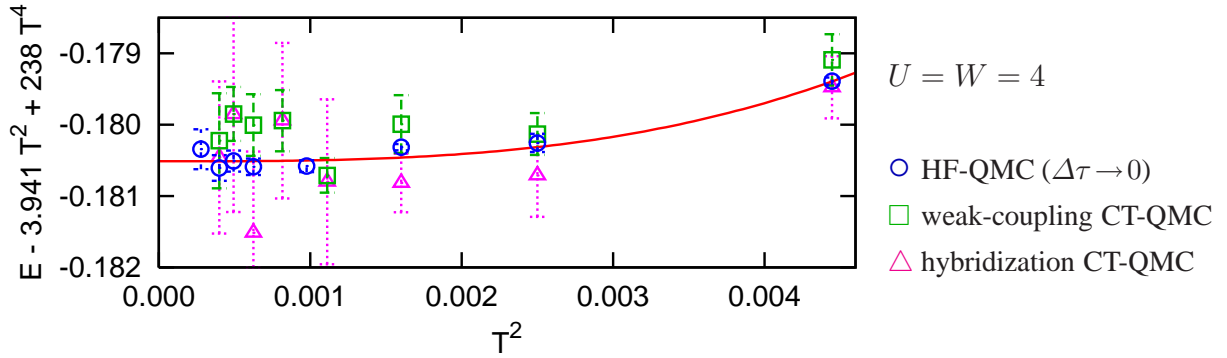


Fig. 16: Comparison of energy estimates of the half-filled Hubbard model obtained within DMFT using different QMC based impurity solvers at fixed computational effort, with (approximate) leading terms subtracted. HF-QMC is competitive, here even more efficient, (only) after careful extrapolation $\Delta\tau \rightarrow 0$ [47].

computing observables is that only runs are included in averages which are close to the solution in comparison to the asymptotic statistical error. For higher precision, the number of sweeps (and not only the number of measurements) must be increased.

Let us, finally, stress that discretization errors can vary greatly between different observables and, due to the DMFT self-consistency, also between points in parameter space. This is clearly seen in Fig. 17: for the kinetic energy, the quadratic contribution to the Trotter error is significant only in the strongly correlated metallic phase; it is negligible in the insulator! This shows that rules for acceptable values of $\Delta\tau$ sometimes mentioned in the literature (e.g. $\Delta\tau U \lesssim 1$) are of limited value; in fact, they cannot follow from the Trotter decomposition since this becomes exact both in the limits of weak and strong coupling. In the case of the double occupancy, the quadratic term is surprisingly small close to the phase transition (and has a maximum in the absolute value at $U \approx 3.5$). However, in both cases the quartic terms are significant and irregular close to the phase transition which will in practice limit the reliability of extrapolations. Interestingly, some of the errors cancel in the total energy which makes numerically exact estimates of this observables particularly accessible for HF-QMC.

5 Conclusion and outlook

Hopefully, these lecture notes give a good first overview of the most important generic aspects of Hirsch-Fye QMC based method for calculating electronic properties within the dynamical mean-field theory. However, they are certainly not complete in any respect: important topics such as the minus-sign problem occurring in general multi-band and cluster DMFT calculations could only be touched upon; also, important literature is missing.

The topics that we would have liked to discuss in an (yet unwritten) section on advanced aspects also include recent methodological developments such as the multigrid HF-QMC approach [45, 48] and our HF-QMC based implementation of real-space dynamical mean-field

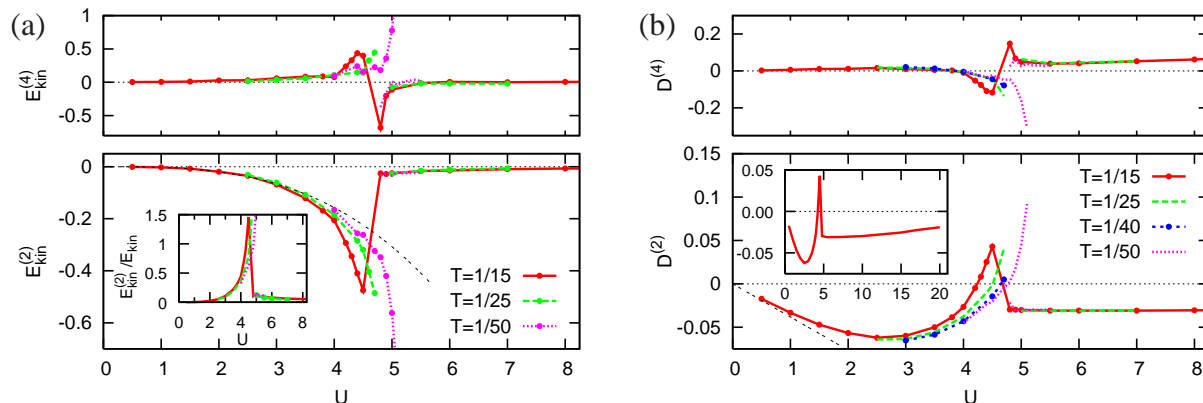


Fig. 17: (a) Coefficients of Trotter errors in HF-QMC estimates of kinetic energy. Inset: relative coefficients. (b) Trotter coefficients for double occupancy [47].

theory (RDMFT) [49, 50] as well as their applications, e.g., in the context of cold atoms on optical lattices. It would also be tempting and potentially fruitful to explore the similarities between the HF-QMC approach and the determinantal QMC method [33] for Hubbard model studies in finite dimensions, in particular its recent application as DMFT impurity solver [51]. However, one has to stop at some point and the material presented so far is probably already overwhelming for students and scientists just entering the field.

This audience should be aware of the fact that the future relevance of the HF-QMC method is controversial; influential groups claim that it has been superseded by the continuous-time QMC methods. These methods are, indeed, conceptually elegant and avoid the systematic discretization bias which can be cumbersome in conventional HF-QMC calculations. We have shown, however, that the HF-QMC method is competitive at least for some classes of problems already in the conventional form [47]; the quasi-continuous time formulation in the multigrid approach eliminates most of the remaining problems. Consequently it seems too early to abandon the general HF-QMC concept.

Acknowledgment

The author thanks E. Gorelik and D. Rost for valuable help in preparing these lecture notes. Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

Appendices

A Multi-band Hubbard models

The applicability of the one-band Hubbard model to d or f electron systems is a priori questionable since the partially filled bands correspond to atomic orbitals which are 5-fold and 7-fold degenerate (for each spin direction), respectively. While bands in a lattice are more complicated than orbitals of isolated atoms, the remaining degeneracy can be inferred from symmetry considerations alone. Often it is useful to consider a cubic representation of the angular part of atomic d orbitals,

$$\begin{aligned} |d_{xy}\rangle &\propto (|2, 2\rangle - |2, -2\rangle), & |d_{yz}\rangle &\propto (|2, 1\rangle + |2, -1\rangle), & |d_{zx}\rangle &\propto (|2, 1\rangle - |2, -1\rangle) \\ |d_{x^2-y^2}\rangle &\propto (|2, 2\rangle + |2, -2\rangle), & |d_{3z^2-r^2}\rangle &\propto |2, 0\rangle, \end{aligned} \quad (46)$$

expressed in terms of eigenfunctions of the angular momentum operator,

$$l^2|l, m\rangle = \hbar^2 l(l+1)|l, m\rangle, \quad l_z|l, m\rangle = \hbar m|l, m\rangle. \quad (47)$$

In lattices with cubic symmetry the five d orbitals are energetically split into the t_{2g} orbitals ($|d_{xy}\rangle, |d_{yz}\rangle, |d_{zx}\rangle$) and the e_g orbitals ($|d_{x^2-y^2}\rangle, |d_{3z^2-r^2}\rangle$), which give rise to one threefold degenerate and one twofold degenerate band, respectively. Lower symmetry can lift the remaining degeneracies; e.g., in the trigonal case the t_{2g} orbitals are further split into one nondegenerate a_{1g} and one twofold degenerate e_g^π band. Thus, it is possible that in some d systems only one band crosses or touches the Fermi surface which then justifies the one-band assumption made in (1) and used for the examples in this lecture. In general, however, the inclusion of several orbitals per site is important. An $SU(2)$ -invariant generalization of the Hubbard model where the interaction is still local but the valence band is degenerate then contains additional coupling terms²¹ [53,54]

$$\begin{aligned} \hat{H}_{m\text{-band}} &= -t \sum_{\langle ij \rangle, \nu\sigma} \hat{c}_{i\nu\sigma}^\dagger \hat{c}_{j\nu\sigma} + U \sum_{i\nu} \hat{n}_{i\nu\uparrow} \hat{n}_{i\nu\downarrow} \\ &\quad + U' \sum_{i; \nu < \nu'; \sigma\sigma'} \hat{n}_{i\nu\sigma} \hat{n}_{i\nu'\sigma'} + J \sum_{i; \nu < \nu'; \sigma\sigma'} \hat{c}_{i\nu\sigma}^\dagger \hat{c}_{i\nu'\sigma'}^\dagger \hat{c}_{i\nu\sigma'} \hat{c}_{i\nu'\sigma}, \end{aligned} \quad (48)$$

where ν, ν' (with $1 \leq \nu \leq m, 1 \leq \nu' \leq m$) are band indices. The exchange term parameterized by the Hund's rule coupling J can be rewritten as

$$\hat{H}_J = -2J \sum_{i, \nu < \nu'} \left(\hat{\mathbf{S}}_{i\nu} \cdot \hat{\mathbf{S}}_{i\nu'} + \frac{1}{4} \hat{n}_{i\nu} \hat{n}_{i\nu'} \right) \quad (49)$$

with $\hat{\mathbf{S}}_{i\nu} = \frac{1}{2} \sum_{\sigma\sigma'} \hat{c}_{i\nu\sigma}^\dagger \boldsymbol{\tau}_{\sigma\sigma'} \hat{c}_{i\nu\sigma}$ being the spin operator for orbital ν at site i and the Pauli matrices $\boldsymbol{\tau}_{\sigma\sigma'}$. In quantum Monte Carlo (QMC) simulations, the spin-flipping terms implicit in (49)

²¹Here, an on-site pair hopping term which only contributes when one orbital is doubly and another singly occupied is neglected [52].

lead to a numeric (minus-sign) problem. Therefore, one here usually replaces the Heisenberg interaction part of (49) by an Ising-type interaction at the cost of breaking the SU(2) symmetry [55, 52]. Since $\hat{s}_{i\nu}^z \hat{s}_{i\nu'}^z = \sigma\sigma' \hat{n}_{i\nu\sigma} \hat{n}_{i\nu'\sigma'} / 4 = (2\delta_{\sigma\sigma'} - 1) \hat{n}_{i\nu\sigma} \hat{n}_{i\nu'\sigma'} / 4$, one can write this modified multi-band Hubbard Hamiltonian as

$$\hat{H}_{m\text{-band}}^z = \sum_{\langle ij \rangle, \nu\sigma} \epsilon_{\mathbf{k}\nu} \hat{c}_{\mathbf{k}\nu}^\dagger \hat{c}_{\mathbf{k}\nu} + U \sum_{i\nu} \hat{n}_{i\nu\uparrow} \hat{n}_{i\nu\downarrow} + \sum_{i;\nu<\nu';\sigma\sigma'} (U' - \delta_{\sigma\sigma'} J) \hat{n}_{i\nu\sigma} \hat{n}_{i\nu'\sigma'}. \quad (50)$$

The interaction U between electrons within each orbital is always larger than the interorbital density-density interaction U' . The smaller exchange coupling J can trigger ferromagnetic and (possibly coexisting) orbital order.

A HF-QMC simulation for this type of models proceeds in principle as in the single-band case; however, one auxiliary field is needed for each of the $m(2m - 1)$ pair interactions.

B Maximum entropy method

Within the DMFT, all single-particle properties can be expressed in terms of the local single-particle spectral function (also called “full” or “interacting” density of states) which is proportional to the imaginary part of the local retarded Green function,

$$A(\omega) = -\frac{1}{\pi} \text{Im} G(\omega + i0^+). \quad (51)$$

The spectral function $A(\omega)$ is accessible experimentally: measured (angular averaged) photoemission spectra (PES) can under certain simplifying assumptions be identified with $A(\omega)$ multiplied by the Fermi function $n_f(\omega - \mu)$. Correspondingly, inverse photoemission spectra or X-ray absorption spectra (XAS) can be identified with $A(\omega)$ multiplied by $1 - n_f(\omega - \mu)$. Furthermore, nonlocal spectral functions and the optical conductivity $\sigma(\omega)$ can, within the DMFT, be calculated from $A(\omega)$.

In QMC calculations, however, the Green function G (and thus the spectral function $A(\omega)$) cannot be directly computed on the real axis. Instead, real-time dynamical information has to be extracted from imaginary-time data $G(\tau)$ (or, equivalently, from the Fourier transformed Matsubara-frequency data $G(i\omega_n)$) via analytic continuation. This is in principle possible through inversion of the spectral representation for $G(\tau)$, i.e.,

$$G(\tau) = \int_{-\infty}^{\infty} d\omega K(\tau, \omega - \mu; \beta) A(\omega), \quad (52)$$

$$K(\tau, \omega; \beta) := \frac{\exp(-\tau\omega)}{1 + \exp(-\beta\omega)}, \quad (53)$$

but poses an ill-conditioned problem since $G(\tau)$ is only measured on a grid $\tau_l = l\Delta\tau$ (where $\Delta\tau = \beta/A$) and since the kernel (53) becomes exponentially small for generic values of τ at large absolute frequencies $|\omega|$ as illustrated in Fig. 18. For a symmetric problem, i.e., symmetric noninteracting DOS and $n = 1$, the integral in (52) can be restricted to positive frequencies when the symmetrized fermion kernel

$$K_s(\tau, \omega; \beta) = K(\tau, \omega; \beta) + K(\tau, -\omega; \beta) \quad (54)$$

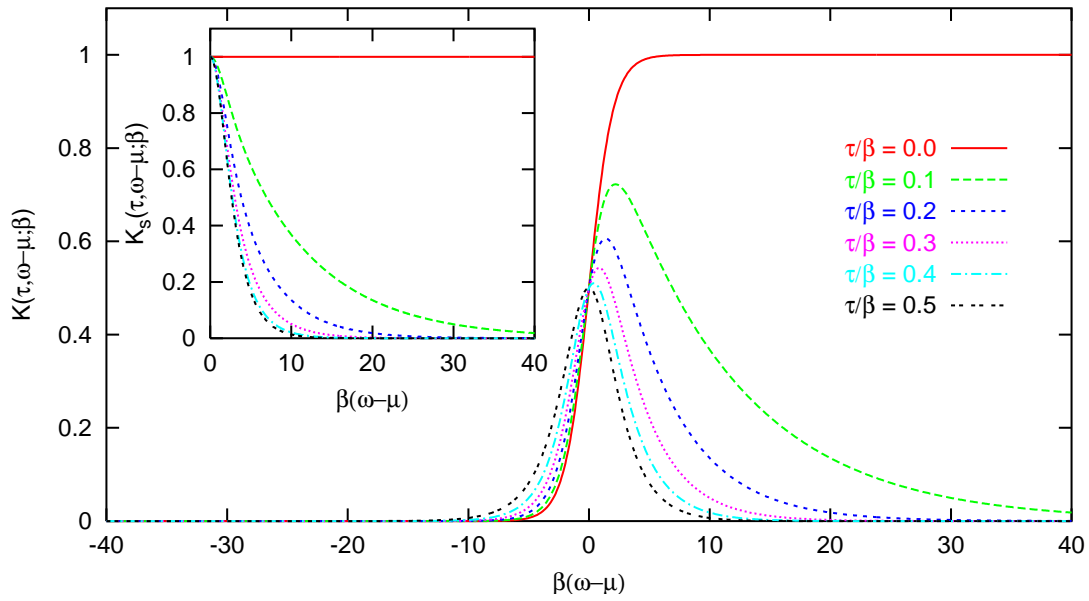


Fig. 18: General fermion kernel (53) for an equidistant set of values of $\tau/\beta \leq 1/2$. Except for $\tau = 0$, large frequencies are suppressed exponentially. Inset: symmetrized fermion kernel (54).

shown in the inset of Fig. 18 is used. We stress that a very small number $\Lambda = \mathcal{O}(10)$ of time slices poses a more serious limitation for obtaining reliable spectra with good resolution than the exponential nature of the kernel since the number of degrees of freedom which can be reliably resolved in a spectrum is obviously much smaller than the number of data points $\{G(l\Delta\tau)\}$ and since there always exists an infinite number of spectra which correspond to the same data. Still, in general, the resolution is much better at smaller frequencies and, according to Nyquist's theorem, essentially no information can be obtained from QMC for $\omega > \Lambda\pi/\beta = \pi/(\Delta\tau)$. Before we address the full analytic continuation problem and introduce the maximum entropy method, we collect some useful relations (denoting the m^{th} derivative as $G^{(m)}$),

$$G(\beta) = n, \quad G(0^+) = 1 - n, \quad (55)$$

$$G^{(m)}(0) + G^{(m)}(\beta) = (-1)^m \langle (\omega - \mu)^m \rangle_{A(\omega)} \quad (56)$$

$$G(\beta/2) \approx \frac{\pi}{\beta} \overline{A(\omega)} \Big|_{|\omega - \mu| \lesssim \pi/\beta}. \quad (57)$$

Since the filling given by (55) is known in the symmetric case, the value of $G(0)$ then provides no useful information which is also seen in the inset of Fig. 18. Equation (56) also shows the loss of high-frequency information from the discretization of imaginary time: for a finite grid the error in estimating derivatives $G^{(m)}$ increases rapidly with order m ; thus, the determination of high order moments $\langle (\omega - \mu)^m \rangle_{A(\omega)}$ of the spectrum is in general an ill-posed problem. At low temperatures, $G(\beta/2)$ gives a hint as to the weight of a quasiparticle peak or the existence of a gap via (57) since its value is proportional to the value of the spectral function near the Fermi energy, averaged over an inverse hyperbolic cosine with width π/β .

First attempts to address the analytic continuation problem for QMC data included least-squares fits, Padé approximants, and regularization (for references, see the pedagogical and concise

review by Jarrell [56]). Least-squares fits of spectra approximated as a set of box functions are inherently unstable. Padé approximations for $G(i\omega_n)$ only work well for very precise data (e.g., in the context of Eliashberg equations), but not for QMC. Regularization of the kernel (53) tends to produce overly smeared-out spectra. What is needed instead is a regularization of the solution $A(\omega)$ that only shows features which are supported by the data, but is as smooth as possible otherwise. This is essentially the idea of the maximum entropy method (MEM) of finding the most probable spectrum compatible with the data.

The MEM is a very general approach for reconstructing continuous, positive semidefinite functions (i.e., densities or spectra) from incomplete, noisy, and possibly oversampled²² data. It has a long history in the context of image reconstruction in such diverse disciplines as radio aperture synthesis, optical deconvolution, X -ray imaging, structural molecular biology, and medical tomography (see [58] and references therein). A preliminary study of its usefulness in the context of the analytic continuation problem [59] was soon followed by applications to spectra [60, 61] and correlation functions [62]. An efficient and general formulation of a MEM algorithm for analytic continuation of (oversampled) QMC data was then given by Gubernatis [63], closely following the general state-of-the-art approach by Bryan [57].

For simplicity, we will first assume that the QMC simulation produces N_d measurements of the observables $G_l \equiv G(l\Delta\tau)$, $0 \leq l < A$ which are independent both in computer time and in imaginary time, i.e., without significant autocorrelation between subsequent measurements or between adjacent imaginary time slices.²³ If we further disregard systematic errors and assume some “true” spectrum A [which corresponds to a discretized Green function G_l via (52)] as well as a Gaussian distribution of statistical errors, the probability distribution for the observed averaged values $\bar{G} \equiv \{\bar{G}_l\}$ (where $\bar{G}_l = \sum_{i=1}^{N_d} G_l^i$) is

$$P(\bar{G}|A) \propto e^{-\frac{1}{2}\chi^2}; \quad \chi^2 = \sum_{l=0}^{A-1} \frac{(\bar{G}_l - G_l)^2}{\sigma_l^2}. \quad (58)$$

Here, σ_l can be estimated from the data alone, $\sigma_l^2 \approx \sum_{i=1}^{N_d} (\bar{G}_l - \bar{G}_l^i)^2 / (N_d(N_d - 1))$. Obviously, the likelihood function $P(\bar{G}|A)$ is not directly helpful; in order to find the most probable spectrum given the measured data, we need the converse probability $P(A|\bar{G})$. These probabilities are related by Bayes’ theorem,

$$P(A|\bar{G}) P(\bar{G}) = P(\bar{G}|A) P(A). \quad (59)$$

Since the data \bar{G} is constant in the search for an optimal A , the associated probability $P(\bar{G})$ drops out of the problem. For the probability $P(A)$ of a spectrum $A(\omega)$ in absence of data, an entropic ansatz is made where prior knowledge can also be incorporated by choosing an

²²Data is called *oversampled* when measurements of different data points (here: for different τ) are not statistically independent, but correlated [57]. In this case, the number of “good” degrees of freedom (entering the error statistics) is reduced.

²³Typically, one of the “measurements” of a set $\{G_l\}$ is generated by binning a macroscopic number of measurements within QMC. Autocorrelation in computer time, i.e., between different bins vanishes in the limit of infinite bin size.

appropriate (positive semidefinite) default model $m(\omega)$,

$$P(A) = e^{\alpha S[A(\omega), m(\omega)]}. \quad (60)$$

Here, α is a numerical parameter while S is a generalized Shannon-Jaynes entropy,

$$S[A, m] = \int d\omega \left(A(\omega) - m(\omega) - A(\omega) \ln \left(A(\omega)/m(\omega) \right) \right). \quad (61)$$

For a constant default model (within some finite frequency range), the entropic form (61) clearly favors smooth spectra. This is also true for a general smooth default model. It also enforces positivity of A and pushes the solution towards the (normalized) default model in absence of data. From (58), (59), and (60), the posterior probability can be read off as

$$P(A|\bar{G}, m, \alpha) = e^{\alpha S[A, m] - \frac{1}{2} \chi^2[\bar{G}, A]}. \quad (62)$$

The balance between a tight match of data and a high entropy is calibrated by the Lagrange parameter α which may be chosen so that $\chi^2 = \Lambda$ (historic MEM). Alternatively, one may use the value of α with the highest probability $P(\alpha|\bar{G}, A, m)$ which can approximately be calculated within the method (classic MEM). Given the QMC data, a default model, a representation of the spectrum (i.e., a possibly inhomogeneous grid of ω_j of frequencies for which A is going to be computed), and a starting guess for α , a simple MEM program thus both searches for the spectrum $\{A(\omega_j)\}$ with maximum probability $P(A|\bar{G}, m, \alpha)$ for given α using, e.g., the Newton-Raphson method and, in an outer loop, searches for the best value of α .

The former procedure can be stabilized using a singular value decomposition (SVD) of the kernel:

$$K = V \Sigma U^T \quad (63)$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_s)$$

where $\sigma_1 > \dots > \sigma_s > 0$.

Here U, V are orthogonal matrices. Typically, most of the singular values σ_i are equal to zero (within machine precision). The columns of U^T , restricted to $\sigma_i \neq 0$, then span the same space as the columns of K . Projecting all related quantities to this new most smaller space (so-called singular space) a stable search can be performed. In practice, the width of the Newton-Raphson steps has to be restricted using a Levenberg-Marquardt-Parameter. An example for the application of this method is shown in Fig. 14.

Alternatively to this deterministic approach, one may also use Markov chain Monte Carlo updates directly in the image space (i.e. on the level of the spectrum) with simulated annealing, as implemented, e.g., by Sandvik. This stochastic approach is however, quite sensitive to the choice of the frequency discretization, requires smoothing runs, and is computationally more costly.

References

- [1] A. Georges, G. Kotliar, W. Krauth, and M. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [2] J.E. Hirsch and R.M. Fye, *Phys. Rev. Lett.* **56**, 2521 (1986)
- [3] W. Metzner and D. Vollhardt, *Phys. Rev. Lett.* **62**, 324 (1989)
- [4] E. Müller-Hartmann, *Z. Phys. B* **74**, 507 (1989)
- [5] M. Jarrell, *Phys. Rev. Lett.* **69**, 168 (1992)
- [6] A. Georges and G. Kotliar, *Phys. Rev. B* **45**, 6479 (1992)
- [7] V. Janiš and D. Vollhardt, *Int. J. Mod. Phys. B* **6**, 731 (1992)
- [8] A.L. Fetter and J.D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971)
- [9] F.A. Berezin, *The method of second quantization* (Academic Press, New York, 1966)
- [10] M. Caffarel and W. Krauth, *Phys. Rev. Lett.* **72**, 1545 (1994)
- [11] H. Keiter and J.C. Kimball, *Phys. Rev. Lett.* **25**, 672 (1970)
- [12] N.E. Bickers, D.L. Cox, and J.W. Wilkins, *Phys. Rev. B* **36**, 2036 (1987)
- [13] T. Pruschke and N. Grewe, *Z. Phys. B* **74**, 439 (1989)
- [14] T. Pruschke, D.L. Cox, and M. Jarrell, *Phys. Rev. B* **47**, 3553 (1993)
- [15] N.E. Bickers, D.J. Scalapino, and S.R. White, *Phys. Rev. Lett.* **62**, 961 (1989)
- [16] N.E. Bickers and D.J. Scalapino, *Ann. Physik* **193**, 206 (1989)
- [17] N.E. Bickers and S.R. White, *Phys. Rev. B* **43**, 8044 (1991)
- [18] K.G. Wilson, *Rev. Mod. Phys.* **47**, 773 (1975)
- [19] H.R. Krishnamurthy, J.W. Wilkins, and K.G. Wilson, *Phys. Rev. B* **21**, 1003 (1980)
- [20] T.A. Costi, A.C. Hewson, and V. Zlatić, *J. Phys. Cond. Matter* **6**, 2519 (1994)
- [21] R. Bulla, T.A. Costi, and T. Pruschke, *Rev. Mod. Phys.* **80**, 395 (2008)
- [22] U. Schollwöck, *Rev. Mod. Phys.* **77**, 259 (2005)
- [23] E. Gull, P. Werner, O. Parcollet, and M. Troyer, *Europhys. Lett.* **82**, 57003 (2008)
- [24] M.J. Rozenberg, X.Y. Zhang, and G. Kotliar, *Phys. Rev. Lett.* **69**, 1236 (1992)

- [25] A. Georges and W. Krauth, Phys. Rev. Lett. **69**, 1240 (1992)
- [26] M. Ulmke, V. Janiš, and D. Vollhardt, Phys. Rev. B **51**, 10411 (1995)
- [27] G. Kotliar, S. Savrasov, K. Haule, V. Oudovenko, O. Parcollet, and C. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)
- [28] H.F. Trotter, Proc. Am. Math. Soc. **10**, 545 (1959)
- [29] M. Suzuki, Prog. Theor. Phys. **56**, 1454 (1976)
- [30] J.E. Hirsch, Phys. Rev. B **28**, 4059 (1983)
- [31] K. Held, *Untersuchung korrelierter Elektronensysteme im Rahmen der Dynamischen Molekularfeldtheorie* (Ph.D. thesis, Universität Augsburg, 1999)
- [32] W. Negele and H. Orland, *Quantum Many-Particle Systems* (Addison-Wesley, New York, 1987)
- [33] R. Blankenbecler, D.J. Scalapino, and R.L. Sugar, Phys. Rev. D **24**, 2278 (1981)
- [34] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953)
- [35] P.K.V.V. Nukala, T.A. Maier, M.S. Summers, G. Alvarez, and T.C. Schulthess, Phys. Rev. B **80** (2009)
- [36] E. Gull, P. Staar, S. Fuchs, P.K.V.V. Nukala, M.S. Summers, T. Pruschke, T.C. Schulthess, and T.A. Maier, Phys. Rev. B **83** (2011)
- [37] X.Y. Zhang, M.J. Rozenberg, and G. Kotliar, Phys. Rev. Lett. **70**, 1666 (1993)
- [38] N. Blümer, *Mott-Hubbard Metal-Insulator Transition and Optical Conductivity in High Dimensions* (Ph.D. thesis, Universität Augsburg, 2003)
- [39] J.J. Deisz, D.W. Hess, and J.W. Serene, in *Recent Progress In Many Body Theories*, vol. 4 (Plenum, New York, 1995)
- [40] C. Knecht, *Numerical and analytical approaches to strongly correlated electron systems* (Ph.D. thesis, Universität Mainz, 2006)
- [41] C. Knecht, N. Blümer, and P. van Dongen, Phys. Rev. B **72**, 081103 (2005)
- [42] M. Potthoff, T. Wegner, and W. Nolting, Phys. Rev. B **55**, 16132 (1997)
- [43] E. Gull, P. Werner, A.J. Millis, and M. Troyer, Phys. Rev. B **76** (2007)
- [44] M. Ulmke, *Phasenübergänge in stark korrelierte Elektronensystemen* (Ph.D. thesis, KFA Jülich, 1995)

-
- [45] N. Blümer, arXiv (2008)
- [46] N. Blümer, *Quanten-Monte-Carlo-Untersuchungen von magnetischen Phasen im Hubbard-Modell mit Nächst-Nachbar-Wechselwirkungen* (Diploma thesis, RWTH Aachen, 1996)
- [47] N. Blümer, Phys. Rev. B **76**, 205120 (2007)
- [48] E. Gorelik and N. Blümer, Phys. Rev. A **80**, 051602 (2009)
- [49] E. Gorelik, I. Titvinidze, W. Hofstetter, M. Snoek, and N. Blümer, Phys. Rev. Lett. **105**, 065301 (2010)
- [50] N. Blümer and E.V. Gorelik, Comp. Phys. Commun. **182**, 115 (2011)
- [51] E. Khatami, C.R. Lee, Z.J. Bai, R.T. Scalettar, and M. Jarrell, Phys. Rev. E **81** (2010)
- [52] K. Held and D. Vollhardt, Eur. Phys. J. B **5**, 473 (1998)
- [53] P.W. Anderson, Phys. Rev. **124**, 41 (1961)
- [54] L. Dworin and A. Narath, Phys. Rev. Lett. **25**, 1287 (1970)
- [55] Y. Motome and M. Imada, J. Phys. Soc. Jap. **66**, 1872 (1997)
- [56] M. Jarrell, in D.J. Scalapino (eds.), *Numerical Methods for Lattice Quantum Many-Body Problems* (Addison Wesley, Reading, 1997)
- [57] R.K. Bryan, Eur. Biophys. J **18**, 165 (1990)
- [58] J. Skilling and R.K. Bryan, Mon. Not. R. Astr. Soc. **211**, 111 (1984)
- [59] R.N. Silver, D.S. Sivia, and J.E. Gubernatis, Phys. Rev. B **41**, 2380 (1990)
- [60] R.N. Silver, J.E. Gubernatis, D.S. Sivia, and M. Jarrell, Phys. Rev. Lett. **65**, 496 (1990)
- [61] M. Jarrell, D.S. Sivia, and B. Patton, Phys. Rev. B **42**, 4804 (1990)
- [62] J. Deisz, M. Jarrell, and D.L. Cox, Phys. Rev. B **42**, 4869 (1990)
- [63] J.E. Gubernatis, M. Jarrell, R.N. Silver, and D.S. Sivia, Phys. Rev. B **44**, 6011 (1991)

10 Continuous-Time Impurity Solvers

Philipp Werner

Theoretical Physics

ETH Zurich, 8093 Zurich, Switzerland

Contents

1	Introduction	2
2	General recipe	4
3	Weak-coupling approach - expansion in the interaction	4
3.1	Monte Carlo configurations	4
3.2	Sampling procedure and detailed balance	6
3.3	Determinant ratios and fast matrix updates	7
3.4	Measurement of the Green's function	8
3.5	Expansion order	8
4	Strong coupling approach - expansion in the impurity-bath hybridization	9
4.1	Monte Carlo configurations	9
4.2	Sampling procedure and detailed balance	11
4.3	Measurement of the Green's function	12
4.4	Generalization - Matrix formalism	13
5	Comparison between the two approaches	14
A	Rubtsov's weak-coupling approach	15

1 Introduction

An impurity model describes an atom or molecule embedded in some host or *bath*, with which it can exchange electrons. This exchange of electrons allows the impurity to make transitions between different quantum states, and leads to a non-trivial dynamics. Therefore, despite the zero dimensional nature (which makes impurity problems computationally much more tractable than fermionic lattice models), their numerical simulation remains a challenging task. Methods such as exact diagonalization or numerical RG, which explicitly treat a finite number of bath states, work well for single orbital models. However, because the number of bath states must be increased proportional to the number of orbitals, the computational effort grows exponentially with system size, and requires severe truncations of the bath already for two orbitals. Monte Carlo methods have the advantage that the bath is integrated out and thus the (infinite) size of the bath Hilbert space does not affect the simulation. While restricted to non-zero temperature, Monte Carlo methods are thus the method of choice for the solution of large multi-orbital or cluster impurity problems.

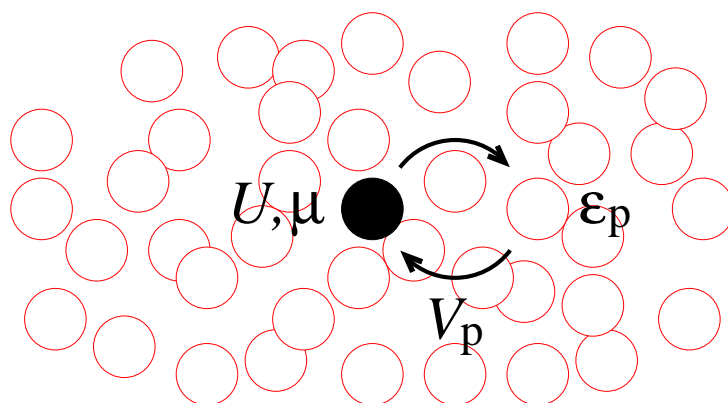


Fig. 1: Schematic representation of a quantum impurity model. Spin up and down electrons on the impurity (black dot) interact with a repulsive energy U and can hop to non-interacting bath levels ϵ_p with transition amplitude V_p^* .

Over the last few years, significant progress has been made (both in terms of efficiency and flexibility) with the development of continuous-time Monte Carlo techniques. This chapter provides an overview of two recently developed, complementary methods: (i) the weak-coupling approach, which scales favorably with system size and allows the efficient simulation of large impurity clusters, and (ii) the strong-coupling approach, which can handle impurity models with strong interactions. The contents of this chapter is based on lecture notes originally written for the Sherbrooke summer school in 2008 [1]. A comprehensive discussion of continuous-time impurity solvers can be found in a recently published review article [2].

For simplicity, we will focus on the single orbital Anderson impurity model (Fig. 1) defined by

the Hamiltonian $H = H_0 + H_U + H_{\text{bath}} + H_{\text{mix}}$ with

$$H_0 = -(\mu - U/2)(n_\uparrow + n_\downarrow), \quad (1)$$

$$H_U = U(n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2), \quad (2)$$

$$H_{\text{bath}} = \sum_{\sigma,p} \epsilon_p a_{p,\sigma}^\dagger a_{p,\sigma}, \quad (3)$$

$$H_{\text{mix}} = \sum_{\sigma,p} (V_p^\sigma d_\sigma^\dagger a_{p,\sigma} + h.c.). \quad (4)$$

Here, $H_0 + H_U \equiv H_{\text{loc}}$ describes the impurity with creation operators d_σ^\dagger , H_{bath} a non-interacting bath of electrons (labeled by quantum numbers p) with creation operators $a_{p,\sigma}^\dagger$, while H_{mix} controls the exchange of electrons between the impurity and the bath. The transition amplitudes V_p^σ are called *hybridizations*.

The impurity model partition function Z is given by

$$Z = \text{Tr} \left[e^{-\beta H} \right], \quad (5)$$

with β the inverse temperature, and $\text{Tr} = \text{Tr}_d \text{Tr}_a$ denotes the trace over the impurity and bath states. By *solving the impurity model* we essentially mean computing the impurity Green's function ($0 < \tau < \beta$)

$$g(\tau) = \langle T d(\tau) d^\dagger(0) \rangle = \frac{1}{Z} \text{Tr} \left[e^{-(\beta-\tau)H} d e^{-\tau H} d^\dagger \right], \quad (6)$$

which we choose to be positive.

Continuous-time Monte Carlo simulation relies on an expansion of the partition function into a series of diagrams and the stochastic sampling of (collections) of these diagrams. We represent the partition function as a sum (or, more precisely, integral) of configurations c with weight w_c ,

$$Z = \sum_c w_c, \quad (7)$$

and implement a random walk $c_1 \rightarrow c_2 \rightarrow c_3 \rightarrow \dots$ in configuration space in such a way that each configuration can be reached from any other in a finite number of steps (*ergodicity*) and that *detailed balance* is satisfied,

$$|w_{c_1}| p(c_1 \rightarrow c_2) = |w_{c_2}| p(c_2 \rightarrow c_1). \quad (8)$$

This assures that each configuration is visited with a probability proportional to $|w_c|$ and one can thus obtain an estimate for the Green's function from a finite number N of measurements:

$$g = \frac{\sum_c w_c g_c}{\sum_c w_c} = \frac{\sum_c |w_c| \text{sign}_c g_c}{\sum_c |w_c| \text{sign}_c} \approx \frac{\sum_{i=1}^N \text{sign}_{c_i} g_{c_i}}{\sum_{i=1}^N \text{sign}_{c_i}} = \frac{\langle \text{sign} \cdot g \rangle_{MC}}{\langle \text{sign} \rangle_{MC}}. \quad (9)$$

The error on this estimate decreases like $1/\sqrt{N}$. If the average sign of the configurations is small and decreases exponentially with decreasing temperature, the algorithm suffers from a *sign problem*.

2 General recipe

The first step in the diagrammatic expansion is to rewrite the partition function as a time ordered exponential using some *interaction representation*. We split the Hamiltonian into two parts, $H = H_1 + H_2$ and define the time dependent operators in the interaction picture as $O(\tau) = e^{\tau H_1} O e^{-\tau H_1}$. We furthermore introduce the operator $A(\beta) = e^{\beta H_1} e^{-\beta H}$ and write the partition function as $Z = \text{Tr}[e^{-\beta H_1} A(\beta)]$. The operator $A(\beta)$ satisfies $dA/d\beta = e^{\beta H_1} (H_1 - H) e^{-\beta H} = -H_2(\beta) A(\beta)$ and can be expressed as $A(\beta) = T \exp[-\int_0^\beta d\tau H_2(\tau)]$.

In a second step, the time ordered exponential is expanded into a power series,

$$\begin{aligned} Z &= \text{Tr} \left[e^{-\beta H_1} T e^{-\int_0^\beta d\tau H_2(\tau)} \right] \\ &= \sum_{n=0}^{\infty} \int_0^\beta d\tau_1 \dots \int_{\tau_{n-1}}^\beta d\tau_n \text{Tr} \left[e^{-(\beta-\tau_n)H_1} (-H_2) \dots e^{-(\tau_2-\tau_1)H_1} (-H_2) e^{-\tau_1 H_1} \right], \end{aligned} \quad (10)$$

which is a representation of the partition function of the form (7), namely the sum of all configurations $c = \{\tau_1, \dots, \tau_n\}$, with $n = 0, 1, \dots$ and $\tau_i \in [0, \beta]$ with weight

$$w_c = \text{Tr} \left[e^{-(\beta-\tau_n)H_1} (-H_2) \dots e^{-(\tau_2-\tau_1)H_1} (-H_2) e^{-\tau_1 H_1} \right] d\tau^n. \quad (11)$$

In the following we will discuss in detail two complementary diagrammatic Monte Carlo algorithms, namely

1. a *weak-coupling* approach, based on an expansion of Z in powers of the interaction U , and on an interaction representation in which the time evolution is determined by the *quadratic* part $H_0 + H_{\text{bath}} + H_{\text{mix}}$ of the Hamiltonian,
2. a *strong-coupling* approach, based on an expansion of Z in powers of the impurity-bath hybridization V , and an interaction representation in which the time evolution is determined by the *local* part $H_0 + H_U + H_{\text{bath}}$ of the Hamiltonian.

3 Weak-coupling approach - expansion in the interaction

The first diagrammatic impurity solver, proposed by Rubtsov *et al.* in 2005 [3], is based on an expansion in powers of the interaction. Here, we will discuss a variant of the weak coupling approach, worked out by Gull *et al.* [4], which combines the weak-coupling expansion with an auxiliary field decomposition. This *continuous-time auxiliary field method* is an adaptation of an algorithm by Rombouts *et al.* [5] for lattice models (the first diagrammatic Monte Carlo algorithm for Fermions) and in some respects similar to the time-honored Hirsch-Fye algorithm [6].

3.1 Monte Carlo configurations

Following Rombouts and collaborators, we define $H_2 = H_U - K/\beta$ and $H_1 = H - H_2 = H_0 + H_{\text{bath}} + H_{\text{mix}} + K/\beta$, with K some non-zero constant. Equation (10) then gives the expression for the partition function after expansion in H_2 , and (11) the weight of a configuration

of n interaction vertices. At this stage, we extend our configuration space by decoupling each interaction vertex using the decoupling formula proposed by Rombouts,

$$-H_2 = K/\beta - U(n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2) = \frac{K}{2\beta} \sum_{s=-1,1} e^{\gamma s(n_\uparrow - n_\downarrow)}, \quad (12)$$

$$\cosh(\gamma) = 1 + (\beta U)/(2K). \quad (13)$$

This formula can easily be verified by checking the four states $|0\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$, and $|\uparrow\downarrow\rangle$. The configuration space is now the collection of all possible *Ising spin configurations* on the imaginary time interval $[0, \beta)$: $c = \{\{\tau_1, s_1\}, \dots, \{\tau_n, s_n\}\}$, $n = 0, 1, \dots$, $\tau_i \in [0, \beta)$, $s_i = \pm 1$. These configurations have weight

$$w_c = \text{Tr} \left[e^{-(\beta-\tau_n)H_1} e^{\gamma s_n(n_\uparrow - n_\downarrow)} \dots e^{-(\tau_2-\tau_1)H_1} e^{\gamma s_1(n_\uparrow - n_\downarrow)} e^{-\tau_1 H_1} \right] \left(\frac{K d\tau}{2\beta} \right)^n. \quad (14)$$

All the operators in the trace are quadratic in c and a , so we can first separate the spin components and then proceed to the analytical calculation of the trace. Introducing $H_1^\sigma = -\mu(n_\sigma - U/2) + \sum_p \epsilon_p a_{p,\sigma}^\dagger a_{p,\sigma} + \sum_p (V_{\sigma,p} c_\sigma^\dagger a_{p,\sigma} + h.c.)$, which is the Hamiltonian of the non-interacting impurity model, the trace in Eq. (14) becomes ($Z_{0,\sigma} = \text{Tr}[e^{-\beta H_1^\sigma}]$)

$$\text{Tr} [\dots] = e^{-K} \prod_\sigma \text{Tr} \left[e^{-(\beta-\tau_n)H_1^\sigma} e^{\gamma s_n \sigma n_\sigma} \dots e^{-(\tau_2-\tau_1)H_1^\sigma} e^{\gamma s_1 \sigma n_\sigma} e^{-\tau_1 H_1^\sigma} \right]. \quad (15)$$

Using the identity $e^{\gamma s \sigma n_\sigma} = e^{\gamma s \sigma} c_\sigma^\dagger c_\sigma + c_\sigma c_\sigma^\dagger = e^{\gamma s \sigma} - (e^{\gamma s \sigma} - 1) c_\sigma c_\sigma^\dagger$, the trace factors can be expressed in terms of non-interacting impurity Green's functions g_0 and evaluated using Wick's theorem. For example, at first order, we find

$$\text{Tr} \left[e^{-(\beta-\tau_1)H_1^\sigma} (e^{\gamma s \sigma} - (e^{\gamma s \sigma} - 1) c_\sigma c_\sigma^\dagger) e^{-\tau_1 H_1^\sigma} \right] = Z_{0,\sigma} (e^{\gamma s \sigma} - g_{0\sigma}(0_+) (e^{\gamma s \sigma} - 1)). \quad (16)$$

For n spins, this expression generalizes to

$$\text{Tr} \left[e^{-(\beta-\tau_n)H_1^\sigma} e^{\gamma s_n \sigma n_\sigma} \dots e^{-(\tau_2-\tau_1)H_1^\sigma} e^{\gamma s_1 \sigma n_\sigma} e^{-\tau_1 H_1^\sigma} \right] = Z_{0,\sigma} \det N_\sigma^{-1}(\{s_i, \tau_i\}), \quad (17)$$

where N_σ is a $(n \times n)$ matrix defined by the location of the decoupled interaction vertices, the spin orientations, and the non-interacting Green's functions:

$$N_\sigma^{-1}(\{s_i, \tau_i\}) \equiv e^{\Gamma_\sigma} - G_{0\sigma}(e^{\Gamma_\sigma} - I). \quad (18)$$

The notation is $e^{\Gamma_\sigma} \equiv \text{diag}(e^{\gamma s_1 \sigma}, \dots, e^{\gamma s_n \sigma})$, $(G_{0\sigma})_{i,j} = g_{0\sigma}(\tau_i - \tau_j)$ for $i \neq j$, $(G_{0\sigma})_{i,i} = g_{0\sigma}(0_+)$. Combining Eqs. (14), (15), (17) and (18) we thus obtain the following weight for the configuration $c = \{\{\tau_1, s_1\}, \dots, \{\tau_n, s_n\}\}$:

$$w_c = e^{-K} \left(\frac{K d\tau}{2\beta} \right)^n \prod_\sigma Z_{0,\sigma} \det N_\sigma^{-1}(\{s_i, \tau_i\}). \quad (19)$$

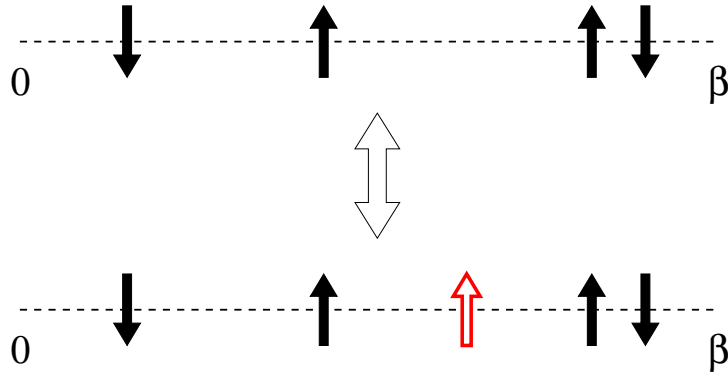


Fig. 2: Local update in the continuous-time auxiliary field method. The dashed line represents the imaginary time interval $[0, \beta)$. We increase the perturbation order by adding a spin with random orientation at a random time. The perturbation order is decreased by removing a randomly chosen spin.

3.2 Sampling procedure and detailed balance

For ergodicity it is sufficient to insert/remove spins with random orientation at random times, because this allows in principle to generate all possible configurations. Furthermore, the random walk in configuration space must satisfy the detailed balance condition (8). Splitting the probability to move from configuration c_i to configuration c_j into a probability to *propose* the move and a probability to *accept* it,

$$p(c_i \rightarrow c_j) = p^{\text{prop}}(c_i \rightarrow c_j)p^{\text{acc}}(c_i \rightarrow c_j), \quad (20)$$

we arrive at the condition

$$\frac{p^{\text{acc}}(c_i \rightarrow c_j)}{p^{\text{acc}}(c_j \rightarrow c_i)} = \frac{p^{\text{prop}}(c_j \rightarrow c_i) |w(c_j)|}{p^{\text{prop}}(c_i \rightarrow c_j) |w(c_i)|}. \quad (21)$$

There is some flexibility in choosing the proposal probabilities. A reasonable choice for the insertion/removal of a spin is the following (illustrated in Fig. 2):

- *Insertion*

Pick a random time in $[0, \beta)$ and a random direction for the new spin:

$$p^{\text{prop}}(n \rightarrow n+1) = (1/2)(d\tau/\beta),$$

- *Removal*

Pick a random spin: $p^{\text{prop}}(n+1 \rightarrow n) = 1/(n+1)$.

For this choice, the ratio of acceptance probabilities becomes

$$\frac{p^{\text{acc}}(n \rightarrow n+1)}{p^{\text{acc}}(n+1 \rightarrow n)} = \frac{K}{n+1} \prod_{\sigma=\uparrow, \downarrow} \frac{|\det(N_\sigma^{(n+1)})^{-1}|}{|\det(N_\sigma^{(n)})^{-1}|}, \quad (22)$$

and the random walk can thus be implemented for example on the basis of the Metropolis algorithm, *i.e.* the proposed move from n to $n \pm 1$ is accepted with probability

$$\min \left[1, \frac{p^{\text{acc}}(n \rightarrow n \pm 1)}{p^{\text{acc}}(n \pm 1 \rightarrow n)} \right]. \quad (23)$$

3.3 Determinant ratios and fast matrix updates

From Eq. (22) it follows that each update requires the calculation of a ratio of two determinants. Computing the determinant of a matrix of size $(n \times n)$ is an $\mathcal{O}(n^3)$ operation (LU decomposition). The important thing to realize is that each insertion or removal of a spin merely changes one row and one column of the matrix N_σ^{-1} . We will now show that it is therefore possible to evaluate the ratio in Eq. (22) in a time $\mathcal{O}(n^2)$ (insertion) or $\mathcal{O}(1)$ (removal).

