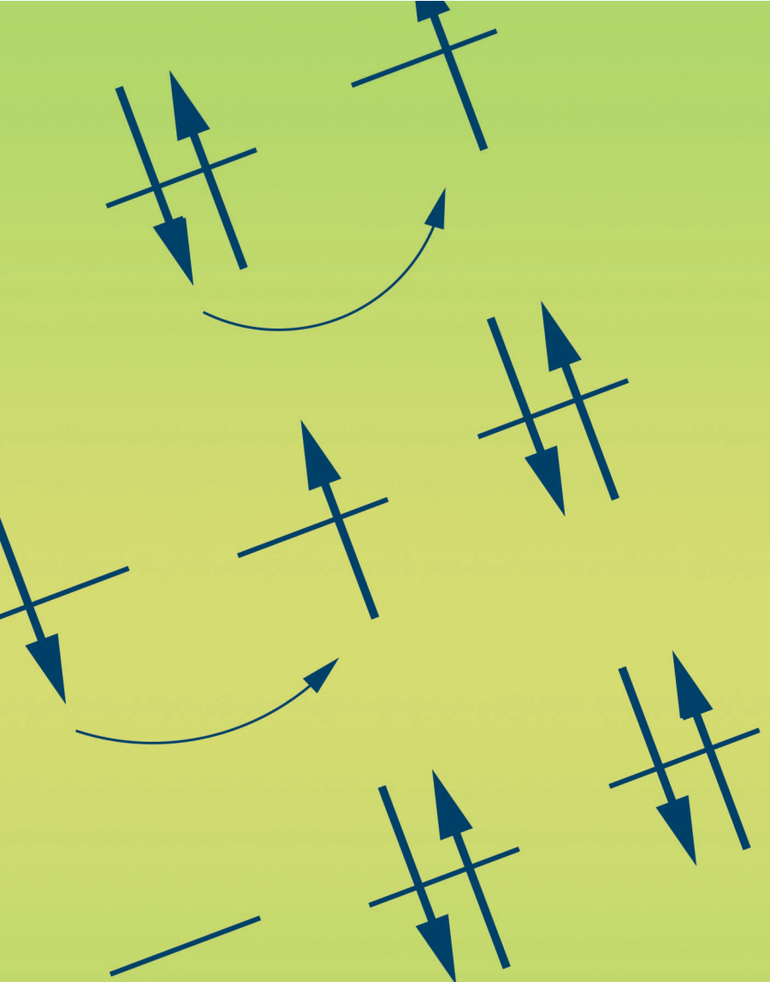
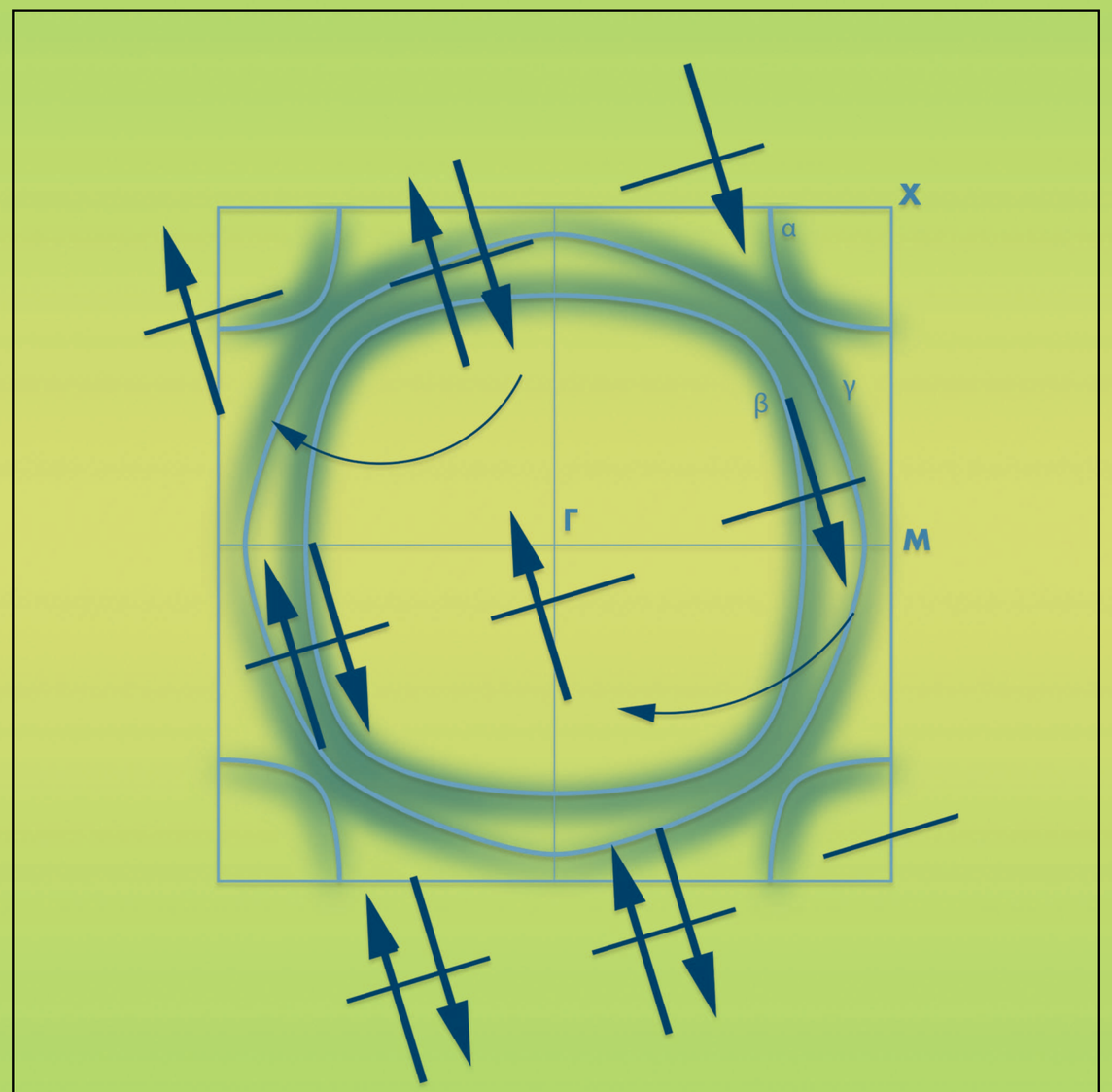


The Physics of Correlated Insulators, Metals, and Superconductors

Eva Pavarini, Erik Koch, Richard Scalettar, and Richard Martin (Eds.)

Modeling and Simulation

7



The Physics of Correlated Insulators, Metals, and Superconductors

Eva Pavarini, Erik Koch, Richard Scalettar, and Richard Martin (Eds.)

JÜLICH
FORSCHUNGSZENTRUM

Modeling and Simulation
Band / Volume 7
ISBN 978-3-95806-224-5

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation Band / Volume 7

Forschungszentrum Jülich GmbH
Institute for Advanced Simulation

Lecture Notes of the Autumn School on
Correlated Electrons 2017

Eva Pavarini, Erik Koch, Richard Scalettar, and Richard Martin (Eds.)

The Physics of Correlated Insulators, Metals, and Superconductors

Autumn School organized by
the Institute of Advanced Simulation
at Forschungszentrum Jülich
25 – 29 September 2017

Schriften des Forschungszentrums Jülich
Reihe Modeling and Simulation Band / Volume 7

ISSN 2192-8525 ISBN 978-3-95806-224-5

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 Simulation

Cover Design: Grafische Medien, Forschungszentrum Jülich GmbH

Printer: Druckerei Schloemer, Düren

Copyright: Forschungszentrum Jülich 2017

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 7

ISSN 2192-8525
ISBN 978-3-95806-224-5



This is an Open Access publication distributed under the terms of the Creative Commons Attribution License 4.0,
which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Contents

Preface

1. Electronic Structure Computation Meets Strong Correlation: Guiding Principles
   Richard Martin
2. Insulator, Metal, or Superconductor: The Criteria
   Richard Scalettar
3. The Insulating State of Matter: A Geometrical Theory
   Raffaele Resta
4. Exchange Mechanisms
   Erik Koch
5. Orbital Physics
   Andrzej Oleś
6. Introduction to the Hubbard Model
   Robert Eder
7. The Finite Temperature Lanczos Method and its Applications
   Peter Prelovšek
8. Gutzwiller Density Functional Theory
   Florian Gebhard
9. Mott transition: DFT+U vs. DFT+DMFT
   Eva Pavarini
10. Path Integrals and Dual Fermions
    Alexander Lichtenstein
11. Dynamical Mean-Field Theory of Disordered Electrons:
    Coherent Potential Approximation and Beyond
    Václav Janiš
12. Interplay of Kondo Effect and RKKY Interaction
    Johann Kroha
13. Kondo Physics and the Mott Transition
    Michele Fabrizio
14. Hund's Metals Explained
    Luca de' Medici
15. Electron-Phonon Coupling
    Rolf Heid
16. Introduction to Superconducting Density Functional Theory
    Antonio Sanna

Index

Preface

A naive distinction between metals and insulators rests on the single-electron picture: completely filled or empty bands characterize insulators while metals have some partially filled bands. Nature, however, offers a much richer variety of behaviors: Mott insulators would be band metals in the absence of electron correlation while strongly-correlated metals behave quasiparticle-like only in the Fermi-liquid regime. Correlated metals and insulators can be distinguished by the gap in the spectral function. Superconductors form a class of their own, they have a single-electron gap but are not insulators.

This year's school addresses the rich physics of correlated insulators, metals, and superconductors. Insulators show complex ordering phenomena involving charge, spin, and orbital degrees of freedom. Correlated metals exhibit non-Fermi-liquid behavior except right at the Fermi surface. Superconductors are dominated by the delicate interplay of coupling bosons and quasiparticles. Along with the phenomena, the models and methods for understanding and classifying them will be explained. The aim of the school is to introduce advanced graduate students and up to the modern approaches for modeling strongly correlated materials and analyzing their behavior.

A school of this size and scope requires support and help from many sources. We are very grateful for all the financial and practical support we have received. The Institute for Advanced Simulation at the Forschungszentrum Jülich and the Jülich Supercomputer Centre provided the major part of the funding and were vital for the organization of the school and the production of this book. The Institute for Complex Adaptive Matter (ICAM) offered travel grants for selected international speakers and participants.

The nature of a school makes it desirable to have the lecture notes available when the lectures are given. This way students get the chance to work through the lectures thoroughly while their memory is still fresh. We are therefore extremely grateful to the lecturers that, despite tight deadlines, provided their manuscripts in time for the production of this book. We are confident that the lecture notes collected here will not only serve the participants of the school but will also be useful for other students entering the exciting field of strongly correlated materials.

We are grateful to Mrs. H. Lexis of the Verlag des Forschungszentrum Jülich and to Mrs. L. Weidener of the Grafische Betriebe for providing their expert support in producing the present volume on a tight schedule. We heartily thank our students and postdocs who helped with proofreading the manuscripts, often on quite short notice: Julian Mußhoff, Esmaeel Sarvestani, Amin Kiani Sheikhabadi, and Qian Zhang.

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

Eva Pavarini, Erik Koch, Richard Scalettar, and Richard Martin

August 2017

1 Electronic Structure Computation Meets Strong Correlation: Guiding Principles

Richard M. Martin
University of Illinois at Urbana-Champaign, Illinois
Stanford University, Stanford, California

Contents

1 Introduction **2**
1.1 Read this first! . . . 2
1.2 Strong correlation: What do we mean? What is the evidence? . . . 3
1.3 Experimental signatures . . . 4
1.4 Fermi liquids, Mott insulators, ... 5

2 The four primary methods for quantitative calculations for materials **5**
2.1 Density functional theory (DFT) . . . 6
2.2 Quantum Monte Carlo . . . 7
2.3 Many-body perturbation theory . . . 7
2.4 Dynamical mean-field theory (DMFT) . . . 7

3 Guiding principles **9**
3.1 Conservations laws and continuity: Normal metals and Fermi liquid theory . . . 10
3.2 The Luttinger theorem and Friedel sum rule . . . 10

4 Instructive models **12**
4.1 The Anderson impurity model and the Kondo effect . . . 12
4.2 The Hubbard and Gutzwiller approximations . . . 13
4.3 Dynamical mean-field theory (DMFT) for the Hubbard model . . . 15

5 A few examples: Data, interpretation, and theory **18**
5.1 Cerium: volume collapse phase transition, heavy fermions, ... 18
5.2 NiO and Mott insulators . . . 21
5.3 V_2O_3 and the Mott transition . . . 23

6 The Mott insulator and topological order **23**

7 A few final remarks **24**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

1.1 Read this first!

Before reading the rest of this chapter, please read the following questions, formulate your answers (or at least your approach to answering them) and think about the relation to the topics of this school. After the school check if you think that your understanding has evolved.

- What does the term correlation mean? The general definition in statistics and the definition(s) used in the fields that are the topics of this school.
- What does "strong correlation" mean? "Strong" relative to what? Do we always have to specify the context in order for the term meaningful?
- What are signatures of correlation? What are the types of experimental information from which we may conclude that a material is strongly correlated. What theoretical findings indicate strong correlation?
- What are examples of materials and phenomena that are often called "strongly correlated" in present-day condensed matter physics? What are the reasons for this designation? Why should we care enough for individuals and groups around the world to devote countless hours and resources for decades?
- Van der Waals interactions are due purely to correlated fluctuations between electrons on different atoms or molecules that are weakly coupled. Is this an example of strong correlations?
- What is a "Mott insulator"? What is the Luttinger theorem and to what extent is it an actual theorem?