The objects which are stored and manipulated during the simulation are, besides the lists of the times $\{\tau_i\}$ and spins $\{s_i\}$, the matrices $N_\sigma = (e^{\Gamma_\sigma} - G_{0\sigma}(e^{\Gamma_\sigma} - I))^{-1}$. Inserting a spin adds a new row and column to N_σ^{-1} . We define the blocks (omitting the σ index)

$$(N^{(n+1)})^{-1} = \begin{pmatrix} (N^{(n)})^{-1} & Q \\ R & S \end{pmatrix}, \quad N^{(n+1)} = \begin{pmatrix} \tilde{P} & \tilde{Q} \\ \tilde{R} & \tilde{S} \end{pmatrix}, \quad (24)$$

where Q , R , S denote $(n \times 1)$, $(1 \times n)$, and (1×1) matrices, respectively, which contain the contribution of the added spin. The determinant ratio needed for the acceptance/rejection probability is then given by

$$\frac{\det(N^{(n+1)})^{-1}}{\det(N^{(n)})^{-1}} = \frac{1}{\det \tilde{S}} = S - [R][N^{(n)}Q]. \quad (25)$$

As we store $N^{(n)}$, computing the acceptance/rejection probability of an insertion move is an $\mathcal{O}(n^2)$ operation. If the move is accepted, the new matrix $N^{(n+1)}$ is computed out of $N^{(n)}$, Q , R , and S , also in a time $\mathcal{O}(n^2)$:

$$\tilde{S} = (S - [R][N^{(n)}Q])^{-1}, \quad (26)$$

$$\tilde{Q} = -[N^{(n)}Q]\tilde{S}, \quad (27)$$

$$\tilde{R} = -\tilde{S}[RN^{(n)}], \quad (28)$$

$$\tilde{P} = N^{(n)} + [N^{(n)}Q]\tilde{S}[RN^{(n)}]. \quad (29)$$

It follows from Eq. (25) that the calculation of the determinant ratio for removing a spin is $\mathcal{O}(1)$, since it is just element \tilde{S} , and from the above formulas we also immediately find the elements of the reduced matrix:

$$N^{(n)} = \tilde{P} - \frac{[\tilde{Q}][\tilde{R}]}{\tilde{S}}. \quad (30)$$

3.4 Measurement of the Green's function

To compute the contribution of a configuration c to the Green's function measurement (6), we insert a creation operator d^\dagger at time 0 and an annihilation operator d at time τ ,

$$g_\sigma^c(\tau) = \frac{1}{w_c} \text{Tr} \left[e^{-(\beta-\tau_n)H_1} e^{\gamma s_n(n_\uparrow - n_\downarrow)} \dots e^{-(\tau_{k+1}-\tau)H_1} d_\sigma e^{-(\tau-\tau_k)H_1} \dots e^{\gamma s_1(n_\uparrow - n_\downarrow)} e^{-\tau_1 H_1} d_\sigma^\dagger \right] \left(\frac{K d\tau}{2\beta} \right)^n$$

with w_c given in Eq. (14). The same steps as in section 3.1 (Wick's theorem) then lead to the expression

$$\begin{aligned} g_\sigma^c(\tau) &= \frac{1}{\det N_\sigma^{-1} \det N_{\bar{\sigma}}^{-1} \det N_{\bar{\sigma}}^{-1}} \det N_{\bar{\sigma}}^{-1} \det \begin{pmatrix} (N_\sigma^{(n)})^{-1} & [g_{0\sigma}(\tau_i)] \\ -[g_{0\sigma}(\tau - \tau_j)(e^{\Gamma_{\sigma j}} - 1)] & g_{0\sigma}(\tau) \end{pmatrix} \\ &= g_{0\sigma}(\tau) + [g_{0\sigma}(\tau - \tau_j)(e^{\Gamma_{\sigma j}} - 1)] N_\sigma^{(n)} [g_{0\sigma}(\tau_i)]. \end{aligned} \quad (31)$$

The second equality follows from Eq. (25) and square brackets denote vectors of length n . To avoid unnecessary and time consuming summations during the Monte Carlo simulations, we only accumulate the quantity

$$S_\sigma(\tilde{\tau}) \equiv \sum_{k=1}^n \delta(\tilde{\tau} - \tau_k) \sum_{l=1}^n [(e^{\Gamma_\sigma} - I) N_\sigma]_{kl} g_{0\sigma}(\tau_l), \quad (32)$$

binning the time points $\tilde{\tau}$ on a fine grid. After the simulation is completed, the Green's function is computed as

$$g_\sigma(\tau) = g_{0\sigma}(\tau) + \int_0^\beta d\tilde{\tau} g_{0\sigma}(\tau - \tilde{\tau}) \langle S_\sigma(\tilde{\tau}) \rangle_{MC}. \quad (33)$$

3.5 Expansion order

It follows from Eq. (10) that

$$\begin{aligned} \langle -H_2 \rangle &= \frac{1}{\beta} \int_0^\beta d\tau \langle -H_2(\tau) \rangle \\ &= \frac{1}{\beta} \frac{1}{Z} \sum_{n=0}^{\infty} \frac{n+1}{(n+1)!} \int_0^\beta d\tau \int_0^\beta d\tau_1 \dots \int_0^\beta d\tau_n \text{Tr} \left[e^{-\beta H_1} T(-H_2(\tau)) (-H_2(\tau_n)) \dots (-H_2(\tau_1)) \right] \\ &= \frac{1}{\beta} \frac{1}{Z} \sum_c n(c) w_c = \frac{1}{\beta} \langle n \rangle, \end{aligned} \quad (34)$$

and because $\langle -H_2 \rangle = K/\beta - U \langle n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2 \rangle$ we conclude that the average perturbation order $\langle n \rangle$ is related to the parameter K and the potential energy by

$$\langle n \rangle = K - \beta U \langle n_\uparrow n_\downarrow - (n_\uparrow + n_\downarrow)/2 \rangle. \quad (35)$$

Increasing K leads to a higher perturbation order (and thus slower matrix updates), but through Eq. (13) also to a smaller value of γ and thus to less polarization of the auxiliary spins. A K of the order 1 appears to work well. We also learn from Eq. (35) that the average perturbation order grows essentially proportional to U (as expected for a weak-coupling method), and proportional to inverse temperature.

4 Strong coupling approach - expansion in the impurity-bath hybridization

The second continuous-time method, which is in many ways complementary to the weak-coupling approach, is based on an expansion of the partition function in powers of the impurity-bath hybridization V . This method has been developed in Ref. [7] and applied to the Anderson impurity model. A more general *matrix* formulation which allows to treat arbitrary impurity models was presented in Refs. [8, 9]. An alternative to the matrix formulation, which we will not touch in this chapter, is the recently proposed *Krylov* method [10].

4.1 Monte Carlo configurations

Here, we decompose the Hamiltonian as $H_2 = H_{\text{mix}}$ and $H_1 = H - H_2 = H_0 + H_U + H_{\text{bath}}$. Since $H_2 \equiv H_2^{d^\dagger} + H_2^d = \sum_{\sigma,p} V_p^\sigma d_\sigma^\dagger a_{p,\sigma} + \sum_{\sigma,p'} V_{p'}^{\sigma*} d_\sigma a_{p',\sigma}^\dagger$ has two terms, corresponding to electrons hopping from the bath to the impurity and from the impurity back to the bath, only even perturbation orders contribute to Eq. (10). Furthermore, at perturbation order $2n$ only the $(2n)!/(n!)^2$ terms corresponding to n creation operators d^\dagger and n annihilation operators d will contribute. We can therefore write the partition function as a sum over configurations $c = \{\tau_1, \dots, \tau_n; \tau'_1, \dots, \tau'_n\}$:

$$Z = \sum_{n=0}^{\infty} \int_0^\beta d\tau_1 \dots \int_{\tau_{n-1}}^\beta d\tau_n \int_0^\beta d\tau'_1 \dots \int_{\tau'_{n-1}}^\beta d\tau'_n \text{Tr} \left[e^{-\beta H_1} T H_2^d(\tau_n) H_2^{d^\dagger}(\tau'_n) \dots H_2^d(\tau_1) H_2^{d^\dagger}(\tau'_1) \right].$$

Since the time evolution of the Anderson model (given by H_1) does not rotate the spin, there is an additional constraint, namely that both for spin up and spin down, there is an equal number of creation and annihilation operators. Taking this into account and writing out the expressions for H_2^d and $H_2^{d^\dagger}$ explicitly, we find

$$\begin{aligned} Z = & \sum_{\{n_\sigma\}} \prod_{\sigma} \int_0^\beta d\tau_1^\sigma \dots \int_{\tau_{n_\sigma-1}^\sigma}^\beta d\tau_{n_\sigma}^\sigma \int_0^\beta d\tau'_1{}^\sigma \dots \int_{\tau'_{n_\sigma-1}{}^\sigma}^\beta d\tau'_{n_\sigma}{}^\sigma \\ & \times \text{Tr} \left[e^{-\beta H_1} T \prod_{\sigma} \sum_{p_1, \dots, p_{n_\sigma}} \sum_{p'_1, \dots, p'_{n_\sigma}} V_{p_1}^\sigma V_{p'_1}^{\sigma*} \dots V_{p_{n_\sigma}}^\sigma V_{p'_{n_\sigma}}^{\sigma*} \right. \\ & \left. d_\sigma(\tau_{n_\sigma}^\sigma) a_{\sigma, p_{n_\sigma}}^\dagger(\tau_{n_\sigma}^\sigma) a_{\sigma, p'_{n_\sigma}}(\tau'_{n_\sigma}{}^\sigma) d_\sigma^\dagger(\tau'_{n_\sigma}{}^\sigma) \dots d_\sigma(\tau_1^\sigma) a_{\sigma, p_1}^\dagger(\tau_1^\sigma) a_{\sigma, p'_1}(\tau'_1{}^\sigma) d_\sigma^\dagger(\tau'_1{}^\sigma) \right]. \quad (36) \end{aligned}$$

Now, because the d and a operate on different spaces and H_1 does not mix the impurity and bath states, we can separate the bath and the impurity and write

$$\begin{aligned}
Z &= Z_{\text{bath}} \sum_{\{n_\sigma\}} \prod_{\sigma} \int_0^{\beta} d\tau_1^{\sigma} \dots \int_{\tau_{n_\sigma-1}^{\sigma}}^{\beta} d\tau_{n_\sigma}^{\sigma} \int_0^{\beta} d\tau_1^{\prime\sigma} \dots \int_{\tau_{n_\sigma-1}^{\prime\sigma}}^{\beta} d\tau_{n_\sigma}^{\prime\sigma} \\
&\times \text{Tr}_d \left[e^{-\beta H_{\text{loc}}} T \prod_{\sigma} d_{\sigma}(\tau_{n_\sigma}^{\sigma}) d_{\sigma}^{\dagger}(\tau_{n_\sigma}^{\prime\sigma}) \dots d_{\sigma}(\tau_1^{\sigma}) d_{\sigma}^{\dagger}(\tau_1^{\prime\sigma}) \right] \\
&\times \frac{1}{Z_{\text{bath}}} \text{Tr}_a \left[e^{-\beta H_{\text{bath}}} T \prod_{\sigma} \sum_{p_1, \dots, p_{n_\sigma}} \sum_{p'_1, \dots, p'_{n_\sigma}} V_{p_1}^{\sigma} V_{p'_1}^{\sigma*} \dots V_{p_{n_\sigma}}^{\sigma} V_{p'_{n_\sigma}}^{\sigma*} \right. \\
&\quad \left. a_{\sigma, p_{n_\sigma}}^{\dagger}(\tau_{n_\sigma}^{\sigma}) a_{\sigma, p'_{n_\sigma}}(\tau_{n_\sigma}^{\prime\sigma}) \dots a_{\sigma, p_1}^{\dagger}(\tau_1^{\sigma}) a_{\sigma, p'_1}(\tau_1^{\prime\sigma}) \right], \quad (37)
\end{aligned}$$

where $Z_{\text{bath}} = \text{Tr}_a e^{-\beta H_{\text{bath}}}$, and $H_{\text{loc}} = H_0 + H_U$. Since the bath is non-interacting, there is a Wick theorem for the bath and $\text{Tr}_a[\dots]$ can be expressed as the determinant of some matrix, whose size is equal to the perturbation order. To find the elements of this matrix, it is useful to consider the lowest perturbation order, $n_\sigma = 1$, $n_{\bar{\sigma}} = 0$. In this case

$$\begin{aligned}
&\sum_{p_1} \sum_{p'_1} V_{p_1}^{\sigma} V_{p'_1}^{\sigma*} \frac{1}{Z_{\text{bath}}} \text{Tr}_a \left[e^{-\beta H_{\text{bath}}} T a_{\sigma, p_1}^{\dagger}(\tau_1^{\sigma}) a_{\sigma, p'_1}(\tau_1^{\prime\sigma}) \right] \\
&= \sum_{p_1} \frac{|V_{p_1}^{\sigma}|^2}{e^{-\epsilon_{p_1} \beta} + 1} \begin{cases} e^{-\epsilon_{p_1}(\beta - (\tau_1^{\sigma} - \tau_1^{\prime\sigma}))} & \tau_1^{\sigma} > \tau_1^{\prime\sigma} \\ -e^{-\epsilon_{p_1}(\tau_1^{\prime\sigma} - \tau_1^{\sigma})} & \tau_1^{\sigma} < \tau_1^{\prime\sigma} \end{cases}. \quad (38)
\end{aligned}$$

Note that $Z_{\text{bath}} = \prod_{\sigma} \prod_p (e^{-\epsilon_p \beta} + 1)$. Introducing the β -antiperiodic *hybridization function*

$$F_{\sigma}(\tau) = \sum_p \frac{|V_p|^2}{e^{-\epsilon_p \beta} + 1} \begin{cases} e^{-\epsilon_p(\beta - \tau)} & \tau > 0 \\ -e^{-\epsilon_p(-\tau)} & \tau < 0 \end{cases}, \quad F_{\sigma}(-i\omega_n) = \sum_p \frac{|V_p|^2}{i\omega_n - \epsilon_p}, \quad (39)$$

which is related to the non-interacting Green's function $G_{0\sigma}$ of Section 3 by $F_{\sigma}(-i\omega_n) = i\omega_n + \mu - U/2 - G_{0\sigma}(i\omega_n)^{-1}$, the first order result becomes $F_{\sigma}(\tau_1^{\sigma} - \tau_1^{\prime\sigma})$. For higher orders, one obtains

$$\begin{aligned}
&\frac{1}{Z_{\text{bath}}} \text{Tr}_a \left[e^{-\beta H_{\text{bath}}} T \prod_{\sigma} \sum_{p_1, \dots, p_{n_\sigma}} \sum_{p'_1, \dots, p'_{n_\sigma}} V_{p_1}^{\sigma} V_{p'_1}^{\sigma*} \dots V_{p_{n_\sigma}}^{\sigma} V_{p'_{n_\sigma}}^{\sigma*} \right. \\
&\quad \left. a_{\sigma, p_{n_\sigma}}^{\dagger}(\tau_{n_\sigma}^{\sigma}) a_{\sigma, p'_{n_\sigma}}(\tau_{n_\sigma}^{\prime\sigma}) \dots a_{\sigma, p_1}^{\dagger}(\tau_1^{\sigma}) a_{\sigma, p'_1}(\tau_1^{\prime\sigma}) \right] = \prod_{\sigma} \det M_{\sigma}^{-1}, \quad (40)
\end{aligned}$$

where M_{σ}^{-1} is a $(n_{\sigma} \times n_{\sigma})$ matrix with elements

$$M_{\sigma}^{-1}(i, j) = F_{\sigma}(\tau_i^{\sigma} - \tau_j^{\prime\sigma}). \quad (41)$$

In the hybridization expansion method, the configuration space consists of all sequences $c = \{\tau_1^{\uparrow}, \dots, \tau_{n_{\uparrow}}^{\uparrow}; \tau_1^{\prime\uparrow}, \dots, \tau_{n_{\uparrow}}^{\prime\uparrow} | \tau_1^{\downarrow}, \dots, \tau_{n_{\downarrow}}^{\downarrow}; \tau_1^{\prime\downarrow}, \dots, \tau_{n_{\downarrow}}^{\prime\downarrow}\}$, of n_{\uparrow} creation and annihilation operators for spin up ($n_{\uparrow} = 0, 1, \dots$), and n_{\downarrow} creation and annihilation operators for spin down ($n_{\downarrow} =$

$0, 1, \dots$). The weight of this configuration is

$$w_c = Z_{\text{bath}} \text{Tr}_d \left[e^{-\beta H_{\text{loc}}} T \prod_{\sigma} d_{\sigma}(\tau_{n_{\sigma}}^{\sigma}) d_{\sigma}^{\dagger}(\tau'_{n_{\sigma}})^{\sigma} \dots d_{\sigma}(\tau_1^{\sigma}) d_{\sigma}^{\dagger}(\tau'_1)^{\sigma} \right] \\ \times \prod_{\sigma} \det M_{\sigma}^{-1}(\tau_1^{\sigma}, \dots, \tau_{n_{\sigma}}^{\sigma}; \tau'_1{}^{\sigma}, \dots, \tau'_{n_{\sigma}}{}^{\sigma}) (d\tau)^{2n_{\sigma}}. \quad (42)$$

The trace factor represents the contribution of the impurity, which fluctuates between different quantum states, as electrons hop in and out. The determinants resum all the bath evolutions which are compatible with the given sequence of transitions.

To evaluate the trace factor, we use the eigenbasis of H_{loc} , which is $|0\rangle$ (energy $E_0 = 0$), $|\uparrow\rangle$, $|\downarrow\rangle$ (energy $E_1 = -\mu$) and $|\uparrow\downarrow\rangle$ (energy $E_2 = U - 2\mu$). In this basis, the time evolution operator $e^{-\tau H_{\text{loc}}} = \text{diag}(e^{-\tau E_0}, e^{-\tau E_1}, e^{-\tau E_1}, e^{-\tau E_2})$ is diagonal while the operators d_{σ} and d_{σ}^{\dagger} will produce transitions between eigenstates with amplitude ± 1 .

Because the time evolution does not flip the spin, the creation and annihilation operators for given spin have to alternate. This allows us to separate the operators for spin up from those for spin down and to depict the time evolution by a *collection of segments* (each segment representing a time interval in which an electron of spin up or down resides on the impurity). At each time, the eigenstate of the impurity follows immediately from the segment representation and we can easily compute the trace factor as (s is a permutation sign)

$$\text{Tr}_d \left[e^{-\beta H_{\text{loc}}} T \prod_{\sigma} d_{\sigma}(\tau_{n_{\sigma}}^{\sigma}) d_{\sigma}^{\dagger}(\tau'_{n_{\sigma}})^{\sigma} \dots d_{\sigma}(\tau_1^{\sigma}) d_{\sigma}^{\dagger}(\tau'_1)^{\sigma} \right] = s \exp \left[\mu(l_{\uparrow} + l_{\downarrow}) - U l_{\text{overlap}} \right], \quad (43)$$

with l_{σ} the total *length* of the segments for spin σ and l_{overlap} the total length of the overlap between up and down segments. The lower panel of Fig. 3 shows a configuration with 3 segments for spin up and two segments for spin down; the time intervals where segments overlap, indicated by gray rectangles, correspond to a doubly occupied impurity and cost a repulsion energy U .

4.2 Sampling procedure and detailed balance

For ergodicity, it is sufficient to insert and remove pairs of creation and annihilation operators (segments or anti-segments) for spin up and down. One possible strategy for inserting a segment is the following: we pick a random time in $[0, \beta)$ for the creation operator. If it falls on an existing segment, the impurity is already occupied and the move is rejected. If it falls on an empty space, we compute l_{max} , the length from this position to the next segment (in the direction of increasing τ). If there are no segments, $l_{\text{max}} = \beta$. The position of the new annihilation operator is then chosen randomly in this interval of length l_{max} (see Fig. 3). If we propose to remove a randomly chosen segment for this spin, then the proposal probabilities are

$$p^{\text{prop}}(n_{\sigma} \rightarrow n_{\sigma} + 1) = \frac{d\tau}{\beta} \frac{d\tau}{l_{\text{max}}}, \quad (44)$$

$$p^{\text{prop}}(n_{\sigma} + 1 \rightarrow n_{\sigma}) = \frac{1}{n_{\sigma} + 1}, \quad (45)$$

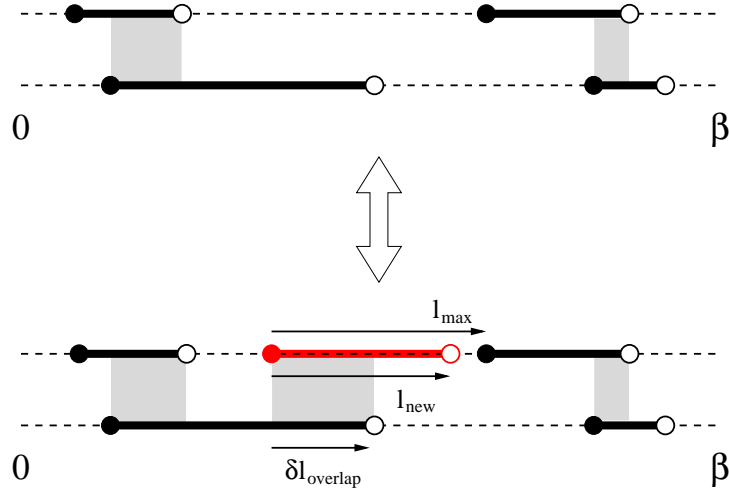


Fig. 3: Local update in the segment picture. The two segment configurations correspond to spin up and spin down. Each segment depicts a time interval in which an electron of the corresponding spin resides on the impurity (the end points are the locations of the operators d^\dagger and d). We increase the perturbation order by adding a segment or anti-segment of random length for random spin. The perturbation order is decreased by removing a randomly chosen segment. (Figure from Ref. [1].)

and the ratio of acceptance probabilities therefore becomes

$$\frac{p^{\text{acc}}(n_\sigma \rightarrow n_\sigma + 1)}{p^{\text{acc}}(n_\sigma + 1 \rightarrow n_\sigma)} = \frac{\beta l_{\text{max}}}{n_\sigma + 1} e^{U l_{\text{new}} - U \delta l_{\text{overlap}}} \frac{|\det(M_\sigma^{(n_\sigma+1)})^{-1}|}{|\det(M_\sigma^{(n_\sigma)})^{-1}|}. \quad (46)$$

Here, l_{new} is the length of the new segment, and $\delta l_{\text{overlap}}$ the change in the overlap. Again, we compute the ratio of determinants using the fast update formulas discussed in Section 3.

4.3 Measurement of the Green's function

The strategy is to create configurations which contribute to the Green's function measurement by decoupling the bath from a given pair of creation and annihilation operators in c . The idea is to write

$$g(\tau) = \frac{1}{Z} \sum_c w_c^{d(\tau)d^\dagger(0)} = \frac{1}{Z} \sum_c w_c^{(\tau,0)} \frac{w_c^{d(\tau)d^\dagger(0)}}{w_c^{(\tau,0)}}, \quad (47)$$

where $w_c^{d(\tau)d^\dagger(0)}$ denotes the weight of configuration c with an additional operator $d^\dagger(0)$ and $d(\tau)$ in the trace factor, and $w_c^{(\tau,0)}$ the complete weight corresponding to the enlarged operator sequence (including enlarged hybridization determinants). Since the trace factors of both weights are identical, and $\det M_c^{-1}$ is a minor of $\det(M_c^{(\tau,0)})^{-1}$, we find

$$\frac{w_c^{d(\tau)d^\dagger(0)}}{w_c^{(\tau,0)}} = \frac{\det M_c^{-1}}{\det(M_c^{(\tau,0)})^{-1}} = (M_c^{(\tau,0)})_{j,i}, \quad (48)$$

with i and j denoting the row and column corresponding to the new operators d^\dagger and d in the enlarged $(M_c^{(\tau,0)})^{-1}$. To transform the sum over c into a sum over configurations $\tilde{c} = \{c, \tau_i, \tau'_j\}$, the new operators must be free to be anywhere on the imaginary time interval, which (due to translational invariance) yields a factor $\frac{1}{\beta}\Delta(\tau, \tau_i - \tau'_j)$, with

$$\Delta(\tau, \tau') = \begin{cases} \delta(\tau - \tau') & \tau' > 0 \\ -\delta(\tau - \tau' - \beta) & \tau' < 0 \end{cases}. \quad (49)$$

Hence, the measurement formula for the Green's function becomes

$$g(\tau) = \frac{1}{Z} \sum_{\tilde{c}} w_{\tilde{c}} \sum_{i,j} \frac{1}{\beta} \Delta(\tau, \tau_i - \tau'_j) (M_{\tilde{c}})_{j,i} = \left\langle \sum_{i,j} \frac{1}{\beta} \Delta(\tau, \tau_i - \tau'_j) M_{j,i} \right\rangle_{MC}. \quad (50)$$

Note that if we let all the integrals run from 0 to β , there is a factor $1/(n!)^2$ in w_c and $1/((n+1)!)^2$ in $w_{\tilde{c}}$, with n the size of M_c . Changing from a sum over c to a sum over \tilde{c} therefore adds a factor $(n+1)^2$ if we restrict the measurement to a specific pair of d^\dagger and d . Equivalently, we can sum over all the $(n+1)^2$ pairs of operators in the enlarged configuration.

4.4 Generalization - Matrix formalism

It is obvious from the derivation in Section 4.1 that the hybridization expansion formalism is applicable to general classes of impurity models. Because the trace factor in the weight (42) is computed exactly, H_{loc} can contain essentially arbitrary interactions (e. g. spin-exchange terms in multi-orbital models), degrees of freedom (e. g. spins in Kondo-lattice models) or constraints (e. g. no double occupancy in the t - J model).

For multi-orbital impurity models with density-density interaction, the segment formalism is still applicable: we have now a collection of segments for each *flavor* α (orbital, spin) and the trace factor can still be computed from the length of the segments (chemical potential contribution) and the overlaps between segments of different flavor (interaction terms).

If H_{loc} is not diagonal in the occupation number basis defined by the d_α^\dagger , the calculation of $\text{Tr}_d[e^{-\beta H_{\text{loc}}} T \prod_\alpha d_\alpha(\tau_{n_\alpha}^\alpha) d_\alpha^\dagger(\tau'_{n_\alpha}^\alpha) \dots d_\sigma(\tau_1^\sigma) d_\sigma^\dagger(\tau'_1^\sigma)]$ becomes more involved. We now have to compute the trace explicitly in some basis of H_{loc} – for example the eigenbasis, in which the time evolution operators $e^{-H_{\text{loc}}\tau}$ become diagonal. The operators d_α and d_α^\dagger are expressed as matrices in this eigenbasis, and the evaluation of the trace factor thus involves the multiplication of matrices whose size is equal to the size of the Hilbert space of H_{loc} . Since the dimension of the Hilbert space grows *exponentially* with the number of flavors, the calculation of the trace factor becomes the computational bottleneck of the simulation, and the matrix formalism is therefore restricted to a relatively small number of flavors ($\lesssim 10$).

An important point, explained in Ref. [9], is the use of *conserved quantum numbers* (typically particle number for spin up and spin down, momentum, ...). If the eigenstates of H_{loc} are grouped according to these quantum numbers, the operator matrices will acquire a sparse block structure, because for example $d_{\uparrow,q}^\dagger$ will connect the states corresponding to quantum numbers $m = \{n_\uparrow, n_\downarrow, K\}$ to those corresponding to $m' = \{n_\uparrow + 1, n_\downarrow, K + q\}$ (if they exist). Checking

the compatibility of the operator sequence with a given starting block furthermore allows one to find the (potentially) contributing quantum number sectors without any matrix multiplications. The evaluation of the trace is thus reduced to a block matrix multiplication of the form

$$\sum_{\text{contr. } m} \text{Tr}_m \left[\dots (O)_{m'',m'} (e^{-(\tau'-\tau)H_{\text{loc}}})_{m'} (O)_{m',m} (e^{-\tau H_{\text{loc}}})_m \right]. \quad (51)$$

5 Comparison between the two approaches

The weak- and strong-coupling methods are in many ways complementary and their respective strengths/weaknesses result from the scaling of the computational effort with interaction strength and system size. For the Anderson impurity model considered in these notes, the U dependence of the average perturbation order is shown in Fig. 4 (these are dynamical mean field theory calculations for a one-band Hubbard model taken from Ref. [11]). In the weak-coupling algorithms, where the average perturbation order is related to the potential energy, one finds a roughly linear increase of the perturbation order with U . In the hybridization-expansion method, the average perturbation order is related to the kinetic energy, and decreases as the interaction strength increases. Thus, in single site models with only density density interactions, where the evaluation of the trace factor in Eq. (42) is cheap, the hybridization expansion method beats the weak coupling method in the regime of strong correlations.

For more complicated models, which require the matrix formalism discussed in section 4.4, the hybridization expansion method scales exponentially with system size, and can only be applied to relatively small systems.¹ Here, the weak-coupling approach – if applicable – becomes the method of choice. Table 1 gives a summary of the different scalings (assuming diagonal hybridization) and indicates which solver is appropriate for which type of problem.

solver	scaling		use for
weak-coupling	β^3	L^3	impurity clusters with density-density interactions and hopping
hybridization expansion (segment formulation)	β^3	L	single site multi-orbital models with density-density interaction
hybridization expansion (matrix formulation)	β	$\exp(L)$	single site multi-orbital models with general U_{ijkl}

Table 1: *Scaling of the different impurity solvers with inverse temperature and system size.*

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

¹While the calculation of the trace over atomic states is time consuming, it also yields useful information about the system (histogram of visited states).

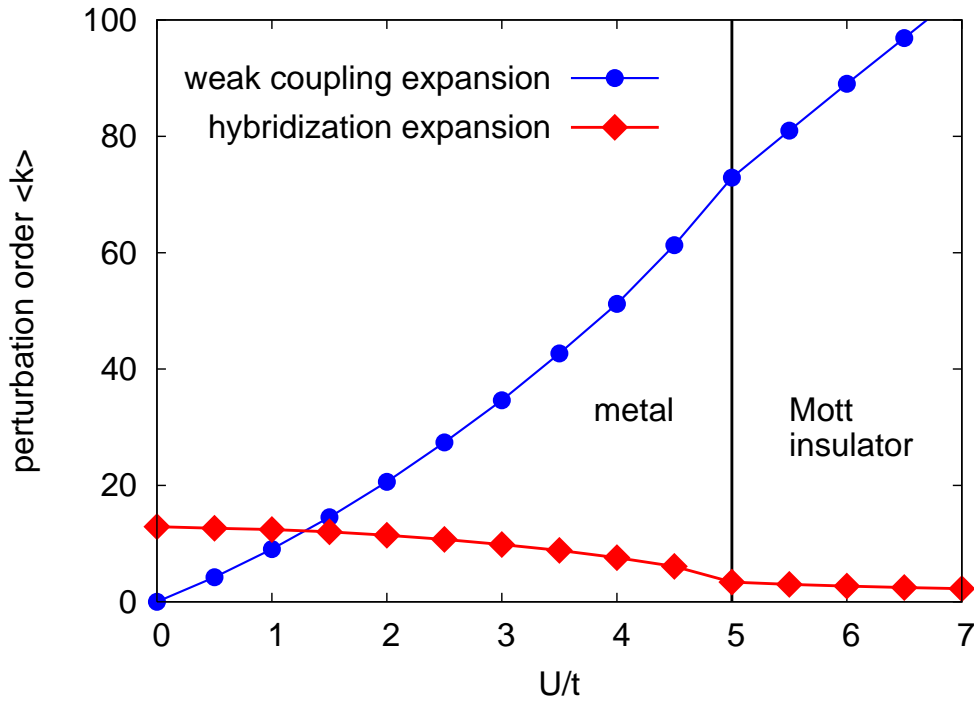


Fig. 4: Average perturbation order for the weak-coupling and strong coupling (hybridization expansion) algorithm. These results correspond to the DMFT solution of the one-band Hubbard model with semi-circular density of states of bandwidth $4t$, and temperature $\beta = 1/T = 30$. The bath is therefore different for each data point. (Figure adapted from Ref. [11].)

Appendices

A Rubtsov's weak-coupling approach

The weak-coupling continuous time impurity solver originally proposed by Rubtsov [3] is based on slightly different definitions of interaction and quadratic terms:

$$H_U = U n_{\uparrow} n_{\downarrow}, \quad (52)$$

$$H_0 = -\mu(n_{\uparrow} + n_{\downarrow}). \quad (53)$$

The method employs an expansion of the partition function in powers of $H_2 = H_U$. Equation (11) then gives the weight of a configuration of n interaction vertices. Since $H_1 = H - H_2 = H_0 + H_{\text{bath}} + H_{\text{mix}}$ is quadratic, we can use Wick's theorem to evaluate the trace. The result is a product of two determinants of $n \times n$ matrices (one for each spin), whose elements are bath Green functions \tilde{g}_0 (here without the chemical potential shift $U/2$) evaluated at the time

intervals defined by the vertex positions:

$$\begin{aligned}\frac{w_c}{Z_0} &= (-Ud\tau)^n \frac{1}{Z_0} \text{Tr} \left[e^{-(\beta-\tau_n)H_1} n_\uparrow n_\downarrow \dots e^{-(\tau_2-\tau_1)H_1} n_\uparrow n_\downarrow e^{-\tau_1 H_1} \right] \\ &= (-Ud\tau)^n \prod_{\sigma} \det \tilde{M}_{\sigma}^{-1},\end{aligned}\quad (54)$$

$$(\tilde{M}_{\sigma}^{-1})_{ij} = \tilde{g}_{0,\sigma}(\tau_i - \tau_j), \quad (55)$$

with $Z_0 = \text{Tr}[e^{-\beta H_1}]$ the partition function of the noninteracting model.

At this point, we encounter a problem. In the paramagnetic phase, where $\tilde{g}_{0,\uparrow} = \tilde{g}_{0,\downarrow}$, the product of determinants is positive, which means that for repulsive interaction ($U > 0$), odd perturbation orders yield negative weights. Except in the particle-hole symmetric case, where one can show that odd perturbation orders vanish, this will result in a severe sign problem. Fortunately, we can solve this sign problem by shifting the chemical potentials for up and down spins in an appropriate way. We rewrite the interaction term as [12]

$$H_U = \frac{U}{2} \sum_s \prod_{\sigma} (n_{\sigma} - \alpha_{\sigma}(s)) + \frac{U}{2} (n_{\uparrow} + n_{\downarrow}) - \frac{U}{4}, \quad (56)$$

$$\alpha_{\sigma}(s) = 1/2 + \sigma s(1/2 + \delta). \quad (57)$$

Here δ is some constant and $s = \pm 1$ an Ising variable. The constant $-U/4$ in Eq. (56) is irrelevant, while the contribution $U(n_{\uparrow} + n_{\downarrow})/2$ can be absorbed into the noninteracting Green function by shifting the chemical potential as $\mu \rightarrow \mu - U/2$. Explicitly, we redefine the bath Green function as $\tilde{g}_{0,\sigma}^{-1} = i\omega_n + \mu - \Delta_{\sigma} \rightarrow g_{0,\sigma}^{-1} = i\omega_n + \mu - U/2 - \Delta_{\sigma}$.

The introduction of an Ising variable s_i at each vertex position τ_i enlarges the configuration space exponentially. A configuration c now corresponds to a collection of Ising spin variables on the imaginary time interval: $c = \{(\tau_1, s_1), (\tau_2, s_2), \dots, (\tau_n, s_n)\}$. The weight of these configurations are

$$\frac{w_c}{Z_0} = (-Ud\tau/2)^n \prod_{\sigma} \det M_{\sigma}^{-1}, \quad (58)$$

$$(M_{\sigma}^{-1})_{ij} = g_{0,\sigma}(\tau_i - \tau_j) - \alpha_{\sigma}(s_i)\delta_{ij}. \quad (59)$$

The Ising variables are in fact not needed to cure the sign problem. They have been introduced to symmetrize the interaction term and prevent ergodicity problems.

Rubtsov's weak-coupling approach is in principle applicable to models with arbitrarily complicated interaction terms. However, the best type of *auxiliary field* representation, which minimizes the sign problem in multi-orbital systems with complicated interaction and correlated hopping terms, is not yet known. For models with density-density interactions, the method is in fact equivalent to the continuous-time auxiliary-field approach discussed in Section 3. It was shown in Ref. [13] that the partition functions for the two weak-coupling methods become identical if the parameters K (for the continuous-time auxiliary field method) and δ (Eq. 57) are related by

$$K = \beta U[(1/2 + \delta) - 1/4]. \quad (60)$$

References

- [1] http://pitp.physics.ubc.ca/confs/sherbrooke/archives/pitpcifar2008_qmc_werner.pdf
- [2] E. Gull, A.I. Lichtenstein, A.J. Millis, A.N. Rubtsov, M. Troyer, and P. Werner, *Rev. Mod. Phys.* **83**, 349 (2011)
- [3] A.N. Rubtsov, V.V. Savkin and A.I. Lichtenstein, *Phys. Rev. B* **72**, 035122 (2005)
- [4] E. Gull, P. Werner, O. Parcollet and M. Troyer, *Europhys. Lett.* **82**, 57003 (2008)
- [5] S.M.A. Rombouts, K. Heyde, and N. Jachowicz, *Phys. Rev. Lett.* **82**, 4155 (1999)
- [6] J.E. Hirsch and R.M. Fye, *Phys. Rev. Lett.* **56**, 2521 (1986)
- [7] P. Werner, A. Comanac, L. de' Medici, M. Troyer and A.J. Millis, *Phys. Rev. Lett.* **97**, 076405 (2006)
- [8] P. Werner and A.J. Millis, *Phys. Rev. B* **74**, 155107 (2006)
- [9] K. Haule, *Phys. Rev. B* **75**, 155113 (2007)
- [10] A.M. Laeuchli and P. Werner, *Phys. Rev. B* **80**, 235117 (2009)
- [11] E. Gull, P. Werner, A.J. Millis, and M. Troyer, *Phys. Rev. B* **76**, 235123 (2007)
- [12] F.F. Assaad and T.C. Lang, *Phys. Rev. B* **76**, 035116 (2007)
- [13] K. Mielson, A. Macridin and M. Jarrell, *Phys. Rev. E* **79**, 057701 (2009)

11 Non-Local Correlation Effects in Solids: Beyond DMFT

A.I. Lichtenstein and H. Hafermann
I. Institut für Theoretische Physik
Universität Hamburg

Contents

1	Introduction	2
2	Cluster DMFT scheme	2
3	Dynamical cluster approximation: general consideration	6
4	Symmetry properties of the cluster scheme	11
4.1	Formalism for the 2-site cluster method	11
4.2	Formalism for the 4-site cluster method	12
5	Long-range correlations: Dual-Fermion approach	14
5.1	Dual-Fermion approach: Exact relations	18
5.2	Self-consistency condition and relation to DMFT	19
5.3	Results for the 2d-Hubbard model	20
5.4	Calculation of susceptibilities	23
5.5	Convergence properties	25
6	Summary and outlook	27

1 Introduction

The tremendous success of Dynamical Mean-Field Theory (DMFT) [1, 2] in understanding the Mott transition in simple model systems shows that the main correlation effects in fermionic lattices have a local character. Moreover realistic investigations of correlated materials within the LDA+DMFT scheme [3–5] also support the idea that the electronic structure of prototype Mott insulators, like V_2O_3 , can be well understood within a local multi-orbital t_{2g} scheme. Nevertheless many interesting correlation effects in solid state physics, such as antiferromagnetic spin fluctuations, superconducting d -wave pairing, and many other phenomena have non-local character. In this Lecture we will discuss different ways to go beyond the DMFT approximation and include non-local correlations. There are two different approaches to non-local effects beyond the DMFT framework: one is based on numerical cluster DMFT extensions while another one is built on an analytical expansion around the local DMFT solution.

2 Cluster DMFT scheme

There are two groups of cluster DMFT extensions, which are formulated in real space (cellular DMFT – CDMFT) or in reciprocal space (Dynamical Cluster Approximation – DCA). We discuss first a simple model for the cluster DMFT scheme in real space which consists of a supercell in a two dimensional square lattice (Fig.1). Lower-case letters will be used for the original lattice vectors (x) and site indices (i, j), while upper-case will be reserved for supercell coordinates (X) and position of atoms in a supercell (I, J). Similarly, for wave vectors in original reciprocal lattice we will use (k) while for the reduced supercell Brillouin zone (K) will be used. The minimal cluster which allows us to investigate both antiferromagnetic (AFM) and superconducting (d -wave) order parameters on an equal footing consists of a 2×2 plaquette in an effective medium (see Fig.1).

The one band Hubbard model on the square lattice reads:

$$H = \sum_{ij\sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}, \quad (1)$$

where t_{ij} are effective hopping parameters and U is the local Coulomb interaction. The exact Green function for the one-band Hubbard model (1) can be written in the following form

$$G(\mathbf{k}, i\omega) = (i\omega + \mu - t(\mathbf{k}) - \Sigma(\mathbf{k}, i\omega))^{-1}, \quad (2)$$

where $\omega = (2n + 1)\pi/\beta$, $n = 0, \pm 1, \dots$ are the Matsubara frequencies, β is the inverse temperature, μ the chemical potential, $t(\mathbf{k})$ the Fourier transform of the hopping parameters t_{ij} , and $\Sigma(\mathbf{k}, i\omega)$ is the non-local self-energy, which contains all information on single-particle correlations.

We can approximate the momentum-dependence of the self-energy in terms of a finite number of basis functions $\phi_i(\mathbf{k})$ [6]

$$\Sigma(\mathbf{k}, i\omega) \approx \sum_{i=1}^N \phi_i(\mathbf{k}) \Sigma_i(\omega) \quad (3)$$

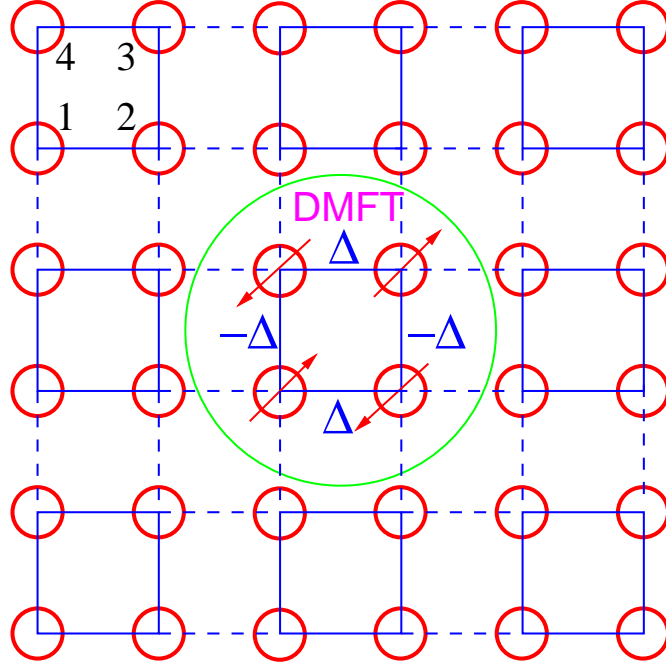


Fig. 1: Schematic representation of the 2×2 supercell with antiferromagnetic and superconducting d -wave order parameters for cluster DMFT.

In the most general scheme we can find $\Sigma_i(\omega)$ as a solution of a fictitious N -site quantum impurity model. Different cluster DMFT schemes differ in the choice of these basis functions. Numerical solutions of generalized multi-site quantum impurity models can be found within the recently developed continuous time Quantum Monte-Carlo scheme [7].

We introduce a "super-site" object as the 2×2 plaquette on a square lattice. The enumeration of the atoms inside the super-site is shown in the Fig. 1. A superspinor $C_I^\dagger = \{c_{I\alpha}^\dagger\}$ where $\alpha = 1, 2, 3, 4$ (including also the spin-indices) defines a super-fermionic operator for the I -th plaquette. The plaquette Green's function for the Hubbard model can be rewritten as

$$G(\mathbf{K}, i\omega) = [(i\omega + \mu) \mathbf{1} - T(\mathbf{K}) - \Sigma(i\omega)]^{-1} \quad (4)$$

where $T(\mathbf{K})$ is the effective hopping supermatrix, and \mathbf{K} are the wave vectors within the reduced Brillouin zone, and $\Sigma(i\omega)$ is the self-energy supermatrix. For simplicity we will write all equations in the nearest-neighbor approximation, which means only one hopping in x - (t_x) and y - (t_y) direction. After supercell Fourier-transform we have the following expression for the supercell hopping matrix:

$$T_{I,J}(\mathbf{K}) = \begin{pmatrix} 0 & t_x K_x^+ & 0 & t_y K_y^+ \\ t_x K_x^- & 0 & t_y K_y^+ & 0 \\ 0 & t_y K_y^- & 0 & t_x K_x^- \\ t_y K_y^- & 0 & t_x K_x^+ & 0 \end{pmatrix} \quad (5)$$

where $K_{x(y)}^\pm \equiv 1 + \exp(\pm i K_{x(y)} a)$, a is the lattice constant, and each elements is a 2×2 matrix in spin space. Within the cluster DMFT approach we introduce the intra-atomic self-energy Σ_0

and inter-atomic self-energies Σ_x , Σ_y as well as the non-local self-energy Σ_{xy} in xy direction, which defines the local self-energy matrix for our 2×2 super-site:

$$\Sigma_{I,J}(i\omega) = \begin{pmatrix} \Sigma_0 & \Sigma_x & \Sigma_{xy} & \Sigma_y \\ \Sigma_x & \Sigma_0 & \Sigma_y & \Sigma_{xy} \\ \Sigma_{xy} & \Sigma_y & \Sigma_0 & \Sigma_x \\ \Sigma_y & \Sigma_{xy} & \Sigma_x & \Sigma_0 \end{pmatrix}$$

For a general $N \times N$ super-site impurity model (simp) the partition function can be written as a functional integral over the $2N$ -component spin and site-dependent spinor Grassmann fields c^* and c :

$$Z = \int \mathcal{D}[c^*, c] e^{-S_{simp}}, \quad (6)$$

where

$$\begin{aligned} S_{simp} = & - \sum_{I,J=0}^N \int_0^\beta d\tau \int_0^\beta d\tau' c_{I\sigma}^*(\tau) [\mathcal{G}_\sigma^{-1}(\tau - \tau')]_{IJ} c_{J\sigma}(\tau') \\ & + \sum_{I=1}^N \int_0^\beta d\tau U n_{I,\uparrow}(\tau) n_{I,\downarrow}(\tau), \end{aligned} \quad (7)$$

where \mathcal{G} is the $N \times N$ matrix of effective bath Green's function for a spin-collinear case.

The main problem of all cluster extension of DMFT is to find an optimal self-consistent way to obtain the bath Green's function matrix in imaginary time $\mathcal{G}_{IJ}(\tau - \tau')$ or in Matsubara space $\mathcal{G}_{IJ}(i\omega)$. In the free-cluster version of the CDMFT scheme [6] which is equivalent to the cellular DMFT method [8] or to the molecular CPA scheme in alloy theory [9] we can use the following prescription. First, we need to integrate out the superlattice degrees of freedom, similarly to the standard DMFT approach, and obtain the local Green's function matrix:

$$G_{IJ}(i\omega) = \sum_{\mathbf{K}} G_{IJ}(\mathbf{K}, i\omega), \quad (8)$$

where the summation runs over the reduced Brillouin zone of the plaquette superlattice.

Next we can write the matrix equation for the bath Green function matrix \mathcal{G} , which describes the effective interactions of the plaquette with rest of crystal. We use the impurity DMFT analogy, which allowed us to account for double-counting corrections for the local self-energy matrix: the bath Green function is not supposed to have any local self-energy contribution, since it comes later from the solution of the effective super-impurity problem (7). Therefore one needs to subtract the local self-energy contribution, which is equivalent to a solution of the following impurity problem, where all super-cites in Fig. 1 have the self-energy contributions, but not the "central-cluster":

$$\mathcal{G}^{-1}(i\omega) = G^{-1}(i\omega) + \Sigma(i\omega), \quad (9)$$

One can solve a complicated many-body problem described by super-impurity action Eq. (7). We can use the numerically exact continuous-time QMC scheme [7] and get the super-impurity

Green's function $G_{IJ}^{simp}(\tau) = -\langle c_{I\sigma}(\tau)c_{J\sigma}^\dagger(0) \rangle_{simp}$. The new cluster-local self-energy is equal to the difference of the inverse input and output Green's functions of this local many-body problem:

$$\Sigma_{new}(i\omega) = \mathcal{G}^{-1}(i\omega) - G_{simp}^{-1}(i\omega). \quad (10)$$

Finally, we can close the CDMFT self-consistent loop for the cluster self-energy $\Sigma_{I,J}(i\omega)$ by using in the next iterations the new self-energy from Eq. (10) in the super-lattice Hamiltonian from Eq. (1). The CDMFT self-consistency condition reads:

$$G_{IJ}^{simp}(i\omega) = G_{IJ}(i\omega). \quad (11)$$

In fact this CDMFT scheme is equivalent to the multi-orbital LDA+DMFT approach [4], where the super-site indices (I, J) play the role of different orbitals (m, m'). A crucial difference is related to the fact, that multi-orbital DMFT does not break the translational symmetry of original lattice, while the standard CDMFT scheme [8, 6] does lower the symmetry of lattice due to the local form of the super-site self-energy Eq. (7). The present "matrix" form of CDMFT with non-periodic self-energy allows us to study multicomponent order parameters (Fig. 1). In this case we have the standard DMFT problem with four "orbital" states per super-site. We use the generalized Gorkov-Nambu technique to analyze the coexistence of magnetic ordering and superconductivity. Let us introduce the superspinor

$$\Psi_I^\dagger(\tau) = \left(c_{I\uparrow}^\dagger, c_{I\downarrow}^\dagger, c_{I\uparrow}, c_{I\downarrow} \right) \quad (12)$$

and the anomalous averages describing the (collinear) antiferromagnetism $\langle c_{I\uparrow}^\dagger c_{J\downarrow} \rangle$ and the superconductivity $\Delta_{IJ} = \langle c_{I\downarrow} c_{J\uparrow} \rangle$.

One may realize that the cellular DMFT approximation is not very suitable for the superconducting d -wave order parameter since Δ is located on the bonds as depicted in Fig. (1). Therefore one can lose half of the superconducting bonds and reduce approximately by a factor of two the HTSC transition temperature. We can also formulate a "periodic" CDMFT scheme by renormalizing the hopping with the cluster self-energy [6].

The effective Hamiltonian defined through the translationally invariant (\mathbf{k} -dependent) self-energy corresponds to the renormalized energy dependent hoppings: $t_x = t + \Sigma_x$, $t_y = t + \Sigma_y$. The functions $\Sigma_0(i\omega)$, $\Sigma_x(i\omega)$, $\Sigma_y(i\omega)$ are found self-consistently within the cluster DMFT scheme [6] and for the d -wave superconducting state $\Sigma_x \neq \Sigma_y$. It is straightforward to generalize this scheme for a next-nearest neighbor hopping as well as the long-range Green function and the self-energy. In this case we can renormalize also the second-nearest hopping: $t_{xy} = t' + \Sigma_{xy}$ for the 2×2 cluster. The local cluster Green matrix in this case is equal to $G_{ij}(i\omega) = \sum_{\mathbf{k}} G_{ij}(\mathbf{k}, i\omega)$, and the summation runs over the original Brillouin zone of the square lattice. Unfortunately we can not prove that this periodic CDMFT scheme is causal. Later we will discuss different ways of obtaining a periodic self-energy within CDMFT.

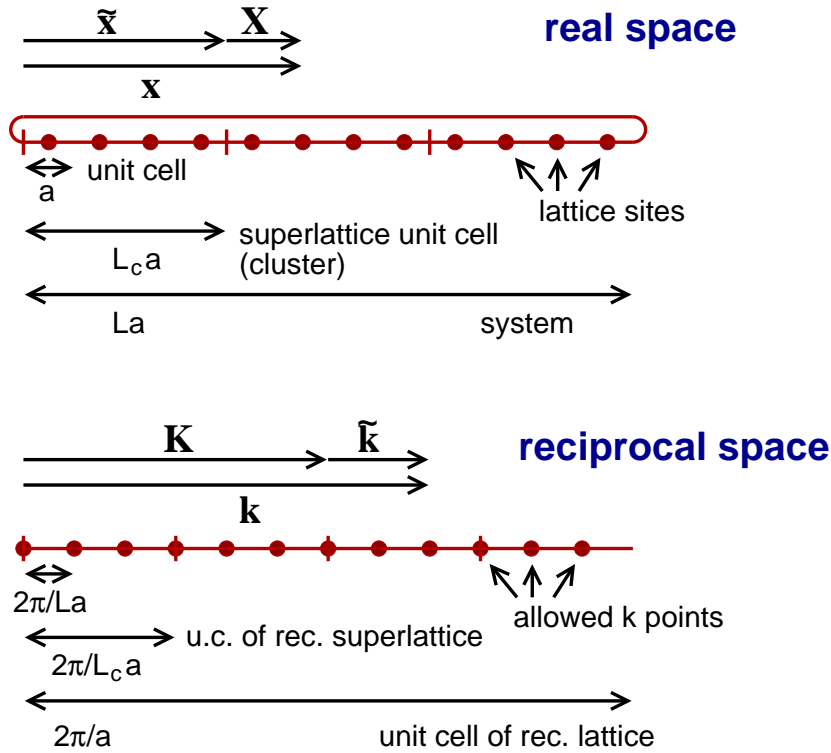


Fig. 2: Decomposition of real-space lattice vectors, $\mathbf{x} = \mathbf{X} + \tilde{\mathbf{x}}$, and reciprocal-space wave vectors, $\mathbf{k} = \tilde{\mathbf{k}} + \mathbf{K}$, for a $D = 1$ dimensional lattice (lattice constant a) with $L = 12$ sites tiled with $L/L_c = 3$ clusters consisting of $L_c = 4$ sites each. \mathbf{x} : original lattice. $\tilde{\mathbf{x}}$: superlattice. \mathbf{X} : sites in a cluster. Reciprocal space: There are L allowed wave vectors \mathbf{k} in the unit cell of the lattice reciprocal to \mathbf{x} , and there are L/L_c allowed wave vectors $\tilde{\mathbf{k}}$ in the unit cell of the lattice reciprocal to the superlattice $\tilde{\mathbf{x}}$. \mathbf{K} are the reciprocal superlattice vectors, $\exp(i\mathbf{K}\tilde{\mathbf{x}}) = 1$. From Ref. [19].

3 Dynamical cluster approximation: general consideration

We start discussion of dynamical cluster approaches in reciprocal space with introducing some notations (see Fig. 2) (for a review, see Ref. [10]). The cluster need not be a physical subsystem of the original lattice [11–13]. We consider a system on a D -dimensional lattice of L sites with periodic boundary conditions and $L \rightarrow \infty$ in the end. The position vector to a site in the lattice is denoted by \mathbf{x} . There are L allowed wave vectors in a unit cell of the reciprocal lattice which are denoted by \mathbf{k} . The lattice is tiled with L/L_c clusters consisting of L_c sites each. Let $\tilde{\mathbf{x}}$ be the position vector of the cluster origin, and \mathbf{X} the position vector of a site in a cluster, referring to the cluster origin. We then have the unique decomposition $\mathbf{x} = \mathbf{X} + \tilde{\mathbf{x}}$. The vectors $\tilde{\mathbf{x}}$ form a superlattice with a unit-cell volume enlarged by the factor L_c . In a unit cell of the reciprocal superlattice there are L/L_c allowed wave vectors $\tilde{\mathbf{k}}$. Its volume is reduced by the factor L_c as compared to the volume of the reciprocal unit cell of the original lattice. For a given \mathbf{k} we have the unique decomposition $\mathbf{k} = \tilde{\mathbf{k}} + \mathbf{K}$ where \mathbf{K} are the vectors of the reciprocal superlattice, i.e. $\exp(i\mathbf{K}\tilde{\mathbf{x}}) = 1$. In the reciprocal unit cell of the original lattice, there are L_c vectors \mathbf{K} . These can also be interpreted as the allowed cluster wave vectors when imposing periodic boundary

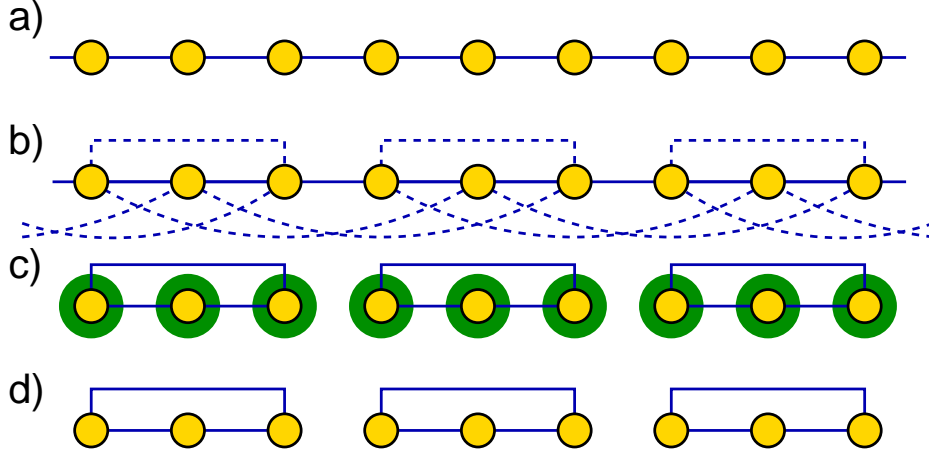


Fig. 3: a) Hubbard 1d-model. b) The original Hubbard model but with a modified one-particle part $t \rightarrow \bar{t}$ which is the starting point for the dynamical cluster approximation (DCA). \bar{t} is invariant under superlattice and cluster translations. c) Reference system generating the DCA. Note that t' has the same translational symmetries as \bar{t} . d) Reference system generating a simplified DCA. From Ref. [19].

conditions on the individual cluster.

In the following we consider the $L \times L$ matrix U with elements

$$U_{x,\mathbf{k}} = \frac{1}{\sqrt{L}} e^{i\mathbf{k}x}, \quad (13)$$

and the $L/L_c \times L/L_c$ matrix V with elements

$$V_{\tilde{x},\tilde{\mathbf{k}}} = \frac{1}{\sqrt{L/L_c}} e^{i\tilde{\mathbf{k}}\tilde{x}}, \quad (14)$$

and the $L_c \times L_c$ matrix W with elements

$$W_{\mathbf{X},\mathbf{K}} = \frac{1}{\sqrt{L_c}} e^{i\mathbf{K}\mathbf{X}}. \quad (15)$$

Notes, that U , V and W are unitary and define Fourier transformations between the respective real and reciprocal spaces. It is obvious, that $U \neq VW = WV$:

$$U_{x,\mathbf{k}} = \frac{1}{\sqrt{L}} e^{i\mathbf{k}x} = \frac{1}{\sqrt{L}} e^{i(\tilde{\mathbf{k}}\mathbf{X} + \tilde{\mathbf{k}}\tilde{x} + \mathbf{K}\mathbf{X})} \neq \frac{1}{\sqrt{L}} e^{i(\tilde{\mathbf{k}}\tilde{x} + \mathbf{K}\mathbf{X})} = V_{\tilde{x},\tilde{\mathbf{k}}} W_{\mathbf{X},\mathbf{K}}. \quad (16)$$

A hopping $t_{x,x'}$ which is invariant under lattice translations \mathbf{x}_0 , i.e. $t_{x+\mathbf{x}_0,x'+\mathbf{x}_0} = t_{x,x'}$, is diagonalized by normal Fourier transformations U : $(U^\dagger t U)_{\mathbf{k}\mathbf{k}'} = t(\mathbf{k}) \delta_{\mathbf{k},\mathbf{k}'}$. By definition, the one-electron spectrum is just Fourier transform of the hopping matrix elements: $\varepsilon_{\mathbf{k}} \equiv t(\mathbf{k})$. A quantity $T_{x,x'}$ which is invariant under superlattice translations $\tilde{\mathbf{x}}_0$ as well as under cluster translations \mathbf{X}_0 (i.e. which is cyclic on the cluster), $T_{x+\tilde{\mathbf{x}}_0,x'+\tilde{\mathbf{x}}_0} = T_{x+\mathbf{X}_0,x'+\mathbf{X}_0} = T_{x,x'}$, is diagonalized by alternative DCA-transformation VW : $(W^\dagger V^\dagger T V W)_{\tilde{\mathbf{k}}\mathbf{K},\tilde{\mathbf{k}}'\mathbf{K}'} = T(\tilde{\mathbf{k}},\mathbf{K}) \delta_{\tilde{\mathbf{k}},\tilde{\mathbf{k}'}} \delta_{\mathbf{K},\mathbf{K}'}$. Following Refs. [15, 16, 19], we introduce a fictitious hopping which corresponds to the real-space formulation of the DCA-scheme:

$$\bar{t} = (VW)U^\dagger t U(VW)^\dagger, \quad (17)$$

which is just the DCA Fourier back-transform of the one-electron spectrum $\varepsilon_{\mathbf{k}}$. For clusters of finite size L_c , the combined Fourier transformation $\mathbf{V}\mathbf{W}$ is different from \mathbf{U} . For $L_c \rightarrow \infty$, however, this becomes irrelevant. With $\varepsilon(\mathbf{k}) = (\mathbf{U}^\dagger \mathbf{t} \mathbf{U})(\mathbf{k})$ we have:

$$\bar{t}_{\mathbf{x}\mathbf{x}'} = \frac{1}{L_c} \sum_{\mathbf{K}} e^{i\mathbf{K}(\mathbf{x}-\mathbf{x}')} \frac{L_c}{L} \sum_{\tilde{\mathbf{k}}} e^{i\tilde{\mathbf{k}}(\tilde{\mathbf{x}}-\tilde{\mathbf{x}'})} \varepsilon(\tilde{\mathbf{k}} + \mathbf{K}). \quad (18)$$

Obviously, $\bar{\mathbf{t}}$ is invariant under superlattice translations as well as under cluster translations (with periodic cluster boundary conditions). The original and the modified system are represented by Fig. 3a, b. The construction of $\bar{\mathbf{t}}$ is such that it exhibits the same translational symmetries as the one-particle parameters \mathbf{t}' of a reference system consisting of isolated clusters tiling the original lattice with periodic boundary conditions, see Fig. 3c, d. Since both, \mathbf{t} and $\bar{\mathbf{t}}$, are invariant under superlattice translations, we can compare $t_{\mathbf{x}\mathbf{x}'}(\tilde{\mathbf{k}}) = (\mathbf{V}^\dagger \mathbf{t} \mathbf{V})_{\mathbf{x}\mathbf{x}'}(\tilde{\mathbf{k}})$ with $\bar{t}_{\mathbf{x}\mathbf{x}'}(\tilde{\mathbf{k}}) = (\mathbf{V}^\dagger \bar{\mathbf{t}} \mathbf{V})_{\mathbf{x}\mathbf{x}'}(\tilde{\mathbf{k}})$. It turns out they are equal up to a phase factor:

$$\begin{aligned} \bar{t}_{\mathbf{x}\mathbf{x}'}(\tilde{\mathbf{k}}) &= \frac{1}{L_c} \sum_{\mathbf{K}} e^{i\mathbf{K}(\mathbf{x}-\mathbf{x}')} \varepsilon(\tilde{\mathbf{k}} + \mathbf{K}) \\ &= \frac{L_c}{L} \sum_{\tilde{\mathbf{x}}\tilde{\mathbf{x}'}} e^{-i\tilde{\mathbf{k}}(\tilde{\mathbf{x}}+\mathbf{x}-\tilde{\mathbf{x}'}-\mathbf{x}')} t_{\tilde{\mathbf{x}}+\mathbf{x},\tilde{\mathbf{x}'}+\mathbf{x}'} \\ &= e^{-i\tilde{\mathbf{k}}(\mathbf{x}-\mathbf{x}')} t_{\mathbf{x}\mathbf{x}'}(\tilde{\mathbf{k}}). \end{aligned} \quad (19)$$

The main idea of the DCA is to restore momentum conservation within the cluster by a rescale the effective hoppings. In CDMFT, the intracluster transform of the dispersion given by the super-cell Fourier sum:

$$\mathbf{t}_{\mathbf{x},\mathbf{x}'}(\tilde{\mathbf{k}}) = \frac{1}{L_c} \sum_{\mathbf{K}} e^{i(\mathbf{K}+\tilde{\mathbf{k}})(\mathbf{x}-\mathbf{x}')} \epsilon_{\mathbf{K}+\tilde{\mathbf{k}}}, \quad (20)$$

while in the DCA, an addition phase factors $e^{i\tilde{\mathbf{k}}\mathbf{x}}$ are excluded using the transform (see Eq. (19)).

$$\bar{t}_{\mathbf{x},\mathbf{x}'}(\tilde{\mathbf{k}}) = \mathbf{t}_{\mathbf{x},\mathbf{x}'}(\tilde{\mathbf{k}}) e^{-i\tilde{\mathbf{k}}(\mathbf{x}-\mathbf{x}')} = \frac{1}{L_c} \sum_{\mathbf{K}} e^{i\mathbf{K}(\mathbf{x}-\mathbf{x}')} \epsilon_{\mathbf{K}+\tilde{\mathbf{k}}}. \quad (21)$$

The intracluster hopping in DCA is therefore given by the intracluster Fourier transform of the dispersion Eq. (21), which is obvious by coarse-graining. This gives the DCA Green's function which is diagonal in cluster Fourier space:

$$G(\mathbf{K} + \tilde{\mathbf{k}}, i\omega) = \frac{1}{i\omega + \mu - \varepsilon(\mathbf{K} + \tilde{\mathbf{k}}) - \Sigma(\mathbf{K}, i\omega)}. \quad (22)$$

The self-energy becomes a piecewise constant function in the k-space [10]. Finally, the self-consistent condition for $\Sigma(\mathbf{K}, i\omega)$ in the DCA-scheme is similar to the CDMFT one Eq. (11):

$$G^{imp}(\mathbf{K}, i\omega) = G(\mathbf{K}, i\omega) \equiv \sum_{\tilde{\mathbf{k}}} G(\mathbf{K} + \tilde{\mathbf{k}}, i\omega). \quad (23)$$

We can also try to 'periodize' the cluster-DMFT scheme [6, 15]. The CDMFT violates translational invariance with respect to the cluster sites. This is obvious for clusters with $L_c \geq 3$, where bulk and surface sites of a cluster may be distinguished. The CDMFT calculations are carried out in the cluster real-space representation (i.e. all quantities are matrices in the cluster sites), since there is no benefit in changing to the cluster k -space representation, which is not diagonal.

Since translational invariance is broken, the lattice quantities are functions of two independent momenta \mathbf{k} and \mathbf{k}' . They can differ by a reciprocal lattice vector \mathbf{Q} , where $Q_i = 0, \dots, (L_c - 1)2\pi/L_c$. The self-energy can be expressed in terms of the cluster self-energy as

$$\Sigma(\mathbf{k}, \mathbf{k}', i\omega) = \frac{1}{L_c} \sum_{\mathbf{Q}} \sum_{\mathbf{X}, \mathbf{X}'} e^{i\mathbf{k}\mathbf{X}} \Sigma_c(\mathbf{X}, \mathbf{X}', i\omega) e^{-i\mathbf{k}'\mathbf{X}'} \delta(\mathbf{k} - \mathbf{k}' - \mathbf{Q}), \quad (24)$$

where the dependence on cluster sites is written explicitly. A translationally invariant solution is obtained by approximating the lattice quantities only by the $\mathbf{Q} = 0$ contribution:

$$\Sigma(\mathbf{k}, i\omega) = \frac{1}{L_c} \sum_{\mathbf{X}, \mathbf{X}'} = e^{i\mathbf{k}(\mathbf{X}-\mathbf{X}')} \Sigma_c(\mathbf{X}, \mathbf{X}', i\omega). \quad (25)$$

Transforming back to real space shows that the lattice quantities for a given distance $\mathbf{x} - \mathbf{x}'$ are obtained as an average over the cluster quantities for the same distance,

$$\Sigma(\mathbf{x} - \mathbf{x}', i\omega) = \frac{1}{L_c} \sum_{\mathbf{X}, \mathbf{X}'} \Sigma_c(\mathbf{X}, \mathbf{X}', i\omega) \delta_{\mathbf{x}-\mathbf{x}', \mathbf{x}-\mathbf{x}'}. \quad (26)$$

Spatial correlations are hence included up to a length determined by the extension of the cluster. Note that Eq. (26) underestimates the nonlocal contributions, in particular for small clusters. Using the shorthand notation $\Sigma_{\mathbf{X}, \mathbf{X}'} = \Sigma(\mathbf{X}, \mathbf{X}')$, one sees that the local self-energy is averaged correctly, $\Sigma(x=0) = (\Sigma_{c00} + \Sigma_{c11})/2$, while the nearest-neighbor self-energy contribution according to (26) would read $\Sigma(x=1) = (1/2)\Sigma_{c10}$, since Σ_{c01} contributes to $\Sigma(x=-1)$. It was therefore suggested to reweigh the terms in the sum [15]. For the above example, $\Sigma(x=1) = \Sigma_{c10}$.

When translational invariance is recovered in this way, the solution of the lattice problem may be viewed as shown in Fig. 4: The lattice is replaced by a lattice of clusters all of which are embedded in a self-consistent bath. The self-energy on a cluster is obtained from the self-consistent solution of the local problem and the intercluster self-energy between sites on neighboring clusters at a distance $\mathbf{x} - \mathbf{x}'$ is artificially set equal to the average of the intracluster self-energy for the same distance. The self-energy for distances exceeding the maximum distance between sites within the cluster is zero.

Following Ref. [17] we can compare the CDMFT and DCA schemes for the linear 3-cite cluster from Fig. (3). Writing the single-electron part of Hamiltonian as the supercell matrix $\mathbf{T}(\tilde{\mathbf{k}})$, the average cluster hopping is given by

$$\mathbf{T}_c = \int d\tilde{\mathbf{k}} \mathbf{T}(\tilde{\mathbf{k}}). \quad (27)$$

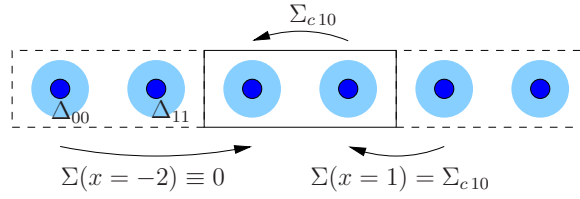


Fig. 4: Illustration of the CDMFT lattice self-energy. The original lattice is replaced by a collection of clusters embedded in a self-consistent bath. The intercluster self-energy $\Sigma(x = 1)$ is approximated by the intracluster self-energy Σ_{c10} for this distance not exceeding the maximal distance between cluster sites and zero otherwise. From the Ref. [14].

The interaction terms are simply those of Eq. (1), restricted to the cluster.

The Hamiltonian $H(\tilde{\mathbf{k}})$ in the reciprocal space of the super-lattice $\{\tilde{\mathbf{x}}\}$ of clusters can be obtained by changing to the basis of fermionic-operators in Eq.(1):

$$\tilde{c}_{\mathbf{X},\sigma}^{\text{CDMFT}}(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{x}}} e^{-i\tilde{\mathbf{k}}\tilde{\mathbf{x}}} c_{\tilde{\mathbf{x}}+\mathbf{X},\sigma}. \quad (28)$$

The resulting quantum cluster approximation is the CDMFT. Alternatively, we can start from the operators in the reciprocal space of the *lattice* to obtain

$$\tilde{c}_{\mathbf{X},\sigma}^{\text{DCA}}(\tilde{\mathbf{k}}) = \sum_{\tilde{\mathbf{x}}} e^{-i\tilde{\mathbf{k}}(\tilde{\mathbf{x}}+\mathbf{X})} c_{\tilde{\mathbf{x}}+\mathbf{X},\sigma} \equiv \sum_{\tilde{\mathbf{x}}} e^{-i(\tilde{\mathbf{k}}\tilde{\mathbf{x}}+\phi(\tilde{\mathbf{k}},\tilde{\mathbf{x}}))} c_{\tilde{\mathbf{x}}+\mathbf{X},\sigma}. \quad (29)$$

The choice of the operators in the two approaches differs just by local phase factors $\phi(\tilde{\mathbf{k}}, \tilde{\mathbf{x}})$ [17]. In the CDMFT this gauge is chosen such that phases appear only in matrix elements involving different clusters. Thus all matrix elements on the cluster are the same as in the original Hamiltonian. The price for retaining the original matrix elements on the cluster is a breaking of the translation-symmetry of the original lattice. The DCA-scheme opts instead to retain this symmetry by distributing the phase change uniformly over the cluster-sites. The price for retaining translation-invariance is that the matrix elements in the cluster Hamiltonian differ from those in the original Hamiltonian. In both cases, CDMFT and DCA, the eigenvalues of $\mathbf{T}(\tilde{\mathbf{k}})$ are identical to the eigenvalues of the non-interacting part of H .

In the CDMFT gauge we have we have for a three-site cluster ($L_c = 3$) in 1-d lattice [17] :

$$\mathbf{T}^{\text{CDMFT}}(\tilde{k}) = t \begin{pmatrix} 0 & 1 & e^{-3i\tilde{k}} \\ 1 & 0 & 1 \\ e^{3i\tilde{k}} & 1 & 0 \end{pmatrix} \quad (30)$$

so that \mathbf{T}_c is the original single-electron Hamiltonian restricted to the cluster:

$$\mathbf{T}_c^{\text{CDMFT}} = \frac{3}{2\pi} \int_{-\pi/3}^{\pi/3} d\tilde{k} \mathbf{T}(\tilde{k}) = t \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}. \quad (31)$$

In the DCA gauge for 3-cite linear cluster we have

$$\mathbf{T}^{DCA}(\tilde{k}) = t \begin{pmatrix} 0 & e^{i\tilde{k}} & e^{-i\tilde{k}} \\ e^{-i\tilde{k}} & 0 & e^{i\tilde{k}} \\ e^{i\tilde{k}} & e^{-i\tilde{k}} & 0 \end{pmatrix}. \quad (32)$$

Now the \mathbf{T}_c matrix is cyclic and has translation symmetry (see Fig.(3 c)), but rescaled hopping matrix elements:

$$\mathbf{T}_c^{DCA} = \frac{3}{2\pi} \int_{\pi/3}^{\pi/3} d\tilde{k} \mathbf{T}(\tilde{k}) = \frac{3\sqrt{3}}{2\pi} t \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}. \quad (33)$$

This effective rescaling of the hopping parameters in DCA-scheme can lead to a problem with investigations of complex band structure effects, such as an extended van Hove singularities [32]. We note also that the similar consideration apply to the variational cluster approach [11], which is based on the self-energy functional theory [18, 19].

4 Symmetry properties of the cluster scheme

Let us discuss a symmetry properties of paramagnetic solution of cluster extension of DMFT in the simple case of 2- and 4-site clusters [20]. In Fig. Fig. (5) the simplest 2-site and 4-site tiling on square and cubic lattices plotted. For each quantity, like Green's function G , self energy Σ , and bath function \mathcal{G} , there are momentum and real-space components labeled by some subscript. In this paper, the real-space component is labeled by a number (0 - on-site, 1 - nearest neighbor, etc.) while the momentum-space sectors labelled by capital letters (S, P, D).

4.1 Formalism for the 2-site cluster method

Now we apply general cluster formalism to specific cases, first to the 2-site cluster in the square lattice. The solution of 2-site impurity problem gives the following matrix Green function:

$$\hat{G}^{imp} = \begin{pmatrix} G_0 & G_1 \\ G_1 & G_0 \end{pmatrix} \quad \hat{\Sigma}^{imp} = \begin{pmatrix} \Sigma_0 & \Sigma_1 \\ \Sigma_1 & \Sigma_0 \end{pmatrix} \quad (34)$$

The partitioning of Brillouin zone in this case is given in Fig(5), so two \mathbf{K} points according to this division is $\mathbf{K}_I = 0$, $\mathbf{K}_{II} = (\pi, \pi)$. We label region I and II or S and P sectors. Corresponding to \mathbf{K}_I and \mathbf{K}_{II} , one gets $\mathbf{R}_0 = 0$ and $\mathbf{R}_1 = (\pm 1, 0)$ or $(0, \pm 1)$. The lattice self energy is related to $\hat{\Sigma}^{imp}$ by

$$\Sigma_{DCA}(\vec{k}, \omega) = \begin{cases} \Sigma_S^{imp} = \Sigma_0 + \Sigma_1 & \text{for } \mathbf{k} \in \text{Region } I(S) \\ \Sigma_P^{imp} = \Sigma_0 - \Sigma_1 & \text{for } \mathbf{k} \in \text{Region } II(P) \end{cases} \quad (35)$$

The partial density of states are

$$D_{S(P)}(\epsilon) = 2 \times \int_{\mathbf{k} \in I(II)} d\mathbf{k} \delta(\epsilon - \epsilon_{\mathbf{k}}) \quad (36)$$

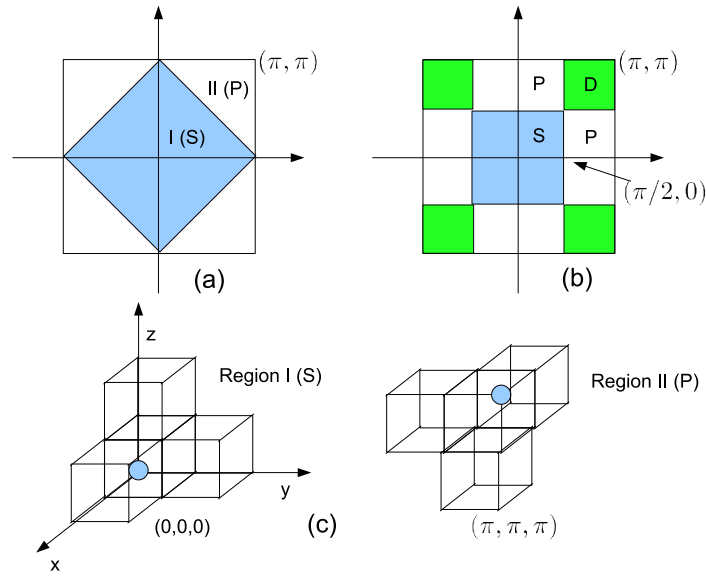


Fig. 5: Partition of the Brillouin zone. (a) 2-site DCA on square lattice. (b) 4-site DCA on square lattice. (c) 2-site DCA on cubic lattice From Ref. [20].

and the self-consistency equation is

$$\begin{aligned} G_0 &= (G_S + G_P)/2 \\ G_1 &= (G_S - G_P)/2 \end{aligned} \quad (37)$$

with

$$G_{S(P)} = \int \frac{D_{S(P)}(\epsilon) d\epsilon}{\omega + \mu - \epsilon_{\mathbf{k}} - (\Sigma_0 \pm \Sigma_1)} \quad (38)$$

4.2 Formalism for the 4-site cluster method

In the 4-site cluster the Brillouin zone is divided into four sectors which are labelled as S, P, and D, as shown in Fig(5). Four \mathbf{K} points are $(0, 0)$ $(\pi, 0)$ $(0, \pi)$ (π, π) leading to four \mathbf{R} as $(0, 0)$ $(1, 0)$ $(0, 1)$ $(1, 1)$. The partial DOS is defined as

$$D_{S(P,D)}^{(4)}(\epsilon) = 4 \times \int_{\mathbf{k} \in S(P,D)} d\mathbf{k} \delta(\epsilon - \epsilon_{\mathbf{k}}) \quad (39)$$

where the superscript (4) is used to distinguish from the partial DOS in 2-site DCA (see Fig(6)). After solving a 4-site impurity cluster problem, in the disordered phase one gets

$$\hat{G}^{imp} = \begin{pmatrix} G_0 & G_1 & G_2 & G_1 \\ G_1 & G_0 & G_1 & G_2 \\ G_2 & G_1 & G_0 & G_1 \\ G_1 & G_2 & G_1 & G_0 \end{pmatrix} \quad \hat{\Sigma}^{imp} = \begin{pmatrix} \Sigma_0 & \Sigma_1 & \Sigma_2 & \Sigma_1 \\ \Sigma_1 & \Sigma_0 & \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_1 & \Sigma_0 & \Sigma_1 \\ \Sigma_1 & \Sigma_2 & \Sigma_1 & \Sigma_0 \end{pmatrix} \quad (40)$$

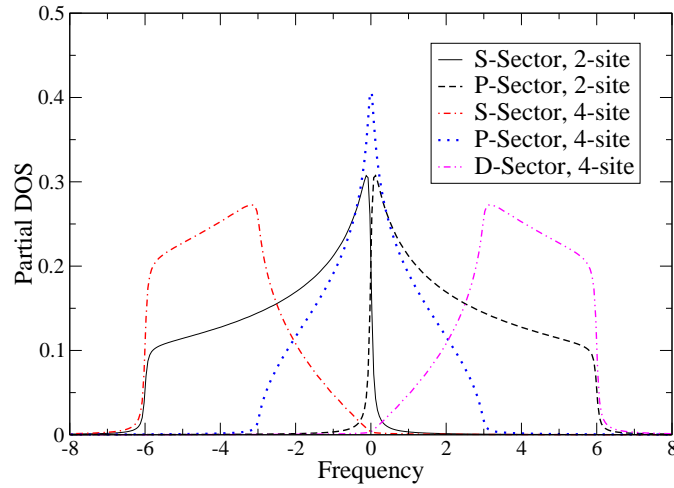


Fig. 6: The PDOS for 2-site and 4-site DCA partitioning on the square lattice with nearest neighbor hopping. The total bandwidth is 12 which corresponds to the hopping $t=1.5$. From Ref. [20].

and the momentum-dependent self energies are

$$\begin{aligned}\Sigma_S &= \Sigma_0 + 2\Sigma_1 + \Sigma_2 \\ \Sigma_P &= \Sigma_0 - \Sigma_2 \\ \Sigma_D &= \Sigma_0 - 2\Sigma_1 + \Sigma_2\end{aligned}\quad (41)$$

and correspondingly the components of lattice Green's functions are

$$G_{S(P,D)} = \int \frac{D_{S(P,D)}(\epsilon)d\epsilon}{i\omega_n + \mu - \epsilon - \Sigma_{S(P,D)}}\quad (42)$$

The self-consistency equations are

$$\begin{aligned}G_0 &= (G_S + 2G_P + G_D)/4 \\ G_1 &= (G_S - G_D)/4 \\ G_2 &= (G_S - 2G_P + G_D)/4\end{aligned}\quad (43)$$

We can compare the DCA-partial DOS with a similar consideration for the cluster DMFT [21]. In this case, one first calculate the proper local matrix of the Green functions Eq. (8) and then transform it to the basis of molecular orbitals (inverse of Eqs. (37) and (43)) in order to obtained partial DOS ρ_m (Fig. (7)). It is clear that the DCA partial DOS overestimate "localization" of the partial sectors orbitals ϕ_m while the CDMFT has larger overlap between different partial DOS with non-local Green function contributions. This can lead to spurious k -selective polarization of correlated orbitals in the DCA-scheme compare to the CDMFT method.

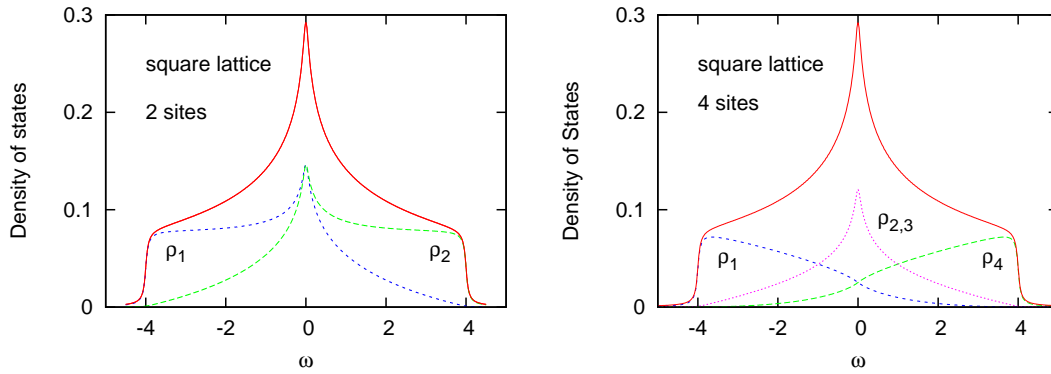


Fig. 7: Total density of states $\rho(\omega)$ and molecular orbital CDMFT-components $\rho_m(\omega)$ for 2-site (left panel) and 4-site clusters (right panel) of square lattice. The total bandwidth is 8, which corresponds to the hopping $t=1$. From Ref. [21].

5 Long-range correlations: Dual-Fermion approach

The shortcomings of cluster DMFT scheme have triggered many efforts to go beyond the mean-field description, while maintaining DMFT as a starting point. The standard DMFT scheme becomes exact in the limit of infinite coordination number z . An expansion in $1/z$, however, leads to difficulties as the action depends in a non-analytic way on the coordination number [22]. Building on earlier work on strong-coupling expansions for the Hubbard model [23–25], a general framework to perform a systematic cumulant expansion around DMFT even considering non-local Coulomb interaction was developed in Ref. [26].

While cluster extension to DMFT breaks translational symmetry of the lattice, the combination of numerical and analytic methods is a promising route for including the effects of long-range correlations. Recent developments have led to approaches which include long-range correlations via straightforward diagrammatic corrections to DMFT [27–29]. Based on earlier suggestions for bosonic fields [30], it was recognized that that a systematic, fully renormalized expansion around DMFT can be formulated in terms of auxiliary fermions [31].

Our goal is to find optimal strong-coupling expansion of the general lattice problem described by the imaginary time action

$$S[c^*, c] = - \sum_{\omega \mathbf{k} \sigma m m'} c_{\omega \mathbf{k} \sigma m}^* \left[(i\omega + \mu) \mathbf{1} - t_{\mathbf{k} \sigma}^{m m'} \right] c_{\omega \mathbf{k} \sigma m'} + \sum_i S_U[c_i^*, c_i]. \quad (44)$$

Here $t_{\mathbf{k} \sigma}$ is the one-electron part of the Hamiltonian, $\sigma = \uparrow, \downarrow$ labels the spin projection, m, m' are orbital indices and c^*, c are Grassmann variables. The index i labels the lattice sites and \mathbf{k} -vectors are quasimomenta. In order to keep the notation simple, it is useful to introduce the combined index $\alpha \equiv \{m\sigma\}$. Translational invariance is assumed for simplicity in the following. For applications it is important to note that the local part of the action, S_U , may contain *any* type of local interaction. The only requirement is that it is local within the multi-orbital atom or cluster.

In order to formulate a perturbation expansion around DMFT, a local quantum impurity problem

is introduced:

$$S_{\text{loc}}[c^*, c] = - \sum_{\omega} \sum_{\alpha\beta} c_{\omega\alpha}^* [(i\omega + \mu)\mathbf{1} - \Delta_{\omega}]_{\alpha\beta} c_{\omega\beta} + S_{\text{U}}[c^*, c], \quad (45)$$

where Δ_{ω} is the hybridization matrix describing the coupling of the impurity to an electronic bath. Apart from the connection to DMFT, another motivation for rewriting the lattice action in this form is to express it in terms of a reference problem that can be solved accurately for an arbitrary hybridization function using the CTQMC methods [7]. Exploiting the locality of the hybridization function Δ_{ω} , the lattice action (44) is rewritten *exactly* by adding and subtracting Δ_{ω} at each lattice site:

$$S[c^*, c] = \sum_i S_{\text{loc}}[c_i^*, c_i] + \sum_{\omega\mathbf{k}\alpha\beta} c_{\omega\mathbf{k}\alpha}^* (t_{\mathbf{k}} - \Delta_{\omega})_{\alpha\beta} c_{\omega\mathbf{k}\beta}. \quad (46)$$

Note that this step leaves the hybridization function unspecified. This will be used later to optimize the approach. The lattice may now be viewed as a collection of impurities, which are coupled through the bilinear term to the right of this equation (see Fig. 8). The effect of spatial correlations enters here and renders an exact solution impossible. A perturbative treatment is desirable, but not straightforward as the impurity action is non-Gaussian and hence there is no Wick theorem. Therefore, dual fermions are introduced in the path integral representation of the partition function from Eq. (6) through the standard Hubbard-Stratonovich transformation

$$\begin{aligned} \exp(c_{\alpha}^* B_{\alpha\beta} (A^{-1})_{\beta\gamma} B_{\gamma\delta} c_{\delta}) = \\ \frac{1}{\det A} \int \mathcal{D}[\gamma^*, \gamma] \exp(-f_{\alpha}^* A_{\alpha\beta} f_{\beta} - f_{\alpha}^* B_{\alpha\beta} c_{\beta} - c_{\alpha}^* B_{\alpha\beta} f_{\beta}). \end{aligned} \quad (47)$$

In order to transform the exponential of the bilinear term in (46), we choose the matrices a, b in accordance with Refs. [31] as

$$A = g_{\omega}^{-1} (\Delta_{\omega} - t_{\mathbf{k}})^{-1} g_{\omega}^{-1}, \quad B = g_{\omega}^{-1}, \quad (48)$$

where g_{ω} is the local, interacting Green function of the impurity problem. With this choice, the lattice action transforms to

$$S[c^*, c, f^*, f] = \sum_i S_{\text{site},i} + \sum_{\omega\mathbf{k}\alpha\beta} f_{\omega\mathbf{k}\alpha}^* [g_{\omega}^{-1} (\Delta_{\omega} - t_{\mathbf{k}})^{-1} g_{\omega}^{-1}]_{\alpha\beta} f_{\omega\mathbf{k}\beta}. \quad (49)$$

Hence the coupling between sites is transferred to a local coupling to the auxiliary fermions:

$$S_{\text{site},i}[c_i^*, c_i, f_i^*, f_i] = S_{\text{loc}}[c_i^*, c_i] + \sum_{\alpha\beta} f_{\omega i\alpha}^* g_{\omega\alpha\beta}^{-1} c_{\omega i\beta} + c_{\omega i\alpha}^* g_{\omega\alpha\beta}^{-1} f_{\omega i\beta}. \quad (50)$$

Since g_{ω} is local, the sum over all states labeled by \mathbf{k} could be replaced by the equivalent summation over all sites by a change of basis in the second term. The crucial point is that the coupling to the auxiliary fermions is purely local and S_{site} decomposes into a sum of local terms.

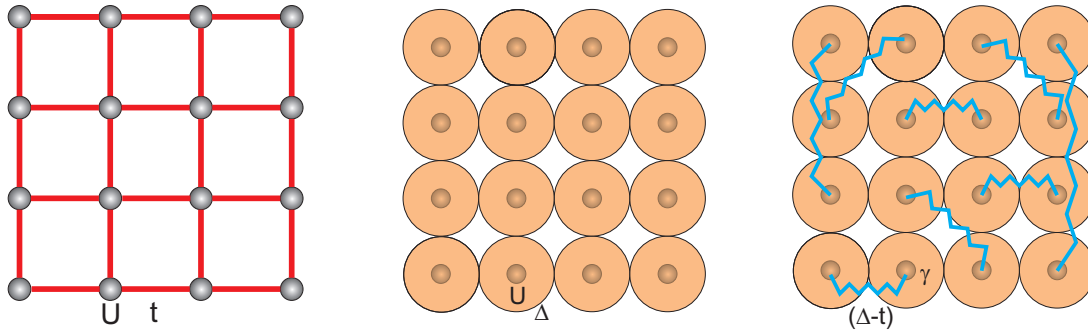


Fig. 8: Construction of the dual fermion approximation: In a first step, the original lattice problem (left) with bonds (red lines) is replaced by a collection of decoupled impurities exerted to an electronic bath, as indicated by the red spheres (middle), finally spatial correlations in the original lattice problem are treated perturbatively over $(\Delta_\omega - t_k)$ (blue wiggly line) with local interaction vertex γ (right).

The lattice fermions can therefore be integrated out from S_{site} for each site i separately. This completes the change of variables:

$$\int \mathcal{D}[c^*, c] \exp(-S_{\text{site}}[c_i^*, c_i, f_i^*, f_i]) = \mathcal{Z}_{\text{loc}} \exp\left(-\sum_{\omega \alpha \beta} f_{\omega i \alpha}^* g_{\omega \alpha \beta}^{-1} f_{\omega i \beta} - V_i[f_i^*, f_i]\right). \quad (51)$$

The above equation may be viewed as the defining equation for the dual potential $V[f^*, f]$. The choice of matrices (48) ensures a particularly simple form of this potential. An explicit expression is found by expanding both sides of Eq. (51) and equating the resulting expressions by order. Formally this can be done up to all orders and in this sense the transformation to the dual fermions is exact. For most applications, the dual potential is approximated by the first non-trivial interaction vertex:

$$V[f^*, f] = \frac{1}{4} \gamma_{1234} f_1^* f_2^* f_4 f_3, \quad (52)$$

where the combined index $1 \equiv \{\omega \alpha\}$ comprises frequency, spin and orbital degrees of freedom. γ is the exact, fully antisymmetric, reducible two-particle vertex of the local quantum impurity problem. It is given by

$$\gamma_{1234} = g_{11'}^{-1} g_{22'}^{-1} [\chi_{1'2'3'4'}^0 - \chi_{1'2'3'4'}^0] g_{3'3}^{-1} g_{4'4}^{-1}, \quad (53)$$

with the two-particle Green function of the impurity being defined as

$$\chi^{1234} = \langle c_1 c_2 c_3^* c_4^* \rangle_{\text{loc}} = \frac{1}{\mathcal{Z}_{\text{loc}}} \int \mathcal{D}[c^*, c] c_1 c_2 c_3^* c_4^* \exp\left(-S_{\text{loc}}[c^*, c]\right). \quad (54)$$

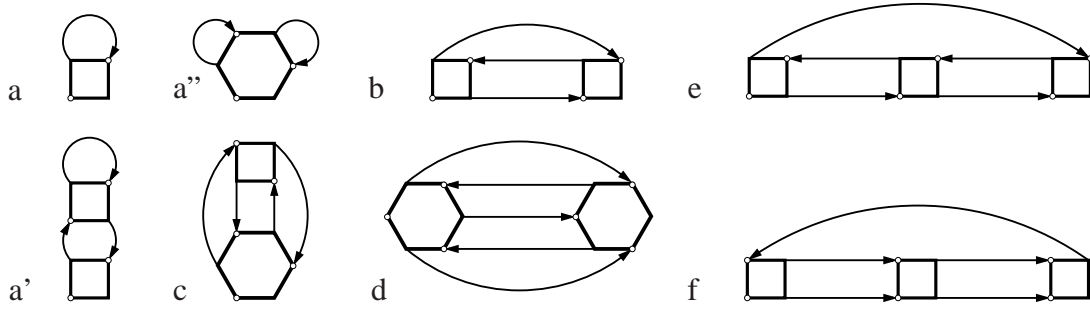


Fig. 9: Diagrams contributing to the dual self-energy Σ^d . Diagrams a), a'), a'') and c) give local, the other ones nonlocal contributions. The three diagrams labeled by a) do not contribute in case the condition (66) is fulfilled. From the Ref. [14].

The disconnected part reads

$$\chi_{1234}^0 = g_{14}g_{23} - g_{13}g_{24}. \quad (55)$$

The single- and two-particle Green functions can be calculated using the CTQMC algorithms [7]. After integrating out the lattice fermions, the dual action depends on the new variables only and reads

$$\tilde{S}[f^*, f] = - \sum_{\omega \mathbf{k} \alpha \beta} f_{\omega \mathbf{k} \alpha}^* [\tilde{G}_{\omega}^0(\mathbf{k})]_{\alpha \beta}^{-1} f_{\omega \mathbf{k} \beta} + \sum_i V_i[f_i^*, f_i]. \quad (56)$$

and the bare dual Green function is found to be

$$\tilde{G}_{\omega}^0(\mathbf{k}) = -g_{\omega} [g_{\omega} + (\Delta_{\omega} - t_{\mathbf{k}})^{-1}]^{-1} g_{\omega}, \quad (57)$$

which involves the local Green function g_{ω} of the impurity model.

Up to now, Eqs. (56), (57) are merely a reformulation of the original problem. In practice, approximate solutions are constructed by treating the dual problem perturbatively. Several diagrams that contribute to the dual self-energy are shown in Fig. 9. These are constructed from the impurity vertices and dual Green functions as lines. The first diagram (a) is purely local, while higher orders contain nonlocal contributions, e.g. diagram b). Inserting the renormalized Green function into diagram a) includes contributions such as the one in a'). In practice, approximations to the self-energy are constructed in terms of skeleton diagrams. The lines shown in Fig. 9 are therefore understood to be fully dressed propagators. The use of skeleton diagrams is necessary to ensure the resulting theory to be conserving in the Baym-Kadanoff sense [33], i. e. it fulfills the basic conservation laws for energy, momentum, spin and particle number. It is an important consequence of the exact transformation (47) that for a theory which is conserving in terms of dual fermions, the result is also conserving in terms of lattice fermions [32]. This allows to construct general conserving approximations within the dual fermion approach. Numerically, the self-energy is obtained in terms of skeleton diagrams by performing a self-consistent renormalization as described below. Once an approximate dual self-energy is found, the result may be transformed back to a physical result in terms of lattice fermions using exact relations.

The action (56) allows for a Feynman-type diagrammatic expansion in powers of the dual potential V . The rules are similar to those of the antisymmetrized diagrammatic technique [34]. Extension of these rules to include generic n -particle interaction vertices is straightforward. Due to the use of an antisymmetrized interaction, the diagrams acquire a combinatorial prefactor. For a tuple of n equivalent lines, the expression has to be multiplied by a factor $1/n!$. As simplest example we can write schematically the first self-energy correction of the diagram a) in Fig. 9 contains a single closed loop:

$$\tilde{\Sigma}_{12}^{(a)} = -T \sum_{34} \gamma_{1324} \tilde{G}_{43}^{\text{loc}} \quad (58)$$

where $\tilde{G}^{\text{loc}} = (1/N) \sum_{\mathbf{k}} \tilde{G}(\mathbf{k})$ denotes the local part of the dual Green function. The second-order contribution represented by diagram b) contains two equivalent lines and one closed loop and hence is \mathbf{k} -dependence:

$$\tilde{\Sigma}_{12}^{(b)}(\mathbf{k}) = -\frac{1}{2} \left(\frac{T}{N} \right)^2 \sum_{\mathbf{k}_1 \mathbf{k}_2} \sum_{345678} \gamma_{1345} \tilde{G}_{57}(\mathbf{k}_1) \tilde{G}_{83}(\mathbf{k}_2) \tilde{G}_{46}(\mathbf{k} + \mathbf{k}_2 - \mathbf{k}_1) \gamma_{6728} \quad (59)$$

In practice, it is more efficient to evaluate the lowest order diagrams in real space and transform back to reciprocal space using the fast Fourier transform.

5.1 Dual-Fermion approach: Exact relations

After an approximate result for the dual self-energy or the dual Green function has been obtained, it has to be transformed back to the corresponding physical quantities in terms of lattice fermions. The fact that dual fermions are introduced through the exact Hubbard-Stratonovich transformation (47) allows to establish exact identities between dual and lattice quantities. Hence the transformation does not involve any additional approximations [14, 31].

The relations between the n -particle cumulants of dual and lattice fermions can be established using the cumulant (linked cluster) technique. To this end, one may consider two different, equivalent representations of the following generating functional:

$$F[J^*, J; L^*, L] = \ln \mathcal{Z}_f \int \mathcal{D}[c^*, c; f^*, f] \exp \left(-S[c^*, c; f^*, f] + J_1^* c_1 + c_2^* J_2 + L_1^* f_1 + f_2^* L_2 \right). \quad (60)$$

Integrating out the lattice fermions from this functional similar to (51) (this can be done with the sources J and J^* set to zero) yields

$$F[L^*, L] = \ln \tilde{\mathcal{Z}}_f \int \mathcal{D}[f^*, f] \exp \left(-\tilde{S}[f^*, f] + L_1^* f_1 + f_2^* L_2 \right). \quad (61)$$

with $\tilde{\mathcal{Z}}_f = \mathcal{Z}/\tilde{\mathcal{Z}}$. The dual Green function and two-particle correlator related with non-local susceptibilities are obtained from (61) by suitable functional derivatives, e.g.

$$\tilde{G}_{12} = -\frac{\delta^2 F}{\delta L_2 \delta L_1^*} \Big|_{L^*=L=0}, \quad \left[\tilde{X} - \tilde{G} \otimes \tilde{G} \right]_{1234} = \frac{\delta^4 F}{\delta L_4 \delta L_3 \delta L_2^* \delta L_1^*} \Big|_{L^*=L=0}, \quad (62)$$

where $G \otimes G$ is the antisymmetrized direct product of Green functions, so that the angular brackets is the connected part of the dual two-particle Green function. Conversely, integrating out the dual fermions from (60) using the HST, one obtains an alternative representation, which more clearly reveals a connection of the functional derivatives with respect to the sources J, J^* and L, L^* . The result is

$$F[J^*, J; L^*, L] = L_1^* [g(\Delta - h)g]_{12} L_2 + \ln \int \mathcal{D}[c^*, c] \exp \left(- S[c^*, c] + J_1^* c_1 + c_2^* J_2 + L_1^* [g(\Delta - t)]_{12} c_2 + c_1^* [(\Delta - t)g]_{12} L_2 \right). \quad (63)$$

In analogy to (62), the cumulants in terms of lattice fermions are obviously obtained by functional derivative with respect to the sources J and J^* with L and L^* set to zero. Applying the derivatives with respect to L, L^* to (63) with $J = J^* = 0$ and comparing to (62), e.g. yields the following identity:

$$\tilde{G}_{12} = -[g(\Delta - t)g]_{12} + [g(\Delta - t)]_{11'} G_{1'2'} [(\Delta - t)g]_{2'2}. \quad (64)$$

Solving for G provides the rule how to transform the dual Green function to the physical quantity in terms of lattice fermions. For higher-order cumulants the additive term in (63) does not contribute and the relation between the two-particle cumulants evaluates to

$$\left[\tilde{X} - \tilde{G} \otimes \tilde{G} \right]_{1234} = [g(\Delta - t)]_{11'} [g(\Delta - t)]_{22'} [X - G \otimes G]_{1'2'3'4'} [(\Delta - t)g]_{3'3} [(\Delta - t)g]_{4'4}, \quad (65)$$

It is apparent that similar relations hold for higher-order cumulants. Note that the transformation only involves single-particle functions. Hence one may conclude that n -particle collective excitations are the same for dual and lattice fermions.

5.2 Self-consistency condition and relation to DMFT

The hybridization function Δ , which so far has not been specified, allows to optimize the starting point of the perturbation theory and should be chosen in an optimal way. The condition of the first diagram (Fig. 9 a) in the expansion of the dual self-energy to be equal to zero at all frequencies fixes the hybridization. This eliminates the leading order diagrammatic correction to the self-energy and establishes a connection to DMFT, which can be seen as follows: Since γ vertex is local, this condition amounts to demanding that the local part of the dual Green function be zero:

$$\sum_{\mathbf{k}} \tilde{G}_{\omega}(\mathbf{k}) = 0. \quad (66)$$

The simplest nontrivial approximation is obtained by taking the leading-order correction, diagram a), evaluated with the bare dual propagator (57). Recalling the expression for the DMFT Green function, Eq. (4), it is readily verified that

$$\begin{aligned} G_{\omega}^{\text{DMFT}}(\mathbf{k}) - g_{\omega} &= [g_{\omega}^{-1} + \Delta_{\omega} - t_{\mathbf{k}}]^{-1} - g_{\omega} \\ &= -g_{\omega} [g_{\omega} + (\Delta_{\omega} - t_{\mathbf{k}})^{-1}]^{-1} g_{\omega} = \tilde{G}_{\omega}^0(\mathbf{k}). \end{aligned} \quad (67)$$

It immediately follows that (66) evaluated with the bare dual Green function is exactly equivalent to the DMFT self-consistency condition:

$$\frac{1}{N} \sum_{\mathbf{k}} \tilde{G}_{\omega}^0(\mathbf{k}) = 0 \quad \Longleftrightarrow \quad \frac{1}{N} \sum_{\mathbf{k}} G_{\omega}^{\text{DMFT}}(\mathbf{k}) = g_{\omega}. \quad (68)$$

Hence DMFT appears as the zero-order approximation in this approach and corrections to DMFT are included perturbatively. A formal relation to DMFT can be established using the Feynman variational functional approach. In this context, DMFT appears as the optimal approximation to a Gaussian ensemble of dual fermions [32].

When diagrammatic corrections are taken into account and the first diagram is evaluated with the dressed propagator \tilde{G} , the condition (66) will in general be violated. It can be reinforced by adjusting the hybridization function iteratively. This corresponds to eliminating an infinite partial series starting from the diagrams labeled by a) in Fig. 9. These contributions are effectively absorbed into the impurity problem. Note that such an expansion is not one around DMFT, but rather around an optimized impurity problem.

The only difference between a DMFT and a DF calculation are the diagrammatic corrections which are included into the dual Green function. To this end, the local impurity vertex γ has to be calculated in addition to the Green function in the impurity solver step.

Since the choice of the hybridization function is not unique, one may replace it by a discrete version $\Delta^{(n)} = \sum_{k=1}^n |V_k|^2 / (i\omega - \epsilon_k)$ for a small number n of bath degrees of freedom, for which the impurity problem can be solved efficiently using exact diagonalization. In this case, the condition (66) cannot be fulfilled in general, but one may require the correction to be *minimal* instead. This results in a variational approach. The corresponding perturbation expansion is considerably more stable than an expansion around the atomic limit, i.e. $\Delta \equiv 0$ [24].

5.3 Results for the 2d-Hubbard model

In the following, we show some illustrative results for the Hubbard model, which is governed by the Hamiltonian (1). Unless otherwise stated, only the two lowest-order diagrams a) and b) of Fig. 9 have been used. It may be considered as a benchmark system for the approach, because the importance of nonlocal correlations is expected to increase by reducing the dimensionality. This is clearly an unfavorable situation for DMFT, which completely neglects spatial correlations.

In order to visualize the nonlocal correlations, the k -dependent self-energy is shown in Fig. (10). The upper panel of Figure 10 presents contour plots for $\text{Im}\Sigma_{\omega=0,k}$ at $U = 1.0$ and $U = 2.0$ (the data are obtained by a polynomial extrapolation from the Matsubara frequencies). The value of $\text{Im}\Sigma_{\omega=0,k}$ grows dramatically as U changes from 1.0 to 2.0. Close to the Mott transition there is a strong k -dependence of Σ . The renormalized dispersion law $\epsilon_k + \text{Re}\Sigma_{\omega=0,k}$ is now also in a qualitative agreement with numerical data, as the lower panel of Figure 10. In these graphs, $\epsilon_k + \text{Re}\Sigma_{\omega=0,k}$ is compared with the reference data for a 10×10 lattice. There is a qualitative difference between the results for $U = 1.0$ and $U = 2.0$: for later case the corrections are quite large so that there is a dependence resembling ϵ_k^{-1} [31, 31].