The purpose of this lecture is to set the tone that interacting, correlated electrons present some of the most difficult challenges in physics. There are no exact solutions except for very few special models; yet there are powerful techniques that are being applied to complex problems. Each method involves approximations and assumptions that may be forgotten or brushed aside when they are applied. This is a perfect example of a field where one should examine critically every aspect of the theory and the methods for calculations.

Modern electronic structure methods successfully determine ground state structures, phonon frequencies, and many other properties to within a few percent for large classes of materials, and they are improving with new ideas, methods, and experience. Yet they may fail miserably to account for some other properties. As calculations become more and more powerful, it is more and more important to formulate the problem clearly and judge what aspects of the results are meaningful. We need fundamental guiding principles to recognize how to use the results to draw well-justified, useful conclusions. In this endeavor, I will give some of my pet principles and examples,[1] and warnings about the care needed to apply them. These are only examples and *you must identify for yourself what is needed in your research!*

[1] Many of the ideas and examples in this chapter are taken from [1], with some added comments. Many additional references are given there.

1.2 Strong correlation: What do we mean? What is the evidence?

Since our goal is to develop theories of materials with strong correlation, we should define carefully what is meant by terms like "correlation" and "strong," and what are the experimental signatures. The standard definition of correlation for fermions is anything beyond Hartree-Fock. However, this is not sufficient for present purposes: Correlation is crucial for *all* materials with more than one electron. The correlation energy is a large part of the total energy that determines structures of materials and related properties. Correlation is even more important for excitations: Hartree-Fock leads to band gaps in insulators much greater than experiment and it is completely wrong for the properties of *any* metal because it leads to a singularity at the Fermi energy due to the long-range Coulomb interaction. Any useful theory of excitations must involve a screened interaction which removes the singularity at the Fermi surface. Screening is one aspect of correlation, and it is one of the successes of the RPA and GW approximations considered below. A model with short-range interactions that is proposed to apply to a real material has some assumption, perhaps hidden, of why a short range interaction is appropriate. In short, correlation *must* be included for a reasonable theory in any material, whether it is metallic Na or Ce, insulating NaCl or NiO.

In addition, the definitions depend upon whether one considers *restricted* or *unrestricted* Hartree-Fock. Often there is a broken symmetry solution that lowers the total energy and captures a large part of the correlation energy. In this case, the question is whether or not the broken symmetry is physically reasonable and how to use such a solution even if there is no such broken symmetry. This brings us immediately to issues of broken symmetry, phase transitions, and order parameters.

For many problems correlation is largest at short range, and the approach is to identify some reduced degrees of freedom on each site that already takes into account local on-site correlation. *The key is that there is a degeneracy or near-degeneracy of the local states, e.g., a spin, weakly coupled to the rest of the system. Then the issues have to do with two aspects of correlation, on each site and between the different sites.* This is the type of problem addressed in this chapter and each example involves identifying some set of local degrees of freedom.[2] The typical systems involved are transition elements with localized atomic-like d and f states illustrated in Fig. 1, which is a rearrangement of the rows of transition elements so that the most localized orbitals are at the upper right and the most extended at the lower left. This provides an intuitive picture of the progression from band-like metallic superconductors colored blue to magnetic systems with localized moments involving $4f$ states colored red. Many of the most interesting elemental solids and compounds contain elements at the boundary where there is the greatest competition between interactions and independent-particle terms.

[2]There are other cases such as the fractional quantum Hall effect, where the single-particle kinetic energy is zero and the interaction is the dominant effect.

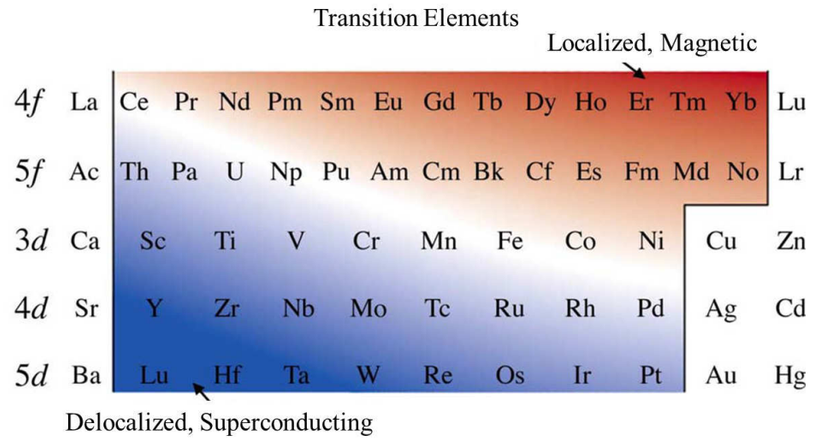


Fig. 1: *Rearranged periodic table of the transition elements showing the trend from delocalized orbitals for elements at the lower left indicated by the blue color to the most localized at the upper right colored red. At the boundary are the elements that exhibit anomalous properties. In compounds the occupation of* d *or* f *orbitals may vary, for example, the extraordinary properties of the copper oxides where the Cu* $3d$ *states are near the boundary. (Similar to figure in [2], originally due to J.L. Smith.)*

1.3 Experimental signatures

So what is the driving force to identity "strongly correlated" materials and what do we want to determine? A general classification of properties of condensed matter is the division between ground state and excitations, where each can reveal essential aspects of correlation. Of course, the qualitative nature of the excitations is determined by the symmetry of the ground state, for example, a gap in an s-wave superconductor, or the magnon dispersion in a magnetic system. This is crucial for understanding but we also want more quantitative information. Consider the example of a high-temperature d-wave superconductor which is certainly often called "strongly correlated." However, it was shown by Kohn and Luttinger [3] many years ago that perturbation theory predicts that, due to repulsive electron-electron interactions, the Fermi liquid state of any metal is unstable to a superconducting state with some pair angular momentum if the temperature is low enough, i.e., the ground state is a superconductor. Clearly symmetry is not enough and we must look further if terms like weak or strong are to be useful.

A theme of this lecture is that a ubiquitous signature of "strongly correlated" materials is *strong variations with temperature*. Temperature dependence is often the most important property for actual applications and it is often the property most directly measured in experiments, but it is difficult to calculate the effects directly. Of course, temperature dependence is due to excitations, but measuring spectra for excitations requires specialized experiments which often are difficult to interpret. However, spectra are more directly related to theoretical methods. High energy features in the spectra are direct evidence for large interactions, but it is the low energy

excitations that determine the temperature dependence, the response to perturbations that lead to phase transitions, and other effects. Prime examples of such behavior involve the elements listed in Fig. 1.

It is *not* my purpose to say that we must use theoretical methods that include temperature directly. This may be very difficult and not needed to understand the observed phenomena. This lecture is devoted to sorting out the issues and using various theoretical methods – with judiciously chosen guiding principles.

1.4 Fermi liquids, Mott insulators, ...

Since this is a school on "The Physics of Correlated Insulators, Metals, and Superconductors," certainly terms like "Fermi liquid," "Mott insulator," etc., will arise and we should be clear what is meant in each case. In particular, it is important to be clear about the meaning of the term *insulator*, and realize that there is not a unique definition of "Mott insulator." There is no hope of forcing everyone to adopt a single convention, and probably it is best not to try to make a single term apply to many complex problems. But we can be clear what is meant in specific cases.

A common feature of the different definitions is an insulator that would be a metal if the electrons did not interact. Even this condition already has problems if we want to identify "correlated" systems. If one allows for a broken symmetry, the *unrestricted Hartree-Fock solution* includes some important aspects of correlation and can lead to an insulating state; in fact, it tends to favor an insulator. There may be additional conditions that are used to define the term "Mott insulator." One is that they are only those that can be insulators without a broken symmetry, i.e., the order is not essential. A yet more restrictive definition is only insulators with no broken symmetry at zero temperature, called quantum "spin liquids" (see, e.g., [4] and papers referred to there), which present deep issues concerning the states of matter. These three definitions exemplify the issues addressed in this lecture: the nature of different states of matter; what experimental signatures show that correlations play an essential role; the power of modern computational methods to make quantitative predictions; the difficulty of making robust conclusions even if we have such powerful methods; and some guiding principles to help.

2 The four primary methods for quantitative calculations for materials

Quantitative calculations for materials are now an essential part of research, a fact made possible by the development of theoretical approaches together with powerful computational methods. There are four complementary approaches each having a crucial step in its genesis in the 1960's, key advances in the 1980's that made possible effective computational algorithms, and active, ongoing developments. The present discussions follow [1, 5] where each approach is described in much more detail.

2.1 Density functional theory (DFT)

It is important to realize that there were two different steps in the development of DFT done in 1964 and 1965.[3] One is the proof by Hohenberg and Kohn [7] that *all* properties of a system of electrons are determined by the ground state density $n(\mathbf{r})$, and in principle, each can be expressed as a functional of the functional of $n(\mathbf{r})$. This would be only a minor observation were it not for the second step by Kohn and Sham [8] to define an *auxiliary system of non-interacting electrons* designed to reproduce that exact ground state density and energy. The Kohn-Sham approach uses independent-particle methods, but it is *not an independent-particle approximation*; instead it is a stroke of genius to reformulate the entire approach to many-body theory to calculate certain properties, in principle exactly, with no guarantee that any other property can be determined by the method. The success of DFT is due to the fact that it is a theory of *interacting, correlated electrons* and it has proven to be possible to find approximate functionals $E_{xc}[n]$ for exchange and correlation contributions to the total ground state energy that are remarkably accurate, even for many systems that are called "strongly correlated." Often the local approximation (LDA) is very good, and improved approximations have made DFT an integral part of research in chemistry and other fields. The work of Car and Parrinello in 1985 [9] set the stage for developments of efficient computational approaches that have made DFT into an indispensable method for realistic calculations.

Extensions of the original formulation illustrate the power and the limitations of the approach. The extension to hybrid functionals and methods like "DFT +U" are often important for systems we call strongly correlated, and the role of such methods are discussed in the examples below. Successful functionals have been developed to describe the van der Waals interaction, which is a non-local effect due to correlation (see, e.g., [10]). The ground state of a superconductor is also described in principle by DFT and functionals that include effects of electron-phonon interaction have been developed (see [11] that cites earlier references). Very soon after the Kohn-Sham paper, Mermin [12] showed that the density and free energy at any temperature are given exactly in terms of a temperature dependent functional $E_{xc}(T)[n]$ for the exchange and correlation energy. However, there has been almost no use of the Mermin functional because it has been so difficult to make useful thermodynamic functionals for exchange and correlation. (Question to the reader: why is this so hard?)

Excitations are not supposed to be given by DFT, even in principle. *It is an independent-particle approximation to use the eigenvalues the Kohn-Sham equations as excitation energies*. This is merely one of many possible independent-particle approximations and it is not justified by any systematic theory. Indeed the results often are badly in error, even for materials as simple as silicon.

[3] An extensive presentation is in [5] as well as in many good references such as the lecture by Jones in volume 3 of this series in 2013 [6].

2.2 Quantum Monte Carlo

There are various ways quantum Monte Carlo (QMC) simulations can be used to treat interacting quantum systems by sampling the many-body wavefunction. One is variational Monte Carlo, where a wavefunction is optimized to provide the lowest possible energy for that type of function. A different approach is diffusion Monte Carlo that is based upon the analogy of the Schrodinger and the diffusion equations. This approach can provide the *exact* solution for the ground state of an interacting system of particles, and the first such calculation was done in 1965 by McMillan for ^{4}He [13]. However, the direct application to fermions does not work because the wavefunction must change sign and cannot be treated as a probability distribution. This has led to the fixed-node approximation where the nodal surface (where the wavefunction changes sign in the $3N$ dimensional space for N particles) is constrained to be the same as that of an optimized many-body variational function. This is the method used in the famous Ceperley-Alder work [14] in 1987 that provided the total energy of the homogeneous electron gas used in the construction of local density approximations to DFT. Other approaches include auxiliary field Monte Carlo and the more recent continuous time [15] methods that are particularly appropriate for embedded systems and are used in DMFT, as discussed below.

2.3 Many-body perturbation theory

A general approach to the theory of excitations in many-electron systems can be developed in terms of a perturbation expansion in the Coulomb interaction. The Green function $G(\mathbf{k},\omega)$ is expressed by a Dyson equation in terms of the non-interacting part $G_0(\mathbf{k},\omega)$ and a self energy $\Sigma(\mathbf{k},\omega)$, which contains the effects of interaction,

$$G^{-1} = G_0^{-1} - \Sigma. \tag{1}$$

The approach used in much current work is the "GW" approximation developed by Hedin [16] in 1965, who applied it to the homogeneous electron gas and gave it the modest name GW to denote that it is lowest-order approximation $\Sigma = GW$, where W is the screened interaction. In condensed matter it is essential to screen the long-range Coulomb interaction, so that W itself is an infinite sum of diagrams. In the 1980's it was possible to develop computational methods to apply GW to materials, and it is now widely used to calculate band gaps in systems where the correlation is relatively weak, such as in semiconductors.[4] Extensive description can be found in [1] and reviews such as [17]. A question addressed here is when it can be a reasonable approximation for more strongly correlated systems with d and f states.

2.4 Dynamical mean-field theory (DMFT)

DMFT is a Green function method that treats the correlation in a way that is very different from GW. Instead of a low-order expansion, the approach is to treat the short range correlation more

[4] It also can describe the asymptotic $1/r^6$ form of the van der Waals interaction, which is a pure correlation effect. The difference from the examples of strong correlation considered here is that perturbation theory works because there is a large gap and no low energy electronic excitations.