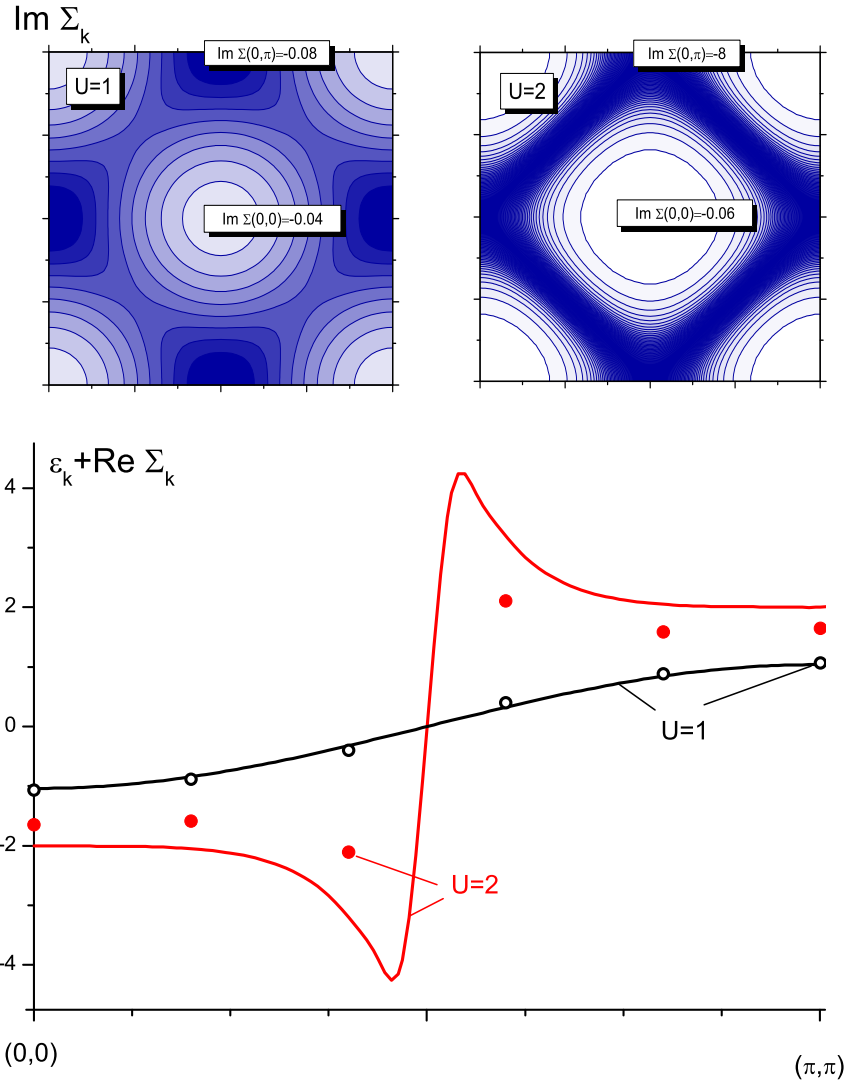


Fig. 10: Momentum dependence for the self-energy function at Fermi energy, obtained with diagram (b) within the translationally-invariant approximation for the undoped Hubbard model. Data are shown at $t = 0.25$, $\beta = 20$, for $U = 1.0$ and $U = 2.0$. Upper panel: contour plots for k -dependence of the imaginary part of the self energy. Lower panel: renormalized dispersion law $\epsilon_k + \text{Re } \Sigma_{\omega=0,k}$, compared with the reference data obtained for 10×10 lattice. From the Ref. [31].

The k -resolved spectral function $A(\mathbf{k}, \omega)$ obtained from maximum-entropy analytical continuation shown in Fig. 11. The DMFT spectral function displays a quasiparticle band, while in the DF calculation, spectral weight is transferred away from the Fermi level. Recalling the nesting condition $\epsilon_{\mathbf{k}+\mathbf{Q}} = -\epsilon_{\mathbf{k}}$ for the antiferromagnetic wave vector $\mathbf{Q} = (\pi, \pi)$, the locus of these features allows to interpret them as shadow bands due to dynamical short-range antiferromagnetic correlations. The strength of these correlations increases as the temperature is lowered.

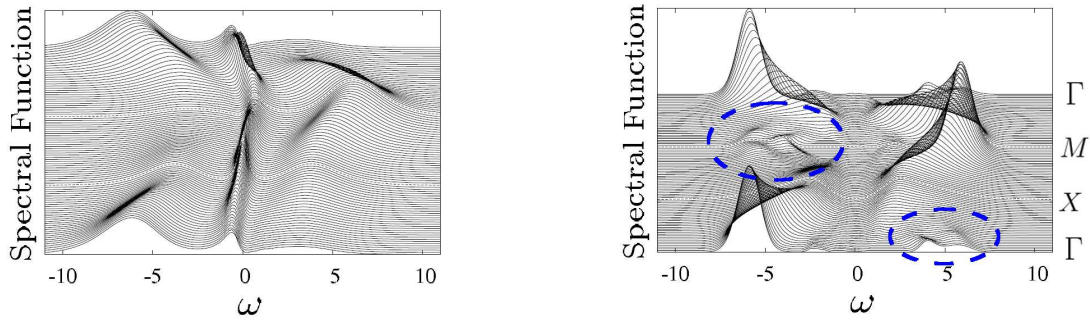


Fig. 11: Spectral function $A(\mathbf{k}, \omega)$ for the 2D Hubbard model at half-filling obtained within DMFT (left) and dual fermion calculations (right) for $U = 8t$ and $T/t = 0.235$. From bottom to top, the curves are plotted along the high-symmetry lines $\Gamma \rightarrow X \rightarrow M \rightarrow \Gamma$. The high-symmetry points $X = (0, \pi)$ and $M = (\pi, \pi)$ are marked by dashed lines. The structures encircled in blue can be attributed to dynamical short-range antiferromagnetic correlations. From the Ref. [14].

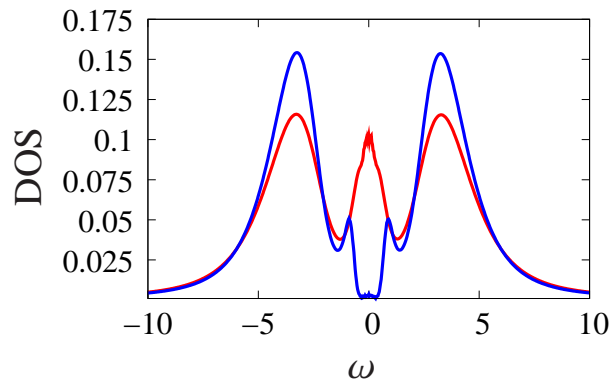


Fig. 12: Metallic and insulating local density of states obtained in the coexistence region of the Mott transition for $U/t = 6.5$ and $T/t = 0.08$. The insulating solution exhibits characteristic peaks at the gap edge. The antiferromagnetic correlations lead to antiferromagnetic-gap-like behavior [35]. The metallic solution exhibits shoulders on the peak at the Fermi level. From the Ref. [14].

A detailed analysis of the phasediagram shows that these correlations lead to a drastic reduction of the critical U from $U_c/t \sim 9.35$ in DMFT down to $U_c/t \sim 6.5$ within the dual fermion calculation. This, as well as the density of states in the coexistence region (Fig. 12) and the slope of the transition lines in the $U - T$ phase diagram below the critical point, which are modified from negative within DMFT to positive [14], is in qualitative agreement with cluster DMFT results [36]. We emphasize that these results cannot be obtained from a straightforward diagrammatic expansion around DMFT as the modification of the Weiss field is essential. This distinguishes the present method from related approaches [28, 29].

$$\Gamma^{d/m}(\mathbf{q}) = \gamma^{(4)} + \gamma^{(4)} \Gamma^{d/m}(\mathbf{q})$$

Fig. 13: Bethe-Salpeter equation for the dual vertex in the electron-hole channel with a local approximation $\Gamma_{irr} = \gamma$ to the irreducible vertex. The solution Γ contains the sum of all ladder diagrams up to infinite order in γ .

$$\chi_0(\mathbf{q}, \Omega) + \tilde{\chi}(\mathbf{q}, \Omega) = \text{loop} + \text{loop} \Gamma^{eh0}$$

Fig. 14: Diagrammatic representation of the susceptibility, Eqs. (71), (72).

5.4 Calculation of susceptibilities

For the calculation of the dual susceptibility, the dual vertex function is first calculated by means of a Bethe-Salpeter equation [37,38] (in the following we write the equations for a single-orbital model for simplicity)

$$\Gamma_{\omega\omega'\Omega}^{\alpha}(\mathbf{q}) = \gamma_{\omega\omega'\Omega}^{\alpha} - \frac{T}{N} \sum_{\omega''} \sum_{\mathbf{k}} \gamma_{\omega\omega''\Omega}^{\alpha} \tilde{G}_{\omega''}(\mathbf{k}) \tilde{G}_{\omega''+\Omega}(\mathbf{k} + \mathbf{q}) \Gamma_{\omega''\omega'\Omega}^{\alpha}(\mathbf{q}). \quad (69)$$

This equation is depicted diagrammatically in Fig. 13. Here the irreducible vertex is approximated by the local irreducible interaction of dual fermions to lowest-order and is hence given by the *reducible* vertex of the impurity model γ (the index '(4)') is omitted in what follows). Here $\alpha = d, m$ stands for the density (d) and magnetic (m) electron-hole channels: $\Gamma^d = \Gamma^{\uparrow\uparrow\uparrow\uparrow} + \Gamma^{\uparrow\uparrow\downarrow\downarrow}$, $\Gamma^m = \Gamma^{\uparrow\uparrow\uparrow\uparrow} - \Gamma^{\uparrow\uparrow\downarrow\downarrow}$. The physical content of the BSE is repeated scattering of particle-hole pairs. In the two channels the particle-hole pair has a definite total spin S and spin projection S_z . The density channel corresponds to the $S = 0$, $S_z = 0$ singlet channel, while Γ^m is the vertex in the $S = 1$, $S_z = 0$ triplet channel. In the magnetic channel, the collective excitations are magnons. The vertex $\Gamma^{\uparrow\downarrow\downarrow\uparrow}$ ($\Gamma^{\downarrow\uparrow\uparrow\downarrow}$) which corresponds to the $S_z = +1(-1)$ spin projection of the $S = 1$ channel must be equal to Γ^m in the paramagnetic state (longitudinal and transverse modes cannot be distinguished).

The BSE may be solved iteratively, starting from the approximation $\Gamma^{(0)} \approx \gamma$. Inserting this into the right-hand-side of Eq. (69) yields a new approximation $\Gamma^{(1)}$. Repeating this step successively generates a sum of all ladder diagrams with $1, \dots, n + 1$ irreducible rungs in the

approximation $\Gamma^{(n)}$. In practice however, the BSE is solved by matrix inversion according to

$$[\Gamma_{\omega\omega'\Omega}^{\alpha}(\mathbf{q})]^{-1} = (\gamma_{\omega\omega'\Omega}^{\alpha})^{-1} + \frac{T}{N} \sum_{\mathbf{k}} \tilde{G}_{\omega}(\mathbf{k}) \tilde{G}_{\omega+\Omega}(\mathbf{k} + \mathbf{q}) \delta_{\omega\omega'}, \quad (70)$$

which corresponds to summing up the infinite series. The vertices are matrices in the fermionic Matsubara frequencies ω, ω' . They are diagonal with respect to Ω and \mathbf{q} , since the center of mass energy and momentum of the particle-hole pair is conserved in scattering processes.

From the vertex, the non-local spin and charge susceptibility is finally obtained as $X = X_0 + X_1$, where

$$X_0(\mathbf{q}, \Omega) = -\frac{T}{N} \sum_{\omega} \sum_{\mathbf{k}} G_{\omega}(\mathbf{k}) G_{\omega+\Omega}(\mathbf{k} + \mathbf{q}) \quad (71)$$

and

$$X_1^{\alpha}(\mathbf{q}, \Omega) = \frac{T^2}{N^2} \sum_{\omega\omega'} \sum_{\mathbf{k}\mathbf{k}'} G_{\omega}(\mathbf{k}) G_{\omega+\Omega}(\mathbf{k} + \mathbf{q}) \Gamma_{\omega\omega'\Omega}^{\alpha} G_{\omega'}(\mathbf{k}') G_{\omega'+\Omega}(\mathbf{k}' + \mathbf{q}). \quad (72)$$

In principle, these relations are valid for dual and lattice fermions. If one is only interested in instabilities, which are signalled by the divergence of the corresponding susceptibility, it is sufficient to consider the dual quantities. The equivalence of two-particle excitations in terms of dual and lattice fermions ensures that the dual and lattice susceptibilities diverge at the same parameters. The lattice susceptibility is obtained using the exact relations between dual and lattice correlation functions (65). In the context of DMFT, the susceptibility is obtained using relations similar to Eqs. (69), (71) and (72) [2]. The momentum dependence of the irreducible vertex is neglected in DMFT. It is further approximated by the irreducible vertex of the impurity model. Recall that the lattice Green function is exactly equal to the DMFT Green function when dual corrections to the self-energy are neglected and the dual Green function fulfills the self-consistency condition (66). Using the relation between the DMFT and bare dual Green function Eq. (67) we can find a simple relations between the bare susceptibilities:

$$\tilde{X}_0(\mathbf{q}, \Omega) = X_0(\mathbf{q}, \Omega) - \chi_0(\Omega). \quad (73)$$

It is an important property of the above equations that under the same conditions the lattice susceptibility calculated within the dual fermion approach is exactly equal to the DMFT susceptibility [14].

As a further illustration, we plot the dynamical susceptibility $\chi(\mathbf{q}, \omega)$ in Fig. 15. It clearly displays the magnon spectrum in the paramagnetic state. The dispersion from spin wave theory is shown for comparison. It is given by the expression [39] $\epsilon(\mathbf{k}) = 2zJS\sqrt{1 - \gamma(\mathbf{k})^2}$ where z is the coordination number, $S = 1/2$ is the spin of the fermions and $\gamma(\mathbf{k}) = \frac{1}{z} \sum_{\text{NN}} e^{i\mathbf{k}\mathbf{r}_{\text{NN}}} = (\cos k_x + \cos k_y) / 2$ for the square lattice. The right panel of Fig. 15 shows a cross-section for the antiferromagnetic wave vector $\mathbf{q}_{\text{AF}} = (\pi, \pi)$ (M -point). The peak is broadened and slightly shifted from zero. Such a behavior is reminiscent of a 2D Heisenberg model at finite temperature, where long-range order with a correlation length $\xi \gg a$ takes place (a is the lattice constant) and a corresponding small energy scale of order Ja/ξ arises [35].

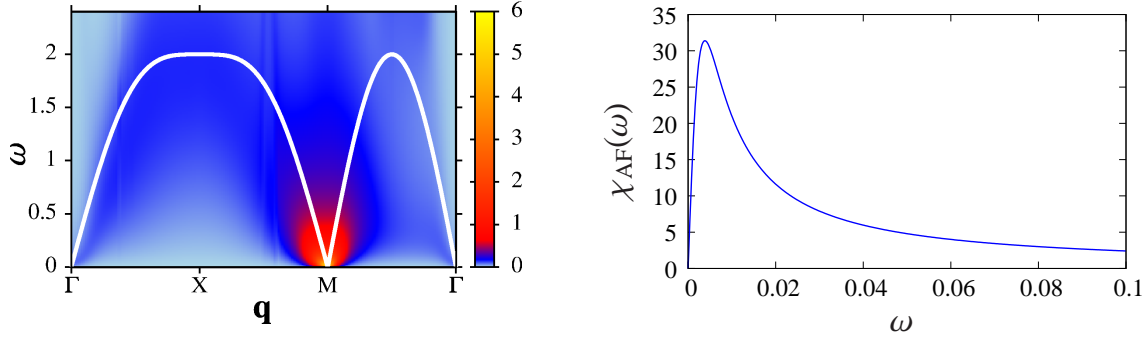


Fig. 15: Left: Dynamical susceptibility $\chi(\mathbf{q}, \omega)$ for $U/t = 4$ and $T/t = 0.19$, obtained from a dual fermion calculation and analytical continuation using Padé approximants. It shows the magnon spectrum in the paramagnetic state. The dispersion from spin wave theory with effective exchange coupling $J = 4t^2/U$ is shown for comparison. Values for $\chi > 6$ are excluded from the colormap to improve the contrast. Right: Cross-section through the peak at the M -point. The displacement from zero is consistent with a small energy scale J/ξ , where ξ is the correlation length (in units of the lattice constant). From the Ref. [14].

5.5 Convergence properties

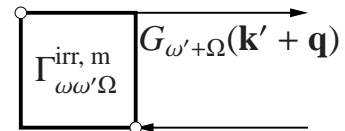
For a perturbative approach, the convergence properties are of paramount importance. For the present theory [41], the vertices appear as a small parameter in the expansion in the weak-coupling limit ($U \rightarrow 0$), because they vanish at least proportionally to U : $\gamma^{(4)} \sim U$, $\gamma^{(6)} \sim U^2, \dots$. On the other hand, for an expansion around the atomic limit ($\Delta \equiv 0$), the dual Green function is small near this limit: For $h_{\mathbf{k}}$ small, the bare dual Green function can be approximated as

$$\tilde{G}_{\omega}^0(\mathbf{k}) \approx g_{\omega} h_{\mathbf{k}} g_{\omega}. \quad (74)$$

This enforces the convergence of the series in the opposite strong coupling limit. In contrast, IPT or FLEX, which operate with the bare interaction U , have to break down at intermediate to large U . In the general case, a fast convergence cannot be proven rigorously. Here we examine the convergence properties numerically in the vicinity of the antiferromagnetic instability (AFI) in the 2D Hubbard model. These can be characterized using the eigenvalue problem derived from the BSE (69).

The matrix is the building block of the particle-hole ladder and may be thought of as the effective two-fermion interaction. For dual fermions, the irreducible vertex is given by the bare dual interaction $\Gamma_{\omega\omega'\Omega}^{\text{irr}, m} = \gamma_{\omega\omega'\Omega}^m = \gamma_{\omega\omega'\Omega}^{\uparrow\uparrow} - \gamma_{\omega\omega'\Omega}^{\uparrow\downarrow}$ in the magnetic channel and \tilde{G} stands for the full dual Green function. Here the focus is on the leading eigenvalues in the vicinity of the AFI and hence $\mathbf{q} = (\pi, \pi)$ and $\Omega = 0$. An eigenvalue of $\lambda_{\text{max}} = 1$ implies a divergence of the ladder sum and hence a breakdown of the perturbation theory.

$$-\frac{T}{N} \sum_{\omega'\mathbf{k}'} \Gamma_{\omega\omega'\Omega}^{\text{irr}, m} \tilde{G}_{\omega'}(\mathbf{k}) \tilde{G}_{\omega'+\Omega}(\mathbf{k} + \mathbf{q}) \phi_{\omega'} = \tilde{\lambda} \phi_{\omega'}. \quad (75)$$



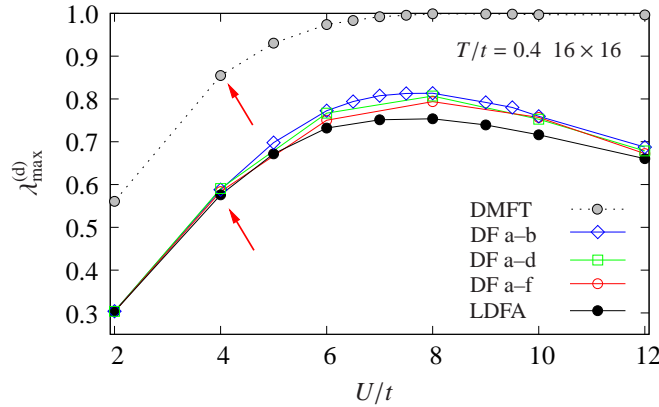


Fig. 16: Leading eigenvalue of the Bethe-Salpeter equation obtained within various approximations in the $\mathbf{q} = (\pi, \pi)$ magnetic channel as a function of the interaction U . λ (λ^d) denotes lattice (dual) fermion eigenvalues. The diagrams included are indicated in the legend (labels are the same as in Fig. 9). The dual perturbation theory converges fast (i.e. the eigenvalues are small) in particular for weak and strong coupling. A straightforward diagrammatic expansion around DMFT breaks down for large U . From the Ref. [41].

The results are displayed in Fig. 16. For weak coupling, the leading eigenvalue is small and implies a fast convergence of the diagrams in the electron-hole ladder. More significantly, the eigenvalues decrease and converge to the same intercept in the large U limit. This nicely illustrates that the dual perturbation theory smoothly interpolates between a standard perturbation expansion at small, and the cumulant expansion at large U , ensuring fast convergence in both regimes. From the figure it is clear that this also improves the convergence properties for intermediate coupling ($U \sim W$). Even here corrections from approximations involving higher-order diagrams remain small, including those from the LDFA. Diagrams involving the three-particle vertex give a negligible contribution.

For a straightforward diagrammatic expansion around DMFT, the building block of the particle-hole ladder is constructed from the irreducible impurity vertex $\gamma_{\omega\omega'\Omega}^{\text{irr},m}$ and DMFT Green functions. As seen in Fig. 16, the corresponding leading eigenvalue (and the effective interaction) is much larger than for dual fermions over the whole parameter range (e.g. at red arrows). When transforming the leading eigenvalue back to lattice fermions, it is close to the DMFT value for these parameters. Hence convergence is enhanced for a perturbation theory in terms of dual fermions. Remarkably, for the intermediate to strong coupling region, a straightforward perturbation theory around DMFT breaks down (since the eigenvalue approaches one), while for a theory in terms of dual fermions, this is not the case. The fact that the leading eigenvalue for dual fermions is smaller is a generic feature. It is also observed away from half-filling and for the electron-electron channel. Note that the interaction in the dual fermion approach is given by the reducible vertex of the impurity. The frequency dependence accounts for the fact that the Coulomb interaction acts on short time scales in this approach. Strong local correlations are effectively separated (and treated non-perturbatively within the solution of the impurity model) from weaker spatial correlations, which are treated diagrammatically.

6 Summary and outlook

Different cluster extensions of the DMFT scheme are very useful tools to describe non-local short-range correlations in solids. The solution of the effective multi-site cluster impurity problem in fermionic bath is a common feature of CDMFT and DCA methods and can be found within numerically exact CTQMC approach. Nevertheless all cluster extension of the DMFT have problems: while CDMFT breaks translational symmetry of crystals, DCA effectively renormalized lattice hopping and makes a step-like momentum dependence of self-energy. We still should find an optimal way of periodization the CDMFT scheme. For the DCA approach one can average over different tiling of the Brillouin zone within the same cluster as was suggested for non-local CPA scheme [9]. For the CDMFT scheme the main problem is to find periodic self-energy solution, which preserve the analytical properties of the lattice Green's function [16]. In this case, even for small clusters one can find a similar solution for both CDMFT and DCA methods.

The dual fermion scheme gives a general framework to include non-local correlations on all scales. The bottleneck of DF-scheme related with finite number of diagrams which one can calculate, and the accuracy of short-range correlations are not as good as in numerically exact cluster solution. There are straightforward generalizations of the single site DF-approach to the cluster dual fermion approach (CDFA) [40] as well as DFCA-scheme [42]. We can think that the cluster DMFT starting point will allow to find a better non-local solutions, which have exact short-range correlations and reasonable long-range correlations. This may be an optimal way to study the complicated non-local effects in solids.

Acknowledgment

Support of the Deutsche Forschungsgemeinschaft through FOR1346 is gratefully acknowledged.

References

- [1] G. Kotliar and D. Vollhardt, *Physics Today* **57**, 53 (2004)
- [2] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [3] V.I. Anisimov, A.I. Poteryaev, M.A. Korotin, A.O. Anokhin, and G. Kotliar, *J. Phys.: Condensed Matter* **9**, 7359 (1997)
- [4] A.I. Lichtenstein and M.I. Katsnelson, *Phys. Rev. B* **57**, 6884 (1998)
- [5] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, and C.A. Marianetti, *Rev. Mod. Phys.* **78**, 865 (2006)
- [6] A.I. Lichtenstein and M.I. Katsnelson, *Phys. Rev. B* **62**, R9283 (2000)
- [7] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner, *Rev. Mod. Phys.* **83**, 349 (2011)
- [8] G. Kotliar, S.Y. Savrasov, G. Palsson, and G. Biroli, *Phys. Rev. Lett.* **87**, 186401 (2001)
- [9] D.A. Rowlands, X.-G. Zhang, and A. Gonis, *Phys. Rev. B* **78**, 115119 (2008)
- [10] T. Maier, M. Jarrell, T. Pruschke, and M.H. Hettler, *Rev. Mod. Phys.* **77**, 1027 (2005)
- [11] M. Potthoff, M. Aichhorn, and C. Dahnken, *Phys. Rev. Lett.* **91**, 206402 (2003)
- [12] A. Fuhrmann, S. Okamoto, H. Monien, and A.J. Millis, *Phys. Rev. B* **75**, 205118 (2007)
- [13] S. Okamoto, A.J. Millis, H. Monien, and A. Fuhrmann, *Phys. Rev. B* **68**, 195121 (2003)
- [14] H. Hafermann: *Numerical Approaches to Spatial Correlations in Strongly Interacting Fermion Systems* (Cuvillier Verlag, Göttingen, Singapore, 2010)
- [15] G. Biroli and G. Kotliar, *Phys. Rev. B* **65**, 155112 (2002)
- [16] G. Biroli, O. Parcollet and G. Kotliar, *Phys. Rev. B* **69**, 205108 (2004)
- [17] E. Koch, G. Sangiovanni, and O. Gunnarsson, *Phys. Rev. B* **78**, 115102 (2008)
- [18] M. Potthoff, *Eur. Phys. J. B* **32**, 429 (2003)
- [19] M. Potthoff and M. Balzer, *Phys. Rev. B* **75**, 125112 (2007)
- [20] C. Lin and A.J. Millis, *Phys. Rev. B* **79**, 205109 (2009)
- [21] A. Liebsch, H. Ishida, J. Merino, *Phys. Rev. B* **78**, 165123 (2008)
- [22] A. Schiller and K. Ingersent, *Phys. Rev. Lett.* **75**, 113 (1995)

-
- [23] S. Pairault, D. Sénéchal, and A.-M.S. Tremblay, *Phys. Rev. Lett.* **80**, 5389 (1998)
- [24] S. Pairault, D. Sénéchal, and A.-M.S. Tremblay, *European Phys. Journal B* **16**, 85 (2000)
- [25] S.K. Sarker, *J. Physics C: Solid State Phys.* **21**, L667 (1988)
- [26] T.D. Stanescu and G. Kotliar, *Phys. Rev. B* **70**, 205112 (2004)
- [27] C. Slezak, M. Jarrell, Th. Maier, and J. Deisz, *J. Phys.: Condens. Matter* **21**, 435604 (2009)
- [28] A. Toschi, A.A. Katanin, and K. Held, *Phys. Rev. B* **75**, 045118 (2007)
- [29] H. Kusunose, *J. Phys. Soc. Japan* **75**, 054713 (2006)
- [30] A.N. Rubtsov, *Phys. Rev. B* **66**, 052107 (2002)
- [31] A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, *Phys. Rev. B* **77**, 033101 (2008)
- [32] A.N. Rubtsov, M.I. Katsnelson, A.I. Lichtenstein, and A. Georges, *Phys. Rev. B* **79**, 045133 (2009)
- [33] G. Baym and L.P. Kadanoff, *Phys. Rev.* **124**, 287 (1961)
- [34] A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinskii: *Methods of Quantum Field Theory in Statistical Physics* (Pergamon Press, New York, 1965)
- [35] V.Y. Irkhin and M.I. Katsnelson, *J. Phys.: Cond. Matter* **3** 3, 6439 (1991)
- [36] H. Park, K. Haule, and G. Kotliar, *Phys. Rev. Lett.* **101**, 186403 (2008)
- [37] A.B. Migdal: *Theory of Finite Fermi Systems and applications to atomic nuclei* (Interscience Publishers, New York, 1967)
- [38] P. Nozières: *Theory of interacting Fermi systems* (Benjamin, New York, 1964).
- [39] A. Auerbach: *Interacting Electrons and Quantum Magnetism* (Springer, New York, 1998).
- [40] H. Hafermann, S. Brener, A.N. Rubtsov, M.I. Katsnelson, A.I. Lichtenstein, *JETP Lett.* **86**, 677 (2007)
- [41] H. Hafermann, G. Li, A.N. Rubtsov, M.I. Katsnelson, A.I. Lichtenstein, and H. Monien, *Phys. Rev. Lett.* **102**, 206401 (2009)
- [42] S.-X. Yang, H. Fotso, H. Hafermann, K.-M. Tam, J. Moreno, T. Pruschke, and M. Jarrell, arXiv:1104.3854

12 Multiple-Scattering Formalism for Correlated Systems: a KKR-DMFT Approach

H. Ebert, J. Minár, and D. Ködderitzsch
Department Chemie – Lehrbereich Phys. Chemie
Ludwig-Maximilians-Universität München

Contents

1	Introduction	2
2	The LSDA+DMFT scheme	4
3	Combination of the LSDA+DMFT with the KKR method	6
3.1	General idea of the KKR method	6
3.2	Solution of the single-site problem	7
3.3	Multiple scattering and Green's function	9
3.4	Electronic structure and relativistic extension	11
3.5	The self-consistency cycle	15
3.6	Treatment of disordered alloys	15
4	Applications	17
4.1	Ground state properties	17
4.2	Electron spectroscopy	22
5	Summary	27

1 Introduction

The combination of the DMFT (dynamical mean field theory) with the LSDA (local spin density approximation) led to a very powerful approach to deal with correlations in solid state materials beyond the capability of plain LSDA [1–3]. Due to this attractive feature the LSDA+DMFT has now been implemented on the basis of many different band structure methods during the last years [4]: first in the linear muffin-tin orbital method in the atomic sphere approximation (ASA-LMTO) [5–7] and then in full-potential LMTO [8, 9]. Common to nearly all of these band structure methods is that they are based on the variational principle representing the electronic structure in terms of Bloch functions $\Psi_{j\vec{k}}(\vec{r})$ and eigen energies $E_{j\vec{k}}$. Corresponding LSDA+DMFT calculations are therefore done in several steps: starting from a standard LSDA calculation the corresponding one-electron Green's function $G_{\text{LSDA}}(\vec{k}, E) = [E - H_{\text{LSDA}}(\vec{k})]^{-1}$ is constructed in reciprocal space. In a next step the DMFT self-energy $\Sigma(E)$, that accounts for correlation effects beyond the LSDA level, is included by solving the Dyson equation $G^{-1}(\vec{k}, E) = G_{\text{LSDA}}^{-1}(\vec{k}, E) - \Sigma(E)$. A Brillouin zone integration of $G(\vec{k})$ gives finally the LSDA+DMFT one-electron Green's function $G(E)$ that enters the DMFT-problem. In the next step the DMFT-problem is solved using $G(E)$ as an input and giving a new self-energy $\Sigma(E)$. The last steps are repeated until self-consistency is reached with respect to the one-electron Green's function $G(E)$ and the self-energy Σ (DMFT self-consistency). This sequence of steps is sketched in Fig. 1. In many cases the calculations are stopped here, i.e. after a one-shot

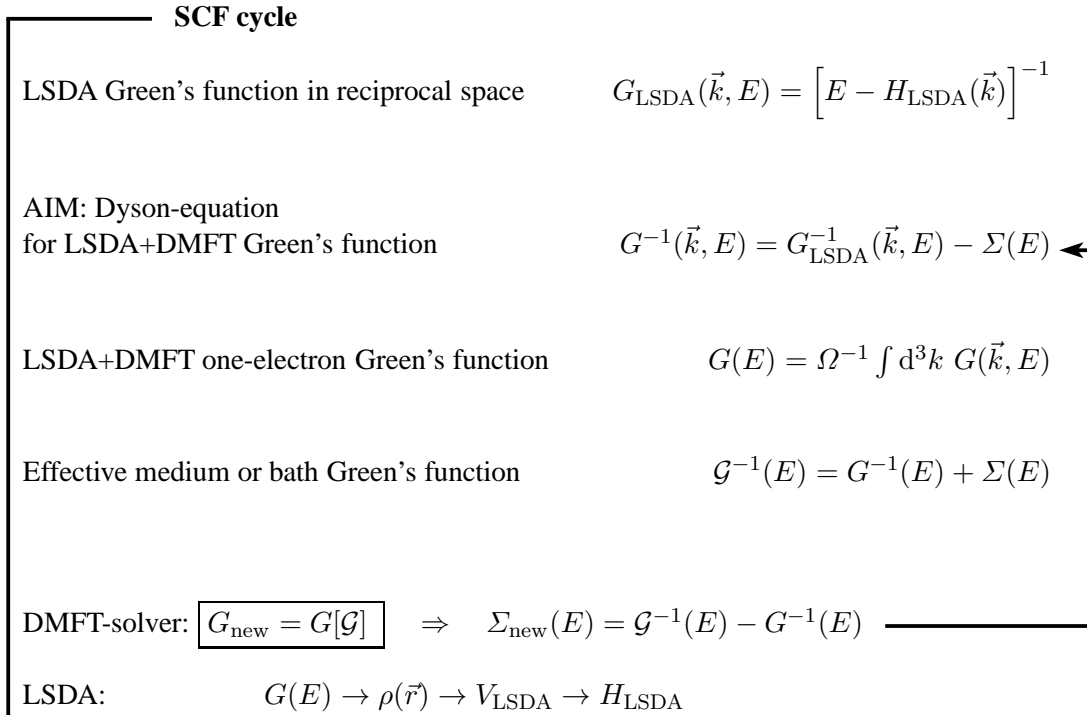


Fig. 1: Scheme for the implementation of the LSDA+DMFT using a standard \vec{k} -space band structure method including the SCF-cycle for the electronic charge $\rho(\vec{r})$ as an outer loop.

inclusion of the DMFT on the basis of a LSDA-calculation. Otherwise the LSDA+DMFT is

performed in a charge self-consistent way by constructing the new charge density from the one-electron Green's function to be used in a standard SCF cycle. To underline the importance of complete LSDA+DMFT self-consistency one should mention that the first successful attempt to combine the DMFT with LSDA charge self-consistency gave an important insight into the long-standing problem of phase diagram and localization in f-electron systems [8, 9] and has been used also to describe correlation effects in half-metallic ferromagnetic materials like NiMnSb [10].

Due to the prominent role of the one-electron Green's function it seems to be highly attractive to implement the LSDA+DMFT scheme using a band structure method that delivers the Green's function directly. This was first done using the LMTO-Green's function method in a non-charge self-consistent way [11]. Subsequently, the so-called EMTO (exact muffin-tin orbital) method was applied in a charge-self-consistent way as a computational framework for the band structure part [12]. In the EMTO approach [13, 14] the one-electron effective potential is represented by the optimized overlapping muffin-tin potential which is considered as the best possible spherical approximation to the full-one electron potential. In essence the one-electron Green's function is evaluated on a complex contour similarly to the screened KKR (Korringa Kohn Rostoker) technique, from which it was derived. In the iteration procedure the LSDA+DMFT Green's function is used to calculate the charge and spin densities. Finally, for the charge self-consistent calculation one constructs the new LSDA effective potential from the spin and charge densities [15], using the Poisson equation in the spherical cell approximation [16]. However, the EMTO-based LSDA+DMFT still follows essentially the scheme sketched above that deals with the DMFT self-energy Σ after having solved the LSDA electronic structure problem. In contrast to this, the KKR implementation [17] follows a natural development in which the self-energy is added directly to the coupled radial differential equations which determine the electronic wave function within a potential well and this way the single-site t -matrix. Because this way also the scattering path operator of multiple scattering theory used to set up the electronic Green's function is determined unambiguously, no further approximations are needed to achieve charge self-consistency.

Representing the electronic structure in terms of the Green's function from the very beginning provides many other advantages: In particular the use of the Dyson equation allows to deal with quite complex systems by connecting the Green's function of a perturbed system with the Green's function of a suitable complementary unperturbed reference system. This gives in particular access to systems without Bloch symmetry [18]. Furthermore, the use of the Coherent Potential Approximation (CPA) alloy theory [19] in combination with Green's function based electronic structure methods allows to deal with substitutional disorder including both diluted impurities and concentrated alloys [20]. The physical condition corresponding to the CPA is simply that a single alloy component embedded in the effective CPA medium should produce no change on the average. A similar philosophy is applied also when dealing with many-body problems for crystals in the framework of the DMFT [21]. Accordingly, the alloy CPA and DMFT can be combined without any conceptual problems [11]. Finally, it should be mentioned

that spectroscopic properties of solids are investigated in a most flexible and powerful way using the Green's function to represent the electronic structure that is probed [22–24]. Thus, an implementation of the LSDA+DMFT using a Green's function based band structure method allows to combine an improved treatment of correlation effects via the DMFT for a wide variety of complex systems with a reliable description of their spectroscopic properties. This is a very attractive feature as it allows on the one-hand side a reliable and detailed interpretation of experimental results and on the other hand to monitor the achievements made by inclusion of correlation effects via the DMFT.

In fact the KKR method of Korringa [25], Kohn and Rostoker [26] is the first band structure method formulated in terms of Green's functions. The obvious advantage of the KKR method lies in the transparent multiple scattering formalism which allows to express the Green's function directly in terms of single-site scattering and geometrical or structural quantities. Thus it seems to be rather natural to combine the DMFT and KKR methods to arrive at a very reliable and flexible band structure scheme that includes correlation effects beyond the standard local density (LSDA) or generalized (GGA) approximations.

2 The LSDA+DMFT scheme

In order to account within LSDA-band structure calculations for correlations an improved hybrid Hamiltonian was proposed by Anisimov et al. [27, 28]. In its most general form such a Hamiltonian is written as

$$H = H_{\text{LSDA}} + H_U - H_{\text{DC}} , \quad (1)$$

where H_{LSDA} stands for the ordinary LSDA Hamiltonian, H_U describes the effective electron-electron interaction and the one-particle Hamiltonian H_{DC} serves to eliminate double counting of the interactions already accounted for by H_{LSDA} .

Using second quantization a rather general expression for H_U is given by:

$$H_U = \frac{1}{2} \sum_{n,ijkl} U_{ijkl}^n \hat{c}_{ni}^\dagger \hat{c}_{nj}^\dagger \hat{c}_{nk} \hat{c}_{nl} , \quad (2)$$

where n runs over all the sites at \vec{R}_n of the crystal and the creation (\hat{c}^\dagger) and annihilation (\hat{c}) operators are defined with respect to some subset of localized orbitals $\phi_i(\vec{r} - \vec{R}_n)$. For the applications presented below these will be the d-orbitals of the transition metals considered. The constants U_{ijkl}^n are matrix elements of the screened Coulomb interaction $v(\vec{r} - \vec{r}')$:

$$U_{ijkl}^n = \int d^3r \int d^3r' \phi_i^\dagger(\vec{r} - \vec{R}_n) \phi_j^\dagger(\vec{r}' - \vec{R}_n) v(\vec{r} - \vec{r}') \phi_k(\vec{r}' - \vec{R}_n) \phi_l(\vec{r} - \vec{R}_n) . \quad (3)$$

The resulting many-particle Hamiltonian can not be diagonalized exactly, thus various methods were developed in the past to find an approximate solution [21]. Among them one of the most promising approaches is to solve Eq. (1) within DMFT that was developed originally to deal with the Hubbard model.

The main idea of the DMFT is to map a periodic many-body problem onto an effective single-impurity problem that has to be solved self-consistently. For this purpose one describes the electronic properties of the system in terms of the one-electron Green's function $\hat{G}(E)$, being the solution of the equation:

$$[E - \hat{H} - \hat{\Sigma}(E)]\hat{G} = \hat{1}, \quad (4)$$

where E is the complex energy and the effective self-energy operator $\hat{\Sigma}(E)$ is assumed to be a single-site quantity for site n :

$$\hat{\Sigma}(E) = \sum_{ij} |\phi_{ni}\rangle \Sigma_{ij}(E) \langle \phi_{nj}|. \quad (5)$$

Fig. 2 shows as an example the spin-dependent self-energy matrix $\Sigma_{ij}(E)$ for ferromagnetic Ni that occurs within Eq. (5). Within DMFT, the self-energy matrix $\Sigma_{ij}(E)$ is a solution of the

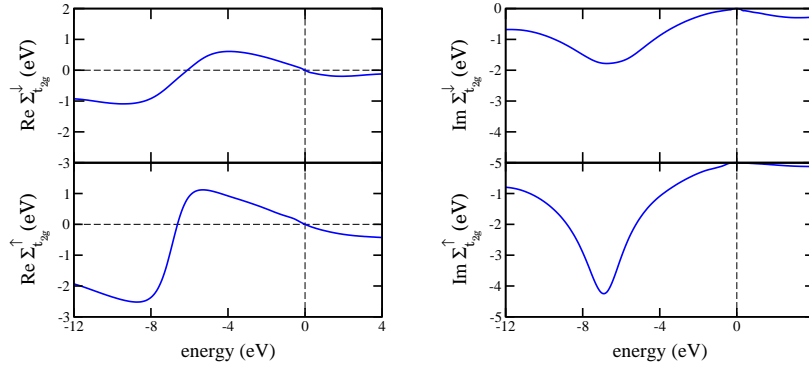


Fig. 2: The spin-dependent self-energy matrix $\Sigma_{ij}(E)$ for ferromagnetic Ni calculated using a DMFT-solver based on the FLEX scheme. Only results for the d -orbitals with t_{2g} -symmetry are shown [17].

many-body problem of an impurity placed in an effective medium. This medium is described by the so called *bath* Green's function matrix \mathcal{G} defined as:

$$\mathcal{G}_{ij}^{-1}(E) = G_{ij}^{-1}(E) + \Sigma_{ij}(E), \quad (6)$$

where the one-electron Green's function matrix $G_{ij}(E)$ is calculated as a projection of $\hat{G}(E)$ onto the impurity site:

$$G_{ij}(E) = \langle \phi_{ni} | \hat{G}(E) | \phi_{nj} \rangle. \quad (7)$$

As the self-energy $\Sigma_{ij}(E)$ depends on the *bath* Green's function $\mathcal{G}_{ij}(E)$ the DMFT equations have to be solved self-consistently. Accordingly, from a technical point of view the problem can be split into two parts. One is dealing with the solution of Eq. (4) and the second one is the effective many-body problem to find the self-energy $\Sigma_{ij}(E)$. The second part can be dealt with in principle by any DMFT-solver as they are presented in the other lectures. To have a reasonable balance between accuracy and computing time within charge self-consistent calculations, one may use for example the perturbative SPTF (spin-polarized T -matrix + FLEX) scheme [29,30].

3 Combination of the LSDA+DMFT with the KKR method

3.1 General idea of the KKR method

The KKR method in its original version [25, 26] was also a \vec{k} -space band structure method that calculates Bloch wave functions $\Psi_{n\vec{k}}(\vec{r})$ together with the associated eigen energies $E_{n\vec{k}}$ by solving the Lippmann-Schwinger equation complementary to the Schrödinger equation for a periodic potential $V(\vec{r})$. The free electron gas was used as the reference system with its corresponding one-electron Green's function $G^0(\vec{r}, \vec{r}', E)$ in real space. This scheme was later extended by various authors [31] to determine the one-electron Green's function $G(\vec{r}, \vec{r}', E)$ of a solid on the basis of the Dyson equation that again uses the free electron gas as the reference system. Obviously, this scheme can be used without major modifications for an implementation of the LSDA+DMFT leading to the following Dyson equation:

$$G(\vec{r}, \vec{r}', E) = G^0(\vec{r}, \vec{r}', E) + \int d^3r'' \int d^3r''' G^0(\vec{r}, \vec{r}'', E) \left[V_{\text{LSDA}}(\vec{r}'') \delta(\vec{r}'' - \vec{r}''') + \Sigma(\vec{r}'', \vec{r}''', E) \right] G(\vec{r}''', \vec{r}', E). \quad (8)$$

Here $V_{\text{LSDA}}(\vec{r})$ is the local, real and energy-independent LSDA-based potential while the self-energy $\Sigma(\vec{r}, \vec{r}', E)$ is non-local, complex and energy-dependent. In the following it is assumed that the self-energy is site-diagonal, i.e. $\Sigma(\vec{r}, \vec{r}', E)$ is non-zero only if \vec{r} and \vec{r}' are within the same atomic cell n . This is well justified in most cases and allows to use the standard KKR approach to deal with Eq. (8). It should be mentioned, however, that taking $\Sigma(\vec{r}, \vec{r}', E)$ to be site-diagonal is not a necessary requirement for a KKR-based implementation. This implies in particular that extensions necessary to deal with a cluster formulation of the DMFT or a \vec{k} -dependent self-energy are possible.

Within the KKR method the Dyson equation (8) is solved by application of multiple scattering theory. This formalism splits the problem into two parts. In a first step the LSDA- or here the LSDA+DMFT-based Schrödinger-like equation is solved for each inequivalent atomic site n in the system (single-site problem). This implies that for a given energy E the corresponding wave functions $\Psi(\vec{r}, E)$ are calculated that in turn can be used to set up the single-site t -matrix $t_{LL'}^n(E)$ in an angular momentum representation ($L = (l, m_l)$). In a second step, the multiple scattering problem for the solid is solved assuming free-electron like propagation of the electrons between the atomic sites n . The corresponding Green's function of the free electron gas is represented by the so-called structure constants $G_{LL'}^{0nn'}(E)$ that contain only structural information on the system. For periodic solids the multiple scattering problem is solved by a Fourier transformation and $G_{LL'}^0(\vec{k}, E)$ depends on the wave vector \vec{k} accordingly. For finite systems, the multiple scattering problem can also be solved directly in real space [18]. Within the KKR-formalism, the solution of the multiple scattering problem is represented by the so-called scattering path operator $\tau_{LL'}^{nn'}(E)$ that describes the transfer of an electronic wave with character L' at site n'

to a wave with character L at the site n with all possible scattering events in between self-consistently taken into account. With the scattering path operator $\tau_{LL'}^{nn'}(E)$ available the Green's function $G(\vec{r}, \vec{r}', E)$ of the system can be set up straight forwardly. The sequence of the various steps of a KKR-implementation are shown in Fig. 3.

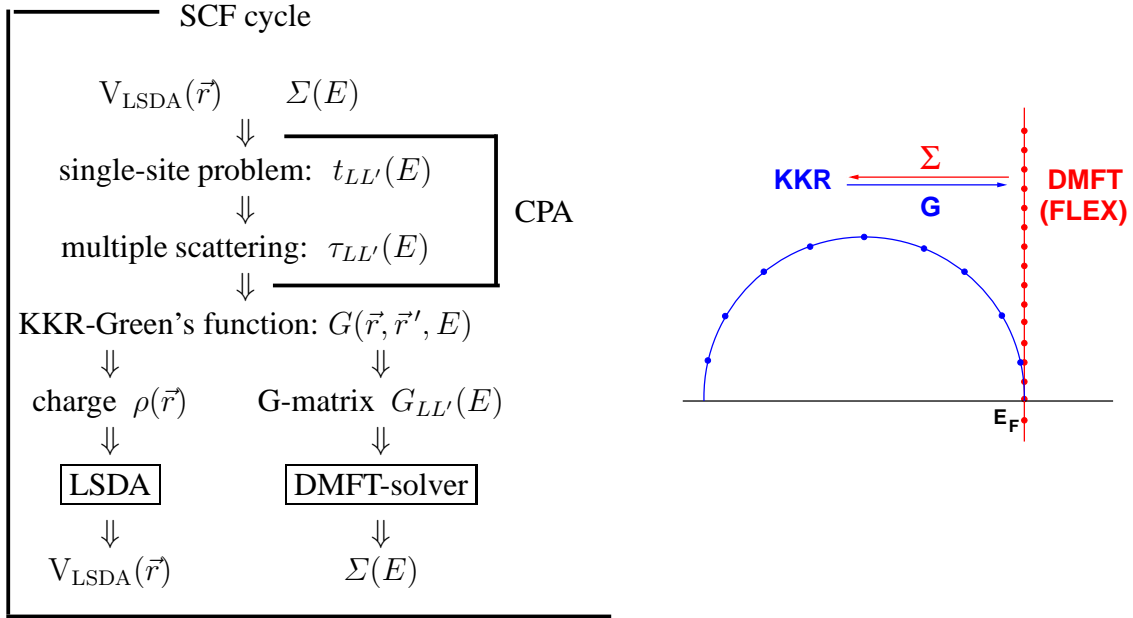


Fig. 3: LSDA+DMFT scheme as implemented on the basis of the KKR formalism. The right part of the figure shows the various energy contours used in case of using a FLEX DMFT-solver (see text).

This brief description of the KKR-formalism makes clear that going from the LSDA to the LSDA+DMFT affects only the single-site but not the multiple scattering problem. This also implies that the DMFT self-energy Σ already enters the scheme when calculating the basis functions. As can be seen from the Dyson equation (8) the LSDA- and DMFT-parts of the electronic structure problem of the solid are accordingly dealt with on the same level.

3.2 Solution of the single-site problem

The solution of the single-site problem including the DMFT self-energy can be worked out following the full-potential description of the KKR formalism [32]. In terms of the wave functions the single-site quasi particle equation to be solved for each spin channel σ reads

$$[-\vec{\nabla}^2 + V^\sigma(r) - E]\Psi(\vec{r}, E) + \int \Sigma^\sigma(\vec{r}, \vec{r}', E) \Psi(\vec{r}', E) d^3r' = 0. \quad (9)$$

In the following the spin index σ is omitted for the moment keeping in mind that for a spin-polarized system described in a non-relativistic way, one has to solve Eq. (9) for each spin channel (see for example Fig. 2) independently. For a particular solution $\Psi_\nu(\vec{r}, E)$ labeled by

the index ν one can start from the ansatz:

$$\Psi_\nu(\vec{r}, E) = \sum_L \Psi_{L\nu}(\vec{r}, E), \quad (10)$$

where the partial waves $\Psi_{L\nu}(\vec{r}, E)$ are chosen to have the same form as the linearly independent solutions for the spherically symmetric potential:

$$\Psi_{L\nu}(\vec{r}, E) = \Psi_{L\nu}(r, E)Y_L(\hat{r}), \quad (11)$$

with $L = (l, m_l)$ standing for the angular momentum and magnetic quantum numbers and $Y_L(\hat{r})$ are spherical harmonics. Inserting the ansatz (10) into the single-site equation (9) and integrating over angle variables leads to the following set of the coupled radial integro-differential equations:

$$\left[\frac{d^2}{dr^2} - \frac{l(l+1)}{r^2} - V(r) + E \right] \Psi_{L\nu}(r, E) = \int d^3r' \Sigma(\vec{r}, \vec{r}', E) \Psi_{L\nu}(\vec{r}', E), \quad (12)$$

where the basis functions $\phi_L(\vec{r}) = \phi_l(r)Y_L(\hat{r})$ will be normalized and suitably chosen d-like wave functions when dealing with transition metals.

For a general non-diagonal self-energy a radial equation similar to Eq. (12) has to be solved to get the so-called left-hand solutions. This implies one has to consider the two adjoint equations [33]:

$$(\hat{H} + \hat{\Sigma} - E)|\Psi\rangle = 0 \quad (13)$$

$$\langle \Psi^\times | (\hat{H} + \hat{\Sigma} - E) = 0, \quad (14)$$

where the superscript \times is used to distinguish the left hand solution $|\Psi^\times\rangle$ from the standard right hand solution $|\Psi\rangle$.