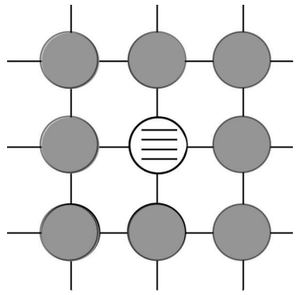


Fig. 2: *Single site embedded in a lattice. For the Anderson impurity model (AIM) this depicts a real impurity with interactions embedded in a metal in which the electrons do not interact. It also represents the many-body problem solved in DMFT in the single site approximation: a site with interactions treated explicitly embedded in a medium with a frequency-dependent self-energy, which is equivalent to an AIM solved self-consistently with the requirement that the self-energy for the surrounding lattice is the same as that calculated for the embedded site.*

accurately and to treat the longer range correlation more approximately or ignore it altogether. This is designed to treat materials with d and f states and models such as the Hubbard model, and it is the main approach considered in the examples in this chapter. There are precedents, both published in 1965, that capture some of the ideas in DMFT. They are still widely-used for intuitive understanding and are described in Sec. 4.2. Two seminal advances have led to the formulation of DMFT: The first is the work of Metzner and Vollhardt [18] and Müller-Hartmann [19] in 1989, who developed the quantum theory of interacting particles in infinite dimensions. The second advance is the recognition by Georges and Kotliar [20] and, independently, by Jarrell [21] in 1992 that an embedded site can be viewed as an auxiliary system equivalent to an Anderson impurity model (AIM). Thus the analytic and computational methods developed for the AIM are directly applicable as solvers for DMFT calculations, and many insights developed over the years can be used to understand the phenomena caused by interactions. Good resources for the concepts and theoretical methods are the review [22] and many lectures in this series, notably in the volumes devoted to DMFT [23, 24] which also describe ongoing developments and realistic applications.

The essence of DMFT is to calculate a Green function and self-energy by a many-body calculation for a site or cluster of sites embedded in a medium that represents the rest of the crystal.[5] For example, the single-site approximation is to carry out the calculation for a site embedded in the rest of the crystal as illustrated in Fig. 2. This approach can be used for a real impurity on a site, which is the Anderson impurity model (AIM) discussed in Sec. 4.1. However, DMFT is *not* an impurity approximation; the calculation is done self-consistently with the atoms in the rest of the crystal (depicted by the solid gray circles in Fig. 2) required to have the same self-energy as that calculated for the central site. The calculation for the embedded site is used only to determine the self-energy for the states on the site, and the final result is a translation-invariant Green function $G(\mathbf{k}, \omega)$ given by Eq. (1), where there is no approximation in the independent-particle $G_0(\mathbf{k}, \omega)$ and the approximation is that the self-energy has only local on-site components, i.e., $\Sigma(\omega)$ is independent of momentum $\mathbf{k}$.

[5] In the literature the term DMFT is often used to denote only the single-site approximation following the traditional definition in statistical mechanics. Here and in [1] DMFT is regarded as a general approach for a site or cluster of sites embedded in a medium. DMFT is designed to give spectra, whereas an alternative approach, Density Matrix Embedding, is designed to give the energies for the ground state [25].

In applications to models with on-site interactions U and hopping matrix elements between sites, DMFT is constructed to be exact in two limits: the non-interacting limit, $U = 0$, and the limit of infinite dimensions $d \to \infty$ where there is no correlation between sites and mean-field theory is exact (see especially [24]). In calculations for materials, interactions W and a small basis set are inputs to the method; this introduces approximations and uncertainties, and usually the parameters are calculated using DFT or many-body perturbation methods. Also there are other less correlated bands that are approximated by some other method like DFT or GW.
There are various ways to do the many-body calculation for the embedded site or cluster. The natural approach is to use the finite temperature formalism with fixed Fermi energy so that the particle number of the site is not fixed. One on the most powerful approaches uses finite temperature Monte Carlo calculations which samples all possible occupations of the central site. This builds upon the work on the AIM which provided the exact solution by Hirsch and Fye [26] for the case of a single-band with spin 1/2, where there is no sign problem. Those methods have been adapted directly and more recently a continuous time algorithm [15] has been developed just for the embedded site or cluster.

3 Guiding principles

All many-body calculations for interacting, correlated electrons – except for a few models that can be solved exactly – involve approximations. How can we justify that the results of such a calculation are meaningful? This is the first question before one addresses the issues of the quantitative accuracy. It is very useful to have some guiding principles that are rooted in deep, fundamental principles like conservation of energy and momentum, symmetry, and topological classification. In addition, there may be guiding principles that are based upon very general arguments that involve theoretical derivations, and one might suspect that the derivations have some loopholes or they break down in some cases. In these cases the guiding principles may be especially fruitful if they are taken seriously: A conclusion that the principle applies is extremely useful and insightful. A demonstration that it is violated may indicate an error or may be truly important; it should not be taken lightly and it may indicate new, here-to-fore unknown possibilities. Let us examine a few guiding principles, and in the following sections illustrate their use.
T=0 vs. T $>$ 0: There are crucial differences between $T = 0$ and $T > 0$. Much of the theory of electronic structure is formulated specifically for perfect crystals at $T = 0$. However, we want to consider both $T = 0$ and $T > 0$, where there is always disorder and finite entropy, and it is crucial to recognize the differences.
Metals vs. Insulators: Strictly the distinction between metal and insulator is well-defined only at zero temperature. At any $T > 0$ there is some conductivity. However, the difference in conductivity may be very large; the effects of thermal disorder are manifested in the temperature dependence and the characteristic behavior as $T \to 0$ is the way metals and insulators are distinguished in practice. We will adopt the terms metal and insulator, with the realization that we must be careful applying arguments valid only at $T = 0$ to problems at $T > 0$, and vice versa.

3.1 Conservations laws and continuity: Normal metals and Fermi liquid theory

Fermi liquid theory is a fabulously successful theory attributed to Landau. It was initially applied to liquid ^{3}He, where the interaction is short range, and later extended to include Coulomb interactions by Silin. As told to me by Alexey Abrikosov, Landau arrived at his proposal by asking himself: "What is conserved?" The fact that conservation laws are unchanged by the interaction suggests continuity. Even though ^{3}He has strong interactions the proposal is that it remains an isotropic liquid with a Fermi surface and low energy excitations that have the same quantum numbers as if it were a system of non-interacting fermions, but with a modified effective mass and susceptibilities. This is not a theory of how to calculate the effective parameters, but rather of relations among them, that are extremely useful and insightful.

The messages that I want to emphasize are the importance of continuity and the separation of the three different aspects of our understanding: the formulation of the theory, explicit mathematical methods to calculate the properties, and proofs (if any) that the results of the calculations are rigorously correct. Some aspects can be argued to be guaranteed if the system evolves continuously from a known state, even if there are no theoretical proofs that the system evolves continuously, and no proofs that the theoretical derivations are foolproof. Systems that are continuously connected to some independent-particle state are termed "normal." The same arguments also apply to insulators where there is a gap. Of course, there are many examples of transitions between states that have different symmetry, and we can ask if each state can be considered to evolve continuously from some known state.[6]

3.2 The Luttinger theorem and Friedel sum rule

The Luttinger theorem states that in a perfect crystal the volume enclosed by the Fermi surface in $\mathbf{k}$-space is the same, independent of interactions. (I will use the term "theorem," but it might be better to call it a sum rule since theorem implies mathematical rigor.) It is sufficient to indicate the sense of the derivation in the original papers by Luttinger and Ward (LW) [27, 28], which is summarized in App. J of [1]. The total number of electrons is given by $d\Omega/d\mu$ where Ω is the thermodynamic potential and μ is the chemical potential. The derivation by LW involves the particular way of summing diagrams to express Ω in terms of the Green function G (which is an infinite sum of diagrams involving the bare G_0) summed over Matsubara frequencies, taking the limit of $T \to 0$ and a partial integration. The derivation uses the fact that the one-particle Green function $G(\mathbf{k}, \omega)$ evaluated at the Fermi energy $\omega = E_F$ (where $E_F = \mu(T = 0)$) is real, and the final result is that $G(\mathbf{k}, E_F)$ changes sign as a function of $\mathbf{k}$ at the Fermi surface, so that the volume is defined by the region where the Green function is negative. The conclusions apply only at $T = 0$ where the Fermi surface is precisely defined and the states at the Fermi energy have infinite lifetime, just as for an independent-particle system.

[6] There can be further distinctions including classification by the topology of the eigenstates as a function of $\mathbf{k}$ where topologically trivial and non-trivial states are each connected to an independent-particle state with the same topology.

As far as I know there are no rigorous proofs that the summations are uniformly convergent, and there are no loopholes in the mathematical derivations.[7] However, we can appeal to the principle of continuity, to rephrase the argument: The theorem should be obeyed so long as the system is normal, i.e., can be continuously connected to some independent-particle system. The lack of a rigorous proof applies to any case, such as Na, Al, or Cu, and it is experiment that shows these systems act like normal Fermi liquids with a Fermi surface that obeys the theorem as $T \to 0$ for the accessible range of temperatures. We already know one example, superconductivity, which is certainly not normal! Any violation of the theorem would herald a different state of matter; in both theory and experiment we should be doubly careful not to treat violations lightly and to search vigorously for exceptions!

It is instructive to consider also the Friedel sum rule [30] on the sum of phase shifts for an impurity in a metal, which is an example of how much can be learned without heavy mathematical calculations. The sum rule follows from the condition that the sum of phase shifts of the states at the Fermi energy due to the impurity equals the number of electrons added (or subtracted) around the impurity compared to the host. Since charges are perfectly screened in a metal, the number of electrons is balanced exactly by the added charge of the nuclei of the impurity relative to the host. The sum rule was originally derived for independent particles; however, the derivations of LW apply here with the difference that the sum is over the phase shifts that are labelled by the point symmetry of the impurity instead of the momentum $\mathbf{k}$ that labels the quantum numbers for a translation symmetry of a crystal [31]. But there is a great difference because the impurity cannot induce a phase transition in an extended solid and the system evolves continuously as the interaction is increased. There are no caveats and we can be confident that the Friedel sum rule is valid no matter how strong are the interactions on the impurity.

Knowing only the sum rule we can draw conclusions about the Kondo effect, one of the classic problem of condenser matter physics described in the next section. At zero temperature resistance caused by an impurity in a metal is determined by the phase shifts. A widely-used model for the Kondo effect and the Anderson impurity model is a one-band model where there is only one phase shift. If the band is chosen to be symmetric, the impurity state is half-filled and it follows that the impurity causes the maximum possible resistance no matter how strong is the interaction or how weak is the coupling to the host. It is not many steps to an understanding of the Kondo effect at least qualitatively.

[7] However, it is tested in the case where there is an exact solution by Lieb and Wu [29] for the one dimensional Hubbard model. There the Luttinger theorem is rigorously satisfied for any particle filling. The excitation spectrum is different from a Fermi liquid, but nevertheless there is a Fermi surface and it is the same independent of interactions, except at $1/2$ filling where there is a gap.

4 Instructive models

4.1 The Anderson impurity model and the Kondo effect

The Kondo effect is the observed temperature dependence of the resistivity of solids containing transition metal impurities.[8] The expected behavior is a decrease as T is lowered, since the resistivity due to dynamical effects such as phonons decreases. It should never increase and at low temperature it reaches a plateau determined by the static disorder. However, the Kondo effect is that there is a resistance minimum after which it increases and saturates for $T \to 0$, contrary to all expectations. This set up the problem of a spin embedded in a metal, which is a strong-coupling problem that defied solution until Wilson's invention of the numerical renormalization group [32]. The problem is that a perturbation series *in the coupling of the spin to the metal* diverges due to the low energy excitations in the metal.

If we step back and note that the spin is actually electrons, the problem can be cast as the Anderson impurity model (AIM) with strong interactions on the impurity site [33]. This is depicted in Fig. 2, where the central site is shown with lines that denote the states of the interacting system on the impurity, while the surrounding lattice is considered as a non-interacting metal. As discussed in the previous section, the principle of continuity and the Friedel sum rule leads to the conclusion that the system must evolve, as the interaction increases, keeping the sum of phase shifts unchanged, so that the resistance at $T = 0$ is constrained. For the simplest problem of a single band there is only one phase shift and the resistance is the maximum possible value at $T = 0$. There are now exact solutions [34] and useful approximations that fully support the analysis.

Anderson [33] provided an instructive analysis illustrated in Fig. 3. Consider the case of a half-filled band with a symmetric density of states and an added state on one site at the Fermi energy so that it is half-filled. If there are no interactions this leads to a density of states on the impurity site that is broadened due to coupling to the continuum of host states as depicted schematically in the lower panel of Fig. 3. If we add an interaction U only on the impurity site, the problem remains symmetric if the energy to remove an electron is $-U/2$ below the Fermi energy and to add a second electron on the site is increased by U, i.e., $U/2$ above the Fermi energy. In this case the impurity state must remain half-filled by symmetry and a restricted Hartree-Fock solution gives exactly the same density of states. However, there can be a broken-symmetry unrestricted solution with one spin mainly occupied and the other mainly empty; for U larger than a critical value this is the lowest energy state as illustrated in the middle panel. As pointed out by Anderson, this is fundamentally incorrect but it heralds the strongly interacting regime where something interesting happens. The exact solution for $T = 0$ is shown schematically at the top in the large U regime; it has both behaviors: the feature at the Fermi energy that is required by the Friedel sum rule and the high energy features that are at energies similar to that given by the unrestricted Hartree-Fock approximation.

[8] Ted Geballe told me that in 1955 he asked John Bardeen to name the most important unsolved problems in condensed matter physics. Bardeen replied, "superconductivity" and, after a characteristic pause, "the resistance minimum" now called the Kondo effect.

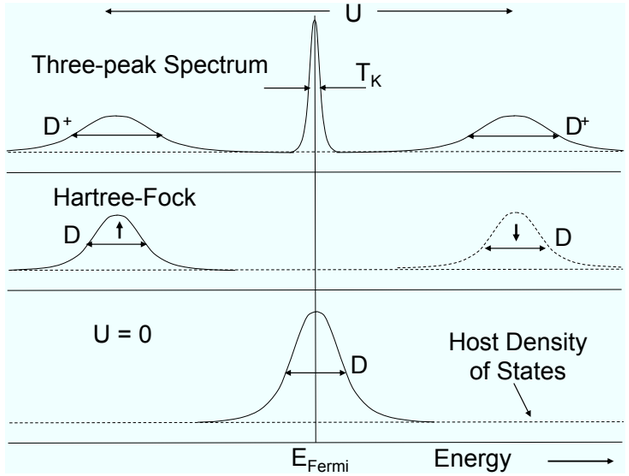


Fig. 3: *Schematic illustration of the spectra for an Anderson impurity in the regime that leads to the Kondo effect. As described in the text, the bottom panel shows the broadened peak for the non-interacting case; the middle illustrates the peaks above and below the Fermi energy for the broken-symmetry unrestricted Hartree-Fock solution in the large U regime; and the top shows the qualitative form of the full spectrum that has a peak at the Fermi energy, which satisfies the Friedel sum rule.*

The spectrum in Fig. 3 already shows how we can understand some important aspects of solids illustrated later. The "three-peak" behavior of the spectral density is a characteristic signature of systems that are often called strongly correlated. We can expect high-energy features due to the strong local interactions and low-energy features that lead to strong temperature dependence and large susceptibility; for the impurity there can be no phase transitions but in a crystal with interactions on every site, the large susceptibility can lead to transitions to various kinds of ordered states. Even though the unrestricted Hartree-Fock calculation is unphysical, it indicates the regime of strong interactions. The temperature dependence is not shown in Fig. 3, but it is similar to that shown later in Fig. 5, which shows that the peak at the Fermi energy disappears as T is increased, another signature of strong interactions.
Notice the "strongly interacting" problem with large U is continuously connected to the non-interacting $U = 0$ state. There is no sharp division between a Mg impurity Na and a Mn impurity in Au. What is the difference? The temperature scale that is denoted by the Kondo temperature!

4.2 The Hubbard and Gutzwiller approximations

Although we have rigorous solutions for the AIM and 1d Hubbard model, we do not yet know whether the conclusions can be used in other problems. Let us try the opposite approach: approximations that are so extreme that they can be solved analytically. The question is what aspects can we use to make confident conclusions about the complicated many-body problem.

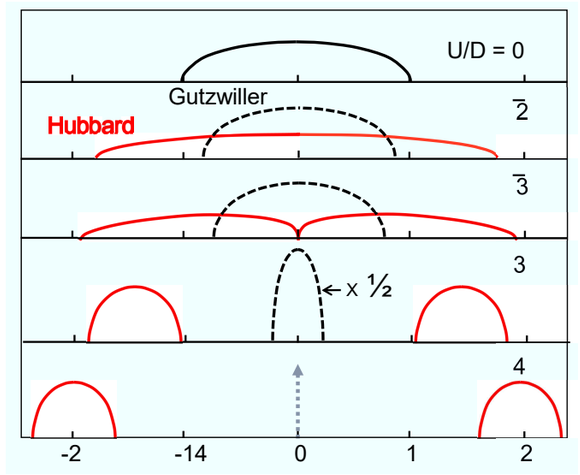


Fig. 4: *Spectral function for a half-filled band with a semicircular density of states shown at the top. The lower four panels show the spectra calculated in the Gutzwiller (dashed lines) and Hubbard (solid lines) approximations as the interaction U increases. The Gutzwiller approximation leads to a narrowed band with a mass (inverse of the band width) that diverges and above a critical U_G there is no solution, which is a model for a metal-insulator transition. In the Hubbard "alloy" approximation the band widens and above U_H a gap opens to form an insulating state.*

In fact the approximations considered in this section were both proposed in 1965 and they are very enlightening for all the methods and examples in the rest of this chapter!
The approximations described here can be applied to any lattice, but there is a special case that allows an analytic solution, the semicircular form for the independent-particle density of states shown at the top of Fig. 4 where

$$\rho(\omega) = -\frac{1}{\pi}\mathrm{Im}G_0(\omega) = \frac{2}{\pi D^2}\sqrt{D^2-\omega^2}, \tag{2}$$

which is non-zero only for $|\omega| < D$ and D is $1/2$ the band width. This is the actual density of states for the Bethe lattice in infinite dimensions so that it can be interpreted as an lattice where each atom has an infinite number of neighbors. Since mean-field theory is exact for an infinite number of neighbors where fluctuations average out, DMFT provides the exact solution, as discussed in the following section.
The approach of Gutzwiller [35] was to propose a variational form for the *ground state wavefunction* that is the same as independent particles (or some other mean-field solution) multiplied by a factor that reduces the probability of double occupation. This is widely used and can be evaluated essentially exactly by Monte Carlo methods. The Gutzwiller approximation is to ignore the correlation of the occupation of electrons on different sites; this leads to equations that can be solved in general numerically and analytically for simple models. Minimization of the energy as a function of U leads to a state with reduced kinetic energy interpreted as a

band narrowing or increased mass. For a one-band Hubbard model the result is that the Fermi surface is unchanged but the mass increases as the density approaches half-filling and exactly at half-filling there is a critical U_c^G where the solution is for infinite mass. This is illustrated for the semicircular density of states at half-filling in Fig. 4. This is widely used as a scenario for a Mott transition as a function of band filling: bands with increasing mass as the density approaches $1/2$ until it diverges signalling an insulator.

Hubbard proposed a different approach in terms of the *Green function* [36]. The propagation of an electron in an interacting system is approximated by assuming the rest of the electrons act as a static array of scatterers, called an "alloy approximation." The further approximation that allows analytic solution is the coherent potential approximation which ignores the correlation between occupation on different sites, which was derived independently by Hubbard and others and is now widely used in real alloys. This leads to a Green function given by Eq. (1) with an on-site self-energy $\Sigma(\omega)$, like in DMFT, but a static mean-field approximation. The result is that the states near the Fermi energy have a lifetime, which may be reasonable for high temperature but is fundamentally incorrect at $T = 0$. On the other hand it leads to a gap for large U as illustrated in Fig. 4 with well-defined band edges. This provides a different scenario for a Mott transition as a function of the interaction: a spectrum that broadens until a gap opens above a critical U_c^H.

Thus we have two widely-used pictures of a Mott transition: one is a band at the Fermi energy that narrows as the interaction increases and the other a broadening of the bands until they are splitting into what are called the upper and lower "Hubbard bands." Each ignores correlation of occupation on different sites and assumes no change of symmetry, and each can be argued to capture a part of the physics.

4.3 Dynamical mean-field theory (DMFT) for the Hubbard model

As discussed in Sec. 2.4, DMFT is an approach to treat interacting electrons in a crystal, but the heart of the many-body calculation is the calculation of the on-site Green function and self-energy for an embedded site that is equivalent to a self-consistent Anderson impurity model (AIM). This section illustrates DMFT by showing a few results for a one-band Hubbard model. It is especially appropriate to consider the semicircular density of states defined in Eq. (2) which allows a direct comparison with the Hubbard and Gutzwiller approximations shown in Fig. 4. Because the single-site DMFT is exact in the limit of infinite dimensions, the result is the exact solution for the Bethe lattice with an infinite number of neighbors. Thus this is an exact solution for a special model and our job is to recognize when the results can help us understand the properties of real complex systems in finite dimensions![9]

[9] In finite dimensions, this is an approximation and an approach to an exact solution requires that the calculations be done for large enough clusters that the correlations between electrons on different sites has converged. For examples like the two-dimensional Hubbard model various size clusters have been treated, which is essential for some important effects [37]. However, this is generally not feasible for calculations for realistic systems with many-bands and all the DMFT calculations in the examples for real materials in the following section are done in the single-site approximation.

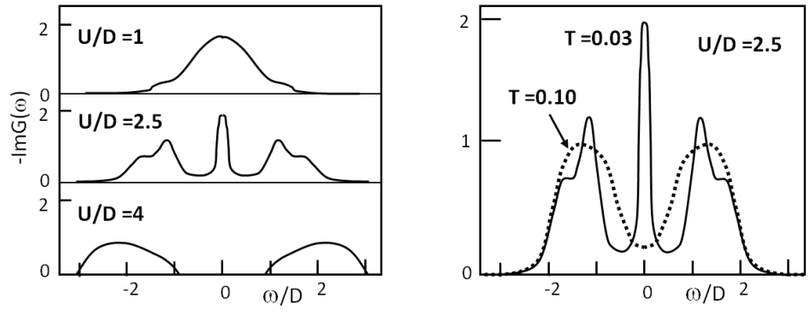


Fig. 5: *Spectral function for a half-filled band with a semicircular density of states calculated using DMFT in the single site approximation. The spectra have features like the Anderson impurity model in Fig. 3 and the approximations for a Bethe lattice in Fig. 4: Gutzwiller (peak at the Fermi level that narrows as the interactions U increases) and Hubbard (spectra that broadens as U increases with weight that decreases at the Fermi level until a gap opens). At the right is shown the effect of temperature: the feature at the Fermi energy at low temperature vanishes at high temperature where the states act like decoupled spins and the spectral weight is transferred to the high energy peaks. Adapted from figures in [22].*

The results of DMFT calculations are shown in Figs. 5 and 6. Here we skip all the details of the calculations and give only a few results. As expected, the spectra shown in the left side of Fig. 5 have the characteristics of those for the AIM shown in Fig. 3, a central peak that narrows as the interaction increases and high energy peaks separated by U. The right side of Fig. 5 shows the temperature dependence for a case where the solution has a narrow central peak that disappears as T increases. This is not a broadening but a true many-body effect. For the AIM the analogous result is the Kondo effect where the resistance decreases at T increases. For a lattice, it is the Fermi surface that disappears, which is illustrated later in Fig. 8.

However, unlike the AIM, the self-consistency leads to a very different consequence: As U increases the central peak disappears at some value of U_c^1 and a gap opens at a value U_c^2. Thus the DMFT result shows both behaviors captured by the Gutzwiller and Hubbard approximations shown in Fig. 4. This provides a scenario for a Mott transition as a function of the interaction that is the exact solution for a model! Whether or not this applies to a real problem in finite dimensions is an issue. Furthermore, if we look more carefully into the solutions for the Hubbard model there is an important issue not considered so far, and a deep issue of physics still to be understood.

The issues are illustrated in Fig. 6. A result of the DMFT calculations is that for some regime of interaction U there are two types of solutions, an insulating solution with a gap for $U > U_c^1$ and a metallic solution with a central peak for all $U < U_c^2$, with $U_c^1 < U_c^2$. In the region from U_c^1 to U_c^2 the two solutions indicate two phases with the same symmetry so that there is a first-order phase transition. Furthermore, as a function of temperature the central peak disappears and the gap fills in until there is a critical point and the solutions merge like the water steam transition. This is shown in the left side of Fig. 6 by the dashed line ending in a critical point, which has the features of a Mott transition.

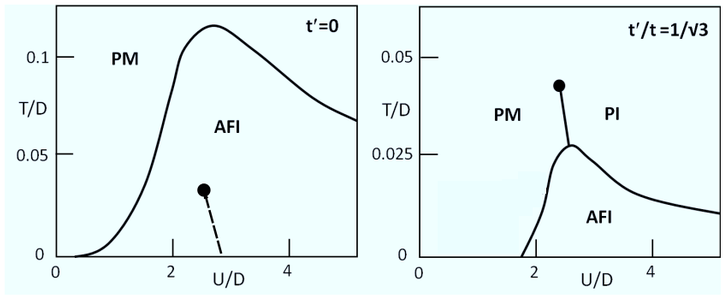


Fig. 6: *Phase diagram for Bethe-lattice in infinite dimensions which has a semi-circular density of states in the limit of infinite dimensions. Left: Results for a Hamiltonian that has only nearest neighbor hopping t; the dashed line ending in a critical point indicates a transition between the two types of spectra in Fig. 5. However, for this model the actual transition is to an antiferromagnetic insulator (AFI) indicated by the solid line. At the right is the phase diagram if a second-neighbor hopping t' is included; the AFI state is not as favorable (note the expanded vertical scale) and there is a range of temperatures where the first-order metal-insulator transition occurs. (For this t' there is also an antiferromagnetic metal phase that is not shown and is not important for our purposes.) Adapted from figures in [22].*

However, there is more that can be learned from the single-site DMFT calculation. The previous calculation is in fact *restricted DMFT where it is required that there is no order of the spins or any other order.* One possibility is an antiferromagnetic state where each site is allowed to have a net spin and it is surrounded by sites with opposite spin. Such a solution is allowed for any lattice that is "bipartite," i.e., that can be divided into A and B sites, where all A sites have only B neighbors and vice versa. The Bethe lattice has this property and we can expect that such a state is favored just as it is in the multitude of other models and actual materials that order antiferromagnetically. This can still be solved by single-site calculation where the A and B sites are related by time reversal symmetry. Indeed the calculations find an antiferromagnetic phase for temperatures and values of U shown by the solid line in the left side of Fig. 6. *The first-order transition is completely eliminated and the only stable phases are a metal with no order and an ordered antiferromagnet.*

Nevertheless, there is a way to uncover the first order transition. A second-neighbor hopping (denoted t') couples an A site to the nearest other A site (the same for B sites) which tends to frustrate the antiferromagnetic order and reduce the temperature range where the antiferromagnetic insulator is the stable phase. It turns out that the independent-particle density of states is still semicircular, but with modified width, and the calculations can still be done within the same model. This is shown in the right side of Fig. 6 (note the reduced temperature scale). Now the first-order transition can actually occur for some temperature range about the antiferromagnetic transition. This is the behavior known from the classic example, V_2O_3, discussed in Sec. 5.3.