In principle these equations can be solved by summing a corresponding Born series. So far, however, the equations have been simplified taking advantage of the following special representation for the self-energy:

$$\begin{aligned} \int d^3r' \Sigma(\vec{r}, \vec{r}', E) \Psi_{L\nu}(\vec{r}', E) &= \sum_L \int d^3r' \Sigma_{L'L}(E) \phi_{L'}^\dagger(\vec{r}) \phi_L(\vec{r}') \Psi_{L\nu}(\vec{r}', E) \\ &\approx \sum_L \Sigma_{L'L}(E) \Psi_{L\nu}(\vec{r}, E). \end{aligned} \quad (15)$$

This way the integro-differential equation Eq. (12) becomes a pure differential equation:

$$\left[\frac{d^2}{dr^2} - \frac{l(l+1)}{r^2} - V(r) + E \right] \Psi_{L\nu}(r, E) = \sum_{L'} \Sigma_{LL'}(E) \Psi_{L'\nu}(r, E). \quad (16)$$

After having solved the set of coupled equations for the wave functions one gets the corresponding single-site t -matrix by introducing the auxiliary matrices a and b [20]:

$$\begin{aligned} a_{L\nu}(E) &= -ipr^2 [h_L^-(pr), \Psi_L^\nu(r)]_r \\ b_{L\nu}(E) &= -ipr^2 [h_L^+(pr), \Psi_L^\nu(r)]_r. \end{aligned} \quad (17)$$

Here $p = \sqrt{E}$ is the momentum, $h_L^\pm(pr)$ are Hankel functions of the first and second kind and $[\dots]_r$ denotes the Wronskian. Evaluating the Wronskians at Wigner-Seitz radii r_{WS} that defines the range of the potential associated with site n one finally has (in matrix notation) [20, 34]:

$$\underline{t}(E) = \frac{i}{2p} [\underline{a}(E) - \underline{b}(E)] \underline{b}^{-1}(E). \quad (18)$$

The regular wave functions Z used to set up the electronic Green's function within the KKR-formalism [35] are obtained by a superposition of the wave functions Ψ_ν according to the boundary conditions at $r = r_{\text{WS}}$:

$$Z_L(\vec{r}, E) = \sum_\nu C_L^\nu \Psi_\nu(\vec{r}) \xrightarrow{r=r_{\text{WS}}} \sum_{L'} j_{L'}(\vec{r}, E) t(E)_{L'L}^{-1} - ip h_L^+(\vec{r}, E). \quad (19)$$

The irregular solutions J_L needed in addition are fixed by the boundary condition

$$J_L(\vec{r}, E) \xrightarrow{r=r_{\text{WS}}} j_L(\vec{r}, E) \quad (20)$$

with the functions j_L being the spherical Bessel functions.

3.3 Multiple scattering and Green's function

Instead of considering the single-site problem for the wave function $\psi_n(\vec{r}, E)$ in terms of the Lippmann-Schwinger equation (9) one can alternatively deal with the single-site Green's function $G^n(\vec{r}, \vec{r}', E)$ in terms of the corresponding Dyson equation with the potential and self-energy restricted to site n (see Eq. (8)). Ignoring the self-energy Σ for the moment and representing all quantities by their associated operators this may be written as [36]:

$$\hat{G}^n(E) = \hat{G}^0(E) + \hat{G}^0(E) \hat{V}^n \hat{G}^n(E). \quad (21)$$

The single-site t -matrix $t_l^n(E)$ used above to account for the matching of the wave functions $\Psi_{L\nu}^n(r, E)$ at the radius $r = r_{\text{WS}}$ is related to a corresponding single-site t -matrix operator $\hat{t}^n(E)$ that allows to write an explicit expression for $\hat{G}^n(E)$ [36]:

$$\hat{G}^n(E) = \hat{G}^0(E) + \hat{G}^0(E) \hat{t}^n(E) \hat{G}^0(E). \quad (22)$$

Completely analogous equations emerge when one is dealing with an array of scatterers [36]:

$$\hat{G} = \hat{G}^0(E) + \hat{G}^0(E) \hat{V} \hat{G}(E) \quad (23)$$

$$= \hat{G}^0(E) + \hat{G}^0(E) \hat{T}(E) \hat{G}^0(E), \quad (24)$$

where all quantities refer now to the total system. Decomposing all scattering processes into sequences of single-site scattering events, represented by single-site t -matrix operators $\hat{t}^n(E)$, and free propagation according to $\hat{G}^0(E)$ in-between, one may decompose the total t -matrix operator $\hat{T}(E)$ accordingly. This central idea of multiple scattering theory is illustrated by Fig. 4. Using the scattering path operator $\hat{\tau}^{nn'}(E)$ introduced by Gyorffy and Stott [37] one

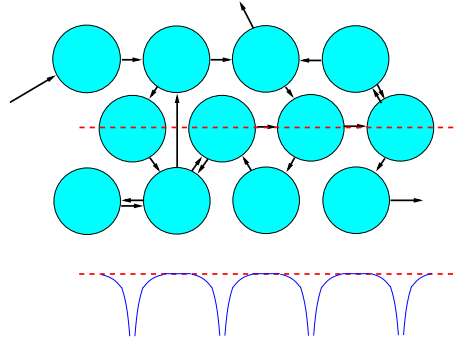


Fig. 4: Central idea of multiple scattering theory: decomposition of electronic motion into scattering at atomic sites and free-electron like propagation in between. The bottom of the figure gives a sketch for the potential along the dashed line.

gets [36, 38]:

$$\hat{T}(E) = \sum_{nn'} \hat{\tau}^{nn'}(E), \quad (25)$$

where $\hat{\tau}^{nn'}(E)$ is defined to transfer an electronic wave incoming at site n' into a wave outgoing from site n with all possible scattering events that may take place in-between in a self-consistent way. Adopting an angular momentum representation as introduced in the previous section, this requirement implies for the corresponding matrix the following equation of motion

$$\underline{\underline{\tau}}^{nn'}(E) = \underline{t}^n(E) \delta_{nn'} + \underline{t}^n(E) \sum_{k \neq n} \underline{\underline{G}}^{0nk}(E) \underline{\underline{\tau}}^{kn'}(E), \quad (26)$$

where the underline indicates matrices with respect to L with $(\underline{\underline{\tau}}^{nn'})_{LL'} = \tau_{LL'}^{nn'}$ etc. For a finite system this equation is solved straight forwardly by a matrix inversion [39]:

$$\underline{\underline{\tau}}(E) = [\underline{t}(E)^{-1} - \underline{\underline{G}}^0(E)]^{-1}, \quad (27)$$

where $\underline{\underline{M}} = [\underline{t}^{-1} - \underline{\underline{G}}^0]$ is the so-called real-space KKR-matrix. The double underline indicates matrices with respect to the angular momentum and site indices, L and n , respectively, with $[\underline{\underline{\tau}}]_{nn'} = \underline{\underline{\tau}}^{nn'}$, $[\underline{\underline{G}}^0]_{nn'} = \underline{\underline{G}}^{0nn'}$ and $[\underline{t}]_{nn'} = \underline{t}^n \delta_{nn'}$. The energy argument has been dropped here. Dealing with a three-dimensional periodic system Eq. (26) can also be solved exactly by Fourier transformation leading to [36, 38]:

$$\underline{\underline{\tau}}^{nn'}(E) = \frac{1}{\Omega_{\text{BZ}}} \int_{\Omega_{\text{BZ}}} d^3k [\underline{t}(E)^{-1} - \underline{\underline{G}}^0(\vec{k}, E)]^{-1} e^{i\vec{k} \cdot (\vec{R}_n - \vec{R}_{n'})}, \quad (28)$$

with the (reciprocal space) structure constants matrix $\underline{\underline{G}}^0(\vec{k}, E)$ being the Fourier transformed of the real-space structure constants matrix $\underline{\underline{G}}^0(E)$ that represent the free-electron like propagation.

Having constructed a set of regular (Z) and irregular (J) solutions of the single-site problem together with the t -matrix and solved the multiple-scattering problem in Eq. (26) subsequently,

the corresponding Green's function is obtained from the expression [35]:

$$G(\vec{r}_n + \vec{R}_n, \vec{r}'_m + \vec{R}_m, E) = \sum_{LL'} Z_L(\vec{r}_n, E) \tau_{LL'}^{nm}(E) Z_{L'}^\times(\vec{r}'_m, E) - \delta_{nm} \sum_L [Z_L(\vec{r}_n, E) J_L^\times(\vec{r}'_n, E) \Theta(r'_n - r_n) + J_L(\vec{r}_n, E) Z_L^\times(\vec{r}'_n, E) \Theta(r_n - r'_n)]. \quad (29)$$

Given the local nature of the many-body solver used within the DMFT approach, the KKR Green's function in Eq. (29) has to be projected accordingly to the matrix $G_{LL'}^{nm}$ (see Eq. (7)). The projection is performed through the following integration:

$$G_{LL'}^{nm}(E) = \sum_{L_1 L_2} \left(\int d^3 r \phi_L^\dagger(\vec{r}) Z_{L_1}(\vec{r}, E) \right) \tau_{L_1 L_2}^{nm}(E) \left(\int d^3 r' Z_{L_2}^\times(\vec{r}', E) \phi_{L'}(\vec{r}') \right) - \delta_{nm} \sum_{L_1} \left[\int d^3 r' \left(\int_0^{r'} d^3 r \phi_L^\dagger(\vec{r}) Z_{L_1}(\vec{r}, E) \right) J_{L_1}^\times(\vec{r}', E) \phi_{L'}(\vec{r}') + \int d^3 r' \left(\int_{r'}^{r_{\text{WS}}} d^3 r \phi_L^\dagger(\vec{r}) J_{L_1}(\vec{r}, E) \right) Z_{L_1}^\times(\vec{r}', E) \phi_{L'}(\vec{r}') \right]. \quad (30)$$

The Green's function matrix $G_{LL'}^{nm}(E)$ (actually $G_{LL'}^{\sigma nm}(E)$ for both spin channels) represents the input into the solution of the effective impurity problem. As the DMFT-approach concentrates on the correlation among electrons of the same angular momentum l only the $l-l$ -subblock of this matrix will be used in the following. For the transition metal systems dealt here this implies that only the d-d-subblock is considered with $\phi_L(\vec{r})$ being appropriate reference wave functions with $l = 2$.

3.4 Electronic structure and relativistic extension

With the Green's function available the most prominent electronic properties of a solid may be expressed and calculated in a straight forward way. For example the density of states (DOS) $n(E)$, the electron density $\rho(\vec{r})$ and the expectation value of a one-electron operator \mathcal{A} may be obtained from the expressions

$$n(E) = -\frac{1}{\pi} \Im \int_V d^3 r G^+(\vec{r}, \vec{r}, E) \quad (31)$$

$$\rho(\vec{r}) = -\frac{1}{\pi} \Im \int^{E_F} dE G^+(\vec{r}, \vec{r}, E) \quad (32)$$

$$\langle \mathcal{A} \rangle = -\frac{1}{\pi} \Im \int^{E_F} dE \int_V d^3 r \mathcal{A} G^+(\vec{r}, \vec{r}, E), \quad (33)$$

where V is the volume associated with the atom at site n . As an example Fig. 5 shows the spin-resolved density of states $n^{\uparrow(\downarrow)}(E)$ calculated for fcc-Ni on the basis of the LSDA and LSDA+DMFT. Obviously, one can note three prominent changes in the DOS when going

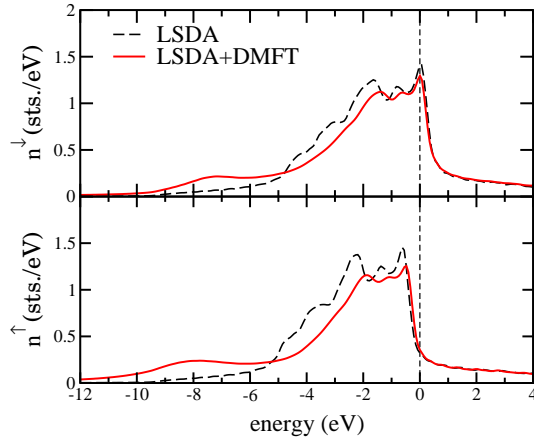


Fig. 5: Spin-resolved density of states $n^{\uparrow(\downarrow)}(E)$ for fcc-Ni calculated via the LSDA and LSDA+DMFT. The upper (lower) panel shows the DOS for the minority (majority) spin [17].

from the LSDA to the LSDA+DMFT: The DOS curves show much less structures and appear smeared-out, the band width shrinks and a satellite peak at around 7 - 8 eV binding energy occurs. All these features can be traced back to the characteristics of the self-energy Σ shown in Fig. 2. In particular one can ascribe the smearing out of the DOS to the imaginary part of Σ while the shrinking of the band width and the satellite is associated with its real part.

To allow for a more detailed discussion of the electronic structure one may introduce the Bloch spectral function $A_B(\vec{k}, E)$ by a Fourier transformation of the real space Green's function $G(\vec{r}, \vec{r}', E)$

$$A_B(\vec{k}, E) = -\frac{1}{\pi N} \sum_{n,m} e^{i\vec{k}(\vec{R}_n - \vec{R}_m)} \Im \int_{\Omega} d^3r \left\langle G(\vec{r} + \vec{R}_n, \vec{r} + \vec{R}_m, E) \right\rangle. \quad (34)$$

For an ordered system dealt with on the basis of the LSDA this leads simply to a set of δ -functions representing the conventional dispersion relation $E_{j\vec{k}}$. Calculating $A_B(\vec{k}, E)$, however, on the basis of the LSDA+DMFT the imaginary part of the self-energy leads to a corresponding broadening, that reflects the finite life time of electronic states with fixed \vec{k} -vector. Corresponding results are shown in Fig. 6 for fcc-Ni. As one notes, the shrinking of the band width and the broadening of the bands take place primarily in the d-band regime.

For the sake of clarity the KKR formulation has been introduced above on a non-relativistic level. For many situations, however, relativistic corrections to the Schrödinger equation play a central role. In fact the interplay of relativistic effects and correlations have been investigated already by various authors on the basis of the LSDA+DMFT [41, 8, 42, 29, 43–45]. Obviously, the most reliable treatment is achieved on the basis of the fully relativistic Dirac formalism. Ignoring the self-energy for the moment this takes the form [46]:

$$\left[\frac{\hbar}{i} c \vec{\alpha} \cdot \vec{\nabla} + \beta m c^2 + \bar{V}(\vec{r}) + \underbrace{\beta \vec{\sigma} \cdot \vec{B}_{\text{eff}}(\vec{r})}_{V_{\text{spin}}(\vec{r})} \right] \Psi(\vec{r}, E) = E \Psi(\vec{r}, E), \quad (35)$$

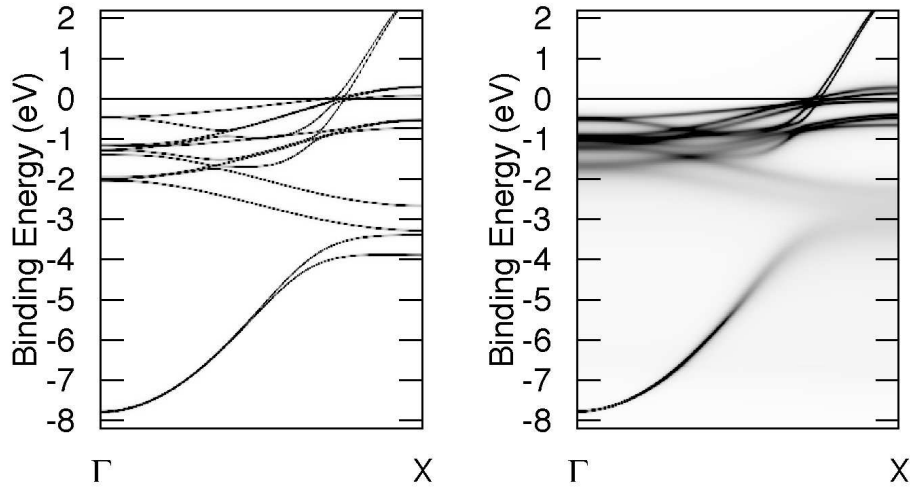


Fig. 6: Band structure of ferromagnetic Ni for \vec{k} along $\overline{\Gamma X}$. Left: LSDA-based dispersion relation $E_{j\vec{k}}$, right: LSDA+DMFT-based Bloch spectral function $A_B(\vec{k}, E)$ [40].

where α_i and β are the standard 4×4 Dirac matrices [47] and $\bar{V}(\vec{r})$ is the scalar Coulomb and spin-averaged part of the exchange-correlation potential. The spin-dependent part $V_{\text{spin}}(\vec{r})$ of the latter one is represented by an effective magnetic field

$$\vec{B}_{\text{eff}}(\vec{r}) = \frac{\delta E_{\text{xc}}[n, \vec{m}]}{\delta \vec{m}(\vec{r})} \quad (36)$$

that is determined by the spin magnetization $\vec{m}(\vec{r})$. Assuming a collinear spin magnetization within an atomic cell one can always choose \hat{z}' such to have:

$$V_{\text{spin}}(\vec{r}) = \beta \sigma_{z'} B_{\text{eff}}(r). \quad (37)$$

To deal with the KKR single-site problem the ansatz

$$\Psi_{\nu}(\vec{r}, E) = \sum_{\Lambda} \Psi_{\Lambda\nu}(\vec{r}, E) = \sum_{\Lambda} \begin{pmatrix} g_{\Lambda\nu}(r, E) \chi_{\Lambda}(\hat{r}) \\ i f_{\Lambda\nu}(r, E) \chi_{-\Lambda}(\hat{r}) \end{pmatrix} \quad (38)$$

is made for the four-component wave function $\Psi_{\nu}(\vec{r}, E)$ in analogy to Eq. (10). Here the partial waves $\Psi_{\Lambda\nu}(\vec{r}, E)$ are constructed using the radial functions, $g_{\Lambda\nu}(r, E)$ and $f_{\Lambda\nu}(r, E)$, of the large and small, respectively, components together with the spin-angular functions [47]

$$\chi_{\Lambda}(\hat{r}) = \sum_{m_s = \pm 1/2} C(l \frac{1}{2} j; \mu - m_s, m_s) Y_l^{\mu - m_s}(\hat{r}) \chi_{m_s} \quad (39)$$

with the Clebsch-Gordon coefficients $C(l \frac{1}{2} j; \mu - m_s, m_s)$ and the short hand notation $\Lambda = (\kappa, \mu)$ and $-\Lambda = (-\kappa, \mu)$ for the relativistic spin-orbit and magnetic quantum numbers, κ and μ , respectively. Obviously, the spin-angular functions $\chi_{\Lambda}(\hat{r})$ are the relativistic counter part to the spherical harmonics occurring in Eq. (10). In particular they are eigen functions to the square of the total angular momentum operator \vec{j}^2 , its z-component j_z as well as the spin-orbit operator

\hat{K} with the eigen values $j(j+1)$, μ and $-\kappa$, respectively, with the quantum numbers restricted by the relations $j = l \pm \frac{1}{2}$, $\mu = -j \dots + j$, $\kappa = l = j + \frac{1}{2}$ for $j = l - \frac{1}{2}$ and $\kappa = -l - 1 = -j - \frac{1}{2}$ for $j = l + \frac{1}{2}$ [47].

Inserting the ansatz in Eq. (38) into the Dirac equation (35) leads to a coupled set of radial equations for the radial functions $g_{\Lambda\nu}(r, E)$ and $f_{\Lambda\nu}(r, E)$ in analogy to Eq. (16). With an extension of Eqs. (16) through (18) the corresponding single-site t -matrix can finally be evaluated [48]. This scheme implies that the spin-polarization i.e. magnetization of a solid is treated on the same footing as all relativistic effects – in particular the spin-orbit coupling. Complementing the LSDA potential $V(\vec{r})$ with the DMFT self-energy Σ , as done in Eq. (12) leads to a coherent treatment of all relativistic and correlation effects as well as spin-polarization when solving the single-site problem. Treatment of the multiple-scattering problem is done again in full analogy to the non-relativistic formalism described by Eqs. (21) through (28).

The relativistic approach sketched above was formulated in the (κ, μ) -representation. Restricting the expansion of the wave functions $\Psi_{\Lambda\nu}$ in Eq. (38) up to the angular momentum $l_{\max} = 2$ the resulting matrices have the dimension $2(l_{\max} + 1)^2 = 18$. Accounting for correlation effects by means of the DMFT scheme for transition metals implies that the corresponding one-electron Green's function matrix $G_{\Lambda\Lambda'}$ for the d-electrons has the dimension 10. To deal with the DMFT the (l, m_l, m_s) -representation is more suitable implying that one has spin-flip components like $G_{m_l m'_l}^{\uparrow\downarrow}$ for the Green's function matrix because of the inclusion of spin-orbit coupling. As a consequence the DMFT self-energy Σ will also be a 10×10 matrix with spin-flip components. As an example Fig. 12 shows corresponding results for the d-states of Ni in fcc-Ni and for the f-states of U in ferromagnetic US. As one can see, the spin-flip components $\Sigma^{\uparrow\downarrow}$ are quite small

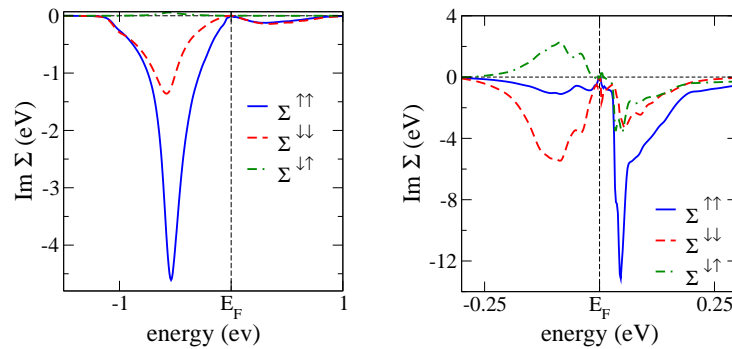


Fig. 7: Spin-dependent self-energy $\Sigma^{m_s m'_s}(E)$ of the d-states of Ni in fcc-Ni (left) and the f-states of U in ferromagnetic US (right) (unpublished).

compared to the spin-diagonal ones $\Sigma^{m_s m_s}$ in the case of Ni. For U on the other hand, they are much larger and comparable to $\Sigma^{m_s m_s}$. This is explained of course by the high atomic number of U and the strong spin-orbit coupling associated with this.

3.5 The self-consistency cycle

A flow chart depicting the self-consistent LSDA+DMFT approach is shown in Fig. 3. The radial equations Eq. (12) provide the set of regular (Z) and (J) irregular solutions of the single-site problem. Together with the t -matrix, the scattering path operator τ and the KKR Green's function is constructed via Eq. (29). To solve the many-body problem the projected impurity Green's function matrix is constructed according to Eq. (30). The LSDA Green's function $G_{LL'}^{nn}(E)$ is calculated on a complex contour (semi circle) which encloses the valence band one-electron energy poles. The Padé analytical continuation is used to map the complex local Green's function $G_{LL'}^{nn}(E)$ on the Matsubara axis which is used when dealing with the many-body problem. For most applications the perturbative SPTF (spin-polarized T -matrix + FLEX) solver of the DMFT problem has been used so far. In fact any DMFT solver could be included which supplies the self-energy $\Sigma(E)$ as a solution of the many-body problem. The Padé analytical continuation is used once more to map back the self-energy from the Matsubara axis to the complex plane, where the new local Green's function is calculated. As was described in the previous sections, the key role is played by the scattering path operator $\tau_{LL'}^{nn}(E)$, which allows to calculate the charge at each SCF iteration and the new potentials that are used to generate the new LSDA Green's function. In practice it turns out that the self-energy converges faster than the charge density. Of course double counting corrections have to be considered explicitly when calculating the total energy. Concerning the self-energy used in the applications presented below the double counting corrections are included when solving the many-body problem (see Ref. [30]).

3.6 Treatment of disordered alloys

Using a local mean-field approximation to treat electron correlations, the corresponding self-energy gets diagonal in the site representation. This allows to use the coherent potential alloy theory (CPA) [19] for the configurational averaging as suggested by Drchal et al. [11]. These authors pointed out that an averaged coherent potential for disordered interacting systems can be constructed using the so-called terminal-point approximation when dealing with disordered alloys, as it was suggested.

Among the electronic structure theories, those based on the multiple scattering formalism are the most suitable to deal with disordered alloys within the CPA. This scheme is considered to be the best theory among the so-called single-site (local) alloy theories that assume complete random disorder and ignore short-range order [20]. Combining the CPA with multiple scattering theory leads to the KKR-CPA scheme, which is applied nowadays extensively for quantitative investigations of the electronic structure and properties of disordered alloys [20,49]. Within the CPA the configurationally averaged properties of a disordered alloy are represented by an auxiliary ordered CPA-medium, which in turn may be described by a corresponding site-diagonal ($n = m$) scattering path operator $\underline{\tau}^{\text{CPA}}$. The corresponding single-site t -matrix $\underline{t}^{\text{CPA}}$ and multiple scattering path operator $\underline{\tau}^{\text{CPA}}$ are determined by the so called CPA-condition:

$$x_A \underline{\tau}^A + x_B \underline{\tau}^B = \underline{\tau}^{\text{CPA}}. \quad (40)$$

Here a binary system A_xB_{1-x} with relative concentrations $x_A = x$ and $x_B = 1 - x$ is considered. The above equation represents the requirement that embedding substitutionally an atom (of type A or B) into the CPA medium should not cause additional scattering as illustrated by Fig. 8. The

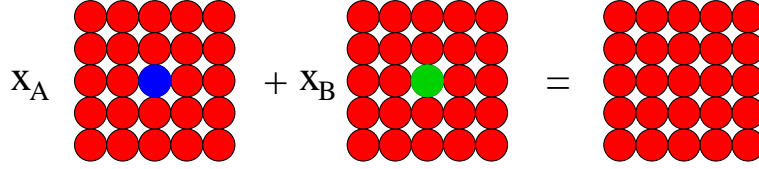


Fig. 8: The major ideas of the CPA: The configurational average over all configurations of a disordered alloy A_xB_{1-x} is represented by an auxiliary CPA medium. Embedding of an A or B atom should not give rise to additional scattering with respect to the CPA medium.

scattering properties of an A atom embedded in the CPA medium, are represented by the site-diagonal component-projected scattering path operator $\underline{\mathcal{T}}^A$

$$\underline{\mathcal{T}}^A = \underline{\mathcal{T}}^{\text{CPA}} \left[1 + (\underline{t}_A^{-1} - \underline{t}_{\text{CPA}}^{-1}) \underline{\mathcal{T}}^{\text{CPA}} \right]^{-1}, \quad (41)$$

where \underline{t}_A and $\underline{t}_{\text{CPA}}$ are the single-site matrices of the A component and of the CPA effective medium. A corresponding equation holds also for the B component in the CPA medium. The coupled sets of equations for $\underline{\mathcal{T}}^{\text{CPA}}$ and $\underline{t}^{\text{CPA}}$ have to be solved iteratively within the CPA cycle. It is obvious that the above scheme can straightforwardly be extended to include the many-body correlation effects for disordered alloys. As was pointed above, within the KKR+DMFT approach the local multi-orbital and energy dependent self-energy $\Sigma_{A(B)}$ is directly included in the single-site matrices \underline{t}^A and \underline{t}^B , respectively. Having solved the CPA equations self-consistently, one has to project the CPA Green's function onto the components A and B by using Eqs. (30) and (41). In Eq. (30) the multiple scattering path operator $\tau_{LL'}^\sigma$ has to be replaced by the component-projected scattering path operator $\tau_{LL'}^{A,\sigma}$ of an A-atom in a CPA medium. The components Green's functions $G_{A(B)}$ are used to construct the corresponding bath Green's functions for which the DMFT self-consistency condition is used according to Eq. (6):

$$\mathcal{G}_{A(B)}^{-1}(E) = G_{A(B)}^{-1}(E) + \Sigma_{A(B)}(E). \quad (42)$$

The many-body solver in turn is used to produce the component specific self-energies $\Sigma_{A(B)}$:

$$\Sigma_{A(B)}(E) = \Sigma_{A(B)}(E) [\mathcal{G}_{A(B)}(E)]. \quad (43)$$

As an example for this Fig. 9 shows the corresponding results of an application to the ferromagnetic alloy system fcc- $\text{Fe}_x\text{Ni}_{1-x}$. As one notes the self-energy for the two alloy components are quite different and show a pronounced concentration dependency. For Fe the results of pure bcc-Fe are given in addition. These demonstrate the strong change of the Fe self-energy that occurs when going from the bcc to the fcc phase with a stronger exchange splitting present for the later one.

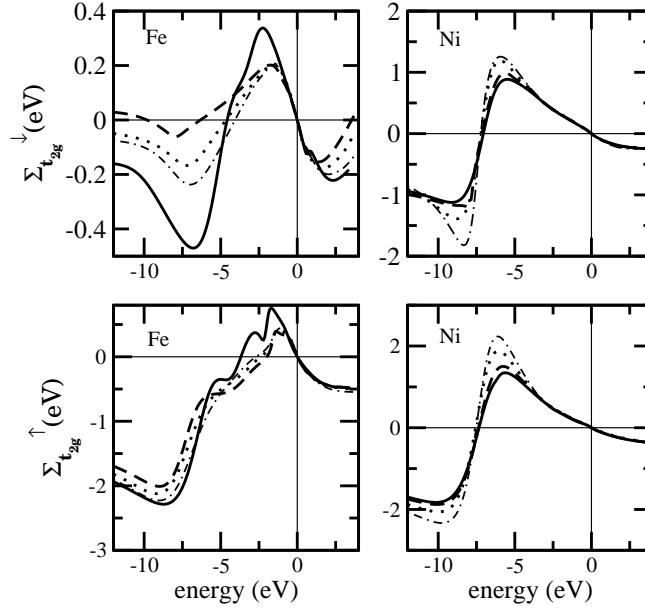


Fig. 9: Left: Concentration dependence of the real part of the spin resolved self-energy for Fe in Fe_xNi_{1-x} . Only results for t_{2g} d-orbitals are shown. Right: Same as in the left panel but for Ni. (dashed: $x=0.1$, dotted: $x=0.4$, dash-dotted: $x=0.75$, full: bcc-Fe) [17].

In the alloy system Fe_xNi_{1-x} both components are influenced by correlation effects in a comparable way. For the alloy system Ni_xPd_{1-x} , on the other hand, correlation effects should play a minor role for Pd due to the broader width of the d-band. In addition Pd as a pure element is non-magnetic and has a much higher atomic number than Ni. As a consequence one can follow simultaneously the variation of correlation effects, exchange splitting and of spin-orbit splitting in this alloy system when the concentration is varied. Fig. 10 shows a sequence of Bloch spectral functions $A_B(\vec{k}, E)$ for Ni_xPd_{1-x} including the pure elements Ni and Pd. As DMFT-corrections are included only for Ni the Bloch spectral function is smeared out only for Ni, while for Pd a conventional dispersion relation $E_{j\vec{k}}$ emerges. For the alloys additional broadening occurs due to the chemical disorder in the system.

4 Applications

4.1 Ground state properties

The KKR-based implementation of the LSDA+DMFT allows the calculation of a large variety of physical properties of the system under investigation. This permits a direct comparison with experimental data and this way to check the theoretical results. Corresponding studies have been made so far in particular for the magnetic moments in various transition metal systems [43, 17, 40, 50]. Using the relativistic KKR formalism the spin and orbital magnetic moments are obtained as expectation values of the operators $\mu_B\beta\Sigma_z$ and $\mu_B\beta l_z$, resp. (see Eq. (33)). A comparison to experiment for the magnetic moments calculated within the LSDA as well as

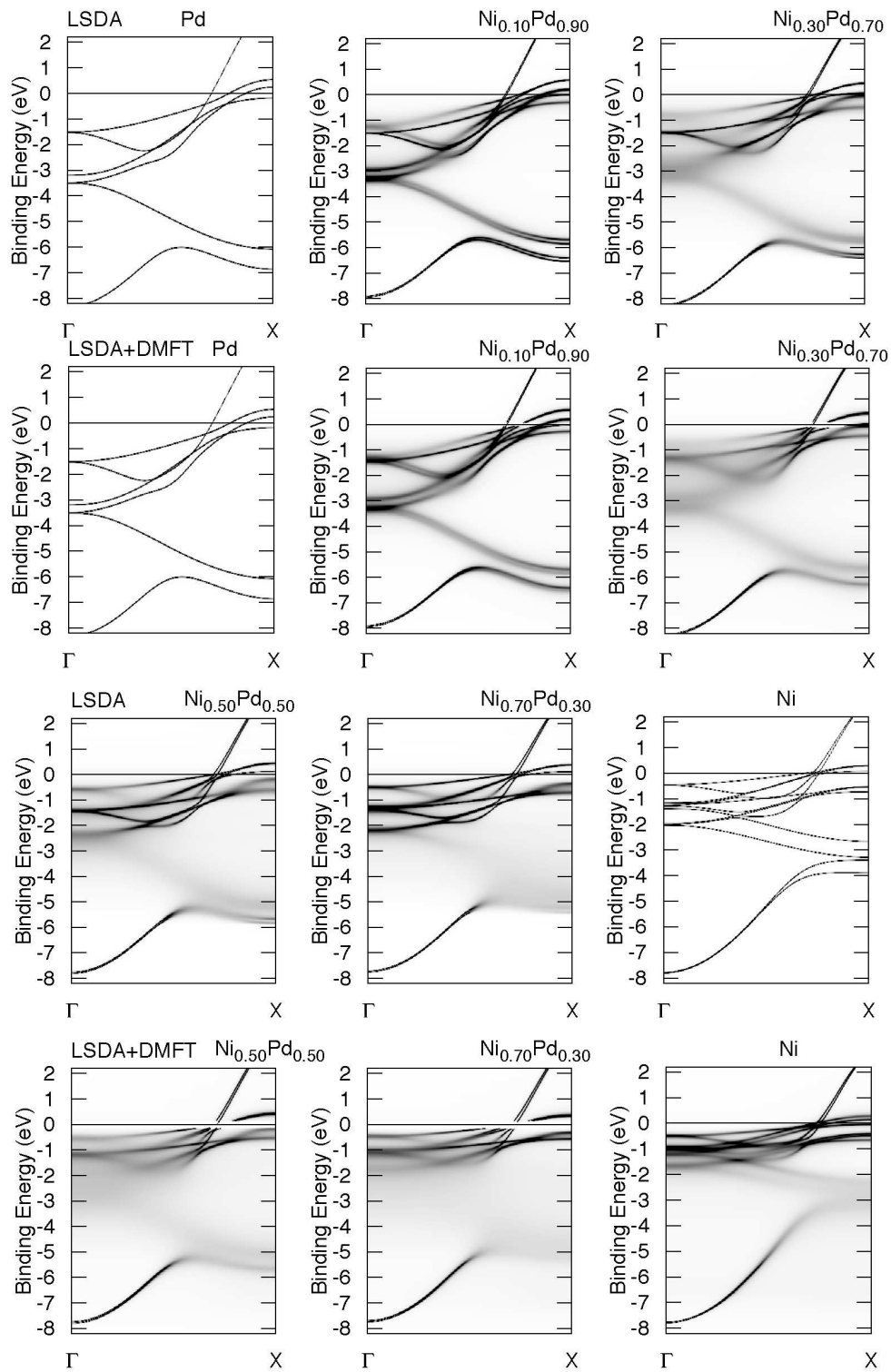


Fig. 10: Spin-polarized Bloch spectral functions for Ni_xPd_{1-x} . Results for LSDA (first and third row) and LSDA+DMFT (second and fourth row) calculations are shown starting from pure Pd (upper panel, left side) to pure Ni (lower panel, right side) [40].

the LSDA+DMFT for bcc Fe, hcp Co and fcc Ni is shown in Fig. 11. The self-energy was parameterized using the values $U = 3$ eV for Co and Ni, and $U = 2$ eV for Fe. As expected, the

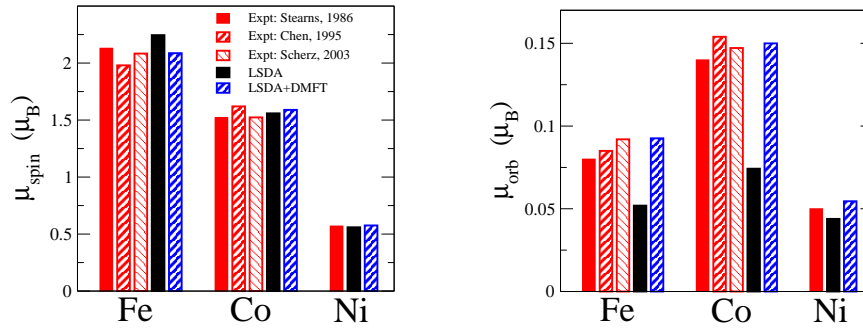


Fig. 11: Spin (left) and orbital (right) magnetic moments in bcc Fe, hcp Co and fcc Ni calculated using LSDA+DMFT (hatched blue bars) compared with plain LSDA calculations (black filled bars) and experimental data (red bars). The corresponding DMFT parameters are $U_{\text{Fe}} = 2$ eV, $U_{\text{Co}} = U_{\text{Ni}} = 3$ eV and $J_{\text{Fe}} = J_{\text{Co}} = J_{\text{Ni}} = 0.9$ eV [43].

LSDA+DMFT approach gives results similar to the orbital polarization (OP) scheme of Brooks [51, 52]: the small orbital splittings imposed by the LSDA+DMFT around the Fermi level have almost no effect on the spin moment, but enhance the orbital moment in an appreciable way.

By construction the dynamical part of the self-energy Σ in the vicinity of the Fermi level behaves like that of a Fermi liquid. Thus it cannot noticeably affect integral quantities as spin and orbital magnetic moments. On the other hand, the applied AMF (around mean field) static double counting which splits the orbitals only slightly at the Fermi level, has no impact on the renormalization of the density of states. As a consequence from Fig. 2, the total DOS curves calculated within LSDA and LSDA+U as well as within LSDA+ Σ (i.e. only the dynamical part of the self energy is used) and LSDA+DMFT are nearly indistinguishable [43]. As the energy shifts of the $(-m, -m)$ and (m, m) matrix elements of the Green's function occur in opposite directions, the total DOS shift for a given spin character appears to be small. As a result, the most affected quantity is the orbital magnetic moment while the change of the spin moment is negligible. At the same time the renormalization of the spectrum is controlled by the dynamical part of the self-energy (see below).

It follows from the various DMFT studies as well as from the DMFT+GW-based calculation [53] that realistic values of U for 3d-TMs are found between 2-3 eV. As it is shown in Fig. 11 this range of U parameters brings both spin and orbital moments into very close agreement with experiment. In the case of Fe the deviation of the orbital moment for U above 2 eV are found to be rather big [43], so that the optimal values of U are confined within 1.5-2 eV. On the other hand, it was already proposed [54] that the local approximation (DMFT) works much better for Ni and Co than for Fe due to relative softness of magnons in the latter case. Recently, the essential non-locality of correlation effects in Fe was also demonstrated experimentally by angle-resolved photo emission [55].

The KKR formalism offers a great flexibility concerning the geometry of the investigated sys-

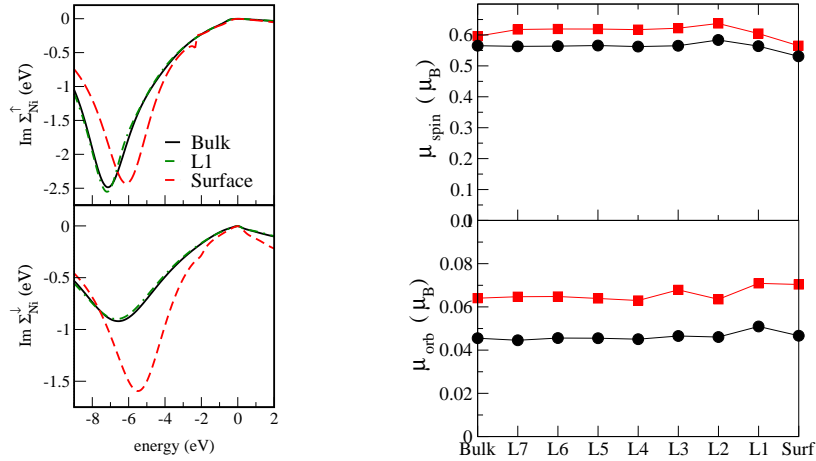


Fig. 12: Left: Spin-dependent self-energy $\Sigma^{\uparrow(\downarrow)}$ for the surface, sub-surface (L1) and bulk-like atomic layers at the (100)-surface of Ni. Right: spin (top) and orbital (bottom) magnetic moments for the individual atomic layers of the (100)-surface of Ni calculated via the LSDA (circles) and LSDA+DMFT (squares) (unpublished).

tem. An example for this is the treatment of surface systems that are usually described in an approximate way by using a film geometry. Within KKR formalism one may consider a number surface near atomic layers on top of a half-infinite bulk-like substrate. As an example Fig. 12 shows some preliminary results for the (100)-surface of Ni for which the 8 top most layers were allowed to relax concerning the charge and potential. As one can see in the left part of the figure the self-energy approaches very rapidly the bulk; already the subsurface layer (L1) has a self-energy very close to that for the bulk. The spin and orbital magnetic moments, however, show an oscillatory behavior very similar to that found within LSDA-calculations.

As described above, a combination of the KKR with the CPA alloy theory allows to deal straight forwardly with disordered alloys. This appealing feature has been exploited so far to study among others the impact of correlations effects on the alloy systems $\text{Fe}_x\text{Ni}_{1-x}$ [17], $\text{Fe}_x\text{Co}_{1-x}$ [43] and $\text{Ni}_x\text{Pd}_{1-x}$ [40]. Corresponding results for the spin and orbital magnetic moments of $\text{Fe}_x\text{Co}_{1-x}$ are shown in Fig. 13. As can be seen from this figure, while the spin magnetic moments for all approaches agree rather well, LSDA+DMFT considerably improves the orbital moments in comparison to plain LSDA calculations in a way similar to the result obtained using the LSDA+OP combined with the CPA [52]. Also in contrast to both the LSDA and LSDA+OP calculations, a better agreement with experimental spin magnetic moments is achieved by LSDA+DMFT within the Fe-rich area of concentrations.

Bulk fcc Ni is a sort of standard test-case for every approach to correlated materials. For this reason it has been chosen for first KKR-based investigations on the influence of correlation effects on the total energy [56]. For this purpose LSDA+DMFT calculations on fcc Ni for various lattice constants starting from $a = 6.2$ a.u. and up to $a = 7.4$ a.u. have been performed. The local DMFT problem was studied for different values of U in the range between 2 and 3 eV, considered acceptable from the results of constrained LSDA calculations [58, 59] and previous

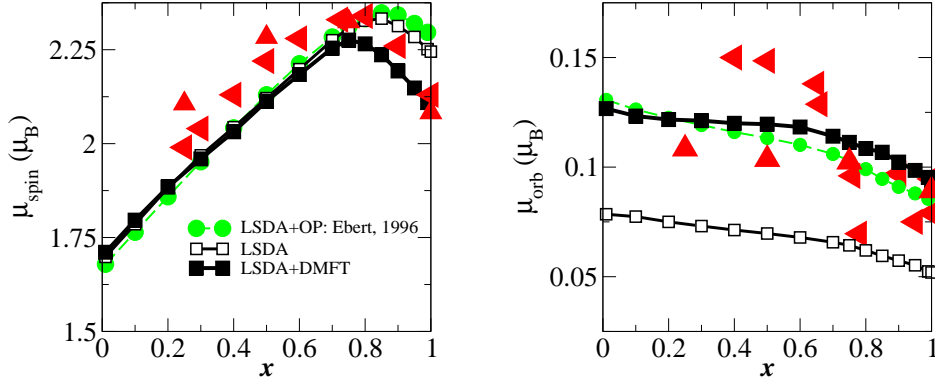
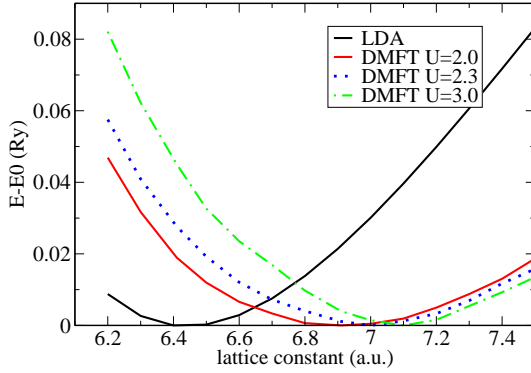


Fig. 13: Spin (left) and orbital (right) total magnetic moments of bcc Fe_xCo_{1-x} alloys calculated via LSDA+DMFT (filled squares), compared to plain LSDA (open squares), LSDA+OP calculations [52] (filled circles), and experimental data (red triangles). The corresponding DMFT parameters are $U_{Fe} = 2$ eV, $U_{Co} = 3$ eV and $J_{Fe} = J_{Co} = 0.9$ eV [43].



$U=2$ eV $J=0.9$ eV	volume (a.u.) ³	bulk mod. (GPa)
LSDA	66.86	280
DMFT (LMTO)	76.20	163
DMFT (KKR)	76.28	171
Expt.	73.52	186

Fig. 14: Left: Energy versus lattice-constant curves for fcc Ni in the LSDA and in the LSDA+DMFT scheme. The zero of the energy of each curve is set to its own minimum value E_0 and three chosen values of U are presented ($T = 400$ K). Right: Computed values of the equilibrium atomic volume V_0 and the bulk modulus B of fcc Ni for the standard LSDA and for the LSDA+DMFT scheme [57].

LSDA+DMFT simulations. The temperature was set as $T = 400$ K and 2048 Matsubara frequencies were used. As for the DFT part, convergence in the LSDA+DMFT total energy was considered acceptable when the changes for subsequent iterations were smaller than 0.1 meV. On the left-hand side of Fig. 14, one can see the total energy curves as functions of the lattice constant for the KKR implementation. The curves have been shifted with respect to their minima, so it is easier to compare them. As observed in previous calculations [60], in LSDA the equilibrium value of the lattice constant is slightly (3%) underestimated with respect to the experimental one. Looking at the curves for the LSDA+DMFT simulations, one immediately notices that the results are strongly dependent on the value of the Hubbard U . Furthermore the best result seems to be obtained for $U = 2$ eV, i.e. for a value smaller than the widely accepted $U = 3$ eV. On the other hand, the curve for $U = 3$ eV seems to overestimate corre-

lation effects. The explanation of these results is in the perturbative nature of the SPTF solver, which tends to overestimate correlation effects in fcc Ni. This was noticed since the first implementation [30], when comparison between LSDA+DMFT results with the SPTF solver and numerically exact Quantum Monte-Carlo solver showed the best agreement for $U = 2$ eV. Furthermore in the already mentioned calculation of the orbital polarization of Ni, it is shown that SPTF with $U = 3$ eV gives too strong a correction to the orbital moment [43]. The table on the right-hand side of Fig. 14, where the equilibrium atomic volume V_0 and the bulk modulus B are given, allows a more quantitative comparison of implementations of the DMFT using the KKR and LMTO respectively, and with previous LSDA studies of fcc Ni [60]. These values of V_0 and B have been computed with polynomial fitting of the energy versus atomic volume curve around the minimum. In addition also fitting through Birch-Murnaghan equation of state [61] was done, leading to almost identical results and confirming the stability of the data.

As for the total energy curves, the best results are obtained for $U = 2$ eV, and one can see that the inclusion of local correlation effects into the LSDA results corrects both the equilibrium atomic volume and the bulk modulus in the right way. While this fact is interesting on its own, one should notice that to have more precise results from a quantitative point of view, a more strict relation between solver, correlated orbitals and values of U is needed. Naturally it would be interesting to repeat these calculations with the numerically exact quantum Monte-Carlo solver to check if better agreement with the experiment can be obtained. Another interesting property can be deduced from the table on the right-hand side of Fig. 14: while the equilibrium atomic volumes are independent of the full self-consistency, the bulk modulus looks to be more strongly influenced. As expected, this discrepancy is proportional to the strength of U .

4.2 Electron spectroscopy

Photo emission is an experimental tool that allows to probe the electronic structure of a solid in a most detailed way. An appropriate theoretical description can be given on the basis of the one-step model that treats the excitation process, the transfer of the photo electron to the surface as well as its escape to the vacuum in a coherent way [22]. Within this framework the photo current intensity is written as:

$$j_{\vec{k}m_s}^{\vec{q}\lambda}(E + \omega) \propto \int d^3r \int d^3r' \left[\mathcal{T}_R \phi_{\vec{k}m_s}^{\text{LEED}}(\vec{r}, E + \omega) \right]^\dagger X_{\vec{q}\lambda}(\vec{r}) \Im G(\vec{r}, \vec{r}', E) X_{\vec{q}\lambda}^\dagger(\vec{r}') \mathcal{T}_R \phi_{\vec{k}m_s}^{\text{LEED}}(\vec{r}', E + \omega). \quad (44)$$

Here the manifold of initial states is represented in terms of the one-electron Green's function $G(\vec{r}, \vec{r}', E)$. Within a fully relativistic approach, the interaction of the electrons with the radiation field is described by the operator $X_{\vec{q}\lambda}$ that involves the electronic current density operator $\vec{j} = -|e|c\vec{\alpha}$ and the polarization vector $\vec{A}_{\vec{q}\lambda}$ of the radiation with frequency ω . Furthermore, the final state is written as a so-called time-reversed LEED-state that is expressed by a Lippmann-Schwinger equation with the free-electron wave function $\Xi_{m_s} e^{i\vec{k}\vec{r}}$ [47] used as a reference:

$$\phi_{\vec{k}m_s}^{\text{LEED}}(\vec{r}, E) = \Xi_{m_s} e^{i\vec{k}\vec{r}} + \int d^3r' G(\vec{r}, \vec{r}', E) V(\vec{r}') \Xi_{m_s} e^{i\vec{k}\vec{r}'}. \quad (45)$$

Implementation of these expressions using the KKR formalism allows, first of all, a proper description of all geometrical aspects of a photo emission experiment [22]. This includes in particular the surface of the sample as well as the directions of the incoming photo and outgoing electron beams. With a proper relativistic treatment of the transition matrix elements all types of dichroism can be dealt with including the spin polarization of the photo electrons.

Eqs. (44) and (45) allow the calculation of photo electron intensities as measured by a spin and angle resolved photo emission (ARPES) experiment. Averaging over the emission angle allows to discuss the spectra of an angle integrated experiment. The results of a corresponding combined experimental and theoretical study on the pure ferromagnetic metals Fe, Co and Ni in the soft X-ray regime ($\hbar\omega = 600$ eV) are shown in Fig. 15 [45]. The upper panel shows the spin-averaged spectra calculated on the basis of the LSDA and LSDA+DMFT in comparison with experiment. Taking into account the influence of the secondary electrons the theoretical results for the VB-XPS spectra are in fairly good agreement with experiment in the case of Fe and Co. For Ni, on the other hand, the LSDA-based calculations lead to a band-width that is much too large. Furthermore they are not able to reproduce the satellite at about 6 eV binding energy. In case of the LSDA+DMFT calculations, on the other hand, the appreciable real part of the self-energy $\Sigma(E)$ gives rise to a corresponding shrinking of the d-band width of Ni (compare with discussion of the DOS in Fig. 5). This leads to a much better agreement of the theoretical VB-XPS spectrum with experiment, as can be seen in Fig. 15. In addition, use of the LSDA+DMFT scheme leads to a pronounced increase of the intensity in the regime of the 6 eV satellite. The bottom row of Fig. 15 gives the spin difference of the photo-current ΔI^+ ,

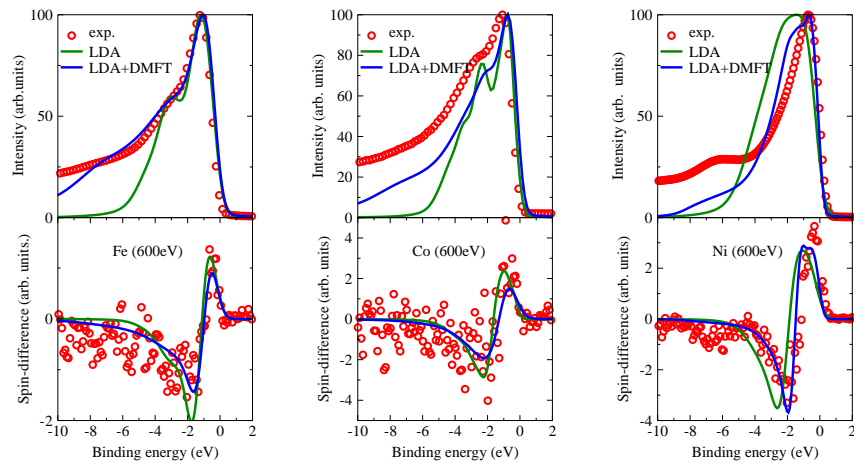


Fig. 15: Top: spin and angle-integrated VB-XPS spectrum of ferromagnetic Fe (left), Co (middle), and Ni (right) for a photon energy of 600 eV. Bottom: spin difference $\Delta I = I_{\uparrow}^+ - I_{\downarrow}^+$ of the photo current for excitation with left circularly polarized radiation. Theory: solid line; experiment: dashed line. The same scale has been used for the intensity and corresponding spin-difference plots [45].

i.e. the difference of the currents of photo-electrons with spin-up and spin-down with respect to the surface normal, for an excitation with left circularly polarized radiation. Because the

polarization analysis of the photo-current is done with respect to an axis that is perpendicular to the spontaneous magnetization \vec{m} (oriented parallel to the surface), ΔI^+ cannot be caused by the exchange splitting of the ground state. In fact ΔI^+ is caused by the Fano-effect. This term denotes the fact that the spin-orbit coupling together with the transition selection rules gives in general rise to a spin-polarized photo electron beam - even in case of non-magnetic solids.

As for the total intensity, the theoretical spin difference ΔI^+ shown in the bottom row of Fig. 15 is found in rather good agreement with experiment in particular for Fe and Co. The sequence for the maximum (minimum) of ΔI^+ of Fe, Co and Ni is found to be 1.3, 2.2 and 2.6 (-1.9, -2.7 and -3.5). Although there are several electronic and structural properties that determine these data, they nevertheless correlate reasonably well with the strength of the spin-orbit coupling parameters of the d-states (66, 85 and 107 meV, respectively) [62] to identify once more the spin-orbit coupling as the source for the observed spin-polarization. As for the standard VB-XPS spectra inclusion of the self-energy $\Sigma(E)$ leads to a substantial improvement for the agreement of the theoretical ΔI^+ spectrum with experiment. As one can see in Fig. 15, the shrinking of the band width is also reflected by the ΔI^+ curves, while their amplitude and shape is only slightly changed.

Performing photo emission experiments in the angle-resolved mode (ARPES) obviously supplies much more detailed information on the electronic structure than angle-integrated experiments. If no specific surface-related features show up in the spectra the observed peaks in an ARPES spectrum may be interpreted qualitatively to reflect the dispersion relation $E_{j\vec{k}}$ via vertical transitions ($\Delta\vec{k} = 0$) to free-electron like final states. Fig. 16 shows corresponding experimental and theoretical ARPES spectra for the (110) surface of Ni with the initial state \vec{k} vector along $\overline{\Gamma Y}$ and different angles of emission. The dotted lines represent the experimental data, whereas the solid lines denote the one-electron approach to the measured spectral function. Obviously, the LSDA-based (top row) calculation completely fails. The energetic positions of the theoretical peaks deviate strongly from the measured ones. Furthermore, the complicated intensity distributions that appear for higher angles of emission could not be reproduced by the LSDA method at all. In contrast, the non-selfconsistent quasi particle calculation based on the three-body scattering formalism of Manghi et al. [63] provides a significant improvement when compared to the measured spectra. For the complete variety of emission angles the energetic peak positions coincide with the experiment within about 0.1 eV.

Only the overall shape of the measured spectral intensities deviate from the calculations because of the neglect of multiple scattering and surface-related effects. In the experiment the different peaks seem to be more broadened and the spectral weight especially for nearly normal emission is shifted by about 0.1 eV to higher binding energies. In addition, it seems that for very high emission angles like 60° an even more complicated peak structure is hidden in the experimental resolution. KKR calculations allowed to go far beyond previous theoretical studies by combining the self-consistent LSDA DMFT method with the one-step based calculation of the corresponding spectral function. The resulting intensity distributions are shown in the lower row of Fig. 16. A first inspection reveals very satisfying quantitative agreement between

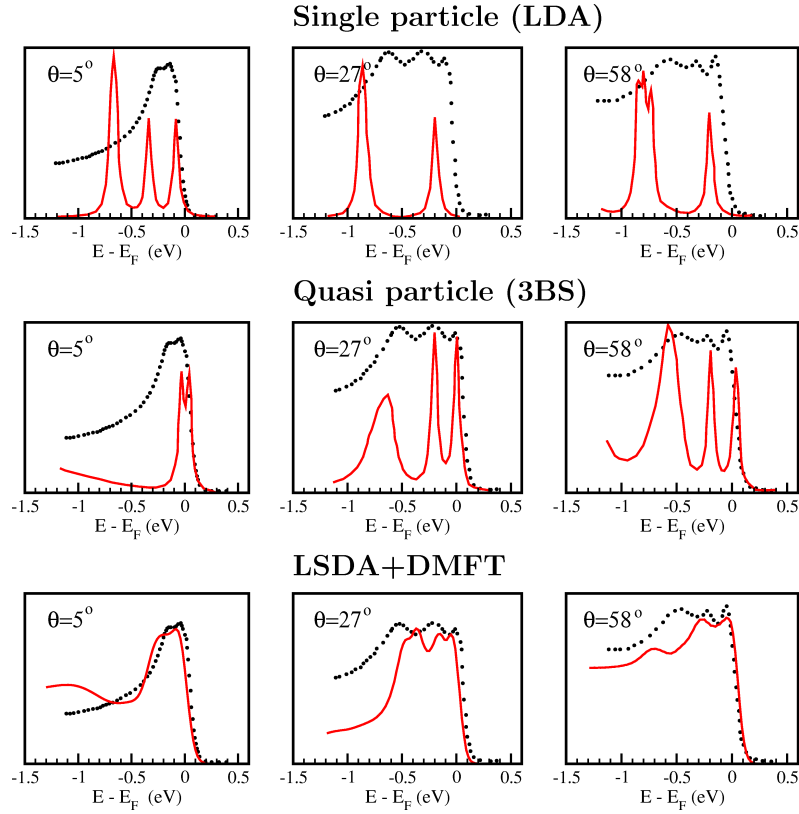


Fig. 16: Spin-integrated ARUPS spectra from Ni(110) along $\overline{\Gamma Y}$ for three different angles of emission. Upper row: comparison between LSDA-based calculation and experiment [63], middle row: comparison between experiment and non-self-consistent quasi particle calculations neglecting matrix element and surface effects [63], lower row: spin-integrated LSDA+DMFT spectra including photo emission matrix elements (this work). Theory: solid red line, experiment: black dots [44].

experiment and theory for all emission angles. For the emission angle 5° the spin-integrated spectrum exhibits a pronounced double-peak structure with binding energies of 0.1 eV and 0.3 eV. The second peak is slightly reduced in intensity which is also in accordance with the experimental findings. Furthermore, the width of the spectral distribution is quantitatively reproduced. The calculated binding energies can be ascribed to the real part of the self-energy that corrects the peak positions due to dynamical renormalization procedure of the quasi particles which is missing in a typical LSDA-based calculation. The relative intensities and the widths of the different peaks, on the other hand, must be attributed to the matrix-element effects which enter the calculations from the very beginning via the one-step model of photo emission. As it has been found for Ni(001) the double-peak structure originates from excitation of the spin-split d bands in combination with a significant amount of surface-state emission [11]. The two spectra calculated for high angles of emission show the more broadened spectral distributions observable from the experimental data. An explanation can be given in terms of matrix-element effects, due to the dominating dipole selection rules. The spin-resolved spectra reveal a variety of d-band excitations in both spin channels, which in consequence lead to the complicated shape of the

spectral distributions hardly to be interpreted in the spin-integrated mode.

Spin-resolved ARPES experiments have been made recently for the elemental ferromagnets Fe, Co and Ni [55, 64]. Fig. 17 displays a comparison between spin-resolved ARPES data and theoretical LSDA+DMFT calculations of Fe(110) along the $\overline{\Gamma\text{N}}$ direction of the bulk Brillouin zone (BZ) for p- and s-polarized photons together with LSDA+DMFT calculations. The k

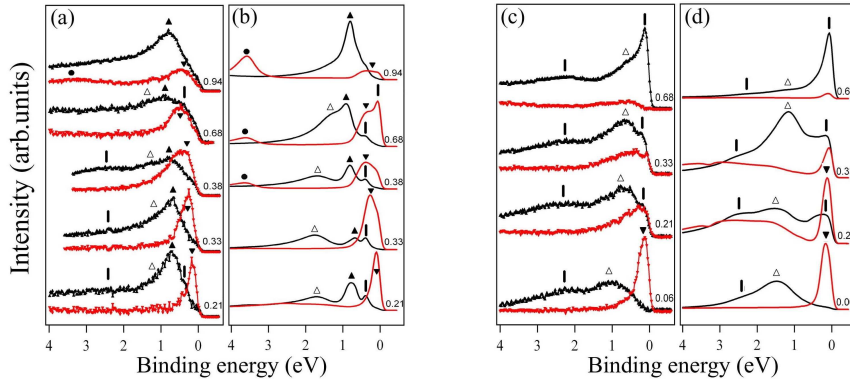


Fig. 17: Spin-resolved ARPES spectra of Fe(110) in normal emission along $\overline{\Gamma\text{N}}$. The curves are labeled by k in units of $\overline{\Gamma\text{N}}=1.55 \text{ \AA}^{-1}$. (a), (c) Experiment [upwards (black) triangles: majority states, downwards (red) triangles: minority states]. (b), (d) LSDA DMFT theory [dark (black) and light (red) lines for majority and minority electrons, respectively]. (a), (b) For p- and (c), (d) for s polarization [55].

values were calculated from the used photon energies ranging from 25eV to 100 eV, using an inner potential $V=14.5$ eV.

Near the Γ point ($k \sim 0.06 \overline{\Gamma\text{N}}$), the intense peak close to the Fermi level corresponds to a $\Sigma_{1,3}^{\downarrow}$ minority surface resonance. Experimentally, its Σ_3^{\downarrow} bulk component crosses the Fermi level at $k \sim 0.33 \overline{\Gamma\text{N}}$, leading to a reversal of the measured spin-polarization and to a strong reduction of the intensity at $k = 0.68 \overline{\Gamma\text{N}}$ in the minority channel, in agreement with the theoretical results (Fig. 17 (b) and (d)). The peak at the binding energy $\text{BE} \sim 0.7$ eV, visible mainly for p-polarization in a large range of wave vectors between Γ and N can be assigned to almost degenerate $\Sigma_{1,4}^{\uparrow}$ bulk-like majority states (Figs. 17 (a) and (b)). For s-polarization (Fig. 17 (c) and (d)), a Σ_3^{\uparrow} feature at $\text{BE} \sim 1.1$ eV dominates the spectrum at the Γ point. For p-polarization its degenerate Σ_1^{\uparrow} states form a shoulder around the same BE. The broad feature around 2.2 eV, visible at various k points, but not at the N point, is related to a majority $\Sigma_{1,3}^{\uparrow}$ surface state (see below). Around the N-point ($0.76 \leq k \leq 1.0$) and at $\text{BE} \geq 3$ eV (Figs. 17 (a)) one observes a Σ_1^{\downarrow} band having strong sp-character. The pronounced difference between its theoretical and experimental intensity distributions can be attributed to the fact that in the calculations only local Coulomb repulsion between d-electrons is considered, without additional lifetime effects for the sp-bands. Finally, one notices that the background intensity of the spectrum at $k=0.66 \overline{\Gamma\text{N}}$, corresponding to a photon energy of 55 eV, is strongly increasing by the appearance of the Fe 3p resonance. Comparing the experimental results from spin-integrated and spin-resolved ARPES measurements with LSDA+DMFT results, one obtains at low BE good agreement for many of

the peak positions. This is also demonstrated in Fig. 18 (a) and (b) where the experimental peak positions are compared with the LSDA+DMFT spectral function. Similar calculations based on the LSDA+3BS scheme are compared with the experimental data in Fig. 17(a) and (c). Since the theoretical calculations do not show big differences, also the LSDA+3BS spectral function agrees well at low BE with the experimental peak positions. On the other hand, quantitative

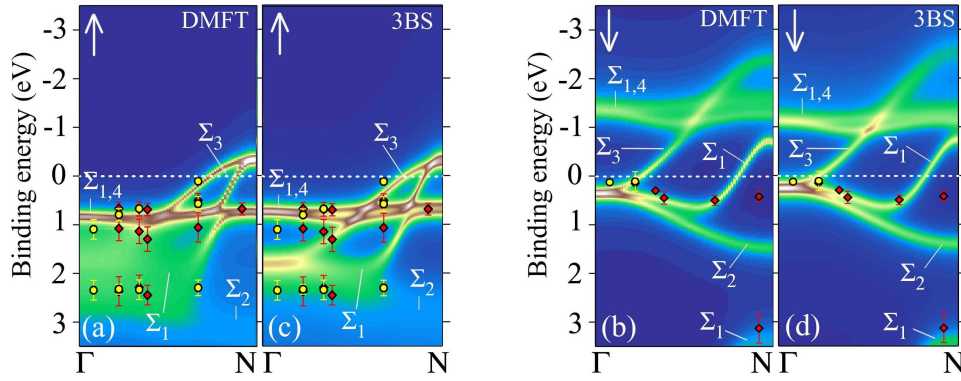


Fig. 18: Spectral functions of Fe(110) and photo emission peak positions obtained from the spin-resolved measurements for different polarizations (Diamonds for horizontal and circles for vertical polarization). Results obtained by LSDA+DMFT ((a), (b)) and by LSDA+3BS ((c), (d)) methods for majority and minority electronic states, respectively [55].

agreement cannot be achieved for higher BE. In particular, the calculated spectral weight near Γ for the $\Sigma_{1,3}^{\uparrow}$ bands is in between the experimental features at 1.2 eV and 2.2 eV. Assuming negligible correlation effects would move the calculated feature to the LSDA value at BE=2.2 eV. Thus the experimental peak at 2.2 eV could be assigned to the bulk $\Sigma_{1,3}^{\uparrow}$ bands. However, a complete neglect of correlation effects in Fe would make the overall comparison between theory and experiment much worse. Thus the experimental peak at BE=2.2 eV is interpreted as a $\Sigma_{1,3}^{\uparrow}$ surface state in agreement with previous experimental and theoretical studies [65]. The theoretical results confirm this view since one clearly observes how changes in the surface barrier potential induce additional shifts in its BE position. Thus, from the data shown in Fig. 18 one can conclude that correlation effects in the calculations using $U=1.5$ eV are underestimated and that a stronger band narrowing is needed to achieve agreement between theory and experiment.

Other investigations on spectroscopic properties using the implementation of the LSDA+DMFT via the KKR method were dealing so far with the magneto-optical Kerr effect, the magnetic X-ray dichroism in X-ray absorption and the magnetic Compton effect. As for the photo emission considered here, it turned out that treatment of correlation effects via the DMFT substantially improves agreement with experiment compared to plain LSDA calculations.

5 Summary

The implementation of the LSDA+DMFT on the basis of the KKR method has been described in some detail. The appealing feature of this approach is that the KKR delivers the one-electron

Green's function directly. It therefore allows to combine the treatment of correlations via the DMFT with calculation of a great variety of physical properties for, in principle, any type of system. As was demonstrated by results for magnetic moments and photo emission spectra of various transition metal systems, this allows in particular a direct comparison with experiment.

Acknowledgment

This work was supported by the Deutsche Forschungsgemeinschaft through FOR 1346, EB-154/18, EB-154/23 and MI-1327/1 and by the BMBF (Project 05KS10WMA).

References

- [1] A.I. Lichtenstein, M.I. Katsnelson, and G. Kotliar, in *Electron Correlations and Materials Properties II*, edited by A. Gonis, N. Kioussis, and M. Ciftan (Kluwer/Plenum, Berlin, 2002), p. 428
- [2] K. Held *et al.*, in *Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, NIC Series*, edited by J. Grotendorst, D. Marx, and A. Muramatsu (NIC, Jülich, 2002), Vol. 10, p. 17
- [3] K. Held, *Adv. Phys.* **56**, 829 (2007)
- [4] G. Kotliar *et al.*, *Rev. Mod. Phys.* **78**, 865 (2006)
- [5] V. I. Anisimov *et al.*, *J. Phys.: Cond. Mat.* **9**, 7359 (1997)
- [6] A. I. Lichtenstein and M. I. Katsnelson, *Phys. Rev. B* **57**, 6884 (1998)
- [7] M. Katsnelson and A. Lichtenstein, *J. Phys.: Cond. Mat.* **11**, 1037 (1999)
- [8] S.Y. Savrasov, G. Kotliar, and E. Abrahams, *Nature* **410**, 793 (2001)
- [9] S.Y. Savrasov and G. Kotliar, *Phys. Rev. B* **69**, 245101 (2004)
- [10] L. Chioncel, M.I. Katsnelson, R.A. de Groot, and A. I. Lichtenstein, *Phys. Rev. B* **68**, 144425 (2003)
- [11] V. Drchal, V. Janiš, and J. Kudrnovský, *Phys. Rev. B* **60**, 15664 (1999)
- [12] L. Chioncel *et al.*, *Phys. Rev. B* **67**, 235106 (2003)
- [13] O.K. Andersen and T. Saha-Dasgupta, *Phys. Rev. B* **62**, R16219 (2000)
- [14] L. Vitos, H.L. Skriver, B. Johansson, and J. Kollar, *Comp. Mater. Science* **18**, 24 (2000)
- [15] A. Schindlmayr, P. García-González, and R. W. Godby, *Phys. Rev. B* **64**, 235106 (2001)
- [16] L. Vitos, *Phys. Rev. B* **64**, 014107 (2001)
- [17] J. Minár *et al.*, *Phys. Rev. B* **72**, 045125 (2005)
- [18] H. Ebert, D. Ködderitzsch, and J. Minár, *Rep. Prog. Phys.* in the press (2011)
- [19] P. Soven, *Phys. Rev.* **156**, 809 (1967)
- [20] J.S. Faulkner, *Progress in Materials Science* **27**, 1 (1982)
- [21] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, *Rev. Mod. Phys.* **68**, 13 (1996)
- [22] J. Braun, *Rep. Prog. Phys.* **59**, 1267 (1996)

- [23] H. Ebert, Rep. Prog. Phys. **59**, 1665 (1996)
- [24] D. Sébilleau *et al.*, J. Phys.: Cond. Mat. **18**, R175 (2006)
- [25] J. Koringa, Physica **13**, 392 (1947)
- [26] W. Kohn and N. Rostoker, Phys. Rev. **94**, 1111 (1954)
- [27] V.I. Anisimov, J. Zaanen, and O.K. Andersen, Phys. Rev. B **44**, 943 (1991)
- [28] V.I. Anisimov, F. Aryasetiawan, and A.I. Lichtenstein, J. Phys.: Cond. Mat. **9**, 767 (1997)
- [29] L.V. Pourovskii, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **72**, 115106 (2005)
- [30] M.I. Katsnelson and A.I. Lichtenstein, European Physics Journal B **30**, 9 (2002)
- [31] R. Harris, J. Phys. C: Solid State Phys. **3**, 172 (1970)
- [32] R. Zeller, J. Phys. C: Solid State Phys. **20**, 2347 (1987)
- [33] E. Tamura, Phys. Rev. B **45**, 3271 (1992)
- [34] H. Ebert and B.L. Györffy, J. Phys. F: Met. Phys. **18**, 451 (1988)
- [35] J.S. Faulkner and G.M. Stocks, Phys. Rev. B **21**, 3222 (1980)
- [36] A. Gonis and W.H. Butler, *Multiple scattering in solids, Graduate Texts in Contemporary Physics* (Springer, Berlin, 1999)
- [37] B.L. Györffy and M.J. Stott, in *Band Structure Spectroscopy of Metals and Alloys*, edited by D.J. Fabian and L.M. Watson (Academic Press, New York, 1973), p. 385
- [38] J. Zablouil, R. Hammerling, L. Szunyogh, and P. Weinberger, in *Electron Scattering in Solid Matter*, Vol. 147 of *A Theoretical and Computational Treatise of Springer Series in Solid-state Sciences* (Springer, Berlin, 2005)
- [39] J.S. Faulkner, J. Phys. C: Solid State Phys. **10**, 4661 (1977)
- [40] J. Braun *et al.*, Phys. Rev. B **82**, 024411 (2010)
- [41] X. Dai *et al.*, Science **300**, 953 (2003)
- [42] L.V. Pourovskii, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **73**, 60506 (2006)
- [43] S. Chadov *et al.*, Europhys. Lett. **82**, 37001 (2008)
- [44] J. Braun *et al.*, Phys. Rev. Lett. **97**, 227601 (2006)
- [45] J. Minár *et al.*, Phys. Rev. Lett. **95**, 166401 (2005)

-
- [46] A.H. MacDonald and S.H. Vosko, *J. Phys. C: Solid State Phys.* **12**, 2977 (1979)
- [47] M.E. Rose, *Relativistic Electron Theory* (Wiley, New York, 1961)
- [48] H. Ebert, in *Electronic Structure and Physical Properties of Solids*, Vol. 535 of *Lecture Notes in Physics*, edited by H. Dreyssé (Springer, Berlin, 2000), p. 191
- [49] H. Ebert and H. Akai, *Mat. Res. Soc. Symp. Proc.* **253**, 329 (1992)
- [50] S. Chadov *et al.*, *J. Phys. D: Appl. Phys.* **42**, 084002 (2009)
- [51] M.S.S. Brooks, *Physica B* **130**, 6 (1985)
- [52] H. Ebert and M. Battochetti, *Solid State Commun.* **98**, 785 (1996)
- [53] S. Biermann, F. Aryasetiawan, and A. Georges, *Phys. Rev. Lett.* **90**, 086402 (2003)
- [54] A.I. Lichtenstein, M.I. Katsnelson, and G. Kotliar, *Phys. Rev. Lett.* **87**, 067205 (2001)
- [55] J. Sánchez-Barriga *et al.*, *Phys. Rev. Lett.* **103**, 267203 (2009)
- [56] I. Di Marco *et al.*, *Eur. Phys. J. B* **72**, 473 (2009)
- [57] I. Di Marco *et al.*, *Phys. Rev. B* **79**, 115111 (2009)
- [58] A.M. Oles and G. Stollhoff, *Phys. Rev. B* **29**, 314 (1984)
- [59] T. Bandyopadhyay and D.D. Sarma, *Phys. Rev. B* **39**, 3517 (1989)
- [60] M. Černý *et al.*, *Phys. Rev. B* **67**, 035116 (2003)
- [61] F.D. Murnaghan, *Proc. Natl. Acad. Sci. USA* **30**, 244 (1944)
- [62] V. Popescu, H. Ebert, B. Nonas, and P.H. Dederichs, *Phys. Rev. B* **64**, 184407 (2001)
- [63] F. Manghi *et al.*, *Phys. Rev. B* **59**, R10409 (1999)
- [64] J. Sánchez-Barriga *et al.*, *Phys. Rev. B* **82**, 104414 (2010)
- [65] H.J. Kim, E. Vescovo, S. Heinze, and S. Blügel, *Surf. Sci.* **478**, 193 (2001)

13 Hedin Equations, GW , $GW+DMFT$, and All That

K. Held, C. Taranto, G. Rohringer, and A. Toschi
Institute for Solid State Physics
Vienna University of Technology, 1040 Vienna, Austria

Contents

1	Introduction	2
2	Hedin equations	4
3	GW approximation	10
3.1	From Hedin equations to GW	10
3.2	GW band gaps and quasiparticles	11
4	$GW+DMFT$	12
5	All of that: <i>ab initio</i> $D\Gamma A$	16
A	Additional steps: equation of motion	18

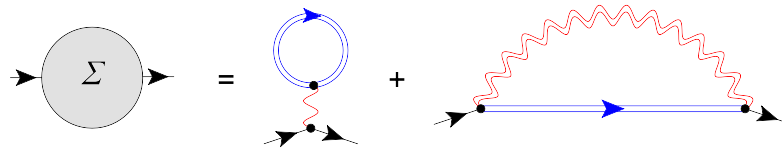


Fig. 1: In *GW*, the self energy is given by the Hartree term plus a Fock-like term which is however in terms of the screened Coulomb interaction W (double wiggled line) instead of the bare Coulomb interaction V (single wiggled line). The interacting Green function G is denoted by a double straight line.

1 Introduction

An alternative to density functional theory [1–3] for calculating materials *ab initio* is the so-called *GW* approach [4, 5]. The name stems from the way the self energy is calculated in this approach: It is given by the product of the Green function G and the screened Coulomb interaction W , see Fig. 1:

$$\Sigma^{\text{GW}}(\mathbf{r}, \mathbf{r}'; \omega) = i \int \frac{d\omega'}{2\pi} G(\mathbf{r}, \mathbf{r}'; \omega + \omega') W(\mathbf{r}, \mathbf{r}'; \omega'). \quad (1)$$

This self energy contribution supplements the standard Hartree term (first diagram in Fig. 1.) Here, \mathbf{r} and \mathbf{r}' denote two positions in real space, ω the frequency of interest; and the imaginary unit i in front of G stems from the standard definition of the real time (or real frequency) Green function and the rules for evaluating the diagram Fig. 1, see e.g. [6].

In *GW*, the screened Coulomb interaction W is calculated within the random phase approximation (RPA) [7]. That is, the screening is given by the (repeated) interaction with independent electron-hole pairs, see Fig. 2. For example, the physical interpretation of the second diagram in the second line of Fig. 2 is as follows: two electrons do not interact directly with each other as in the first term (bare Coulomb interaction) but the interaction is mediated via a virtual electron-hole pair (Green function bubble). Also included are repeated screening processes of this kind (third term etc.). Because of the virtual particle-hole pair(s), charge is redistributed dynamically, and the electrons only see the other electrons through a screened interaction. Since this RPA-screening is a very important contribution from the physical point of view and since, at the same time, it is difficult to go beyond, one often restricts oneself to this approximation when considering screening, as it is done in *GW*.

For visualizing what kind of electronic correlations are included in the *GW* approach, we can replace the screened interaction in Fig. 1 by its RPA expansion Fig. 2. This generates the Feynman diagrams of Fig. 3. We see that besides the Hartree and Fock term (first line), additional diagrams emerge (second line). By the definition that correlations are *what goes beyond Hartree-Fock*, these diagrams constitute the electronic correlations of the *GW* approximation. They give rise to quasiparticle renormalizations and finite quasiparticle lifetimes as well as to renormalizations of the gaps in band insulators or semiconductors.

It is quite obvious that a restriction of the electronic correlations to only the second line of Fig. 3 is not sufficient if electronic correlations are truly strong such as in transition metal

$$\begin{aligned}
 \text{---} \text{ } \text{---} &= \text{---} \text{ } \text{---} + \text{---} \text{ } \text{---} \\
 &= \text{---} \text{ } \text{---} + \text{---} \text{ } \text{---} + \text{---} \text{ } \text{---} + \dots
 \end{aligned}$$

Fig. 2: The screened interaction W (double wiggled line) is calculated from the bare Coulomb interaction V (single wiggled line) and corrections which describe screening processes. In GW , the screening is given by the random phase approximation, i.e., only bubble diagrams in a geometric series (second line) are considered. This geometric series can be generated from a single bubble connected to the screened interaction (first line). If we start with $W = V$ on the right hand side of the first line, we will generate the second term of the second line, and by further iterations obtain the whole series (whole second line).

$$\begin{aligned}
 \Sigma &= \text{---} \text{ } \text{---} + \text{---} \text{ } \text{---} \\
 &+ \text{---} \text{ } \text{---} + \text{---} \text{ } \text{---} + \dots
 \end{aligned}$$

Fig. 3: Substituting the screened interaction of Fig. 2 into the self energy diagram Fig. 1 generates besides the Hartree term (which is independently considered) the Fock term (first line), and some electronic correlations beyond (second line).

oxides or f -electron systems. The GW approximation cannot describe Hubbard side bands or Mott-Hubbard metal-insulator transitions. Its strength is for weakly correlated electron systems and, in particular, for semiconductors. For these, extended sp^3 orbitals lead to an important contribution of the non-local exchange. This is not well included (underrated) in the local exchange-correlation potential V_{xc} of the local density approximation (LDA) and overrated by the bare Fock term. The GW exchange is “in between” in magnitude and energy dependence, which is also important, as we will see later. If it is, instead, the local correlation-part of V_{xc} which needs to be improved upon, as in the aforementioned transition metal oxides or f -electron systems, we need to employ dynamical mean field theory (DMFT) [8–10] or similar many-body approaches.

Historically, the GW approach was put forward by Hedin [4] as the simplest approximation to the so-called Hedin equations. In Section 2, we will derive these Hedin equations from a Feynman-diagrammatical point of view. Section 3.1 shows how GW arises as an approximation to the Hedin equations. In Section 3.2, we will briefly present some typical GW results for materials, including the aforementioned quasiparticle renormalizations, lifetimes, and band gap enhancements. In Section 4, the combination of GW and DMFT is summarized. Finally, as a prospective outlook, *ab initio* dynamical vertex approximation ($D\Gamma A$) is introduced in Section 5 as a unifying scheme for all that: GW , DMFT and non-local vertex correlations beyond.

2 Hedin equations

In his seminal paper [4], Hedin noted, when deriving the equations bearing his name: "The results [i.e., the Hedin equations] are well known to the Green function people". And indeed, what is known as the Hedin equations in the bandstructure community are simply the Heisenberg equation of motion for the self energy (also known as Schwinger-Dyson equation) and the standard relations between irreducible and reducible vertex, self energy and Green function, polarization operator and screened interaction. While Hedin gave an elementary derivation with only second quantization as a prerequisite, we will discuss these from a Feynman diagrammatic point of view as the reader/student shall by now be familiar with this technique from the previous chapters/lectures of the Summer School. Our point of view gives a complementary perspective and sheds some light to the relation to standard many body theory. For Hedin's elementary derivation based on functional derivatives see [4] and [5].

Let us start with the arguably simplest **Hedin equation**: the well known *Dyson equation*, Fig. 4, which connects self energy and Green function:

$$\begin{aligned} G(11') &= G^0(11') + G(12)\beta\Sigma(22')G^0(2'1') \\ &= G^0(11') + G^0(12)\beta\Sigma(22')G^0(2'1') + G^0(13)\beta\Sigma(33')G^0(3'2)\beta\Sigma(22')G^0(2'1') + \dots \end{aligned} \quad (2)$$

Here, we have introduced a short-hand notation with 1 representing a space-time coordinate (\mathbf{r}_1, τ_1) also subsuming a spin if present; employ Einstein summation convention; G and G^0 denote the interacting and non-interacting ($V = 0$) Green function¹, respectively. Here and in the following part of this Section we will consider the Green function in imaginary time with Wick rotation $t \rightarrow -i\tau$, and closely follow the notation of [11]².

In terms of Feynman diagrams, the Dyson equation means that we collect all one-particle *irreducible* diagrams, i.e., all diagrams that do not fall apart into two pieces if one Green function line is cut, and call this object Σ . All Feynman diagrams for the interacting Green function are then generated simply by connecting the one-particle irreducible building blocks Σ by Green function lines in the Dyson equation (second line of Fig. 4). This way, no diagram is counted twice since all additional diagrams generated by the Dyson equation are one-particle *reducible* and hence not taken into account for a second time. On the other hand all diagrams are generated: the irreducible ones are already contained in Σ and the reducible ones have by definition

¹Please recall the definition of the Green function with Wick time-ordering operator T for τ_1 and τ_1' :

$$G(11') \equiv -\langle \mathcal{T} c(1)c(1')^\dagger \rangle. \quad (3)$$

$$\equiv -\langle \mathcal{T} c(1)c(1')^\dagger \rangle \Theta(\tau_1 - \tau_{1'}) + \langle \mathcal{T} c(1')^\dagger c(1) \rangle \Theta(\tau_{1'} - \tau_1) \quad (4)$$

The first term of Eq. (4) describes the propagation of a particle from $1'$ to 1 (and the second line the corresponding propagation of a hole). Graphically we hence denote $G(11')$ by a straight line with an arrow from $1'$ to 1 (The reader might note the reverse order in $G(11')$; we usually apply operators from right to left).

²Note that in [11] summations are defined as $(1/\beta) \int_0^\beta d\tau$ or in Matsubara frequencies \sum_m ; Fourier transformations are defined as $G(\tau) = \sum_m e^{-i\nu_m \tau} G(\nu_m)$. The advantage of this definition is that the equations have then the same form in τ and ν_m . Hence, we employ this notation in this Section and in Section 3. In Sections 1 and 4, the more standard definitions [6] are employed, i.e., $\int_0^\beta d\tau; (1/\beta) \sum_m$. This results in some factors β (inverse temperatur), which can be ignored if one only wants to understand the equations.

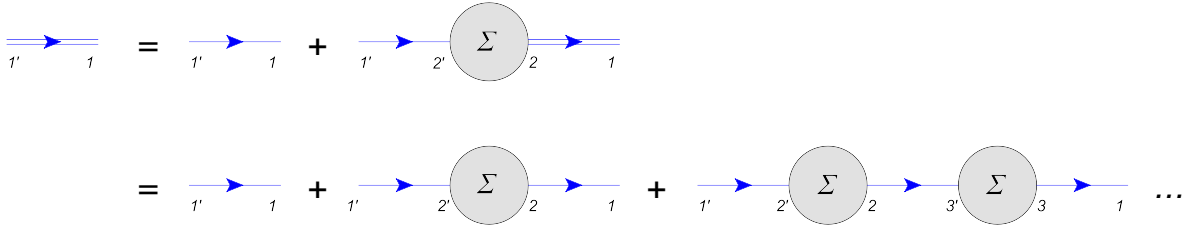


Fig. 4: One of the five Hedin equations is the well-known Dyson equation, connecting the interacting Green function G (double line), non-interacting Green function G^0 (single line) and self energy Σ .

the form of the second line of Fig. 4.

Let us note that the Dyson equation (2) can be resolved for

$$G(11') = \left(\left[(G^0)^{-1} - \beta \Sigma \right]^{-1} \right)_{(11')} \quad (5)$$

with a matrix inversion in the spatial and temporal coordinates. It is invariant under a basis transformation from \mathbf{r}_1 to, say, an orbital basis or from time to frequency (note in a momentum-frequency basis, the G and Σ matrices are diagonal).

A second equation of the five Hedin equations actually has the same form as the *Dyson equation* but with the Green function and interaction changing their role. It relates the *screened Coulomb interaction* W to the polarization operator P , see Fig. 5, which is a generalization of Fig. 2 to arbitrary polarizations. As W we simply define (sum) all Feynman diagrams which connect to the left and right side by interactions V . Physically, this means that we consider, besides the bare interaction, also all more complicated processes involving additional electrons (screening). Similar as for the Dyson equation (2), we collect all Feynman diagrams which do not fall apart into a *left* and a *right* side by cutting one *interaction line* V , and call this object P . From P , we can generate all diagrams of W connecting left and right side by a geometric series (second and third line of Fig. 5) with a repeated application of P and V (second line of Fig. 5). As for the Dyson equation, we generate all Feynman diagrams (in this case for W) and count none twice this way.

Mathematically, Fig. 5 translates into a **second Hedin equation**

$$W(11'; 22') = V(11'; 22') + W(11'; 33')P(3'3; 4'4)V(44'; 22') . \quad (6)$$

Note, that in a general basis the two particle objects have four indices: an incoming particle $2'$ and hole 2 , and an outgoing particle $1'$ and hole 1 , with possible four different orbital indices. In real space two of the indices are identical $\mathbf{r}_1 = \mathbf{r}_{1'}$ and $\mathbf{r}_2 = \mathbf{r}_{2'}$.³

Next, we turn to the polarization operator P which can be related to a vertex Γ^* . This is the standard relation between two particle Green functions (or response functions) and the vertex

³ This is obvious for the bare Coulomb interaction $V(\mathbf{r}_1, \mathbf{r}'_1; \mathbf{r}_2, \mathbf{r}'_2) = V(\mathbf{r}_1, \mathbf{r}_2)\delta(\mathbf{r}_1 - \mathbf{r}'_1)\delta(\mathbf{r}_2 - \mathbf{r}'_2)$ with $V(\mathbf{r}_1, \mathbf{r}_2) = \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_1 - \mathbf{r}_2|}$. As one can see in Fig. 5 this property is transferred to W for which hence only a polarization with two \mathbf{r} 's needs to be calculated in real space.

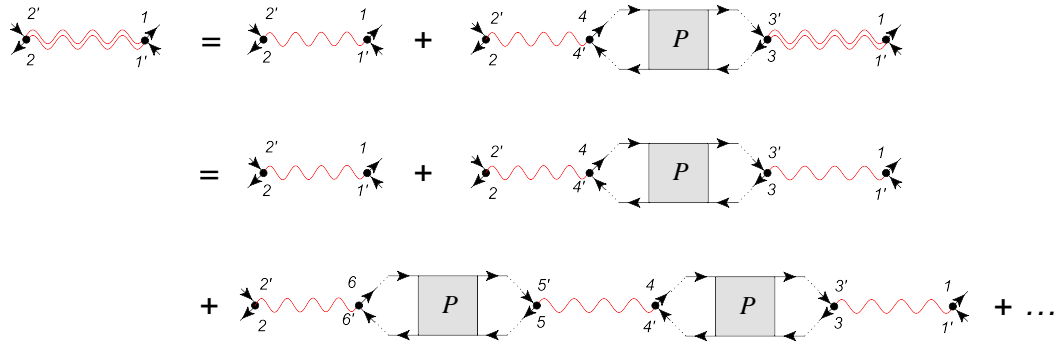


Fig. 5: A second of the five Hedin equations is an analog of the Dyson equation, Fig. 4 but for the Coulomb interaction. It relates the screened Coulomb interaction W (double wiggled line) with the bare Coulomb interaction V and the polarization operator P . As the Dyson equation can be considered as the defining equation for Σ , this second Hedin equation effectively defines what P is.

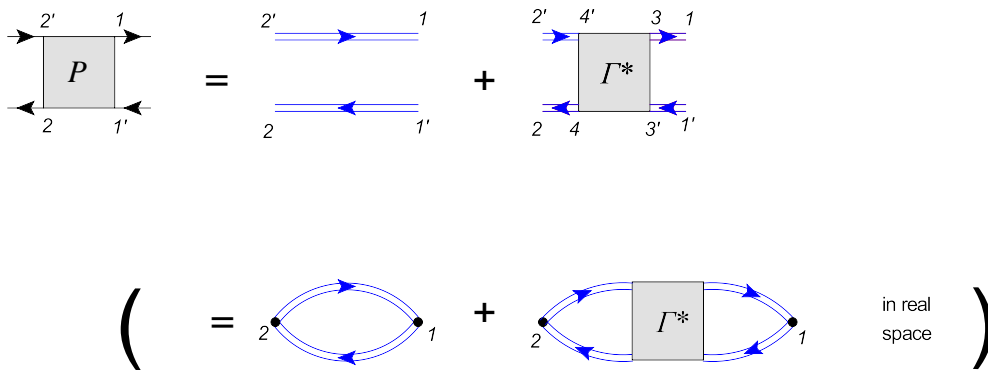


Fig. 6: A third Hedin equation relates the polarization operator P to two separated Green functions (“bubble” term) plus vertex (Γ^*) corrections. This is the standard relation between two particle Green functions and fully reducible vertex Γ . However since the polarization operator cannot include interaction-reducible diagrams Γ has to be replaced by Γ^* (see text). Second line: In terms of real space or momentum (but not in an orbital representation) two indices can be contracted to a single one (cf. footnote 3).

and represents a **third Hedin equation**. That is P is given by the simple connection of left and right side by two (separated) Green functions plus vertex corrections, see Fig. 6:

$$P(11'; 22') = \beta G(12')G(21') + \beta G(13)G(3'1')\Gamma^*(33'; 44')\beta G(4'2')G(24). \quad (7)$$

Note, in real space $2 = 2'$ and $1 = 1'$ (cf. footnote 3) so that working with a two index object, see second line of Fig. 6, is possible (and was done by Hedin), the inverse temperature β ($k_B \equiv 1$) arises from the rules for Feynman diagrams in imaginary times/frequencies, see [11]. Let us keep in mind, that in P or Γ^* not all Feynman diagrams are included: those diagrams, that can be separated into left and right by cutting a single interaction line have to be explicitly excluded, see Fig. 7. This is the reason why we put the symbol $*$ to the vertex Γ^* ; indicating that some diagrams of the full vertex Γ are missing.

Having introduced the two-particle vertex Γ^* , the **fourth Hedin equation** is obtained by relating this vertex to another object, the particle-hole irreducible vertex Γ_{ph}^* . This relation is the

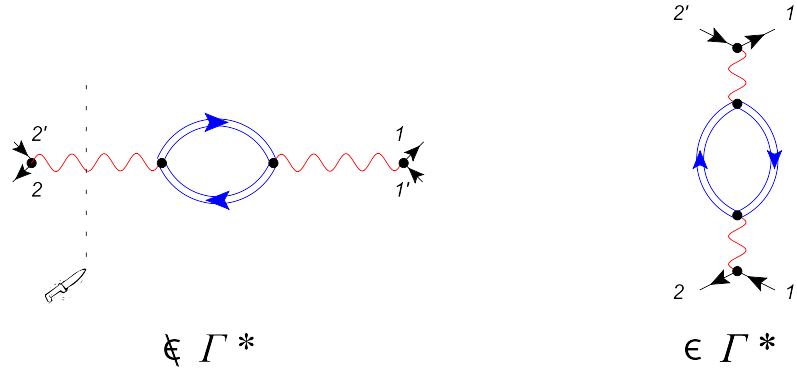


Fig. 7: Left: A Feynman diagram that is **not** part of Γ^* (or P) since cutting a single interaction line separates the diagram into left and right part. Right: This Feynman diagram is included in Γ^* (or P).

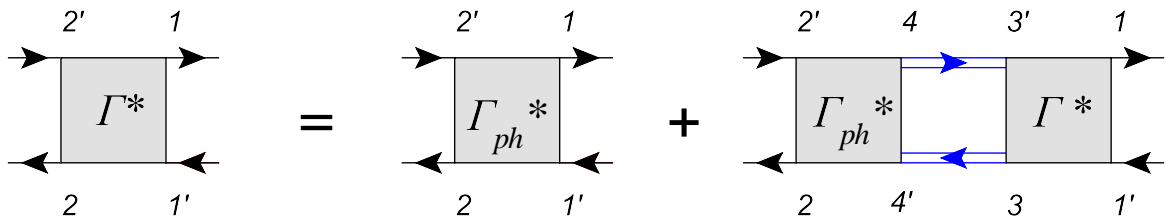


Fig. 8: A fourth Hedin equation is the Bethe-Salpeter equation between the irreducible Γ_{ph}^* and particle-hole reducible vertex Γ^* . The particle-hole irreducible vertex Γ_{ph}^* collects the Feynman diagrams which cannot be separated into left and right part by cutting two Green function lines. All diagrams are then generated by the Bethe-Salpeter equation.

standard Bethe-Salpeter equation. As in the Dyson equation, we define the vertex Γ^* as the irreducible vertex Γ_{ph}^* plus a geometric series of repetitions of Γ_{ph}^* connected by two Green functions, see Fig. 8:

$$\Gamma^*(11'; 22') = \Gamma_{ph}^*(11'; 22') + \Gamma^*(11'; 33')\beta G(3'4)G(4'3)\Gamma_{ph}^*(44'; 22') \quad (8)$$

Here, the particle-hole irreducible vertex collects all Feynman diagrams which cannot be separated by cutting *two* Green function lines into a left and right (incoming and outgoing) part. The Bethe-Salpeter equation then generates all vertex diagrams by connecting the irreducible building blocks with two Green function lines, in analogy to the Dyson equation for the one-particle irreducible vertex Σ .

Since in the polarization operator P (and the corresponding reducible vertex vertex Γ^*) diagrams which connect left and right by only one bare Coulomb interaction line V are however excluded, we have to explicitly take out this bare Coulomb interaction V from the vertex Γ^* , i.e., we have the standard particle-hole vertex Γ_{ph} (i.e., all particle-hole irreducible diagrams) minus the bare Coulomb interaction V diagram, see Fig. 9:

$$\Gamma_{ph}^*(11'; 22') = \Gamma_{ph}(11'; 22') - V(11'; 22') \quad (9)$$

The bare V and any combinations of Γ_{ph}^* and V are then generated in the screening equation (6), Fig. 5.

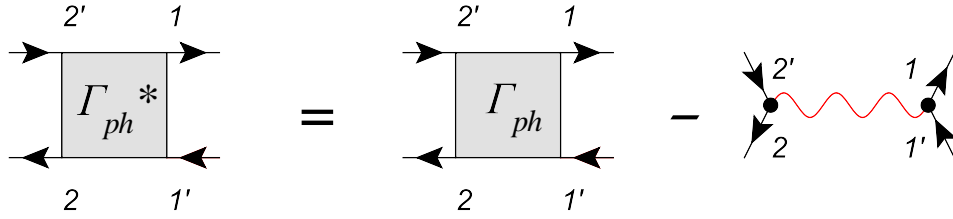


Fig. 9: We have to exclude the bare Coulomb interaction V from the particle-hole vertex Γ_{ph} since such contributions are already considered in Fig. 5.

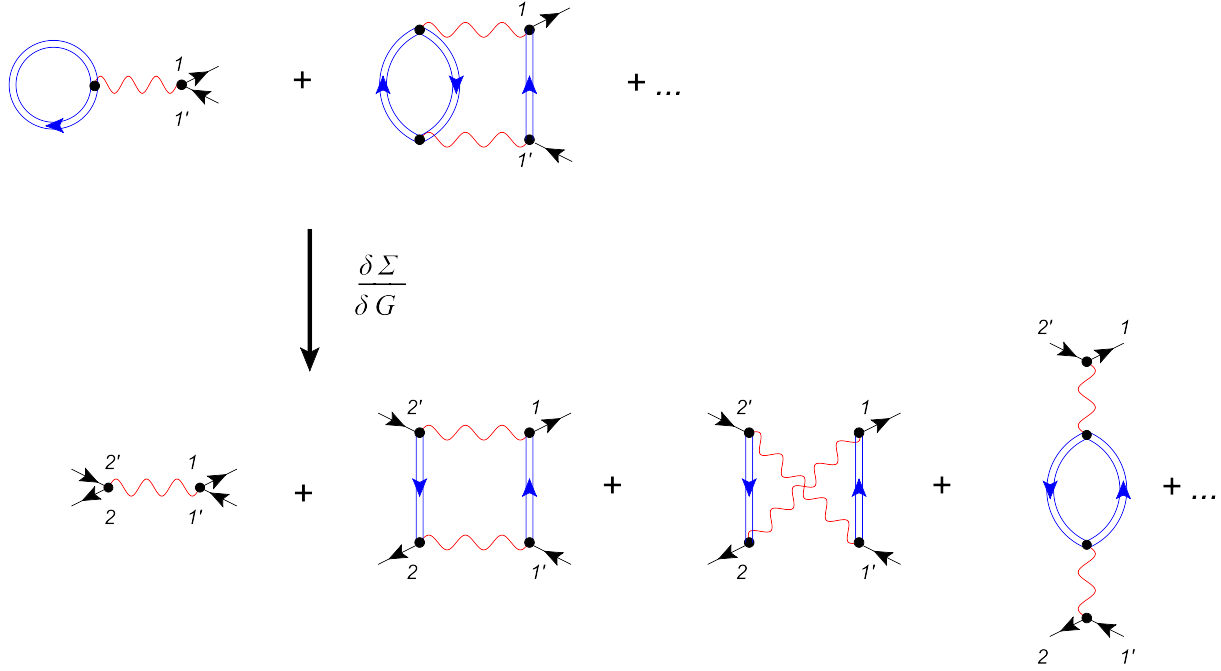


Fig. 10: By hands of selected Feynman diagrams we illustrate that differentiation of Σ w.r.t. G yields the particle-hole irreducible vertex.

In this fourth equation, Hedin directly expresses Γ_{ph}^* as the derivative of the self-energy w.r.t. the Green function [5] (respectively V [4]). This is a standard quantum field theoretical relation

$$\Gamma_{ph}(11'; 22') = \frac{\delta \Sigma(11')}{\delta G(2'2)}, \quad (10)$$

which in terms of Feynman diagrams follows from the observation that differentiation w.r.t. G means removing one Green function line, see Fig. 10 and Ref. [11]. If we, as Hedin, consider the self energy without Hartree term $\Sigma_{\text{Hartree}}(11') = V(11'; 22')G(2'2)$ we obtain the vertex Γ_{ph}^* instead of Γ_{ph} :

$$\Gamma_{ph}^*(11'; 22') = \frac{\delta[\Sigma(11') - \Sigma_{\text{Hartree}}(11')]}{\delta G(2'2)} \quad (11)$$

Note, the derivative of the Hartree term w.r.t. $G(2'2)$ yields $V(11'; 22')$, i.e., precisely the term not included in Γ_{ph}^* , see first diagram of Fig. 10.

The **fifth Hedin equation** is the *Heisenberg equation of motion* for the self energy, which follows from the derivative of the Green function (3) w.r.t. τ_1 , i.e., the time-part of the coordinate

1. Let us start with the Heisenberg equation for the Heisenberg operator $c(1)$ ($\hbar \equiv 1$):

$$-\frac{\partial c(\mathbf{r}_1, \tau_1)}{\partial \tau_1} = [c(\mathbf{r}_1, \tau_1), H] \quad (12)$$

and a general Hamiltonian of the form

$$H = H_0(11')c(1)^\dagger c(1') + \frac{1}{2}V(11'; 22')c(1)^\dagger c(1')c(2)^\dagger c(2'). \quad (13)$$

From Eq. (12), we obtain the Heisenberg equation of motion for the Green function:⁴

$$-\frac{\partial G(11')}{\partial \tau_1} = \delta(1 - 1') + H_0(12')G(2'1') - V(13'; 22')\langle \mathcal{T} c(3')c(2)^\dagger c(2')c(1')^\dagger \rangle. \quad (14)$$

The last term on the right hand side of Eq. (14) is by definition the self energy times the Green function. Hence this combination equals the interaction V times a two-particle Green function. The two particle Green function in turn is, in analogy to Fig. 6, given by the bubble term (two Green function lines; there are actually two terms of this: crossed and non-crossed) and vertex correction with the full (reducible) vertex Γ . Besides the Hartree term, the self energy is given by (see Fig. 11):

$$\Sigma(11') = -V(13'; 22')\beta G(4'2)G(2'4)G(3'3)\Gamma(31'; 44') - V(12'; 21')G(2'2) \quad (15)$$

Note Eq. (15) can be formulated in an alternative way (see Appendix A for a detailed calculation): Instead of expressing this correlation part of the self energy by the bare interaction and the full vertex we can take out the bare interaction line V and any particle-hole repetitions of V from the vertex, i.e., take Γ^* instead of Γ . Consequently we need to replace V by W to generate the same set of all Feynman diagrams:

$$\Sigma(11') = -W(13'; 22')\beta G(4'2)G(2'4)G(3'3)\Gamma^*(31'; 44') - W(12'; 21')G(2'2) \quad (16)$$

see second line of Fig. 11.⁵

The five equations Eq. (2), (6), (7), (8), (16) correspond to Hedin's equations (A13), (A20), (A24), (A22), (A23), respectively [4] [or to equations (44), (46), (38), (45), and (43), respectively, in [5)]. This set of equations is exact; it is equivalent to the text book quantum field

⁴Note, the first term here is generated by the time derivative of the Wick time ordering operator, the second term stems from $[c(1), H_0(22')c(2)^\dagger c(2')]$ and the third one from

$$[c(1), \frac{1}{2}V(33'; 22')c(3)^\dagger c(3')c(2)^\dagger c(2')]$$

employing $[A, BC] = B[A, C] + [A, B]C$ and the Fermi algebra $\{c(1'), c(1)^\dagger\} \equiv c(1')c(1)^\dagger + c(1)^\dagger c(1') = \delta(1 - 1')$, $\{c(1), c(1')^\dagger\} = \{c(1)^\dagger, c(1')\} = 0$.