Finally, we arrive at a deep issue that opens the door (or not) to new states and phenomena that are qualitatively different from normal states of matter. If the antiferromagnetic order can be eliminated could the insulating state extend all the way to $T = 0$? Further discussions are postponed to later.

5 A few examples: Data, interpretation, and theory

5.1 Cerium: volume collapse phase transition, heavy fermions, ...

Experimental facts and interpretations. Cerium is the first transition element in the lanthanide series. In various compounds it exhibits behavior that is readily identified as due to one electron in the localized, atomic-like $4f$ shell (f^1) (for example, in magnetic Ce_2O_3 where Ce has a formal valence $+3$) or an empty $4f$ shell (f^0) (as in non-magnetic CeO_2, formal valence $+4$). As an impurity in metallic La it leads to the Kondo effect expected for a localized f^1 embedded in a metal. Some compounds such as $CeIrIn_5$ are "heavy fermion" materials [42] with masses (inversely proportional the band widths) orders of magnitude larger than ordinary metallic bands, similarly to the Kondo effect for impurities where the energy scale can vary over orders of magnitude for different impurities and hosts.

For many decades it has been known that elemental Ce has a phase diagram as a function of temperature and pressure with a first-order transition between the high volume γ- and low volume α-phases, each with the same fcc structure. There is an anomalously large 15% volume change at room temperature and a critical point at higher temperature analogous to the water-steam transition. The γ phase has a temperature-dependent magnetic susceptibly accounted for by weakly coupled f^1 spins, whereas the α phase appears to be inert magnetically. The first interpretation was the natural one: a transition between magnetic f^1 states and non-magnetic f^0 with an electron transferred to the weakly interacting metallic bands. However, more recent experiments showed that the $4f$ occupation in the α phase is also close to f^1, for example, as indicated by the photoemission and inverse photoemission data shown at the top of Fig. 7, where the sidebands below and above the Fermi energy indicate the strong interactions $U \approx 7$ eV and f states in both phases, with bands at the Fermi energy in the α phase like the three peak structure for the Anderson impurity model in Fig. 3.

Guiding principles for theory. We are faced with a strongly interacting problem of a lattice of localized f states coupled to delocalized metallic bands. It is certainly very difficult to explain in detail all the properties of cerium and its compounds.[10] The purpose of the present discussion is to examine what we can understand from experiment alone and what we can believe from DFT calculations that are simple to do now-a-days, but which are clearly inadequate in some respects. Furthermore we want the reasoning to apply, a least qualitatively, not just for one material but for many different cases. The same reasoning is also invaluable in judging what can be believed from a heavy many-body calculation, which certainly involves assumptions and approximations.

A great success of DFT calculations using standard functionals like LDA is that for Ce the calculations find an equilibrium volume near that for the α phase and a $4f$ occupation of ap-

[10] Two mechanisms have been proposed to explain the α-γ transition caused by the large interactions for electrons in the $4f$ states. In one proposal the primary effect is a "Mott transition" of the f states; the d states do not play an essential role. The other proposal is a "Kondo volume collapse" in which the primary driving force is the coupling of the f states to the band-like d states; this is a simplified version of the results of the DMFT calculations described here.

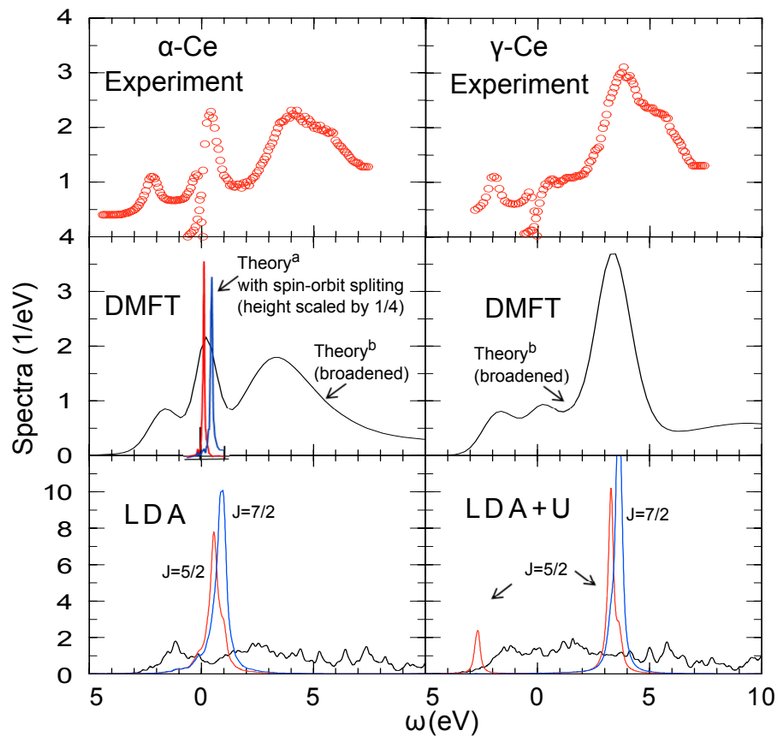


Fig. 7: *Electron removal (photoemission) and addition (inverse photoemission) spectra for α (left) and γ (right) phases of Ce. Top: Measured spectra at room temperature [38] that is mainly the Ce $4f$ component. Middle: Single-site DMFT calculations for the $4f$ states with parameters from LDA. For the α phase (left) theory "b" [39, 40] is broadened by the experimental resolution and includes the $6s$–$6p$–$5d$ states. Theory "a" [41] shows the high-resolution peaks near the Fermi energy including spin-orbit splitting. Note widths are $\approx 10\times$ narrower than the LDA peaks in the bottom panel. For the γ phase (right) the calculation finds upper and lower Hubbard bands with no peak at the Fermi energy. Bottom: The density of states for static mean-field calculations: LDA for a non-magnetic solution in the α phase and unrestricted LDA+U for a magnetically ordered state in the γ phase. The peaks are $4f$ and the broad spectra are bands formed from $6s$–$6p$–$5d$ states. Each static mean-field calculation describes an aspect of the spectrum, but cannot explain the three-peak spectrum found in the DMFT. Similar to figure in [1] provided by A.K. McMahan except "theory a" modified from figure provided by K. Haule.*

proximately one. Such a calculation is *guaranteed* to have partially-filled $4f$ bands at the Fermi energy, as illustrated in the bottom left panel of Fig. 7, which satisfy the Luttinger theorem including the f states. Quantitatively, the bands are narrow, but yet broad enough that there would be only band-like paramagnetism, consistent with what is observed in the α phase. However, the LDA calculation cannot explain the peaks well above and below the Fermi energy that indicate a large interaction.

What can we conclude about the results of a heavy many-body calculation even before it is done? So long as the system does not have a transition to an ordered magnetic state and the temperature is much lower than the characteristic energy scales, continuity implies that the Luttinger theorem is obeyed and *any correct many-body calculation* must give bands that are well-defined at the Fermi energy, even though they may be much narrower than the LDA bands and the spectrum is very different as one goes away from the Fermi energy. This is illustrated by the results shown in Fig. 7 for DMFT calculations done in the single-site approximation.

What about the γ phase? Another success of DFT calculations is that if one allows for spin polarization there is another solution with larger volume close to that observed for the γ phase at room temperature. The spectrum is illustrated in the lower right panel of Fig. 7, which shows the result of an "LDA+U" calculation for a magnetically ordered state. The spectrum is qualitatively like the experiment with apparently no $4f$ bands at the Fermi energy, similar to the unrestricted Hartree-Fock approximation in Fig. 3. Is this inconsistent with the Luttinger theorem? Before we do a many-body calculation, what do we expect to be the different possible results? If the temperature is high compared to some characteristic energy scale (the scale of interactions between spin on different sites), we do not expect there to be bands at the Fermi energy! Indeed, this is the result of the single-site DMFT calculation for the model in Fig. 5, where the central peak vanishes at high temperature, and the same behavior is found in the DMFT calculation for γ-Ce shown in the middle right panel of Fig. 7.

Consider now the heavy fermion material $CeIrIn_5$ for which DMFT results [43] are shown in Fig. 8. From experiment such as specific heat measurements we know the characteristic energy scale is very low, somewhat above 10 K and there is no magnetic order at 10 K. The figure shows the huge effects of interactions: At room temperature the bands at the Fermi energy are almost like an LDA calculation at $T = 0$ *with the f states artificially removed*. This mimics the fact that the f states are essentially decoupled at this temperature and result is a "small fermi surface" that does not include f electrons. At low temperature the bands are like an LDA calculation with the f states included, i.e., a "large Fermi surface," but they are greatly renormalized as shown in the right side of Fig. 8 where the scale is expanded by a factor of 100.

What can we conclude? None of the calculations definitively establishes the mechanisms for the behaviors and there is no theoretical proof that the systems do not order in some way. But if we take from experiment that there is no transition to some ordered phase at the relevant temperatures, then we can have a qualitative picture without doing a heavy many-body calculation. Furthermore, since there is no order, the single-site approximation is reasonable and it appears to be an appropriate starting point for quantitative understanding. The fact that low energy scales emerge implies that susceptibilities are large such that there may be transformations to other states at lower temperature, which is found in many heavy fermion materials [42]. Finally, a lesson from the previous sections is that if we seek the behavior at $T = 0$, we expect the system to be either a Fermi liquid or have a transition to an ordered state, and we should be very careful about using $T = 0$ arguments at $T > 0$ and vice versa.

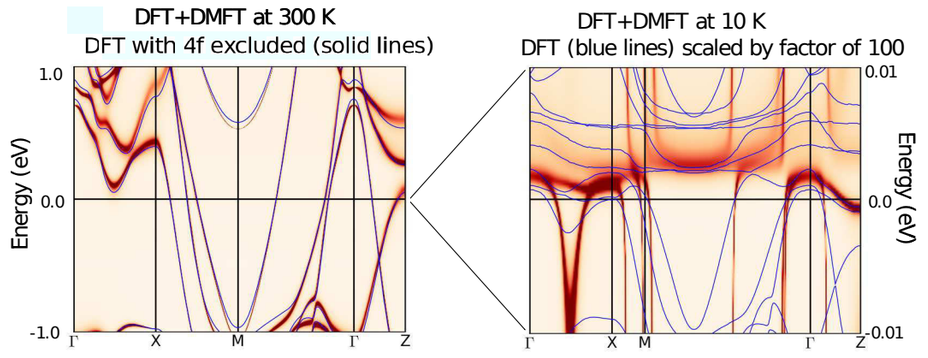


Fig. 8: *Momentum resolved spectra at 300 K (left) and 10 K (right) calculated using single-site DMFT. As discussed in the text at 300 K the $4f$ weight at the Fermi energy is very small, but a very narrow $4f$ band (note the greatly expanded scale energy in the right figure!) emerges in analogy to the Kondo effect and the model DMFT calculation shown on the right side of Fig. 6. From [43].*

5.2 NiO and Mott insulators

NiO is the original "Mott insulator" as identified by de Boer and Verwey [48] and pointed out by Mott and Peierls [49] in 1937. Below the Curie temperature of 525 K, it is an antiferromagnetic insulator with a gap around 4 eV, which can be seen from the spectra shown in the top panel of in Fig. 9. It satisfies the narrower condition for a Mott insulator: If there is no order there should be fractionally filled bands that would indicate a metal, but in fact it acts like an insulator with a gap that does not depend on the magnetic order and remains above the Curie temperature (which is found in experiments not shown in the figure). The structure is fcc and there are no complications about which d states are involved (unlike the other classic case V_2O_3).

Guiding principles. This case is simpler than the various problems for Ce compounds since NiO is an insulator with a gap, and it allows us to ask simple direct questions. What about the Luttinger theorem when there is no broken symmetry? Since this is a "Mott insulator" can we trust any results from methods such as the GW approximation, based upon perturbation theory? The theoretical results shown in Fig. 9 suggest that we have a rather good understanding of the basic issues even if there is much yet to be understood. As in many systems, DFT calculations are very successful in predicting the structure. At the bottom is the spectra that shows the result from an LDA calculation for the ordered antiferromagnetic structure; there is a gap, but it is very small and not visible in the figure since the spectra are broadened to compare with the experiment. In addition to the gap being too small, the relative positions of the oxygen p and the Ni d states are wrong. Improved DFT calculations using DFT+U (a generalized gradient functional GGA + U in this case) functionals open the gap but cannot give both the magnitude of the gap and the positions of the p and d in agreement with experiment. The HSE hybrid density and Hartree-Fock functional leads to much improved spectra. An example of GW calculation is shown for "one shot" G_0W_0 starting from the HSE wavefunctions and eigenvalues. Similar results are found using a self-consistent GW method that is independent of the starting point.