⁵Note, Hedin defines as a "vertex" Λ [5] a combination of Γ^* and two Green function lines. Because he works in real space (or momentum) coordinates only at a common point $2 = 2'$ needs to be considered (in the same way as in the second line of Fig. 6). Hedin also adds a "1" in form of two δ -functions:

$$\Lambda(11'; 2) = \Gamma^*(11'; 33')\beta G(3'2)G(32) + \delta(1' - 2)\delta(2 - 1). \quad (17)$$

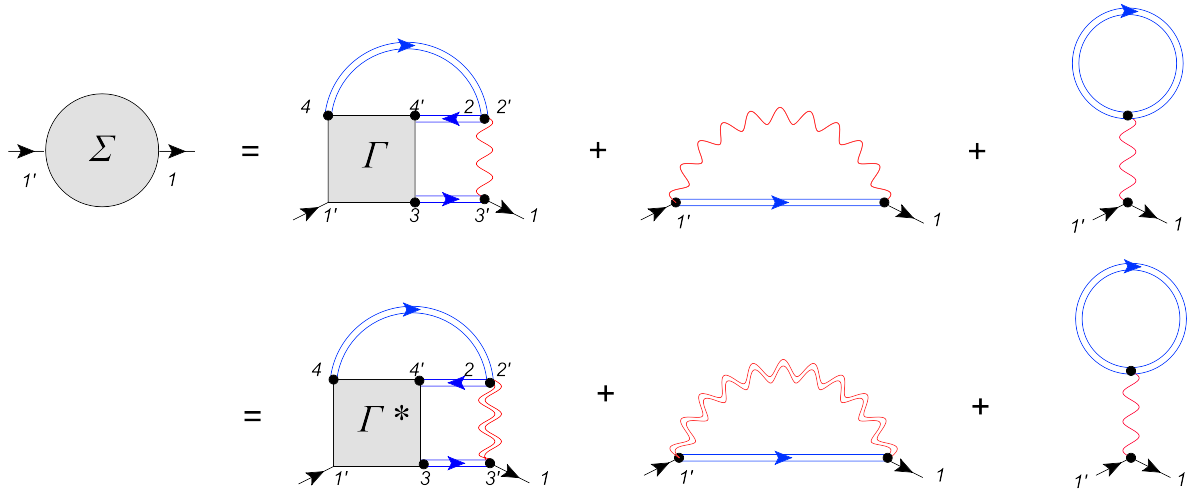


Fig. 11: The fifth Hedin equation is the Heisenberg equation of motion, which connects the one-particle Green function with the two particle Green function or as shown in the figure (and as employed) the self energy with the vertex.

theory [6, 11] relations between $\Gamma, \Gamma_{\text{ir}}, \Sigma$ and G ; but it contains additional equations since W and P are introduced. The advantage is that, this way, one can develop much more directly approximations where the screened Coulomb interaction plays a pronounced role such as in the GW approach.

3 GW approximation

3.1 From Hedin equations to GW

The simplest approximation is to neglect the vertex corrections completely, i.e., to set $\Gamma_{\text{ph}}^* = 0$.⁶ Then the Bethe-Salpeter equation (8) yields

$$\Gamma^* = 0. \quad (18)$$

The polarization in Eq. (7) simplifies to the bubble

$$P^{\text{GW}}(11'; 22') = \beta G(12')G(21'). \quad (19)$$

The screened interaction in Eq. (6) is calculated with this simple polarization

$$W(11'; 22') = V(11'; 22') + W(11'; 33')P^{\text{GW}}(3'3; 4'4)V(44'; 22'). \quad (20)$$

The self energy in the Heisenberg equation of motion (16) simplifies to Fig. 2, i.e.,

$$\Sigma^{\text{GW}}(11') = -W(12'; 21')G(2'2) \quad (21)$$

(plus Hartree term).

From this, the Green function is obtained via the Dyson equation (2):

$$G(11') = G^0(11') + G(12)\beta\Sigma^{\text{GW}}(22')G^0(2'1') \quad (22)$$

These five (self-consistent) equations constitute the GW approximation.

⁶Note this violates the Pauli principle, see last paragraph of Appendix A.

3.2 GW band gaps and quasiparticles

While the five GW equations above are meant to be solved self-consistently, most calculations hitherto started from a LDA bandstructure calculation⁷ and calculated from the LDA polarization (or dielectric constant) a screened interaction W_0 which in turn was used to determine the self energy with the Green function G_0 from the LDA: $\Sigma = iG_0W_0$.

Such calculations are already pretty reliable for semiconductor band gaps, which are underestimated in the LDA. Due to the energy(frequency)-dependence of Σ bands at different energies are under the influence of differently strong screened exchange contributions. In semiconductors, it turns out that the conduction band is shifted upwards in an approximately rigid way. The valence band is much less affected so that the GW band gap increases in comparison to the LDA gap. This effect can be mimicked by a so-called scissors operator, defined as cutting the density functional theory (DFT) bandstructure between valence and conduction band and moving the conduction band upwards. Cutting LDA bandstructures by a pair of scissors and rearranging them yields the GW bandstructure within an error of 0.1 eV for Si and 0.2 eV for GaAs [17].

More recently, self-consistent GW calculations became possible. Many of these calculations employ an approximation of Schilfgaard and Kotani [18, 19] where instead of the frequency dependent GW self energy $\Sigma_{nn'}(\omega, \mathbf{q})$ a frequency-independent Hermitian operator

$$\bar{\Sigma}_{nn'} = \text{Re}[\Sigma_{nn'}(\epsilon_q, \mathbf{q}) + \Sigma_{n'n}(\epsilon_q, \mathbf{q})]/2 \quad (23)$$

is constructed in the basis n, n' employed in the GW/LDA algorithm. This self energy operator has the advantage that (as in LDA) we can remain in a one-particle description and employ the Kohn-Sham equations with Hermitian operator $\bar{\Sigma}_{nn'}$ to recalculate electron densities and Bloch eigenfunctions.

The band gaps of this self-consistent approach are slightly larger than experiment, see open triangles of Fig. 12. This can be improved upon and band gaps can be calculated very reliably if additional to GW (some) vertex corrections are taken into account. In Fig. 12, the inclusion of electron-hole ladder diagrams (visualized on the right hand side of Fig. 12) results in the filled triangles with band gaps being within a few percent of the experimental ones. As in other areas of many-body theory, doing the self-consistency without including vertex corrections does not seem to be an improvement w.r.t. the non-self-consistent GW since self-consistency and vertex corrections compensate each other in part. Full GW calculations beyond the Schilfgaard and Kotani one-particle-ization (23) have only been started and applied to simple systems such as molecules [14, 13] and simple elements [15, 16].

Besides this big success to overcome a severe LDA/DFT shortcoming for semiconductor gaps, GW or G_0W_0 calculations also show a quasiparticle renormalization of the bandwidth. For alkali metals, electronic correlation are expected to be weak. Nonetheless experiments observe e.g. in Na a band narrowing (of the occupied bands) of 0.6 eV [22] compared to the nearly free electron theory. While GW [23] yields such a band narrowing, it is quantitatively with 0.3 eV

⁷An alternative, in particular for f electron systems, is to use LDA+ U as a starting point [12].

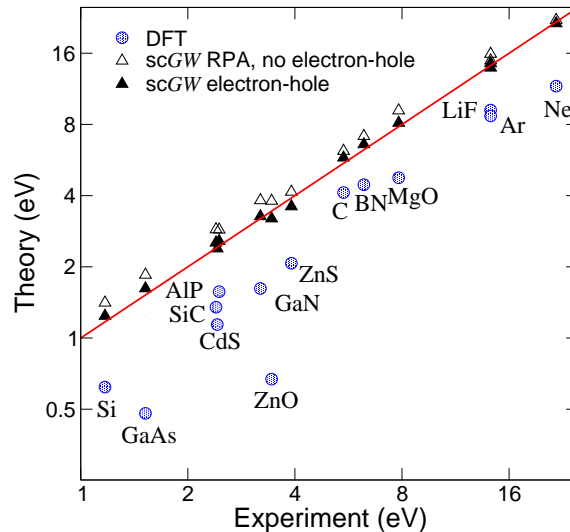


Fig. 12: *Theoretical vs. experimental band gap of DFT, self-consistent GW (scGW) without and with (some) vertex corrections (electron-hole ladder diagrams) (reproduced from [21]).*

only half as large as in experiment [23]. Of the more strongly correlated transition metals, Ni is best studied: here, the occupied d bandwidth is 1.2 eV smaller in experiment than LDA and there is a famous satellite peak at -6eV in the spectrum [24]. While GW [25] yields a band-narrowing of 1 eV which is surprisingly good (see Fig. 13), the satellite is missing. In fact, it can be identified as a (lower) Hubbard band whose description requires the inclusion of strong local correlation. This is possible by DMFT; and indeed the satellite is found in LDA+DMFT [26] and GW +DMFT calculations [27], see next section.

Besides the mentioned band-narrowing which is associated with a reduced quasiparticle weight or effective mass enhancement (related to the real part of the self energy), there is also the imaginary part of the GW self energy, which corresponds to a scattering rate. For Ag the GW scattering rate is reported to be in close agreement with the experimental one obtained from two-photon photoemission [28].

4 GW +DMFT

Since GW yields bandstructures similar to LDA (with the improvements for semiconductors discussed in the previous section) substituting the LDA part in LDA+DMFT by GW is very appealing from a theoretical point of view: Both approaches GW and DMFT are formulated in the same many-body framework, which does not only has the advantage of a more elegant combination, but also overcomes two fundamental problems of LDA+DMFT: (i) The screened Coulomb interaction employed for d - d or f - f interactions in DMFT can be straight forwardly calculated via W ; one does not need an additional constrained LDA approximation [29–31] to this end; (ii) the double counting problem, i.e., to subtract the LDA/DFT contribution of the local d - d or f - f interaction which is included a second time in DMFT, can be addressed in a rigorous manner since for GW +DMFT we actually know which Feynman diagram is counted

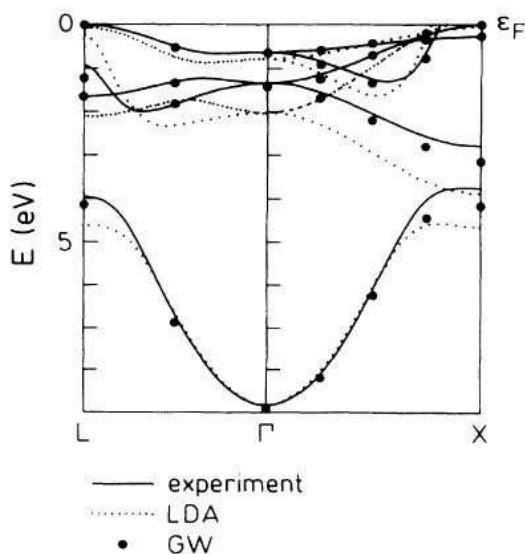


Fig. 13: Experimental, LDA and GW bandstructure of Ni (reproduced from [25]).

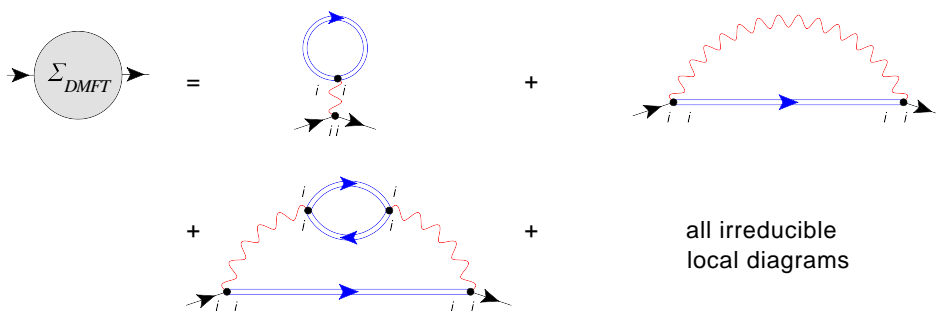


Fig. 14: The DMFT self energy is calculated from the local contribution of all (one-particle irreducible) Feynman diagrams.

twice.

Biermann *et al.* [27] proposed GW+DMFT, which they discuss from a functional integral point of view: a GW functional and a local impurity functional are added; the derivatives yield the mixed GW+DMFT equations. From a Feynman diagrammatic point of view, this corresponds to adding the GW self energy, Fig. 1 and the DMFT self energy which is just given by the local contribution of all (one-particle irreducible) Feynman diagrams, see Fig. 14. From these, the *local* screened exchange GW and the doubly counted Hartree term need obviously to be explicitly subtracted for not counting any diagram twice.

This results in the algorithm Fig. 15. Here, we leave the short-hand notation of Section 2 and 3.1 with 1, 2 since GW is diagonal in ω and \mathbf{k} and DMFT is diagonal in ω and site indices. Let us briefly discuss the GW+DMFT algorithm step-by-step; for more details see [32]:

- In most GW calculations, the starting point is a conventional LDA calculation (or another suitably chosen generalized Kohn-Sham calculation), yielding an electron density $\rho(\mathbf{r})$, bandstructure $\epsilon^{\text{LDA}}(\mathbf{k})$ and also an LDA Green function $G_{\mathbf{k}}(\omega)$ (the latter is calculated as

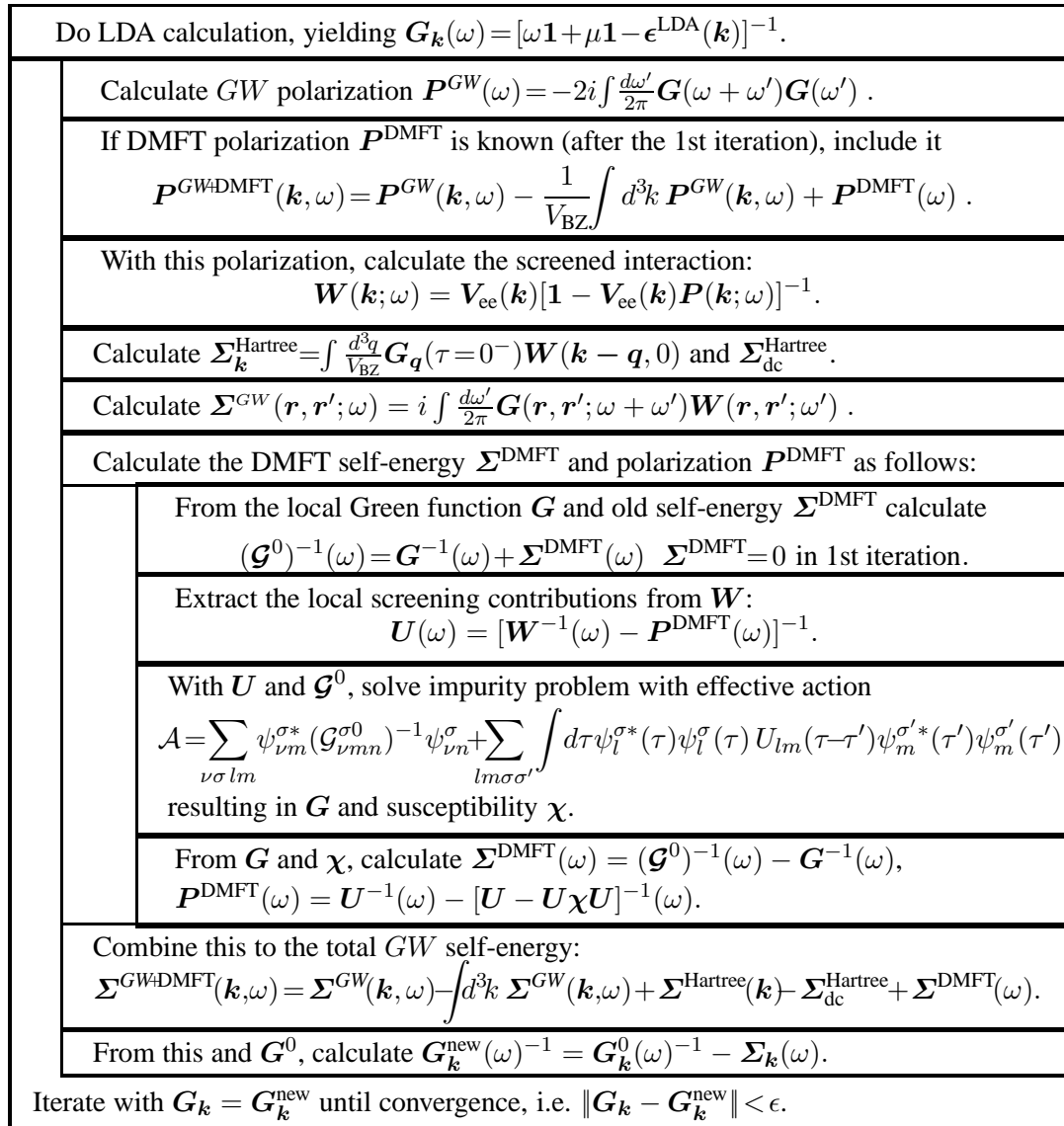


Fig. 15: Flow diagram of the $GW+DMFT$ algorithm (reproduced from [32]).

in the first line of Fig. 15, where bold symbols denote an (orbital) matrix representation).

- From this Green function, the independent particle polarization operator \mathbf{P}^{GW} is calculated convoluting two Green functions (2nd line of flow diagram Fig. 15). Note there is a factor of 2 for the spin.
- From the polarization operator in turn, the local polarization has to be subtracted since this can (and has to) be calculated more precisely within DMFT, which includes more than the RPA bubble diagram (after the first DMFT iteration).
- Next, the screened interaction \mathbf{W} is calculated from the bare Coulomb interaction \mathbf{V}_{ee} and the overall polarization operator $\mathbf{P}^{GW+DMFT}$.
- Now, we are in the position to calculate the GW self energy. The first term is the Hartree

diagram, which can be calculated straight forwardly in imaginary time τ , yielding Σ^{Hartree} and the corresponding local contribution $\Sigma_{\text{dc}}^{\text{Hartree}}$, which we need to subtract later to avoid a double counting as it is also contained in the DMFT.

- The second diagram is the exchange from Fig. 1 which has the form G times W for the GW self energy.
- This GW self energy has to be supplemented by the local DMFT self energy, which together with the DMFT polarization operator is calculated in the following four steps:
 1. The non-interacting Green function \mathcal{G}^0 which defines a corresponding Anderson impurity model is calculated.
 2. The local (screened) Coulomb interaction $U(\omega)$ has to be determined *without* the local screening contribution, since the local screening will be again included in the DMFT. That is we have to “unscreen” W for these contributions.
 3. The Anderson impurity model defined by \mathcal{G}^0 and $U(\omega)$ has to be solved for its interacting Green function $G(\omega)$ and two-particle charge susceptibility χ . This is numerically certainly the most demanding step.
 4. From this $G(\omega)$ and \mathcal{G}^0 , we obtain a new DMFT self energy $\Sigma(\omega)$ and from the charge susceptibility a new DMFT polarization operator.
- All three terms of the self-energy have now to be added; and the local screened exchange and Hartree contribution need to be subtracted to avoid a double counting.
- From this $GW+DMFT$ self energy we can finally recalculate the Green function and iterate until convergence.

The flow diagram already shows that the $GW+DMFT$ approach is much more involved than LDA+DMFT. However, it has the advantage that the double counting problem is solved and also the Coulomb interaction is calculated *ab initio* in a well defined and controlled way. Hence, no *ad hoc* formulas or parameters need to be introduced or adjusted.

For defining a well defined interface between GW and DMFT a particular problem is that GW is naturally formulated in real or \mathbf{k} space and is presently implemented, e.g., in the LMTO [5] or PAW basis [20]. However, on the DMFT side we do need to identify the interacting *local* d - or f -orbitals on the sites of the transition metal or rare earth/lanthanoid sites, respectively. The switching between these two representations is non-trivial. It can be done by a downfolding [33, 34] or a projection onto Wannier orbitals, e.g., using maximally localized Wannier orbitals [35, 37] or a simpler projection onto the d (or f) part of the wave function within the atomic spheres [36, 38]. However, not only the one-particle wave functions and dispersion relation need to be projected onto the interacting subspace but also the interaction itself. To approach the latter, a constrained random phase approximation (cRPA) method has been proposed [39, 7] and improved by disentangling the d (or f)-bands [40]. The latter improvement now actually

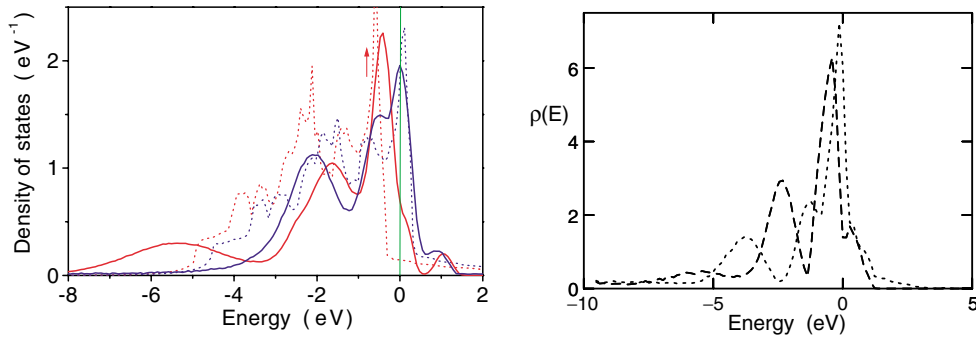


Fig. 16: Spectrum (\mathbf{k} -integrated) of Ni [left: LDA+DMFT (solid lines), spinpolarized LDA (LSDA) (dotted lines); right: GW +DMFT]. The two lines represent the minority- and majority-spin spectrum respectively. At roughly -6eV , a satellite peak is clearly visible in the majority-spin spectrum (reproduced from [26] and [27], respectively).

allows us to do cRPA in practice. For the calculation of the two-particle polarization operators and interactions, Aryasetiawan *et al.* [41] even proposed to use a third basis: the optimal product basis.

On the DMFT side, the biggest open challenges are to actually perform the DMFT calculations with a frequency dependent Coulomb interaction $U(\omega)$ and to calculate the DMFT charge susceptibility or polarization operator.

As the fully self-consistent GW +DMFT scheme is a formidable task, Biermann *et al.* [27] employed a simplified implementation for their GW +DMFT calculation of Ni, which is actually the only GW +DMFT calculation hitherto: For the DMFT impurity problem, only the local Coulomb interaction between d orbitals was included and its frequency dependence was neglected $\mathbf{W}(\omega) \approx \mathbf{W}(0)$. Moreover, only one iteration step has been done, calculating the inter-site part of the self energy by GW with the LDA Green function as an input and the intra-site part of the self energy by DMFT (with the usual DMFT self-consistency loop). The GW polarization operator \mathbf{P}^{GW} was calculated from the LDA instead of the GW Green function. This is, actually, common practice even for conventional GW calculations which are often of the G_0W_0 form (see Section 3.2).

Fig. 16 (right panel) shows the GW +DMFT \mathbf{k} -integrated spectral function of Ni which is similar to LDA+DMFT results (left panel). Both approaches yield a satellite peak at $\approx -6\text{eV}$.

5 All of that: *ab initio* D Γ A

From the Hedin equations, it seems to be much more natural to connect (i) the GW physics of screened exchange and (ii) strong, local correlations on the two-particle level than on the one particle level as done in GW +DMFT. In the Hedin equations, the natural starting point is the two-particle (particle-hole) irreducible vertex. A generalization of DMFT to n -particle correlation functions is the dynamical vertex approximation (D Γ A) [42] which approximates the

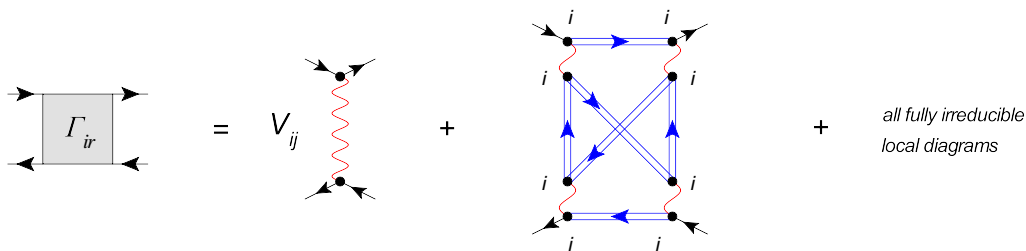


Fig. 17: In *ab initio* D Γ A we take as the fully irreducible vertex the bare Coulomb interaction plus all local vertex corrections (only one such diagram is exemplarily shown) [44].

n -particle fully irreducible⁸ vertex Γ_{ir} by the corresponding local contribution of all Feynman diagrams. For $n = 1$ the one-particle irreducible vertex is the self energy so that D Γ A yields the DMFT. For $n = 2$, we obtain non-local correlations on all length scales and can calculate, e.g., the critical exponents of the Hubbard model [43].

Recently, some of us have proposed to use this D Γ A *ab initio* for materials calculation [44]. The fully irreducible vertex Γ_{ir} is then given by the bare Coulomb interaction, which possibly is non-local, and all higher order local Feynman diagrams, see Fig. 17. From Γ_{ir} the full (reducible) vertex is calculated via the parquet equations [11]. The calculation of the local part of Γ_{ir} only requires us to calculate the two-particle Green functions of a single-site Anderson impurity model, which is well doable even for realistic multi-orbital models. For the parquet equations on the other hand, there has been some recent progress [45].

As a simplified version of *ab initio* D Γ A one can restrict oneself to a subset of the three channels of the parquet equations, as was done in [42,43]. In this case one has to solve the Bethe-Salpeter equation with the particle-hole irreducible vertex Γ_{ph} instead of the parquet equations with the fully irreducible vertex Γ_{ir} . That is, our approximation to the Hedin equations is to take the local Γ_{ph}^* (all Feynman diagrams given by the local Green function and interaction) in the Hedin equation (8). In practice, one solves an Anderson impurity model numerically to calculate Γ_{ph}^* . Full and simplified version of *ab initio* D Γ A contain the diagrams (and physics) of GW, DMFT as well as non-local correlations which are responsible for (para-)magnons, (quantum) criticality and “all that”.

Acknowledgment

Support of the Austrian Science Fund (FWF) through I597 (Austrian part of FOR 1346 with the Deutsche Forschungsgemeinschaft as lead agency) is gratefully acknowledged.

⁸Fully irreducible means, cutting any two Green function lines does not separate the diagram into two parts. It is even more restrictive (less diagrams) than the particle-hole irreducible vertex (whose diagrams can be reducible e.g. in the particle-particle channel).

Appendices

A Additional steps: equation of motion

In this appendix a detailed explanation is given how to derive the Hedin equation of motion for Σ , i.e., equation (16), from the standard equation of motion (15). In a first step Γ is expressed in terms of Γ^* . In order to keep the notation simple, the arguments of all functions are omitted and the functions are considered as operators.

The starting point of the calculations are the Bethe-Salpeter equations for Γ and Γ^* :

$$\begin{aligned}\Gamma &= \Gamma_{\text{ph}} + \Gamma\beta G G \Gamma_{\text{ph}} \implies \Gamma = \Gamma_{\text{ph}}(1 - \beta G G \Gamma_{\text{ph}})^{-1} \\ \Gamma^* &= \Gamma_{\text{ph}}^* + \Gamma_{\text{ph}}^* \beta G G \Gamma^* \implies \Gamma_{\text{ph}}^* = \Gamma^*(1 + \beta G G \Gamma^*)^{-1}.\end{aligned}\quad (24)$$

Using equation (10), i.e., $\Gamma_{\text{ph}} = \Gamma_{\text{ph}}^* + V$, one gets:

$$\Gamma = \left(\Gamma^*(1 + \beta G G \Gamma^*)^{-1} + V \right) \frac{1}{1 - \beta G G \left(\Gamma^*(1 + \beta G G \Gamma^*)^{-1} + V \right)} \quad (25)$$

Multiplying this equation with $1 = (1 + \beta G G \Gamma^*)^{-1}(1 + \beta G G \Gamma^*)$ and using the standard relations for operators, $A^{-1}B^{-1} = (BA)^{-1}$ leads to:

$$\Gamma = \left(\Gamma^*(1 + \beta G G \Gamma^*)^{-1} + V \right) \underbrace{V^{-1}V}_1 \left(1 - \underbrace{(\beta G G + \beta G G \Gamma^* \beta G G)}_P V \right)^{-1} \left(1 + \beta G G \Gamma^* \right), \quad (26)$$

where $1 = V^{-1}V$ was inserted and the definition for the polarization operator, equation (7). Now one can use the second Hedin equation (6), which can be rewritten as $W = V(1 - PV)^{-1}$. Inserting this relation into the equation for Γ , one arrives at the following result:

$$\Gamma = \left(\Gamma^*(1 + \beta G G \Gamma^*)^{-1} \right) V^{-1} W \left(1 + \beta G G \Gamma^* \right) + W \left(1 + \beta G G \Gamma^* \right). \quad (27)$$

Another formulation of the second Hedin equation gives $V^{-1} = W^{-1} + P = W^{-1} + (1 + \beta G G \Gamma^*)\beta G G$. Replacing V^{-1} by this expression gives:

$$\begin{aligned}\Gamma &= \Gamma^* + \Gamma^* \beta G G W (1 + \beta G G \Gamma^*) + W (1 + \beta G G \Gamma^*) \\ &= \Gamma^* + \Gamma^* \beta G G W + \Gamma^* \beta G G W \beta G G \Gamma^* + W \beta G G \Gamma^* + W.\end{aligned}\quad (28)$$

This equation shows how the full Γ is related to the Γ^* . Diagrammatically this relation is shown in Fig. 18.

In the next step Γ as given in equation (28) is inserted into equation (15), yielding

$$\begin{aligned}\Sigma &= -GV\beta G G \Gamma - GV \\ &= -GV \underbrace{(\beta G G + \beta G G \Gamma^* \beta G G)}_P W - GV \left(\underbrace{(\beta G G + \beta G G \Gamma^* \beta G G)}_P W + 1 \right) \beta G G \Gamma^* - GV.\end{aligned}\quad (29)$$

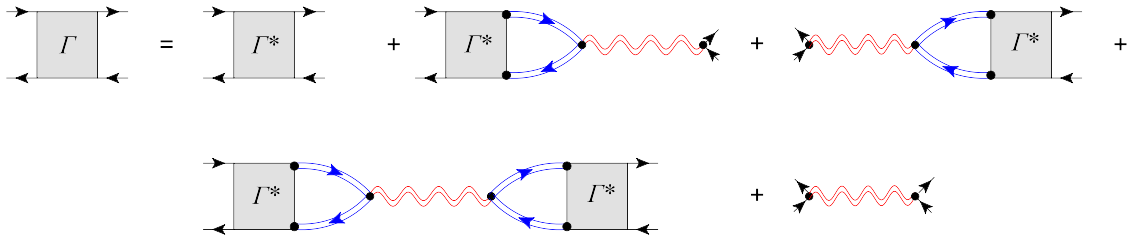


Fig. 18: Relation between Γ and Γ^* in terms of Feynman diagrams.

From the second Hedin equation it follows that $VPW = W - V$. Inserting this relation into Equation (29) yields:

$$\begin{aligned} \Sigma &= -G(W - V) - G(W - V)\beta G G \Gamma^* - GV\beta G G \Gamma^* - GV = \\ &= -GW - GW\beta G G \Gamma^*, \end{aligned} \quad (30)$$

which is exactly equation (16).

Let us also, at this point, mention that Γ should satisfy an important relation:

$$\Gamma(11'; 22') = -\Gamma(12'; 21') \quad (31)$$

This relation is known as **crossing symmetry** (see e.g. [11], equation 7.5) and is simply a consequence of the Pauli-principle: Exchanging two identical fermions leads to a $-$ sign in the wave function. The screened interaction W , however, does not fulfill this crossing symmetry. Therefore, setting $\Gamma^* = 0$, as it is done in the GW -approximation leads to $\Gamma = W$ (see Fig. 18) which violates this crossing symmetry, i.e. it violates the Pauli-principle.

References

- [1] P. Hohenberg and W. Kohn, Phys. Rev. **136** B864 (1964)
- [2] R. O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61** 689 (1989)
- [3] See Lecture notes of P. Blöchl.
- [4] L. Hedin, Phys. Rev. A **139** 796 (1965)
- [5] For a review see, F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61** 237 (1998)
- [6] A. A. Abrikosov, L. P. Gorkov and I. E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Dover, New York, 1963)
- [7] For more details on RPA, see Lecture notes of F. Aryasetiawan
- [8] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62** 324 (1989)
- [9] A. Georges, G. Kotliar, W. Krauth and M. Rozenberg, Rev. Mod. Phys. **68** 13 (1996)
- [10] For more details on DMFT, see Lecture notes of M. Kollar
- [11] N. E. Bickers, “Self consistent Many-Body Theory of Condensed Matter” in *Theoretical Methods for Strongly Correlated Electrons CRM Series in Mathematical Physics Part III*, Springer (New York 2004)
- [12] H. Jiang and R. Gomez-Abal and P. Rinke and M. Scheffler, Phys. Rev. Lett. **102**, 126403 (2009); Phys. Rev. B **82**, 045108 (2010)
- [13] A. Stan, N. E. Dahlen, and R. van Leeuwen, J. Chem. Phys. **130**, 114105 (2009)
- [14] C. Rostgaard, K. W. and Jacobsen, and K. S. Thygesen Phys. Rev. B **81**, 085103 (2010)
- [15] W.-D. Schöne and A. G. Eguiluz, Phys. Rev. Lett. **81**, 1662 (1998)
- [16] A. Kutepov, S. Y. Savrasov, G. and Kotliar, Phys. Rev. B **80**, 041103 (2009)
- [17] R. W. Godby, M. Schlüter and L. J. Sham, Phys. Rev. B **37** 10159 (1988)
- [18] S. V. Faleev, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **93**, 126406 (2004)
- [19] A. N. Chantis, M. van Schilfgaarde, and T. Kotani, Phys. Rev. Lett. **96**, 086405 (2006)
- [20] M. Shishkin and G. Kresse, Phys. Rev. B **74**, 035101 (2006)
- [21] M. Shishkin, M. Marsman, and G. Kresse, Phys. Rev. Lett. **99**, 246403 (2007)
- [22] I.-W. Lyo and W. E. Plummer, Phys. Rev. Lett. **60**, 1558 (1988)

- [23] J. E. Northrup, M. S. Hybertsen, and S. G. Louie, Phys. Rev. Lett. **59** 819 (1987); Phys. Rev. B **39**, 8198 (1989)
- [24] S. Hüfner, G. K. Wertheim, N. V. Smith, and M. M. Traum, Solid State Comm. **11** 323 (1972)
- [25] F. Aryasetiawan, Phys. Rev. B **46**, 13051 (1972)
- [26] A. I. Lichtenstein, M. I. Katsnelson, and G. Kotliar, Phys. Rev. Lett. **87** 67205 (2001)
- [27] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90** 086402 (2003)
- [28] R. Keyling, W.-D.Schöne, and W. Ekardt, Phys. Rev. B **61**, 1670 (2000)
- [29] P. H. Dederichs, S. Blügel, R. Zeller and H. Akai, Phys. Rev. Lett. **53** 2512 (1984)
- [30] A. K. McMahan, R. M. Martin and S. Satpathy, Phys. Rev. B **38** 6650 (1988)
- [31] O. Gunnarsson, O. K. Andersen, O. Jepsen and J. Zaanen, Phys. Rev. B **39** 1708 (1989)
- [32] K. Held, Advances in Physics **56**, 829 (2007)
- [33] O. K. Andersen, T. Saha-Dasgupta, R. W. Tank, C. Arcangeli, O. Jepsen and G. Krier, In *Lecture notes in Physics*, edited by H. Dreyse (Springer, Berlin, 1999)
- [34] O. K. Andersen, T. Saha-Dasgupta, S. Ezhov, L. Tsetseris, O. Jepsen, R. W. Tank and C. A. G. Krier, Psi-k Newsletter # **45** 86 (2001), http://psi-k.dl.ac.uk/newsletters/News_45/Highlight_45.pdf
- [35] N. Marzari and D. Vanderbilt, Phys. Rev. B **56** 12847 (1997)
- [36] V.I. Anisimov, D.E. Kondakov, A.V. Kozhevnikov, I.A. Nekrasov, Z.V. Pchelkina, J.W. Allen, S.-K. Mo, H.-D. Kim, P. Metcalf, S. Suga, A. Sekiyama, G. Keller, I. Leonov, X. Ren, D. Vollhardt, Phys. Rev. B **71**, 125119 (2005)
- [37] J. Kunes, R. Arita, P. Wissgott, A. Toschi, H. Ikeda, and K. Held, Comp. Phys. Comm. **181**, 1888 (2010)
- [38] M. Aichhorn, L. Pourovskii, V. Vildosola, M. Ferrero, O. Parcollet, T. Miyake, A. Georges, and S. Biermann Phys. Rev. B **80**, 085101 (2009)
- [39] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann and A. I. Lichtenstein, Phys. Rev. B **70**, 195104 (2004)
- [40] T. Miyake, F. Aryasetiawan, M. Imada, arXiv:0906.1344
- [41] F. Aryasetiawan, S. Biermann and A. Georges, In *Proceedings of the conference on "Coincidence Studies of Surfaces, Thin Films and Nanostructures"*, edited by A. Gonis (Wiley, New York, 2004)

- [42] A. Toschi, A. A. Katanin and K. Held, *Phys. Rev. B* 75, 045118 (2007); *Prog. Theor. Phys. Suppl.* 176, 117 (2008); *Phys. Rev. B* 80, 075104 (2009)
- [43] G. Rohringer, A. Toschi, A. A. Katanin and K. Held arxiv.org/abs/1104.1919
- [44] A. Toschi, G. Rohringer, A. A. Katanin, K. Held arxiv.org/abs/1104.2118
- [45] S.-X. Yang, H. Fotsos, J. Liu, T. A. Maier, K. Tomko, E. F. D'Azevedo, R. T. Scaletar, T. Pruschke, M. Jarrell, *Phys. Rev. E* 80, 046706 (2009)

14 Challenges from Experiment

L. Hao Tjeng

Max Planck Institute for Chemical Physics of Solids

Dresden

Contents

1	Introduction	2
1.1	Strongly correlated materials	2
1.2	Electronic structure and synchrotron based spectroscopies	2
1.3	Challenges for theory	3
2	Spin-state transition in LaCoO_3	4
2.1	Introduction	4
2.2	Experimental	4
2.3	Results: XAS	5
2.4	Analysis	5
2.5	Results: MCD	7
2.6	Discussion	9
3	Electronic structure of NiO	9
3.1	Introduction	9
3.2	Experimental	11
3.3	Results	12
3.4	Analysis	13
3.5	Non-local screening	15
3.6	Multi-site analysis	15
3.7	Conclusion	16

1 Introduction

1.1 Strongly correlated materials

One of the most intriguing aspects of transition-metal and rare-earth materials is the wide variety and richness of their physical properties. These are often quite spectacular and include phenomena such as metal-insulator transitions (MIT), colossal magneto-resistance (CMR) behavior, and superconductivity [1, 2]. It is well accepted by now that the interplay between electron correlation and band formation effects is somehow responsible for these unexpected and novel phenomena. Although conceptually clean and beautiful, theoretical simplifications in terms of, for instance, a Heisenberg model or a single band Hubbard model turn out to be inadequate [3–5]. It now becomes more and more clear that a full identification of the relevant charge, orbital and spin degrees of freedom of the metal ions involved is needed to understand the intricate balance between band formation and electron-correlation effects. An important example is the manganates [6], where orbital ordering and charge distribution [7, 8] of the Mn ions play an important role for its CMR behavior. For the newly synthesized layered cobaltates, it is the spin-state transitions that are thought to govern the CMR and MIT phenomena [9, 10]. Another example is V_2O_3 [11–13], for which it has been discovered that the orbital occupations across the various MITs change dramatically, leading to a switch of the nearest-neighbor spin-spin correlations so that in turn the effective band widths are strongly modified.

An important characteristic determining the charge, orbital and spin degrees of freedom is the presence of the so-called atomic-like multiplet interactions on the transition-metal and rare-earth ions [14, 15]. The ground state and near ground state properties could depend very much on the intricate details of these local Coulomb and exchange interactions, and how they play out in the presence of crystal or ligand fields as well as band formation. It is not trivial to describe these multiplet effects using ab-initio mean-field theories. LDA+U type of approaches, for example, utilize occupation number operators of the type $U_{ij}d_i^\dagger d_i d_j^\dagger d_j$, but these do not capture the spin and orbital flip processes contained in the $U_{ijkl}d_i^\dagger d_j^\dagger d_k^\dagger d_l$ terms of the full Hamiltonian [16], leading to major errors in the total energy level diagram for both the ground state problem as well as for the excited states. In this respect an approach based on the LDA+DMFT is perhaps more promising and hopefully its implementation can be made in the very near future. It is also not trivial for real materials to determine experimentally their energy levels associated with the multiplet interactions. One often needs synchrotron based spectroscopies in order to have sufficient dynamic range in energy to map out the excitations as well as to have well defined matrix elements to quantitatively analyze the spectra.

1.2 Electronic structure and synchrotron based spectroscopies

Electron spectroscopies are powerful tools to unravel the basic electronic structure of materials. Use can be made of the extremely large dynamic range in energy that these forms of spectroscopies can cover. By studying excitation spectra in the energy range from several eV up to several hundreds of eV, one can obtain direct information about the 'bare' electrons, e.g. the

charge, spin, and orbital state of the ions that make up the correlated material. By measuring the excitation spectra in the vicinity of the chemical potential with ultra-high resolution, one can find directly the momentum dependent behavior of the 'dressed' electrons, i.e., quasi particles. In the case of photoemission or inverse photoemission, we should stress that it is not the excitation energy that determines whether the measurement probes the high or low energy scale physics. Instead, it is the resolution with which the measurement is carried out, since this determines the quality of the observed features, and thus of the measured low energy scale phenomena if the measurement is focussed on the region near the chemical potential.

There are several forms of electron spectroscopies, and which one to apply depends very much on the material and the problem that needs to be solved. While photoemission and inverse photoemission are well established techniques [17], the use of synchrotron radiation based high energy spectroscopies such as linearly and circularly polarized soft-X-ray absorption spectroscopy is relatively new [18–20,22,21]. This type of spectroscopy has been developed into full maturity only in the last 15 years, both in terms of instrumentation as well as in terms of a quantitative theoretical analysis of the spectra, which are often dominated by multiplet structures. These spectra are not only element specific, but above all, their multiplet structures are extremely sensitive to the charge, spin and orbital state of the ion due to the very effective dipole selection rules associated with the K ($1s \rightarrow 2p$), $L_{2,3}$ ($2p \rightarrow 3d$), and $M_{4,5}$ ($3d \rightarrow 4f$) transitions for the oxygen, transition metal and rare-earth ions, respectively. In fact, the specificity to the initial state symmetry is so large, that one does not need a very good energy resolution in order to measure which of the possible initial state symmetries is occupied, since different symmetries lead to completely different multiplet structured spectra. In other words, an experimental energy resolution of 500 meV is often enough to distinguish initial states that are different by not more than a few meV or less in energy, making the technique extremely valuable for the study of phase transitions which usually occur in the temperature range up to roughly room temperature ($kT = 25$ meV).

1.3 Challenges for theory

In the following, we will address two long standing topics in the research field of strongly correlated transition metal oxides, namely the spin-state transition in LaCoO_3 and the electronic structure of NiO . Synchrotron based spectroscopic experiments are presented. The results on the first topic show how crucial it is to take the full multiplet theory into account in order to be able to understand the magnetic properties of a partially filled t_{2g} material, while the results from the second topic provide a clear cut example how on-site multiplet structures could be affected by band formation. These experimental results illustrate the challenges faced by ab-initio based theories, namely to describe accurately both the (near) ground state properties and excitation spectra of strongly correlated transition metal oxides. We hope that LDA+DMFT and its extensions will provide a further major contribution in this research field.

2 Spin-state transition in LaCoO_3

2.1 Introduction

LaCoO_3 shows a gradual non-magnetic to magnetic transition with temperature, which has been interpreted originally four decades ago as a gradual population of high spin (HS, $t_{2g}^4 e_g^2$, $S = 2$) excited states starting from a low spin (LS, t_{2g}^6 , $S = 0$) ground state [23–30]. This interpretation continued to be the starting point for experiments carried out up to roughly the first half of the 1990's [31–34]. All this changed with the theoretical work in 1996 by Korotin *et al.*, who proposed on the basis of local density approximation + Hubbard U (LDA+U) band structure calculations, that the excited states are of the intermediate spin (IS, $t_{2g}^5 e_g^1$, $S = 1$) type [35]. Since then many more studies have been carried out on LaCoO_3 with the majority of them [36–49] claiming to have proven the presence of this IS mechanism. In fact, this LDA+U work is so influential [50] that it forms the basis of most explanations for the fascinating properties of the recently synthesized layered cobaltate materials, which show giant magneto resistance as well as metal-insulator and ferro-ferri-antiferro-magnetic transitions with various forms of charge, orbital and spin ordering [51, 52]. There have been several attempts made since 1996 in order to revive the LS-HS scenario [53–57], but these were overwhelmed by the above mentioned flurry of studies claiming the IS mechanism [36–49]. Moreover, a new investigation using inelastic neutron scattering (INS) has appeared in [58] making again the claim that the spin state transition involves the IS states. In the following we describe our soft X-ray absorption spectroscopy (XAS) and magnetic circular dichroism (MCD) experiments at the Co- $L_{2,3}$ edge [59] to evaluate the validity of the various scenarios.

2.2 Experimental

Single crystals of LaCoO_3 have been grown by the traveling floating-zone method in an image furnace. The magnetic susceptibility was measured using a Quantum Design vibrating sample magnetometer (VSM), reproducing the data reported earlier [41]. The Co- $L_{2,3}$ XAS measurements were performed at the Dragon beamline of the National Synchrotron Radiation Research Center (NSRRC) in Taiwan with an energy resolution of 0.3 eV. The MCD spectra were collected at the ID08 beamline of the European Synchrotron Radiation Facility (ESRF) in Grenoble with a resolution of 0.25 eV and a degree of circular polarization close to 100% in a magnetic field of 6 Tesla. Clean sample areas were obtained by cleaving the crystals *in-situ* in chambers with base pressures in the low 10^{-10} mbar range. The Co- $L_{2,3}$ spectra were recorded using the total electron yield method (TEY). O- K XAS spectra were collected by both the TEY and the bulk sensitive fluorescence yield (FY) methods, and the close similarity of the spectra taken with these two methods verifies that the TEY spectra are representative for the bulk material. A CoO single crystal is measured *simultaneously* in a separate chamber to obtain relative energy referencing with better than a few meV accuracy, sufficient to extract reliable MCD spectra.

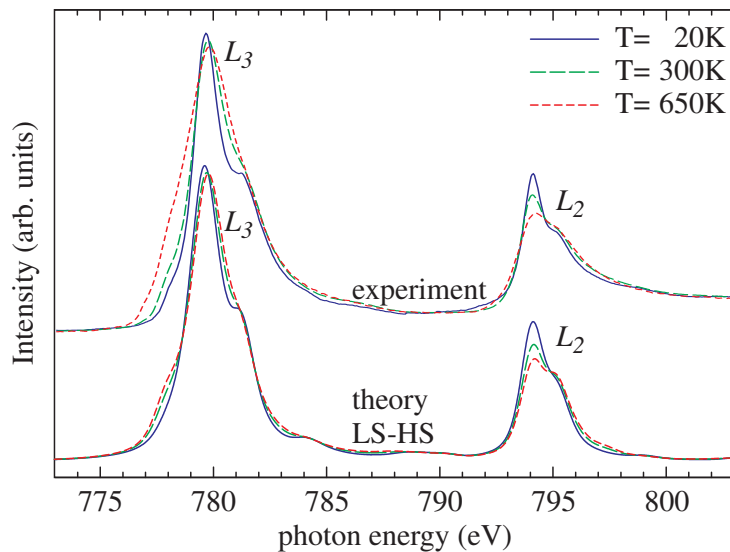


Fig. 1: Experimental Co- $L_{2,3}$ XAS spectra taken from LaCoO_3 at various temperatures between 20 and 650 K, together with the corresponding theoretical isotropic spectra calculated using a CoO_6 cluster in the LS-HS scenario. For clarity, only the 20, 300 and 650 K spectra are shown.

2.3 Results: XAS

Fig. 1 shows the set of Co- $L_{2,3}$ XAS spectra of LaCoO_3 taken for a wide range of temperatures. The set is at first sight similar to the one reported earlier [61], but it is in fact essentially different in details. First of all, our set includes a low temperature (20 K) spectrum representative for the LS state, and second, our spectra do not show a pronounced shoulder at 777 eV photon energy which is characteristic for the presence of Co^{2+} impurities [62]. The extended temperature range and especially the purity of the probed samples provide the required sensitivity for the spin-state related spectral changes.

The spectra are dominated by the Co $2p$ core-hole spin-orbit coupling which splits the spectrum roughly in two parts, namely the L_3 ($h\nu \approx 780$ eV) and L_2 ($h\nu \approx 796$ eV) white lines regions. The line shape of the spectrum depends strongly on the multiplet structure given by the Co $3d$ - $3d$ and $2p$ - $3d$ Coulomb and exchange interactions, as well as by the local crystal fields and the hybridization with the O $2p$ ligands. Unique to soft X-ray absorption is that the dipole selection rules are very effective in determining which of the $2p^5 3d^{n+1}$ final states can be reached and with what intensity, starting from a particular $2p^6 3d^n$ initial state ($n=6$ for Co^{3+}) [20, 21]. This makes the technique extremely sensitive to the symmetry of the initial state, e.g., the spin state of the Co^{3+} [52].

2.4 Analysis

We now simulate the spectrum of a Co^{3+} ion in the LS state using the successful configuration interaction cluster model that includes the full atomic multiplet theory and the hybridization with the O $2p$ ligands [20–22]. The CoO_6 cluster is taken to have the octahedral symmetry and the parameters are the same as the ones which successfully reproduce the spectrum of LS

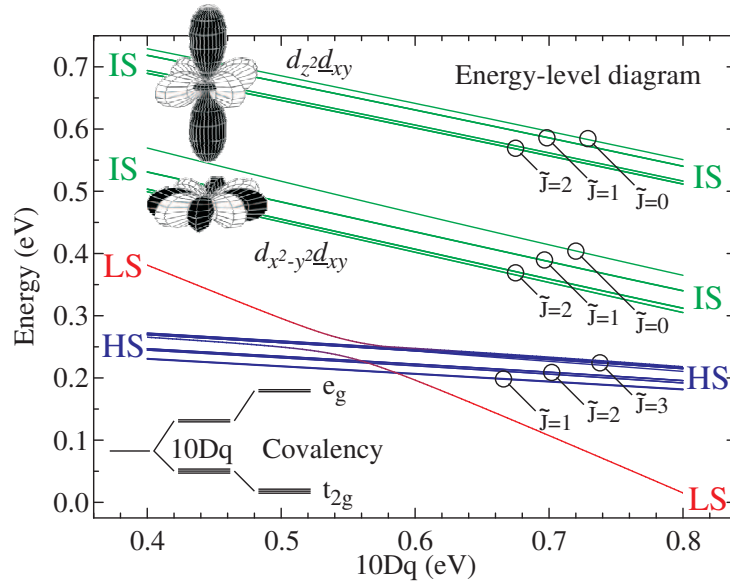


Fig. 2: Energy level diagram of a CoO_6 cluster [63] as a function of the ionic part of the crystal field splitting $10Dq$.

EuCoO_3 [52,63]. The result with the ionic part of the crystal field splitting set at $10Dq = 0.7$ eV is shown in Fig. 1 and fits well the experimental spectrum at 20 K.

Next we analyze the spectra for the paramagnetic phase. We use the same cluster keeping the O'_h symmetry, and calculate the total energy level diagram as a function of $10Dq$, see Fig. 2. We find that the ground state of the cluster is either LS or HS (and never IS) with a cross-over at about $10Dq = 0.58$ eV [64]. We are able to obtain good simulations for the spectra at all temperatures, see Fig. 1, provided that they are made from an incoherent sum of the above mentioned LS cluster spectrum calculated with $10Dq = 0.7$ eV and a HS cluster spectrum calculated with $10Dq = 0.5$ eV. It is not possible to fit the entire temperature range using one cluster with one particular temperature-independent $10Dq$ value for which the ground state is LS-like and the excited states HS-like. Moreover, each of these two $10Dq$ values have to be sufficiently far away from the LS-HS crossover point to ensure a large enough energy separation between the LS and HS so that the two do not mix due to the spin-orbit interaction. Otherwise, the calculated low temperature spectrum, for instance, will disagree with the experimental one. All this indicates that LaCoO_3 at finite temperatures is an inhomogeneous mixed spin state system.