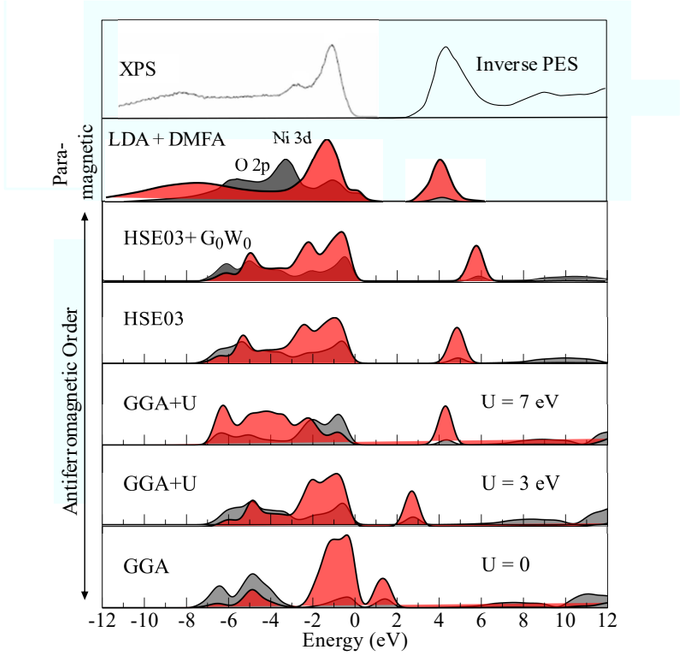


Fig. 9: *Spectra for electron removal and addition in NiO. Top: experimental results for* $3d$ *weight from [44] and [45]. The other panels are calculated spectra projected on the O-*p *(grey) and Ni-*$3d$ *(red) components. The lower four panels: DFT calculations with different functionals, above these: GW calculation, all for the antiferromagnetic state [46]. The top theoretical results are the spectra for the paramagnetic state calculated using single-site DMFT (called DMFA in the figure) with parameters from DFT calculations [47]. Similar to a figure in [1].*

These methods are not chosen to fit NiO; the same approximations give improved results for semiconductors, large classes of oxides, and other systems. The conclusion is that DFT and GW calculations can be very useful and describe systems like NiO *so long as we restrict ourselves to the ordered state at low temperature.*

The gap and insulating behavior of NiO remain for temperatures above the Curie temperature where there is no ordered moment. In this case *any* $T = 0$ *DFT or GW-type calculation must lead to a metal*; this is not shown but must occur by electron counting for partially filled bands that originate from atomic-like d states. Why do these methods not work for NiO above the Curie temperature where there is no order? The simple answer is that temperature has a large effect because the large on-site interactions lead to local moments that persist above the Curie temperature. Approaches that take the disordered moments into account can explain the gap in the high temperature state. A simple approach is a static approximation like the Hubbard alloy approximation that opens a gap like that shown in Fig. 4. A greatly improved approach is the DMFT calculation for the dynamically disordered system shown in Fig. 9. This is the result within the single-site approximation which means that correlations between electrons on a site are included but correlations between different are sites treated only as an average mean-field, which is justified at high enough temperature.

5.3 V_2O_3 and the Mott transition

Whereas NiO is described as a Mott insulator, V_2O_3 is the classic example of what has been termed a Mott transition. The famous phase diagram [50] is not shown here but the basic points are illustrated in the phase diagram in the right side of Fig. 6. Experimentally the different parts of the phase diagram are accessed by applying pressure and doping with Cr. Pressure squeezes the atoms together resulting in larger independent-particle hopping terms and a decrease in the relative importance of interactions, which favors the metal. At low temperature and low pressure V_2O_3 is an antiferromagnetic insulator, like the right side of Fig. 6. Applying pressure corresponds to moving to the left in the figure until there is a transition to a paramagnetic metal; these are only two states observed at low temperature. With Cr doping there is also a first-order transition in the paramagnetic phase at higher temperature, like the phase line in Fig. 6 that ends in a critical point. It is this transition that has long been taken as the classic example of a Mott transition with no change of symmetry. Because of the complicated crystal structure and possible ordering of the electronic states, there have been many models proposed to explain the behavior. But for our purposes the only relevant point is that there exists a transition with a change in the spectrum and conductivity from a form expected for an insulator to that for a metal in some range of temperature and pressure. However, the two states are not really distinct since they merge continuously above the critical point.

Does this violate the Luttinger theorem? Just as the analysis for NiO, it does not because the transition is only observed at high temperature. It is important to understand the reasons for insulating-like and metallic-like behavior, but it should be done with the understanding that it is not inconsistent with the theorem unless the transition actually continues all the way to zero temperature, which leads up to the issues in the following section.

6 The Mott insulator and topological order

Several times in the previous sections there were references to deep issues in physics that are brought to light by relentlessly pursuing the question of whether or not a Mott insulator with no order can exist at zero temperature. Recall that the difference between a metal and an insulator is precisely defined only at $T = 0$, and the Fermi surface and Luttinger theorem are precisely defined only at $T = 0$. An insulating state with fractional occupation and no order $T = 0$ (termed a quantum spin-liquid; see, e.g., [4] and papers referred to there) would violate the original statement of Luttinger theorem. There is now growing evidence that such a state would have some form of topological order[11] and would be a state of matter with quantum order not described by any classical order parameter [51]. There is a well-known example, superconductivity, and it would be an extraordinary advance to discover other states of matter with quantum order. For states with topological order, there is not a continuous connection to

[11] Topological insulators that have been of great interest recently are band insulators with a Hamiltonian $H(\mathbf{k})$ that has non-trivial topology as a function of $\mathbf{k}$ in the Brillouin zone. For a Mott insulator there may be related properties, but this has not been worked out to my knowledge.

an independent-particle system, and many works using different approaches find that at a transition to a topologically ordered state the volume enclosed by the Fermi surface can change by a discrete amount. This leads to a possible new formulation of the Luttinger theorem that extends the original derivation to systems with topological order. Perhaps the most direct conclusions related to the Luttinger theorem are due to Sachdev and coworkers who have argued that any state that does not satisfy the original Luttinger theorem must have topological order. (See, e.g., [52].)

7 A few final remarks

It is nice to appreciate the prescient work of the 1960's that are the basis of much of the work today and still provide very valuable insights, along with the advances of the 1980's and early 1990's that have provided new ideas and methods that are the basis for actual calculations done today. As the calculations have become more and more powerful we have reached the point where quantitative theory is an essential part of research on actual materials in actual experimental conditions. Despite great successes in many cases, there are failures in others. In addition, present day theories often use different methods for different regimes of correlation, for example, the band-like behavior for the materials at the lower left in Fig. 1 and the extreme localized local-moment behavior for materials in the upper right. It is the materials on the borderline that have interesting, anomalous behavior; they present the greatest challenges because correlation plays an especially important role. At such points we need guiding principles to keep the ship afloat and not be caught on the rocks.

Consider the example presented in Sec. 5.1 of cerium in the α and γ phases, where temperature is an essential ingredient in the complete picture. It is a luxury to be able to carry out a heavy many-body calculation like DMFT that can span the range of behaviors within one method, but the main point of the discussion in Sec. 5.1 is that we can understand the qualitative behavior using methods like DFT if we use clear guiding principles like the Luttinger theorem together with well-established properties of models such as the spectra of the Anderson impurity model and the temperature dependence in the Kondo effect. The same guiding principles help clarify the limitations of methods like the single-site approximation (or other approximations) in DMFT and help us understand the results.

The guiding principles emphasized here have many tentacles that reach into the very essence of the goal of making robust theories and conclusions that can be trusted. In all the cases described here, whether real materials or models, there is no rigorous proof that the calculations are the final answer. It is very difficult to prove that one has found the global minimum free energy state, i.e., there can be no state with a different type of order that is lower in free energy. Nevertheless, there are many very useful lessons to be learned and interesting results, if they are used judiciously. The combination of experiment, theory, and computational methods, each held to high standards, can provide solid understanding and even open doors to possible new states of matter.

References

[1] R.M. Martin, L. Reining, and D.M. Ceperley: *Interacting Electrons: Theory and Computational Approaches* (Cambridge University Press, 2016)

[2] K.T. Moore and G. van der Laan, Rev. Mod. Phys. **81**, 235 (2009)

[3] W. Kohn and J.M. Luttinger, Phys. Rev. **118**, 41 (1960)

[4] M.R. Norman, Rev. Mod. Phys. **88**, 041002 (2016)

[5] R.M. Martin: *Electronic Structure: Basic Theory and Methods* (Cambridge University Press, 2004)

[6] E. Pavarini, E. Koch, and U. Schollwöck (eds.): *Emergent Phenomena in Correlated Matter* Modeling and Simulation, Vol. 3 (Forschungszentrum Jülich, 2013) `http://www.cond-mat.de/events/correl13`

[7] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964)

[8] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965)

[9] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985)

[10] D.C. Langreth, B.I. Lundqvist, S.D. Chakarova-Käck, V.R. Cooper, M. Dion, P. Hyldgaard, A. Kelkkanen, J. Kleis, L. Kong, S. Li, P.G. Moses, E. Murray, A. Puzder, H. Rydberg, E. Schröder, and T. Thonhauser, Journal of Physics: Condensed Matter **21**, 084203 (2009)

[11] M. Lüders, M.A.L. Marques, N.N. Lathiotakis, A. Floris, G. Profeta, L. Fast, A. Continenza, S. Massidda, and E.K.U. Gross, Phys. Rev. B **72**, 024545 (2005)

[12] N.D. Mermin, Phys. Rev. **137**, A1441 (1965)

[13] W.L. McMillan, Phys. Rev. **138**, A442 (1965)

[14] D.M. Ceperley and B.J. Alder, Phys. Rev. B **36**, 2092 (1987)

[15] E. Gull, A.J. Millis, A.I. Lichtenstein, A.N. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011)

[16] L. Hedin, Phys. Rev. **139**, A796 (1965)

[17] F. Aryasetiawan and O. Gunnarsson, Rep. Prog. Phys. **61**, 237 (1998)

[18] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)

[19] E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989)

[20] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)

[21] M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992)

[22] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)

[23] E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.):
The LDA+DMFT approach to strongly correlated materials
Modeling and Simulations, Vol. 1 (Forschungszentrum Jülich, 2011)
`http://www.cond-mat.de/events/correl11`

[24] E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.):
DMFT at 25:Infinite Dimensions
Modeling and Simulation, Vol. 4 (Forschungszentrum Jülich, 2014)
`http://www.cond-mat.de/events/correl14`

[25] G. Knizia and G.K.-L. Chan, Phys. Rev. Lett. **109**, 186404 (2012)

[26] R.M. Fye and J.E. Hirsch, Phys. Rev. B **38**, 433 (1988)

[27] J.M. Luttinger and J.C. Ward, Phys. Rev. **118**, 1417 (1960)

[28] J.M. Luttinger, Phys. Rev. **119**, 1153 (1960)

[29] E. Lieb and F.Y. Wu, Phys. Rev. Lett. **20**, 1445 (1968)

[30] J. Friedel, Phil. Mag. **43**, 153 (1952)

[31] R.M. Martin, Phys. Rev. Lett. **48**, 362 (1982)

[32] K.G. Wilson, Rev. Mod. Phys. **47**, 773 (1975)

[33] P.W. Anderson, Phys. Rev. **124**, 41 (1961)

[34] N.A., K. Furuya, and J.H. Lowenstein, Rev. Mod. Phys. **55**, 331 (1983)

[35] M.C. Gutzwiller, Phys. Rev. **137**, A1726 (1965)

[36] J. Hubbard, Royal Society of London Proceedings Series A **281**, 401 (1964)

[37] E. Gull, O. Parcollet, P. Werner, and A.J. Millis, Phys. Rev. B **80**, 245102 (2009)

[38] M. Grioni, P. Weibel, D. Malterre, Y. Baer, and L. Du‘o, Phys. Rev. B **55**, 2056 (1997)

[39] K. Held, A.K. McMahan, and R.T. Scalettar, Phys. Rev. Lett. **87**, 276404 (2001)

[40] A.K. McMahan, K. Held, and R.T. Scalettar, Phys. Rev. B **67**, 075108 (2003)

[41] K. Haule, C.-H. Yee, and K. Kim, Phys. Rev. B **81**, 195107 (2010)

[42] G.R. Stewart, Rev. Mod. Phys. **56**, 755 (1984)

[43] H.C. Choi, B.I. Min, J.H. Shim, K. Haule, and G. Kotliar,
Phys. Rev. Lett. **108**, 016402 (2012)

[44] J. Weinen: *Hard X-Ray Photoelectron Spectroscopy: New Opportunities for Solid State Research*. Ph.D. thesis, (Universität zu Köln, 2015)

[45] G.A. Sawatzky and J.W. Allen, Phys. Rev. Lett. **53**, 2339 (1984)

[46] C. Rödl, F. Fuchs, J. Furthmüller, and F. Bechstedt, Phys. Rev. B **79**, 235114 (2009)

[47] J. Kuneš, I. Leonov, M. Kollar, K. Byczuk, V. Anisimov, and D. Vollhardt,
Eur. Phys. J. Special Topics **180**, 5 (2009)

[48] J.H. de Boer and E.J.W. Verwey, Proc.Phys. Soc. **49**, 59 (1937)

[49] N.F. Mott and R. Peierls, Proc. Phys. Soc. London, Ser. A **49**, 72 (1937)

[50] D.B. McWhan, A. Menth, J.P. Remeika, W.F. Brinkman, and T.M. Rice,
Phys. Rev. B **7**, 1920 (1973)

[51] X.-G. Wen: *Quantum Field Theory of Many-Body Systems*
(Oxford University Press, 2004)

[52] S. Sachdev and K. Yang, Phys. Rev. B **73**, 174504 (2006)

2 Insulator, Metal, or Superconductor: The Criteria

Richard T. Scalettar
University of California
Physics Department, Davis CA 95616

Contents

1 Introduction **2**

2 A brief introduction to tight-binding Hamiltonians; Metals and band insulators **3**

3 Antiferromagnetic and charge density wave insulators **6**

4 Anderson and Mott insulators **7**

5 Formal definitions **12**

6 Applications of formal theory **15**

7 Conductivity and spectral function **16**

8 Conclusions **19**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

In the United States (and perhaps also around the world) we have a joke about some students' tendency to try to understand physics by memorizing equations: In comprehending electric circuits, we say, it is important that such students completely master Ohm's three laws for current flow in a metal,

$$V = IR \qquad I = V/R \qquad R = V/I \ . \tag{1}$$

In this chapter, we shall present the subtle relations between resistance, voltage and current, and come to grips with the equations and the deep concepts governing metallic and insulating behavior, and their extension to superconductors. We will see that there is considerably more depth to the field than Ohm's Three Laws, as represented by Eq. (1).

The difference between metals, insulators, and superconductors can be precisely defined, and illustrated, within the framework of tight-binding Hamiltonians (TBH). That will be our primary language here. In addition to developing some analytic approaches to the solution of these Hamiltonians, and hence their characterization into distinct charge transport categories, a good fraction of the material will involve a discussion of how to implement the concepts and equations in precise calculational frameworks, including exact diagonalization and Quantum Monte Carlo (QMC).

The organization of this chapter is as follows. We first describe, in a rather qualitative way, the different types of insulators (band, Anderson, and Mott) which can arise. Our criteria for insulating behavior will focus on the appearance of a gap in the single particle energy levels (band insulator), the appearance of localized eigenfunctions in the presence of disorder (Anderson insulator), or the possibility that interactions between electrons are so strong that motion of electrons is inhibited (Mott insulator). The first two cases can be addressed with some precision with simple calculations, but the latter is much more challenging. In fact, it is fair to say that a full understanding of Mott insulating behavior has not yet been achieved, an especially unfortunate state of affairs since out of Mott insulators many of the most interesting new materials and novel physics develops.

The second part of the chapter develops a more formal set of mathematical criteria for distinguishing metals, insulators, and superconductors, one which focussed directly on the current-current correlation function (and hence, in a sense, can be viewed as a proper treatment of the quantities in Eq. (1) !) This closely follows the discussion of Scalapino, White, and Zhang in Ref. [1]. These criteria will be shown to give sensible results both in simple analytic treatments and also with QMC methods. In the latter case, disorder can also be included, along with interactions.

The final section will outline alternative approaches to distinguishing metallic, insulating, and superconducting behavior which involve an approximate formula for the conductivity and an examination of the single particle spectral function.

2 A brief introduction to tight-binding Hamiltonians Metals and band insulators

Tight-Binding Hamiltonians (TBH) allow for a simplified description of electrons in a solid, which complements methods like density functional theory. Rather than calculate the wave functions in continuum space, one instead focuses on a collection of discrete sites or orbitals which the electrons can occupy and, between which, make transitions. We will assume the student has some familiarity with second quantization, which forms the language of TBHs. We begin with the simplest TBH

$$\hat{H} = -t \sum_{\langle \mathbf{j},\mathbf{l} \rangle \sigma} \left(\hat{c}^\dagger_{\mathbf{j}\sigma} \hat{c}_{\mathbf{l}\sigma} + \hat{c}^\dagger_{\mathbf{l}\sigma} \hat{c}_{\mathbf{j}\sigma} \right) - \mu \sum_{\mathbf{j}} \left(\hat{n}_{\mathbf{j}\uparrow} + \hat{n}_{\mathbf{j}\downarrow} \right) \ . \tag{2}$$

$\hat{H}$ consists of a kinetic energy term which describes the destruction of a fermion of spin σ on site $\mathbf{l}$, via the operator $\hat{c}_{\mathbf{l}\sigma}$, and its re-creation on site $\mathbf{j}$, via the operator $\hat{c}^\dagger_{\mathbf{j}\sigma}$; and a chemical potential term. The creation and destruction operators obey anticommutation relations $\{\hat{c}^\dagger_{\mathbf{j}\sigma}, \hat{c}^\dagger_{\mathbf{l}\sigma'}\} = \{\hat{c}_{\mathbf{j}\sigma}, \hat{c}_{\mathbf{l}\sigma'}\} = 0$ and $\{\hat{c}_{\mathbf{j}\sigma}, \hat{c}^\dagger_{\mathbf{l}\sigma'}\} = \delta_{\mathbf{j}\mathbf{l}}\,\delta_{\sigma\sigma'}$, which guarantee that they describe fermionic particles. As one consequence, the number operators $\hat{n}_{\mathbf{j}\sigma} = \hat{c}^\dagger_{\mathbf{j}\sigma}\hat{c}_{\mathbf{j}\sigma}$ can take only the values $0, 1$.

The symbol $\langle \mathbf{j}, \mathbf{l} \rangle$ in Eq. (2) denotes the collection of pairs of sites between which the hopping of electrons is allowed. Very commonly, this is restricted to the near neighbor sites of some periodic lattice, for example a one-dimensional chain, two dimensional square, triangular, or honeycomb lattice, etc. Because there are no interactions, the two spin species $\sigma = \uparrow, \downarrow$ can, for the moment, be considered independently. We will define the density ρ to be the number of fermions per lattice site.

For most of this chapter, we will assume periodic boundary conditions. In this situation, the translation invariance of the geometry suggests that going to momentum space will simplify our understanding. Indeed, if we introduce

$$\hat{c}^\dagger_{\mathbf{k}\sigma} = \frac{1}{\sqrt{N}} \sum_{\mathbf{j}} e^{+i\mathbf{k}\cdot\mathbf{j}}\, \hat{c}^\dagger_{\mathbf{j}\sigma} \qquad\qquad \hat{c}^\dagger_{\mathbf{j}\sigma} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{j}}\, \hat{c}^\dagger_{\mathbf{k}\sigma} \ , \tag{3}$$

the Hamiltonian Eq. (2) becomes diagonal: rather than destruction on one spatial site being partnered with creation on a *different* spatial site, creation and destruction processes only occur between *identical* momenta. It is worth emphasizing that the new 'momentum creation and destruction operators' obey the same anti-commutation relations of the original operators in real space, so that each of the momenta states $\mathbf{k}$ can be occupied by at most one fermion of each spin species.

Let's consider, for concreteness, a one dimensional chain. The explicit calculation is (ignoring

the chemical potential term)

$$\begin{aligned}\hat{H} &= -t\sum_{j,\sigma}\left(\,\hat{c}^{\dagger}_{j\,\sigma}\hat{c}_{j+1\,\sigma}+\hat{c}^{\dagger}_{j+1\,\sigma}\hat{c}_{j\,\sigma}\,\right)\\ &= -\frac{t}{N}\sum_{j,\sigma}\sum_{k}\sum_{k'}\left(\,e^{-ik\,j}\hat{c}^{\dagger}_{k\,\sigma}e^{+ik'\,(j+1)}\hat{c}_{k'\,\sigma}+e^{-ik\,(j+1)}\hat{c}^{\dagger}_{k\,\sigma}e^{+ik'\,j}\hat{c}_{k'\,\sigma}\,\right)\\ &= -\frac{t}{N}\sum_{k}\sum_{k'}\sum_{j\,\sigma}e^{+i(k'-k)\,j}\left(e^{+ik'}+e^{-ik}\right)\hat{c}^{\dagger}_{k\,\sigma}\hat{c}_{k'\,\sigma}\;. \end{aligned} \tag{4}$$

If we use the orthogonality relation $\sum_j e^{+i(k'-k)\,j} = N\delta_{k\,k'}$ (which is also employed in the inversion of the site to momentum transformation of Eq. (3)) we obtain the Hamiltonian in momentum space

$$\hat{H} = \sum_{k\,\sigma} -2t\cos k\;\hat{c}^{\dagger}_{k\,\sigma}\hat{c}_{k\,\sigma}\;. \tag{5}$$

The structure of Eq. (5) is quite general, that is, also correct in higher dimension and on different lattice structures. For an arbitrary TBH,

$$\hat{H} = \sum_{\mathbf{k}\sigma}\varepsilon_{\mathbf{k}}\,\hat{n}_{\mathbf{k}\,\sigma} \qquad\qquad \hat{n}_{\mathbf{k}\,\sigma} = \hat{c}^{\dagger}_{\mathbf{k}\,\sigma}\hat{c}_{\mathbf{k}\,\sigma}\,. \tag{6}$$

As noted earlier, $\hat{H}$ is diagonal in the momentum indices, so that a state characterized by the occupation of certain momenta is an eigenstate of $\hat{H}$ with an energy equal to the sum of the corresponding $\varepsilon_{\mathbf{k}}$. This is, obviously, not true of position occupation number states. Different lattice geometries are encapsulated in the specific dispersion relation $\varepsilon_{\mathbf{k}}$. Summarizing, then, when viewed in momentum space there is a single, continuous, 'energy band' which is, at $T = 0$ occupied by two, spin $\uparrow$ and $\downarrow$, fermions for all $\varepsilon_{\mathbf{k}} < \mu$. Such a model is always metallic, except, at zero temperature T, in the trivial limits where μ is below the lowest level in the band, i.e., when there are no fermions on the lattice ($\rho = 0$) or when μ is above the highest level in the band, i.e., when every level is occupied ($\rho = 2$).

A more interesting situation arises when multiple energy bands are present. This can occur in a variety of ways. Again focussing on a one dimensional chain, consider an additional staggered potential $\Delta\sum_j(-1)^{\,j}\,n_{j\sigma} = e^{i\pi j}n_{j\sigma}$ in the Hamiltonian. When one goes to momentum space the staggered potential mixes momenta k and $k+\pi$:

$$\Delta\sum_{j}(-1)^{j}\,c^{\dagger}_{j}c_{j} = \Delta\frac{1}{N}\sum_{j}\sum_{k}\sum_{p}\;e^{i\pi j}\,e^{-ikj}\,c^{\dagger}_{k}\,e^{+ipj}\,c_{p}\;=\Delta\sum_{k}c^{\dagger}_{k}c_{k+\pi}\;. \tag{7}$$

(We have used the orthogonality relation $\sum_j e^{+i(p+\pi-k)\,j} = N\delta_{k\,p+\pi}$ again.) Now, going to momentum space has not fully diagonalized the Hamiltonian: the wavevectors k and $k+\pi$ mix. Using the forms already written down for the hopping term,

$$H = \sum_{k}\begin{pmatrix} c^{\dagger}_{k} & c^{\dagger}_{k+\pi}\end{pmatrix}\begin{pmatrix} -2t\cos k & \Delta \\ \Delta & -2t\cos\left(k+\pi\right)\end{pmatrix}\begin{pmatrix} c_{k} \\ c_{k+\pi}\end{pmatrix}, \tag{8}$$

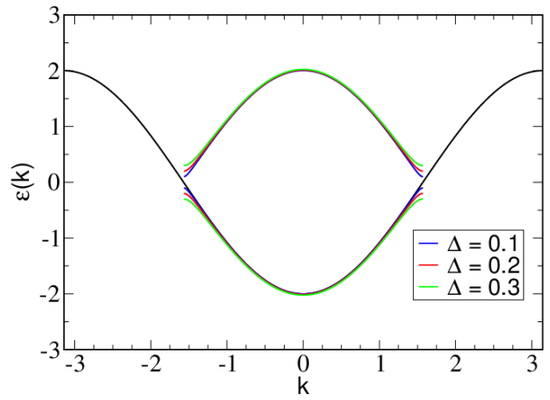


Fig. 1: *The dispersion relation of a one dimensional non-interacting TBH before (black) and after (red, blue, green) a staggered potential $(-1)^j \hat{n}_j$ is added. The staggered potential opens a gap near $k = \pm\pi/2$ and leads to insulating behavior at half-filling.*

where the k sum is over the reduced Brillouin zone $-\pi/2 < k < \pi/2$, so as not to overcount the modes.

This structure is not restricted to a one dimensional chain, but will arise for any 'bipartite' lattice (that is, one whose sites divide into two sets $\mathcal{A}$ and $\mathcal{B}$ such that the neighbors of $\mathcal{A}$ belong only to $\mathcal{B}$ and *vice-versa*. In this general situation, momenta $\mathbf{k}$ and $\mathbf{k} + \vec{\pi}$ mix. One must still do a final diagonalization of the 2x2 matrices in Eq. (8). The allowed energy levels are

$$E_{\mathbf{k}} = \pm\sqrt{\varepsilon_{\mathbf{k}}^2 + \Delta^2} \ , \tag{9}$$

where $\mathbf{k}$ ranges only over the reduced Brillouin zone containing only one of each pair $\mathbf{k}$ and $\mathbf{k} + \vec{\pi}$. The dispersion relation of Eq. (9) has a gap 2Δ separating the positive and negative $E_{\mathbf{k}}$. The system is insulating, not just in the trivial limits when there are no electrons on the lattice or when all sites are fully occupied, but also at half-filling $\rho = 1$, which occurs when the chemical potential $-2\Delta < \mu < +2\Delta$. See Fig. 1

One way of diagnosing such a band insulator is by computing $\rho(\mu)$. Within an energy band, the density ρ increases as the chemical potential μ is raised. However, for μ in the gap, ρ is constant. This plateau in $\rho(\mu)$ reflects a vanishing of the electronic compressibility $\kappa = \partial\rho/\partial\mu = 0$. We will see that this criterion for insulating behavior applies also to interaction-driven situations, but not to the disorder-induced Anderson insulator.