The temperature dependence has been fitted by taking different ratios of LS and HS states contributing to the spectra. The extracted HS percentage as a function of temperature is shown in Fig. 3a. The corresponding effective activation energy is plotted in Fig. 3b. It increases with temperature and varies between 20 meV at 20 K to 80 meV at 650 K, supporting a recent theoretical analysis of the thermodynamics [57]. Here we would like to point out that these numbers are of the order $k_B T$ and reflect total energy differences which include lattice relaxations [57] as sketched in the inset of Fig. 3b. Without these relaxations, we have for the LS state ($10Dq = 0.7$ eV) an energy difference of at least 50 meV between the LS and the HS as

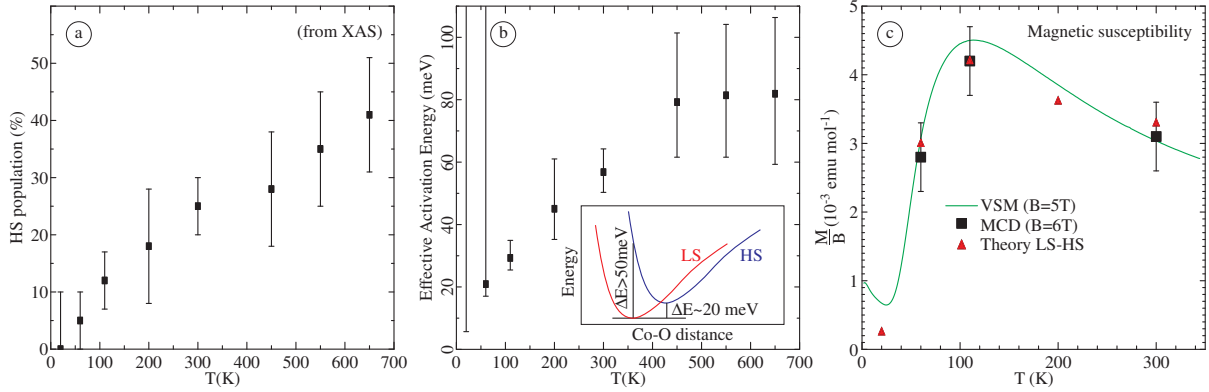


Fig. 3: (a) The percentage of the HS population as obtained from the XAS data. (b) Corresponding effective activation energy between the LS and the lowest HS state. The inset sketches the role of lattice relaxations. (c) Magnetic susceptibility measured by VSM (solid line), calculated from the cluster (red triangles) using the HS population of (a), and extracted from MCD data (black squares) of Fig. 4.

shown in Fig. 2. In such a frozen lattice, the energy difference is larger than $k_B T$. It is also so large that the ground state is indeed highly pure LS as revealed by the 20 K spectrum.

To check the validity of our analysis, we calculate the magnetic susceptibility using the CoO_6 cluster and the HS occupation numbers from Fig. 3 as derived from the XAS data. The results are plotted in Fig. 3c (red triangles) together with the magnetic susceptibility as measured by the VSM (solid line). We can observe clearly a very good agreement: the magnitude and its temperature dependence is well reproduced. This provides another support that the spin-state transition is inhomogeneous and involves lattice relaxations. A homogeneous LS-HS model, on the other hand, would produce a much too high susceptibility if it is to peak at 110 K [33, 34, 36, 41]. In addition, it is crucial to realize that the Van Vleck contribution to the magnetic susceptibility strongly depends on the intermixing between the LS and HS states. It is precisely this aspect which also sets the condition that the energy separation between the LS and HS states in the cluster should be larger than 50 meV, otherwise the calculated Van Vleck contribution would already exceed the experimentally determined total magnetic susceptibility at low temperatures. This in fact is a restatement of the above mentioned observation that the low temperature spectrum is highly pure LS.

2.5 Results: MCD

To further verify the direct link between the spectroscopic and the VSM magnetic susceptibility data, we carried out MCD experiments on LaCoO_3 at 60, 110 and 300 K, i.e. in the paramagnetic phase, using a 6 Tesla magnet. Fig. 4 shows XAS spectra taken with circularly polarized soft-X-rays with the photon spin parallel and antiparallel aligned to the magnetic field. The difference in the spectra using these two alignments is only of the order of 1%, but can nevertheless be measured reliably due to the good signal to noise ratio, stability of the beam, and the accurate photon energy referencing. The difference curves are drawn in the middle of Fig. 4 with a

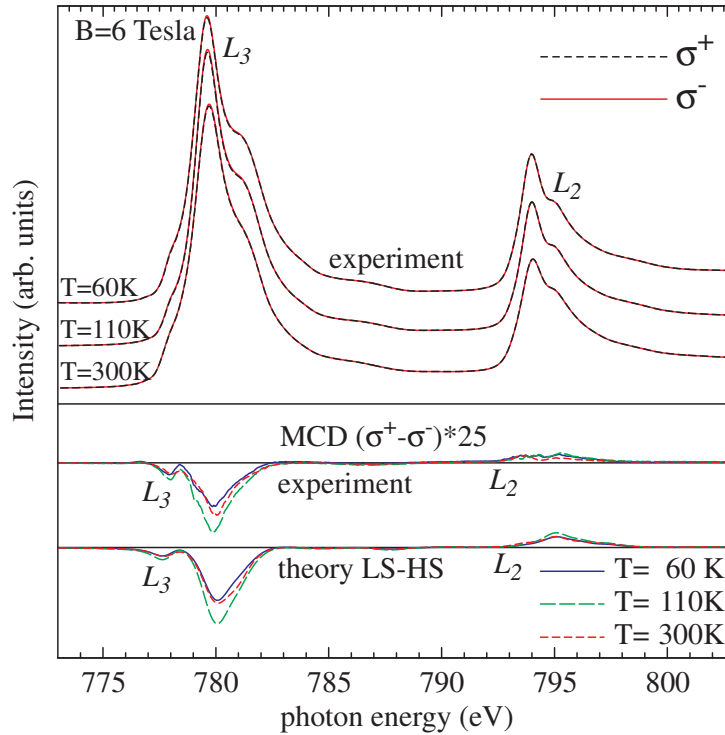


Fig. 4: Top curves: experimental Co- $L_{2,3}$ XAS spectra taken from LaCoO_3 at 60, 110, and 300 K using circularly polarized X-rays with the photon spin aligned parallel (black dotted line, σ^+) and antiparallel (red solid line, σ^-) to the 6 Tesla magnetic field. Middle curves: experimental MCD spectra defined as the difference between the two spin alignments. Bottom curves: theoretical MCD spectra calculated in the LS-HS scenario.

magnification of 25x. Hereby we have subtracted a small signal due to the presence of about 1.5% Co^{2+} impurities. We also plotted the simulated MCD spectra from the cluster model within the LS-HS scenario, and we can clearly observe a very satisfying agreement with the experiment. Alternatively, using the MCD sum-rules developed by Thole and Carra et al. [65, 66], we can extract directly the orbital (L_z) and spin ($2S_z$) contributions to the induced moments without the need to do detailed modeling [67]. This result normalized to the applied magnetic field is plotted in Fig. 3c, and we can immediately observe the close agreement with the VSM data.

An important aspect that emerges directly from the MCD experiments, is the presence of a very large induced orbital moment: we find that $L_z/S_z \approx 0.5$. This means that the spin-orbit coupling (SOC) must be considered in evaluating the degeneracies of the different levels, as is done for the energy level diagram in Fig. 2. Let us discuss the consequences for the HS state. We see that the 15-fold degenerate (3-fold orbital and 5-fold spin) HS state is split by the SOC. A t_{2g} electron has a pseudo orbital momentum of $\tilde{L}=1$ [68] which couples with the spin to a pseudo total momentum of $\tilde{J}=1, 2, \text{ or } 3$. The $\tilde{J}=1$ triplet is the lowest in energy and we find from our cluster that this state has $L_z=0.6$ and $S_z=1.3$, in good agreement with the experimental $L_z/S_z \approx 0.5$. Realizing that this state is a triplet with a spin momentum (S_z) so close to 1, it is no wonder that many studies incorrectly interpreted this state as an IS state. Its expectation

value for the spin ($\langle S^2 \rangle = S(S + 1)$) is however very close to 6 and the formal occupation numbers of the d_{z^2} and the $d_{x^2-y^2}$ orbitals are both equal to 1. This state is clearly a HS state and should not be confused with an IS state. We find a g -factor of 3.2, in good agreement with the values found from ESR [54, 56] and INS data [60].

2.6 Discussion

We have shown so far that the spin state transition in LaCoO_3 is in very good agreement with a LS–HS picture. The question now remains if it could also be explained within a LS–IS scenario. For that we first have to look what the IS actually is. The IS state has one hole in the t_{2g} shell and one electron in the e_g shell. Due to the strong orbital dependent Coulomb interactions, the strong-Jahn-Teller states of the type $d_{z^2}\underline{d}_{xy}$ and their x, y, z -cyclic permutations have much higher energies than the weak-Jahn-Teller $d_{x^2-y^2}\underline{d}_{xy}$ plus cyclic permutations. Here the underline denotes a hole (see Fig. 2). These weak-Jahn-Teller states indeed form the basis for the orbital ordering scheme as proposed for the IS scenario by Korotin *et al.* [35]. However, these real-space states do not carry a large orbital momentum, and are therefore not compatible with the values observed in the MCD measurements. Likewise, the strong Jahn-Teller-like local distortions in the IS state proposed by Maris *et al.* [46] would lead to a quenching of the orbital momentum. We therefore can conclude that the IS scenarios proposed so far have to be rejected on the basis of our MCD results. Moreover, an IS state would lead in general to a much larger van Vleck magnetism than a HS state. This is related to the fact that the LS state couples directly to the IS via the SOC, while the HS is not. To comply with the measured low temperature magnetic susceptibility, the energy difference between the LS and IS has to be 150 meV at least, making it more difficult to find a mechanism by which the maximum of the susceptibility occurs at 110 K. Finally, within the LS-IS scenario, we were not able to find simulations which match the experimental XAS and MCD spectra.

To summarize, we provide unique spectroscopic evidence that the spin state transition in LaCoO_3 can be well described by a LS ground state and a triply degenerate HS excited state, and that an inhomogeneous mixed-spin-state system is formed. The large orbital momentum revealed by the MCD measurements invalidates existing LS-IS scenarios. A consistent picture has now been achieved which also explains available magnetic susceptibility, ESR and INS data.

3 Electronic structure of NiO

3.1 Introduction

NiO is a benchmark system in solid state physics. It crystallizes in the NaCl structure, has a partially filled $3d$ shell ($\text{Ni}^{2+} d^8$), and is an antiferromagnetic insulator with a Néel temperature of 523 K [69]. It was pointed out early on by de Boer and Verwey [70] that many of the properties of the $3d$ transition metal compounds do not agree with the predictions of band theory, e.g., standard band theory predicts NiO to be metallic. A qualitative explanation was proposed

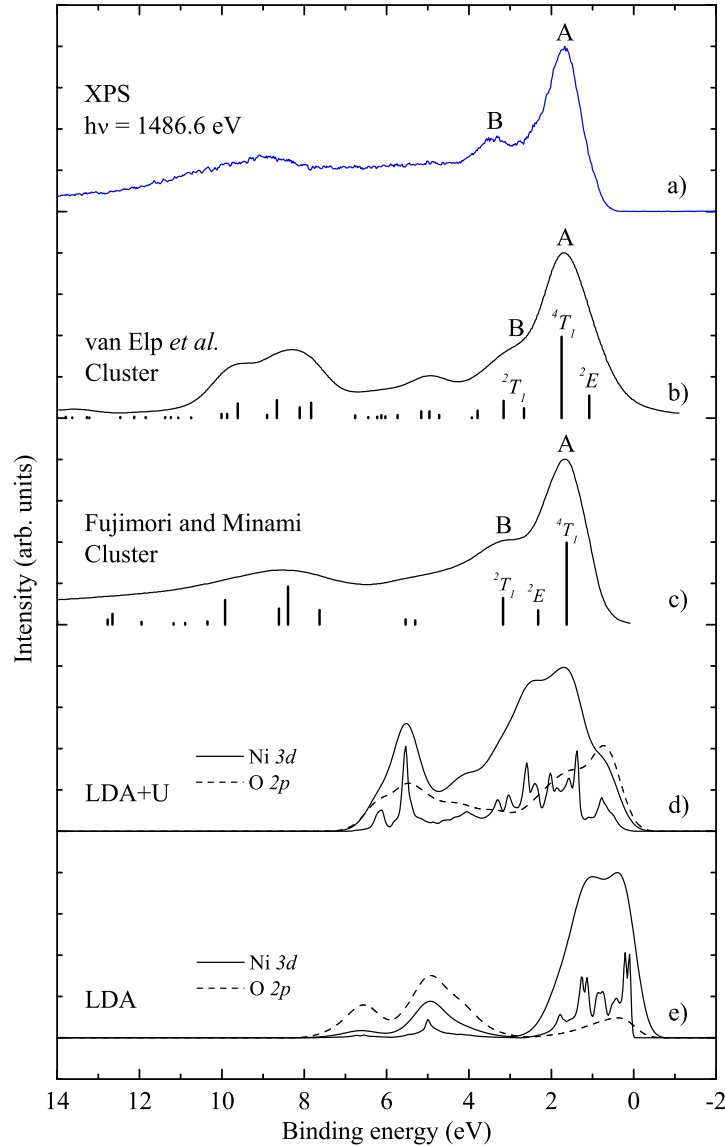


Fig. 5: Valence band XPS (1486.6 eV) spectrum of an *in situ* cleaved NiO single crystal. The results of two single-site cluster calculations (reproduced from Refs. 79 and 80), and LDA and LDA+ U calculations are also included for comparison.

in terms of the Mott-Hubbard model [71,72] in which the on-site Ni 3d-3d Coulomb interaction plays a decisive role.

An early *ab initio* attempt to fix the shortcoming of band theory was to treat NiO as a Slater insulator in which the doubling of the unit cell allows for the existence of a gap [73–75]. However, the calculated gap of about 0.2 eV [73] turned out to be much too small: A combined photoemission (PES) and bremsstrahlung-isochromat (BIS) spectroscopy study showed that the band gap is 4.3 eV [76] and established thereby the correlated nature of NiO. The inclusion of a self-interaction-correction (SIC) or Hubbard U term to the density-functional formalism may provide a justification for the magnitude of the experimental band gap [77,78].

Yet, one of the most direct methods to critically test the accuracy of the different approaches,

is to determine the excitation spectrum associated with the introduction of an extra particle into the system [81]. Curve (a) in Fig. 5 displays the valence band X-ray photoemission spectrum (XPS, $h\nu = 1486.6$ eV) of an *in situ* cleaved NiO single crystal. This spectrum represents essentially the Ni $3d$ spectral weight since the photoionization cross section of the O $2p$ is relatively small [82]. One can clearly observe from curve (e) in Fig. 5 that the Ni $3d$ density of states calculated by band theory (in the local density approximation, LDA) does not match at all: It has a Fermi cut-off and the line shape is completely different. The inclusion of the Hubbard U in the calculations (LDA+U) does not solve the line shape problem, see curve (d). All this demonstrates the shortcomings of mean field theories to describe spectra associated with the fundamental one-particle Green's function of the system [78, 83].

A completely different approach is to give up the translational symmetry of the system in order to focus on the local correlations and, especially, the dynamics of the propagation of the injected particle. Curve (c) of Fig. 5 shows the Ni $3d$ spectral weight from an early cluster configuration-interaction calculation by Fujimori and Minami [80], which also includes the full atomic multiplet theory. The agreement with the experimental spectrum is extremely good. Nevertheless, a later cluster calculation by van Elp *et al.* [79] arrived at a less satisfactory result: Peak B has almost disappeared in the calculation, see curve (b). The prime motivation to use a different set of model parameters here is to infer that the first ionization state is low spin (2E) [84] rather than the Hund's rule high spin (4T), analogous to the case of Zhang-Rice singlets in the cuprates [85, 86]. Recent developments combining LDA with dynamical mean field [87–90] or GW approaches [91] yield Ni $3d$ spectral weights which deviate in important details from the experimental spectrum. These discrepancies between the experiment and the later theoretical simulations [79, 87–90] do not provide confidence that one has made progress in understanding the nature of the first ionization state.

The issues that we need to address now are threefold. First of all we have to establish whether the XPS valence band spectrum in Fig. 5 is truly representative for bulk NiO. There are reports in the literature claiming that certain satellite peaks in the Ni $2p$ spectrum are due to surface effects [92–95]. Second, we have to determine to what extent a single-site many body approach can be utilized to describe the electronic structure of NiO for which band formation is also essential. Third, we need to identify the nature of the first ionization state in the framework of a local *ansatz*. To this end we measured the valence band of NiO utilizing the more bulk-sensitive hard X-ray photoelectron spectroscopy (HAXPES) and we investigated experimentally the electronic structure of NiO impurities in MgO [96].

3.2 Experimental

The XPS data ($h\nu = 1486.6$ eV) on *in situ* cleaved NiO single crystals were recorded using a Vacuum Generators twin crystal monochromator Al- K_α source and an Scienta SES-100 analyzer, with an overall energy resolution set to 0.35 eV. The HAXPES data ($h\nu = 6500$ eV) were taken at the Taiwan beamline BL12XU of SPring-8 in Hyogo, Japan using an MBS A-1HE analyzer. The overall energy resolution was set to 0.35 eV. The NiO impurity in MgO system

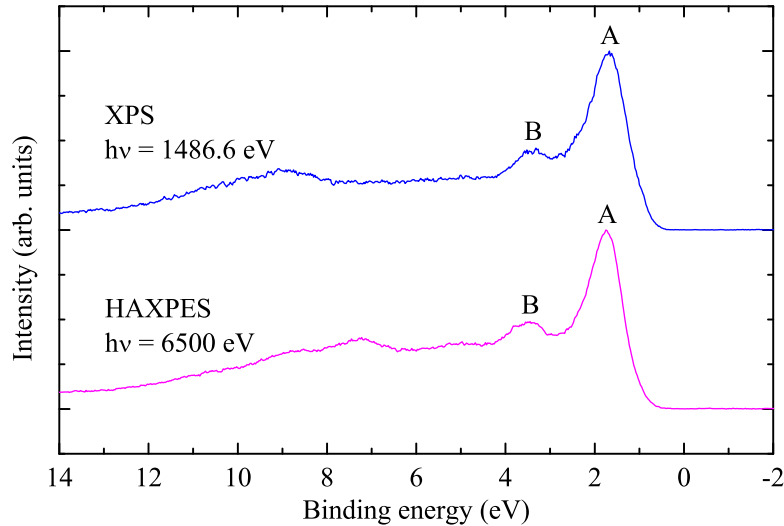


Fig. 6: Valence band photoemission spectra of an *in situ* cleaved NiO single crystal recorded using 1486.6 eV (XPS) and 6500 eV (HAXPES) photons.

was prepared *in situ* as 10-20 nm thin films on polycrystalline Ag by means of molecular beam epitaxy. The measurements were performed at the 11A1 Dragon beamline of the NSRRC in Hsinchu, Taiwan. The photoemission spectra were recorded at the Cooper minimum of Ag $4d$ ($h\nu = 140$ eV) [97] using a Scienta SES-100 analyzer with an overall energy resolution set at about 0.15 eV.

3.3 Results

In Fig. 6 we show the valence band photoemission spectra of a freshly cleaved NiO bulk crystal, taken with a photon energy of 1486.6 eV (XPS) and 6500 eV (HAXPES). By increasing the photon energy we increase also the kinetic energy of the outgoing photoelectron and, thus, also the inelastic mean free path. One can estimate that the probing depth is then enhanced from about 15 Å to roughly 80 Å [98]. We observe that the spectra are very similar. We, thus, conclude that the XPS data as displayed in Figs. 5 and 6 is representative for the NiO bulk material and that the contribution of surface effects [92–95] can be safely neglected. To be specific: Peak B is intrinsic for bulk NiO. We would like to note that increasing the photon energy from 1486.6 eV to 6500 eV does not alter much the Ni $3d$ character of the spectrum. The O $2p$ photoionization cross section relative to that of the Ni $3d$ remains very small, it changes from 1/13 to only 1/10 [82], meaning that peak B truly belongs to the Ni $3d$ spectral weight and not to the O $2p$ [89].

The valence band spectrum of the $\text{Ni}_{0.05}\text{Mg}_{0.95}\text{O}$ impurity system is shown in Fig. 7 together with the spectrum of an MgO reference thin film grown simultaneously under identical oxygen and substrate conditions. The $\text{Ni}_{0.05}\text{Mg}_{0.95}\text{O}$ film (and also the MgO film) was capped by 2 monolayers of MgO in order to prevent the surface termination to have an effect on the local electronic structure of the Ni impurity. The inset in the figure displays the Ni $L_{2,3}$ X-ray absorption spectra of the $\text{Ni}_{0.05}\text{Mg}_{0.95}\text{O}$ and the NiO bulk. The spectra are essentially identical,

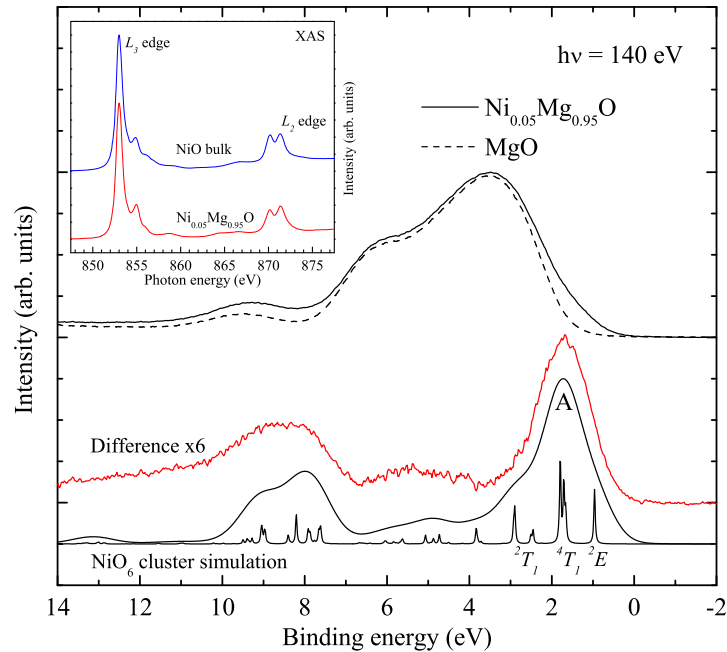


Fig. 7: Extraction of the NiO impurity valence band photoemission spectrum: Valence band spectra of $\text{Ni}_{0.05}\text{Mg}_{0.95}\text{O}$ and an MgO reference, together with the resulting difference spectrum. Also included is the result of a single-site NiO_6 configuration-interaction cluster calculation. The inset shows the Ni $L_{2,3}$ X-ray absorption spectrum of the $\text{Ni}_{0.05}\text{Mg}_{0.95}\text{O}$ together with that of bulk NiO.

verifying that the Ni in the $\text{Ni}_{0.05}\text{Mg}_{0.95}\text{O}$ has very similar local surrounding (NiO_6 octahedra) as in the bulk.

The valence band spectra of the $\text{Ni}_{0.05}\text{Mg}_{0.95}\text{O}$ and MgO systems are normalized to their O $2s$ core level intensities. Both are dominated by the O $2p$ valence band, yet, there are clear differences between them due to the presence or absence of the 5% NiO impurity. The difference spectrum multiplied by a factor of 6 is given by the red curve in Fig. 7. The line shape remains the same for films with lower Ni concentrations, as is the case for that of the Ni $2p$ [99]. This curve represents essentially the Ni $3d$ spectral weight of the NiO impurity since the photoionization cross section of the Ni $3d$ is an order of magnitude larger than that of the O $2p$ at the photon energy used [82]. Remarkable is that it is different from the spectrum of bulk NiO as shown in Figs. 5 and 6. The impurity spectrum lacks specifically peak B which is prominently present in the bulk spectrum.

3.4 Analysis

To interpret and understand the impurity spectrum, we have performed simulations using the well-proven configuration-interaction cluster model which includes the full atomic multiplet theory [22, 20, 21]. The simulations have been carried out for a NiO_6 cluster using the program XTLS 8.3 [22].

The bottom curve in Fig. 7 shows the Ni $3d$ one-electron removal spectrum from the cluster

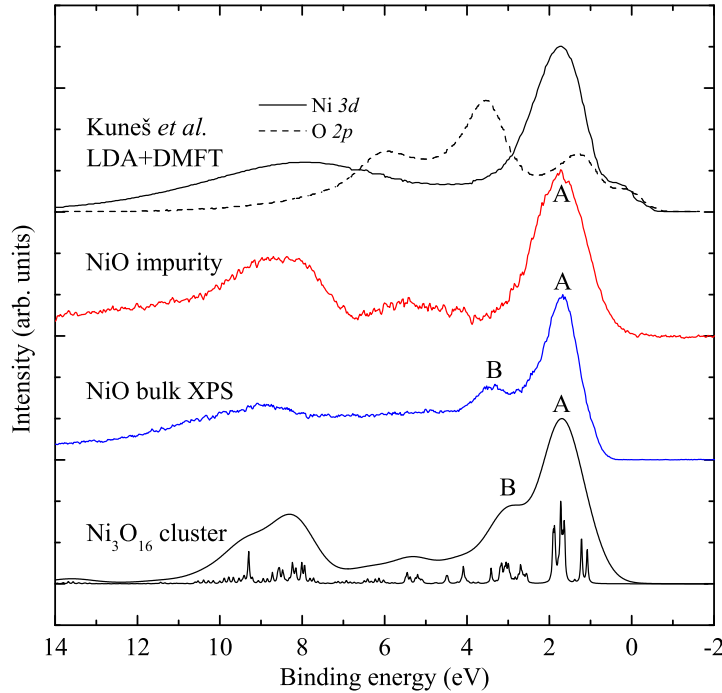


Fig. 8: Comparison of the valence band photoemission spectra of bulk NiO and NiO impurity in MgO. Also included is the simulated Ni 3d and O 2p spectral weights of the NiO valence band from a LDA+DMFT calculation (reproduced from Ref. 89) and the result of a three-site Ni_3O_{16} configuration-interaction cluster calculation.

calculation. The agreement with the experiment is very satisfactory. In order to achieve this, we have started the calculations by using parameter values which were suggested from earlier studies on NiO [79, 22, 100, 101]. We then fine-tune the parameters describing the octahedral crystal and ligand fields, and also the difference between the Hubbard U and the O 2p-Ni 3d charge transfer energy [102]. The crucial issue here is to obtain a main line (peak A) without having another feature appearing at about 2 eV higher energies (peak B) as was the case in the simulations by Fujimori and Minami [80] and by van Elp *et al.* [79]. This has implications for the energetics of the states making up the valence band as we explain in the following.

A detailed look at the cluster calculations displayed in Fig. 5 shows that peak A is given by the 4T_1 final state of the Ni 3d⁷ multiplet structure while peak B is due to the 2T_1 . Avoiding the appearance of peak B means that the energy splitting between these two states must be made smaller, e.g., 1 eV or less. This is what we have done in our simulation in Fig. 7, using different but equally reasonable parameter values [102]. The consequences for the physics are quite far reaching. Given the fact that various X-ray absorption studies find an effective octahedral crystal and ligand field splitting of about 1.65 eV [100, 101], i.e., the splitting between the isospin 2T_1 and 2E states, we arrive at the conclusion that the 2E must be lower in energy than the 4T_1 by 0.65 eV or more. This is what we read from our results in Fig. 7. In other words, our impurity study provides the spectroscopic evidence that the first ionization state has a compensated-spin character rather than the Hund's rule high-spin. This in turn justifies that the ground state of a hole doped NiO system may indeed be low-spin in nature [84].

We now return to the problem of the bulk NiO valence band spectrum. Fig. 8 shows the Ni $3d$ spectral weight taken with XPS and compares it with the spectra of the NiO impurity and of the single-site LDA+DMFT calculation [89]. One can clearly observe that peak B is absent in the impurity as well as in the Ni $3d$ spectral weight of the single-site calculation. In fact, one could infer that the calculation reproduces quite well the impurity spectrum, with perhaps some discrepancies due to the incomplete implementation of the multiplet structure of the on-site Coulomb interactions. Yet, the discrepancy with the bulk spectrum strongly suggests that the origin of peak B must be sought in non-local correlations, i.e., effects which cannot be included in a single-site approach.

3.5 Non-local screening

Our suggestion is that peak B is due to non-local screening processes involving the formation of low-energetic coherent many body states on neighboring NiO clusters, which are of the 2E type as we have shown above. The mechanism is analogous as proposed earlier for the Ni $2p$ core level spectrum of bulk NiO [103], but the application of it for the valence band is only valid for local states which are relatively stable against band formation. This may not be applicable for the 2E state, which is a state in which a hole is injected in an e_g orbital starting from the $3d^8 {}^3A_2$ ground state [101]. This hole can be expected to readily propagate in the lattice since the hopping between the Ni $3d(e_g)$ and O $2p(\sigma)$ orbitals are rather large [79, 103], yet, it may leave behind an energetically costly wake of wrong spins in the antiferromagnetic lattice. In any case, it would not be meaningful to describe its band formation as a low-energy screening process involving neighboring 2E states [104].

However, for the main peak of the bulk NiO spectrum, i.e., the 4T_1 state, we infer that we can make a meaningful approximation by using the coherent 2E screening model. The 4T_1 consists of a hole injected into the t_{2g} orbital, and its ability to move is rather limited since the overlap between the Ni $3d(t_{2g})$ and O $2p(\pi)$ is small. One could consider the 4T_1 as a localized quasi-core state. We then can invoke the non-local screening process as follows: After the creation of the 4T_1 state, an e_g electron from a neighboring NiO cluster hops onto the Ni site, leaving behind a coherent 2E hole state on that neighbor. A sketch for this process is given in Fig. 9. These two states are energetically almost degenerate [103], and the Ni $3d(e_g)$ and O $2p(\sigma)$ hybridization between them is then strong enough to produce two peaks: Not only the main peak A but also the satellite peak B.

3.6 Multi-site analysis

To confirm our assignments, Tanaka [96] has performed a Ni_3O_{16} cluster calculation consisting of three edge-shared NiO_6 octahedra. While all the O $2p$ and Ni $3d$ orbitals are included for the NiO_6 octahedron in the center where the photo-excitation takes place, those on the other parts of the cluster are replaced by a reduced basis set using the method in Ref. [105]. The results are displayed in Fig. 8 and demonstrate the presence of both peaks A and B. Note that the same parameters have been used as for the single-site calculation which produces only peak

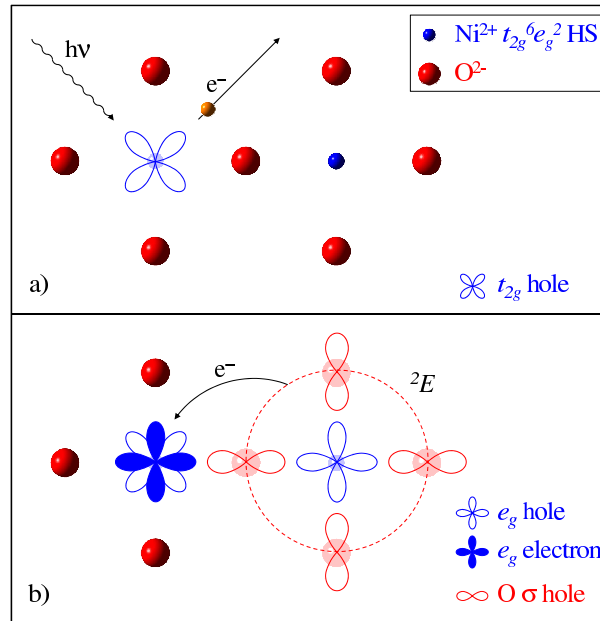


Fig. 9: Non-local screening in valence band photoemission on NiO: (a) Creation of the atomic-like (quasi-core) 4T_1 hole state by the photoemission process. (b) Screening by a next nearest neighbor NiO₆ cluster producing a coherent low-energetic 2E hole state there.

A, see Fig. 7. We also note that the energy difference between peaks A and B is somewhat smaller and the intensity of peak B is slightly larger than those of the experiments. This can be explained by the fact that the number of neighboring Ni sites is only two in the Ni₃O₁₆ cluster: the energy difference will increase and the intensity of peak B will decrease for a larger number of neighboring sites [105].

3.7 Conclusion

To summarize: We have succeeded to determine reliably the Ni 3d valence band spectra representative for bulk NiO as well as for NiO as an impurity system. From the impurity data we are able to extract the local electronic structure and the correlations herein, thereby establishing firmly the compensated-spin character of the first ionization state. By comparing the bulk with the impurity system, we were able to identify features in the bulk NiO spectrum which are caused by screening processes involving local quasi-core valence band states and non-local low-energetic many body states.

Acknowledgement

The research is supported by the Deutsche Forschungsgemeinschaft through SFB608 and FOR 1346. We gratefully acknowledge the NSRRC, ESRF, and SPring-8 staff for providing us with beamtime. We would like to thank L. Hamdan for her skillful technical and organizational assistance and J. Weinen for his assistance in preparing this manuscript.

References

- [1] for a review see: *Electronic Conduction in Oxides* by N. Tsuda, K. Nasu, A. Yanase, and K. Siratori, Springer Series in Solid-State Sciences 94, (Springer Verlag, Berlin 1991)
- [2] for a review see: M. Imada, A. Fujimori, and Y. Tokura, *Rev. Mod. Phys.* **70**, 1039 (1998)
- [3] R.J. Birgenau and M.A. Kastner, *Science* **288**, 437 (2000)
- [4] Y. Tokura and N. Nagaosa, *Science* **288**, 462 (2000)
- [5] J. Orenstein and A.J. Millis, *Science* **288**, 468 (2000)
- [6] for a review see: A.P. Ramirez, *J. Phys.: Conden. Matter* **9**, 8171 (1997)
- [7] D.I. Khomskii and G.A. Sawatzky, *Solid State Commun.* **102**, 87 (1997)
- [8] T. Mizokawa and A. Fujimori, *Phys. Rev. B* **51**, 12880 (1995); *Phys. Rev. B* **54**, 5368 (1996); *Phys. Rev. B* **56**, R493 (1997)
- [9] A.A. Taskin, A.N. Lavrov, and Yoichi Ando, *Phys. Rev. Lett.* **90**, 227201 (2003)
- [10] A. Maignan, C. Martin, D. Pelloquin, N. Nguyen, and B. Raveau, *J. Solid State Chem.* **142**, 247 (1999)
- [11] J.-H. Park, L.H. Tjeng, J.W. Allen, C.T. Chen, P. Metcalf, J.M. Honig, F.M.F. de Groot, and G.A. Sawatzky, *Phys. Rev. B* **61**, 11506 (2000)
- [12] S.Y. Ezhov, V.I. Anisimov, D.I. Khomskii, and G.A. Sawatzky, *Phys. Rev. Lett.* **83**, 4136 (1999)
- [13] F. Mila, R. Shiina, F.-C. Zhang, A. Joshi, M. Ma, V.I. Anisimov, and T.M. Rice, *Phys. Rev. Lett.* **85**, 1714 (2000)
- [14] C. J. Ballhausen, *Introduction to Ligand Field Theory* (McGraw-Hill, New York, 1962)
- [15] S. Sugano, Y. Tanabe, and H. Kamimura, *Multiplets of Transition-Metal Ions in Crystals* (Academic, New York, 1970)
- [16] M.W. Haverkort, PhD thesis, University of Cologne (2005), chapter 1, arXiv:cond-mat/0505214
- [17] for a review see: 'Photoelectron Spectroscopy' by S. Hüfner, Springer Series in Solid-State Sciences 82, (Springer Verlag, Berlin 1996)
- [18] for a review see: 'Unoccupied Electronic States', 'Fundamentals for XANES, EELS, IPS, and BIS', edited by J.C. Fuggle and J.E. Inglesfield, Topics in Applied Physics, Vol. 69 (Springer Verlag, Berlin 1992)

-
- [19] C. T. Chen and F. Sette, *Phys. Scr.* **T31**, 119 (1990)
- [20] See review by F. M. F. de Groot, *J. Electron Spectrosc. Relat. Phenom.* **67**, 529 (1994)
- [21] See review in the Theo Thole Memorial Issue,
J. Electron Spectrosc. Relat. Phenom. **86**, 1 (1997)
- [22] A. Tanaka and T. Jo, *J. Phys. Soc. Jpn.* **63**, 2788 (1994)
- [23] R.R. Heikes, R.C. Miller, and R. Mazelsky, *Physica* **30**, 1600 (1964)
- [24] G. Blasse, *J. Appl. Phys.* **36**, 879 (1965).
- [25] C. S. Naiman *et al.*, *J. Appl. Phys.* **36**, 1044 (1966)
- [26] G. H. Jonker, *J. Appl. Phys.* **37**, 1424 (1966)
- [27] J.B. Goodenough and P. M. Racciah, *J. Appl. Phys. Suppl.* **36**, 1031 (1965).
- [28] P.M. Racciah and J.B. Goodenough, *Phys. Rev.* **155**, 932 (1967)
- [29] J.B. Goodenough, in *Progress in Solid State Chemistry*, edited by H. Reiss (Pergamon, Oxford, 1971), Vol. 5.
- [30] V.G. Bhide, D.S. Rajoria, G. Rama Rao, and C.N.R. Rao, *Phys. Rev. B* **6**, 1021 (1972)
- [31] K. Asai *et al.*, *Phys. Rev. B* **50**, 3025 (1994)
- [32] M. Itoh, I. Natori, S. Kubota, and K. Motoya, *J. Phys. Soc. Jpn.* **63**, 1486 (1994)
- [33] M. Itoh, M. Sugahara, I. Natori, and K. Motoya, *J. Phys. Soc. Jpn.* **64**, 3967 (1995)
- [34] S. Yamaguchi, Y. Okimoto, H. Taniguchi, and Y. Tokura, *Phys. Rev. B* **53**, R2926 (1996)
- [35] M. A. Korotin *et al.*, *Phys. Rev. B* **54**, 5309 (1996)
- [36] T. Saitoh *et al.*, *Phys. Rev. B* **55**, 4257 (1997); *ibid.* **56**, 1290 (1997)
- [37] S. Stølen *et al.*, *Phys. Rev. B* **55**, 14103 (1997)
- [38] K. Asai *et al.*, *J. Phys. Soc. Jpn.* **67**, 290 (1998)
- [39] J. Okamoto *et al.*, *Phys. Rev. B* **62**, 4455 (2000)
- [40] P. Ravindran *et al.*, *J. Appl. Phys.* **91**, 291 (2002)
- [41] C. Zobel *et al.*, *Phys. Rev. B* **66**, 020402(R) (2002)
- [42] P. G. Radaelli and S.-W. Cheong, *Phys. Rev. B* **66**, 094408 (2002)
- [43] T. Vogt, J.A. Hriljac, N.C. Hyatt, and P. Woodward, *Phys. Rev. B* **67**, 140401(R) (2003)

- [44] I.A. Nekrasov, S.V. Streltsov, M.A. Korotin, and V.I. Anisimov, Phys. Rev. B **68**, 235113 (2003)
- [45] D. Louca and J.L. Sarrao, Phys. Rev. Lett. **91**, 155501 (2003)
- [46] G. Maris *et al.*, Phys. Rev. B **67**, 224423 (2003).
- [47] A. Ishikawa, J. Nohara, and S. Sugai, Phys. Rev. Lett. **93**, 136401 (2004)
- [48] M. Magnuson *et al.*, Europhys. Lett. **68**, 289 (2004)
- [49] K. Knizek, P. Novak, and Z. Jirak, Phys. Rev. B **71**, 054420 (2005)
- [50] The Korotin Phys. Rev. B 1996 paper has at present been cited for already more than 170 times.
- [51] See List of References in Hu *et al.*
- [52] Z. Hu *et al.*, Phys. Rev. Lett. **92**, 207402 (2004)
- [53] M. Zhuang, W. Zhang, and N. Ming, Phys. Rev. B **57**, 10705 (1998)
- [54] S. Noguchi, *et al.*, Phys. Rev. B **66**, 094404 (2002)
- [55] T. Kyômen, Y. Asaka, and M. Itoh, Phys. Rev. B **67**, 144424 (2003)
- [56] Z. Ropka and R.J. Radwanski, Phys. Rev. B **67**, 172401 (2003)
- [57] T. Kyômen, Y. Asaka, and M. Itoh, Phys. Rev. B **71**, 024418 (2005)
- [58] D. Phelan *et al.*, Phys. Rev. Lett. **96**, 027201 (2006).
- [59] M.W. Haverkort, Z. Hu, J.C. Cezar, T. Burnus, H. Hartmann, M. Reuther, C. Zobel, T. Lorenz, A. Tanaka, N.B. Brookes, H.H. Hsieh, H.-J. Lin, C.T. Chen, and L.H. Tjeng, Phys. Rev. Lett. **97**, 176405 (2006)
- [60] Z. Podlesnyak *et al.*, submitted to Phys. Rev. Lett. (LE10336/Podlesnyak)
- [61] M. Abbate *et al.*, Phys. Rev. B **47**, 16124 (1993)
- [62] S. I. Csiszar *et al.*, Phys. Rev. Lett. **95**, 187205 (2005)
- [63] CoO₆ cluster parameters [eV]: $\Delta=2.0$, $pd\sigma=-1.7$, $U_{dd}=5.5$, Slater integrals 80% of Hartree Fock values.
- [64] In a pure ionic model, the LS-HS cross-over occurs at $10Dq \approx 2.2$ eV.
- [65] B.T. Thole, P. Carra, F. Sette, and G. van der Laan, Phys. Rev. Lett. **68**, 1943 (1992)
- [66] P. Carra, B.T. Thole, M. Altarelli, and X. Wang, Phys. Rev. Lett. **70**, 694 (1993)

- [67] The sumrules gives numbers for L_z and $S_z + \frac{7}{2}T_z$; $T_z = 0$ for the HS state as justified from cluster calculations; the number of $3d$ holes is about 3.5.
- [68] A. Abragam and B. Bleaney, *Electron paramagnetic resonance of transition ions* (Clarendon, Oxford, 1970)
- [69] W.L. Roth, Phys. Rev., **110**, 1333 (1958)
- [70] J.H. de Boer and E.J.W. Verwey, Proc. Phys. Soc. London, **49**, 59 (1937)
- [71] N.F. Mott, Proc. Phys. Soc. London, Sect. A, **62**, 416 (1949)
- [72] J. Hubbard, Proc. R. Soc. London, Ser. A, **276**, 238 (1963)
- [73] T. Oguchi et al., Phys. Rev. B, **28**, 6443 (1983)
- [74] K. Terakura et al., Phys. Rev. Lett., **52**, 1830 (1984)
- [75] K. Terakura et al., Phys. Rev. B, **30**, 4734 (1984)
- [76] G.A. Sawatzky and J.W. Allen, Phys. Rev. Lett., **53**, 2339 (1984)
- [77] A. Svane and O. Gunnarsson, Phys. Rev. Lett., **65**, 1148 (1990)
- [78] V.I. Anisimov et al., Phys. Rev. B, **44**, 943 (1991)
- [79] J. van Elp et al., Phys. Rev. B, **45**, 1612 (1992)
- [80] A. Fujimori and F. Minami, Phys. Rev. B, **30**, 957 (1984)
- [81] C.-O. Almbladh and L. Hedin, *Handbook on Synchrotron Radiation*, edited by E. E. Koch, Vol. 1b (North-Holland, Amsterdam, 1983) p. 607
- [82] M.B. Trzhaskovskaya et al., At. Data Nucl. Data Tables, **77**, 97 (2001)
- [83] V.I. Anisimov et al., Phys. Rev. B, **48**, 16929 (1993)
- [84] P. Kuiper et al., Phys. Rev. Lett., **62**, 221 (1989)
- [85] F.C. Zhang and T.M. Rice, Phys. Rev. B, **37**, 3759 (1988)
- [86] H. Eskes and G.A. Sawatzky, Phys. Rev. Lett., **61**, 1415 (1988)
- [87] X. Ren et al., Phys. Rev. B, **74**, 195114 (2006)
- [88] J. Kuneš et al., Phys. Rev. Lett., **99**, 156404 (2007)
- [89] J. Kuneš et al., Phys. Rev. B, **75**, 165115 (2007)
- [90] Q. Yin et al., Phys. Rev. Lett., **100**, 066406 (2008)

- [91] H. Jiang et al., *Phys. Rev. B*, **82**, 045108 (2010)
- [92] L. Sangaletti et al., *Solid State Commun.*, **103**, 421 (1997)
- [93] L. Soriano et al., *Phys. Rev. B*, **75**, 233417 (2007)
- [94] I. Preda et al., *Phys. Rev. B*, **77**, 075411 (2008)
- [95] R.J.O. Mossaneck et al., *Chem. Phys. Lett.*, **501**, 437 (2011)
- [96] T. Haupricht, J. Weinen, A. Tanaka, R. Gierth, S.G. Altendorf, Y.-Y. Chin, T. Willers, J. Gegner, H. Fujiwara, F. Strigari, A. Hendricks, D. Regesch, Z. Hu, Hua Wu, K.-D. Tsuei, Y.F. Liao, H.H. Hsieh, H.-J. Lin, C.T. Chen, and L.H. Tjeng (unpublished)
- [97] S.L. Molodtsov et al., *Phys. Rev. Lett.*, **85**, 4184 (2000)
- [98] C.J. Powell and A. Jablonski, *Surf. Interface Anal.*, **29**, 108 (2000)
- [99] S. Altieri et al., *Phys. Rev. B*, **61**, 13403 (2000)
- [100] D. Alders et al., *Phys. Rev. B*, **57**, 11623 (1998)
- [101] M.W. Haverkort et al., *Phys. Rev. B*, **69**, 020408 (2004)
- [102] Parameters used for the calculation of the NiO₆ cluster (in eV): $\Delta=5.0$, $U_{dd}=6.5$, $10Dq=0.5$, $T_{pp}=0.8$, $V(e_g)=2.2$, $V(t_{2g})=V(e_g)/2$, ζ see Ref. 22, Slater integrals 70% of Hartree-Fock values
- [103] M.A. van Veenendaal and G.A. Sawatzky, *Phys. Rev. Lett.*, **70**, 2459 (1993)
- [104] M. Taguchi et al., *Phys. Rev. Lett.*, **100**, 206401 (2008)
- [105] A. Tanaka, *J. Phys. Soc. Jpn.*, **68**, 3493 (1999)

Index

A

adiabatic connection, 2.14, 6.3
analytic continuation, 9.29
Anderson impurity model (AIM), 3.10,
6.8, 8.18, 9.3, 10.2, 11.15
angle-resolved photoemission (ARPES),
5.6, 12.23
atomic limit, 3.7
atomic orbitals, 6.32
atomic sphere approximation (ASA), 2.21
atomic units, 6.32
augmented plane-wave method (APW),
2.19
autocorrelation, 9.10, 9.31
auxiliary (binary) field, 9.7, 10.4

B

band narrowing, 3.8
band theory, 3.2
bath Green function, 8.18, 11.4
bath symmetries, 8.26
Bethe lattice, 5.8
Bloch spectral function, 12.17
Born-Oppenheimer approximation, 3.2

C

Car-Parrinello molecular-dynamics (CPMD),
2.25
charge self-consistency, 6.19
charge transferability, 2.24
cluster DMFT, 8.22, 11.2
cluster methods, 8.20, 11.5
constrained
local density approximation (cLDA),
7.4, 12.20
random-phase approximation (cRPA),
7.9
continuous-time auxiliary field method,
10.4, 10.16
correlated hopping, 3.14
correlation energy, 2.8
Coulomb interaction, 6.10, 6.12
coupling-constant integration 2.14, 6.3
crystal field, 3.3, 6.10

D

density
conditional, 2.5
two-particle, 2.4
density matrix
one-particle, 2.3
density-density interaction, 3.12, 6.12
density-functional theory (DFT), 2.3,
6.2
detailed balance, 9.11, 10.6, 10.11
diagrammatic technique, 11.18
dielectric function, 7.5
distributed-memory, 8.16
double-counting correction, 6.10, 6.18
downfolding, 6.15
dynamical cluster approximation (DCA),
8.22
dynamical mean-field theory (DMFT),
1.10, 5.2, 5.12, 6.8
Dyson equation, 1.12, 5.4, 5.13

E

energy transferability, 2.24
entangled bands, 4.15, 7.16
exchange-correlation (xc)
energy, 2.8
hole, 2.5, 6.4
potential energy, 2.5
exchange energy, 2.8
exchange matrix, 7.14
extended Hubbard model, 3.14

F

Fano effect, 12.24
fast matrix updates, 10.7
Fermi gas, 3.7
Fourier transform, 9.14
functional, 2.6

G

gauge transformation, 4.7, 8.22
Gaunt coefficients, 6.35
generalized gradient approximation (GGA),
2.13
ghost states, 8.8

H

Hartree energy, 2.5
Hartree-Fock, 1.2, 3.16, 6.17, 6.20
Heisenberg model, 3.9
high spin (HS), 14.4
hole function, 2.5
Holstein model, 3.14
Hubbard U , 6.12, 7.3
Hubbard bands, 3.8, 5.6
Hubbard I approximation, 3.19
Hubbard matrix, 7.14
Hubbard model, 1.4, 3.5, 5.2, 6.7, 8.11
Hubbard-Stratonovich transformation, 9.6
hybrid functional, 2.14
hybridization function, 8.22, 9.4, 10.10
hybridization sum-rule, 8.22

I

importance sampling, 9.9
impurity model, 3.10, 6.8, 8.18, 9.3, 10.2, 11.15
impurity system, 14.11
infinite dimensions, 1.7, 5.6
intermediate spin (IS), 14.4
Ising model, 1.6

K

kinetic energy functional, 2.7
KKR method, 2.19, 12.6
KKR-CPA, 12.15
Kohn gauge, 8.22
Kohn-Sham energy, 2.9
Kohn-Sham orbitals, 2.7
Kondo Hamiltonian, 3.10
Krylov space, 8.5

L

Lanczos method, 8.5
Landau Fermi liquid, 5.6
LDA+DMFT, 6.21, 12.4
LDA+ U , 6.17
Lehmann representation, 5.3, 8.15
limit of infinite dimensions, 1.7
linear augmented plane-wave method (LAPW), 2.20
linear muffin-tin orbital method (LMTO), 2.20

linear response function, 7.5
local-density approximation (LDA), 2.12, 6.4
local self-energy, 5.11
local spin-density approximation (LSDA), 2.12, 6.4
localized orbitals, 4.3
low spin (LS), 14.4
LSD, 2.12

M

magnetic circular dichroism, 14.7
Markov process, 9.9, 9.13
matrix formalism, 10.13
Matsubara frequencies, 9.3
maximally-localized Wannier functions, 4.13
maximum entropy method, 9.29
mean-field theory, 1.5, 3.15, 9.4
meta GGA, 2.14
metal-insulator transition, 1.13, 3.8, 5.6, 6.9, 14.2
Metropolis transition rule, 9.11
moment expansion, 8.11
Monte Carlo, 9.8, 10.3, 11.17
Mott insulator, 3.4, 6.5
muffin-tin potential, 2.19
multi-orbital DMFT, 11.5
multiple scattering theory, 12.6
multiplets, 14.2

N

natural orbitals, 2.4
 N th-order muffin-tin orbital method (NMTO), 6.11
non-local screening, 14.15
Nyquist theorem, 9.15, 9.29

P

Padé analytic continuation, 9.30, 12.15
path integral, 5.4, 11.15
photoelectron spectroscopy, 14.3, 14.11
polarization function, 7.7
probability distribution, 9.9, 9.32
product basis set, 7.12, 7.23
projector augmented-wave method (PAW), 2.19, 2.25
projectors, 4.15

pseudopotential, 2.18, 2.21

Q

quantum-impurity solver, 6.9, 8.18, 9.5,
10.2

quantum Monte Carlo, 9.4, 9.14, 9.28,
10.2

R

Racah parameters, 6.36

random-phase approximation (RPA), 7.4

real harmonics, 6.13, 6.32

recursion method, 8.2

Rubtsov method, 10.15

S

scaling of hopping amplitudes, 5.7

screened interaction, 7.4

segment picture, 10.11

self-interaction, 6.4

semi-local form, 2.23

separable form, 2.24

shared-memory, 8.15

sign problem, 9.9, 9.28

skeleton expansion, 5.10, 5.13, 5.17

Slater determinant, 6.2

Slater integral, 3.13

Slater-Koster two-center integrals, 6.14,
6.33

slave-boson theory, 3.20

soft-X-ray absorption, 14.3

spectral function, 3.6, 8.8

spin-state transition, 14.4

spline interpolation, 9.16

SPTF solver, 12.15

steepest-descent, 8.3

strong-coupling ctqmc, 10.9

strongly correlated materials, 6.5, 14.2

sum-rule, 8.22

super-exchange, 6.8, 6.26

susceptibility, 11.24

T

t - J model, 3.9

thermal average, 9.3

Trotter-Suzuki decomposition, 9.6

W

Wannier function, 4.3, 4.7, 6.10, 6.16,
7.12

weak-coupling ctqmc, 10.4

Z

Zhang-Rice singlet, 14.11

1. The LDA+DMFT approach to strongly correlated materials

Lecture Notes of the Autumn School 2011 Hands-on LDA+DMFT
edited by E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (2011)
ca. 420 pages
ISBN: 978-3-89336-734-4