In the discussion above we generated multiple bands and a band gap through an additional staggered potential. One could also generalize the original TBH, Eq. (2), so that several fermionic species are present. One can, for example, allow two orbitals (and associated operators $\hat{c}$ and $\hat{d}$) on every site of a square lattice,

$$\begin{aligned}\hat{H} = &- t \sum_{\langle \mathbf{j},\mathbf{l}\rangle\,\sigma} (\hat{c}^\dagger_{\mathbf{j}\sigma}\hat{c}_{\mathbf{l}\sigma} + \hat{c}^\dagger_{\mathbf{l}\sigma}\hat{c}_{\mathbf{j}\sigma}) - t \sum_{\langle \mathbf{j},\mathbf{l}\rangle\,\sigma} (\hat{d}^\dagger_{\mathbf{j}\sigma}\hat{d}_{\mathbf{l}\sigma} + \hat{d}^\dagger_{\mathbf{l}\sigma}\hat{d}_{\mathbf{j}\sigma}) \\ &- t' \sum_{\mathbf{j}\sigma} (\hat{d}^\dagger_{\mathbf{j}\sigma}\hat{c}_{\mathbf{j}\sigma} + \hat{c}^\dagger_{\mathbf{j}\sigma}\hat{d}_{\mathbf{j}\sigma}) - \mu \sum_{\mathbf{j}} (\hat{n}^d_{\mathbf{j}\uparrow} + \hat{n}^d_{\mathbf{j}\downarrow} + \hat{n}^c_{\mathbf{j}\uparrow} + \hat{n}^c_{\mathbf{j}\downarrow}) \ . \end{aligned} \tag{10}$$

Each of the individual fermionic species associated with operators $\hat{c}$ and $\hat{d}$ hops on near-neighbor sites. However, the two types of fermions are also allowed to interconvert on the same site of the lattice, with hopping parameter t'.
Once again going to momentum space, the mixing of the two fermionic species leads to

$$\begin{aligned}\hat{H} &= \sum_{\mathbf{k}\,\sigma} \varepsilon_{\mathbf{k}} \hat{c}^{\dagger}_{\mathbf{k}\,\sigma} \hat{c}_{\mathbf{k}\,\sigma} + \sum_{\mathbf{k}\,\sigma} \varepsilon_{\mathbf{k}} \hat{d}^{\dagger}_{\mathbf{k}\,\sigma} \hat{d}_{\mathbf{k}\,\sigma} + t' \sum_{\mathbf{k}\,\sigma} \left(\hat{d}^{\dagger}_{\mathbf{k}\,\sigma} \hat{c}_{\mathbf{k}\,\sigma} + \hat{c}^{\dagger}_{\mathbf{k}\,\sigma} \hat{d}_{\mathbf{k}\,\sigma} \right) \\ &= \sum_{k} \begin{pmatrix} c^{\dagger}_{\mathbf{k}} & d^{\dagger}_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} \varepsilon_{\mathbf{k}} & t' \\ t' & \varepsilon_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} c_{\mathbf{k}} \\ d_{\mathbf{k}} \end{pmatrix} . \end{aligned} \tag{11}$$

The final 2x2 rotation yields the energy levels,

$$E^{\pm}_{\mathbf{k}} = -2t \left(\cos k_x + \cos k_y \right) \pm t' \ . \tag{12}$$

This band structure is somewhat more rich than that which arises from a staggered potential, where a gap opens for *any* nonzero Δ. Here, instead, the bands overlap for $t' < 4\,t$ and the system is always metallic (except for $\rho = 0$ or $\rho = 2$). However, it can be made insulating at $\rho = 1$ if $t' > 4\,t$. The TBH of Eq. (10) is sometimes used to describe 'bilayer' geometries, where c and d label two different spatial layers, as opposed to distinct orbitals.
The considerations of this section have described the simplest type of metal-insulator transition: Fermions which are noninteracting, on a translationally invariant lattice such that the placement of the chemical potential either within one of the energy bands (a metal) or in a gap between them (insulator).

3 Antiferromagnetic and charge density wave insulators

Insulating behavior which is closely connected, from the viewpoint of mathematical structure, to that of the previous section arises when interactions are included within mean-field theory (MFT). Consider the most simple type of TBH interaction, a repulsion between spin up and spin down fermions on the same spatial site. Together with the kinetic energy of Eq. (2) we obtain the Hubbard Hamiltonian,

$$\hat{H} = -t \sum_{\langle \mathbf{j},\mathbf{l} \rangle\,\sigma} \left(\hat{c}^{\dagger}_{\mathbf{j}\,\sigma} \hat{c}_{\mathbf{l}\,\sigma} + \hat{c}^{\dagger}_{\mathbf{l}\sigma} \hat{c}_{\mathbf{j}\,\sigma} \right) - \mu \sum_{\mathbf{j}} \left(\hat{n}_{\mathbf{j}\uparrow} + \hat{n}_{\mathbf{j}\downarrow} \right) + U \sum_{\mathbf{j}} \hat{n}_{\mathbf{j}\uparrow} \hat{n}_{\mathbf{j}\downarrow} \ . \tag{13}$$

The MFT approximation consists of recasting the interaction term in Eq. (13) as,

$$U \sum_{\mathbf{j}} \left(\hat{n}_{\mathbf{j}\uparrow} \langle \hat{n}_{\mathbf{j}\downarrow} \rangle + \hat{n}_{\mathbf{j}\uparrow} \langle \hat{n}_{\mathbf{j}\downarrow} \rangle - \langle \hat{n}_{\mathbf{j}\uparrow} \rangle \langle \hat{n}_{\mathbf{j}\downarrow} \rangle \right) \ . \tag{14}$$

It is clear that if the fermionic occupations possess an antiferromagnetic (AF) pattern, $\langle \hat{n}_{\mathbf{j}\uparrow} \rangle = \rho + (-1)^{\mathbf{j}}\, m$ and $\langle \hat{n}_{\mathbf{j}\downarrow} \rangle = \rho - (-1)^{\mathbf{j}}\, m$, on a bipartite lattice, then a staggered potential similar to that described by Eq. (7) is present. As a consequence of this 'spin density wave' (SDW), a band gap opens and 'Slater' insulating behavior arises, in direct analogy of the argument leading up to Eq. (8).

Although this MFT treatment of the Slater insulator is indeed close to that of a staggered potential, it is worth emphasizing that AF can also arise away from $\rho = 1$. Of course, it is necessary to determine whether the *ansatz* for $\langle n_{\mathbf{j}\,\sigma}\rangle$ in which the occupations vary spatially actually lowers the free energy for nonzero m. The answer will depend, in general, on U and ρ and can be used to generate the MFT phase diagram. The result for the $d = 2$ square lattice is given in [2]. Note that it is also possible that a 'ferromagnetic' *ansatz* $\langle \hat{n}_{\mathbf{j}\uparrow}\rangle = \rho + m$ and $\langle \hat{n}_{\mathbf{j}\downarrow}\rangle = \rho - m$, lowers the energy. This is quite a bit less likely to lead to an insulating gap since, as discussed above within the context of a bilayer model, a large order parameter m is required to introduce a gap between energy bands which are rigidly shifted, whereas a gap immediately opens for any staggered potential amplitude Δ.

A similar type of insulator arises when fermions interact with local phonon (oscillator) modes $\hat{p}_{\mathbf{j}}, \hat{q}_{\mathbf{j}}$, rather than with each other, *e.g.* in the Holstein model

$$\hat{H}_{\rm Holstein} - t \sum_{\langle \mathbf{j},\mathbf{l}\rangle\,\sigma} \left(\hat{c}^{\dagger}_{\mathbf{j}\sigma}\hat{c}_{\mathbf{l}\sigma} + \hat{c}^{\dagger}_{\mathbf{l}\sigma}\hat{c}_{\mathbf{j}\,\sigma} \right) + \frac{1}{2}\sum_{\mathbf{j}} \left(\hat{p}^2_{\mathbf{j}} + \omega^2 \hat{q}^2_{\mathbf{j}} \right) + \lambda \sum_{\mathbf{j}} \left(\hat{n}_{\mathbf{j}\uparrow} + \hat{n}_{\mathbf{j}\uparrow} \right) \hat{q}_{\mathbf{j}} \ . \qquad (15)$$

One can get a preliminary understanding of its physics by ignoring the phonon kinetic energy and considering only static ionic displacements. On a bipartite lattice, an oscillating set of displacements $\langle\, \hat{q}_{\mathbf{j}}\,\rangle = q_0(-1)^{\mathbf{j}}$ opens a gap in the fermion dispersion relation precisely as with a staggered potential associated with an AF spin pattern. Unlike the latter case, however, the resulting densities of up and down spin are in phase, leading to a charge density wave (CDW) as opposed to a SDW. At half-filling, the lowering of the electronic energy from $\varepsilon_{\mathbf{k}}$ to $E_{\mathbf{k}} = -\sqrt{\varepsilon^2_{\mathbf{k}} + \Delta^2}$ favors non-zero values of q_0. Against this competes the increase in the potential energy $\omega_0^2\, q_0^2/2$. Which effect dominates depends on the phonon frequency ω_0, the electron-phonon coupling λ, the dimensionality of the lattice, and, of course, a proper treatment of quantum fluctuations of the phonons.

4 Anderson and Mott insulators

In this section we combine a discussion of two distinct types of insulator, those arising from disorder and those arising from strong repulsive interactions.

Anderson insulators develop from randomness in a tight binding Hamiltonian. We begin our discussion by considering a one dimensional TBH with a single site at the chain center $N/2$ with a lower energy than all the others

$$\hat{H} = -t \sum_{j,\sigma} \left(\hat{c}^{\dagger}_{j\,\sigma}\hat{c}_{j+1\,\sigma} + \hat{c}^{\dagger}_{j+1\,\sigma}\hat{c}_{j\,\sigma} \right) - \mu_0\, c^{\dagger}_{N/2\,,\sigma} c_{N/2\,,\sigma} \ . \qquad (16)$$

Figure 2(top left) shows the eigenenergies, obtained numerically by diagonalizing $\hat{H}$. Since translation invariance is broken by the impurity, we no longer label the eigenvalues with a momentum index k. Nevertheless, all but one of the eigenvalues form a band, which looks very much like the $\varepsilon(k) = -2t\cos(k)$ in the absence of the impurity ($\mu_0 = 0$). However, there is one extremal eigenvalue split off from all the others, which we have placed at $n = 512$. This

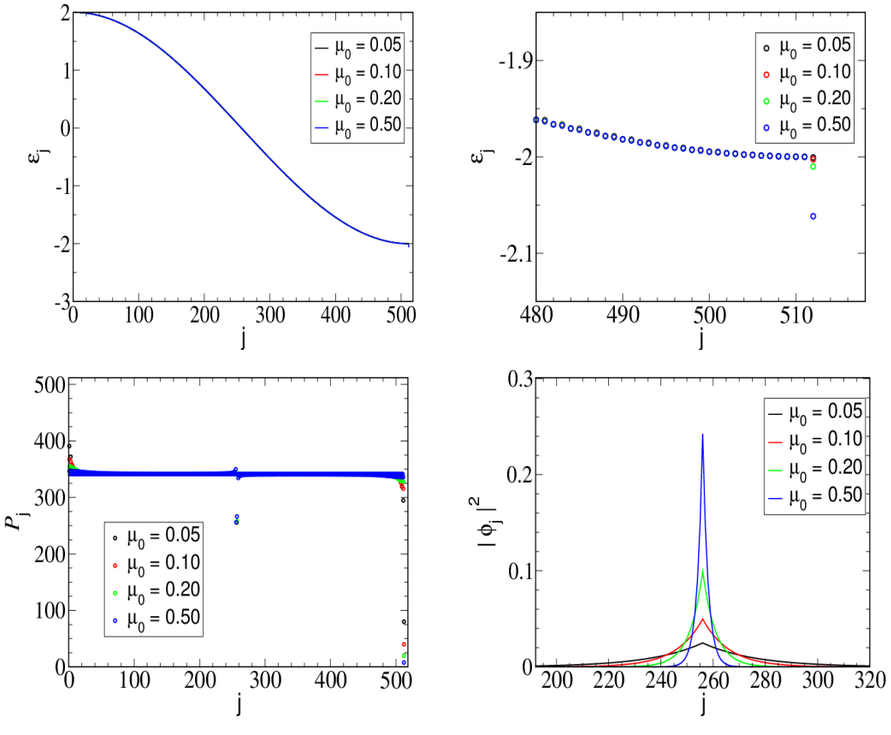


Fig. 2: *Top Left: Eigenspectrum of the TBH Eq.* (16) *for a chain of length* $N = 512$ *and impurity depths* $\mu_0 = 0.05, 0.10, 0.20, 0.50$*. Shown over their full range, the energy levels are indistinguishable from each other and from those of Fig. 1. Top Right: A blow-up of the eigenspectrum allows the resolution of the impurity level split off below the energy band. Bottom Left: Participation ratios of all eigenvectors are of order the number of sites* N*, except for the single, localized mode. Bottom Right: The square of the components of the associated eigenfunctions, in the vicinity of the defect at* $N/2 = 256$*. As the impurity depth* μ_0 *decreases, the eigenfunctions are less localized. Since momentum is no longer a good quantum number in the presence of the breaking of translation invariance by the defect, the horizontal axes in the top row are labeled by the eigenvalue index* j *rather than* k*.*

separation is clear in the blow up of Fig. 2(top right). Figure 2(bottom right) plots the square of the amplitude of the components $\left|\phi_j\right|^2$ of the localized eigenfunctions. They are seen to be sharply peaked at $N/2$.

A useful way to characterize the spatial extent of an eigenfunction with components ϕ_j (which we assume are normalized to $\sum_j \left|\phi_j\right|^2 = 1$) is via the participation ratio $\mathcal{P}$

$$\mathcal{P}^{-1} = \sum_j \left|\phi_j\right|^4. \tag{17}$$

If the eigenfunction is fully localized on a single site j_0, that is, if $\phi_j = \delta(j, j_0)$, it is easy to see $\mathcal{P}^{-1} = 1$ and hence $\mathcal{P} = 1$. On the other hand, if the eigenfunction is completely delocalized $\phi_j = 1/\sqrt{N}$ we have $\mathcal{P}^{-1} = 1/N$ and hence $\mathcal{P} = N$. By considering other cases one can

be convinced that, roughly speaking, $\mathcal{P}$ measures the number of sites in the lattice where ϕ_j is "large." Figure 2(bottom left) plots these participation ratios. They are all of the order of the lattice size N (meaning the states are delocalized) except for a single mode ($j = 512$) which is localized.

This problem is formally very similar to that of localization of vibrations of a harmonic chain with a single mass m' or spring k' which differs from all the others, in the sense that the solution of both reduces to diagonalization of the same type of matrices. In the case of vibrations, it is interesting to note that localization occurs only for a defect mass which is *lighter* than all the others. This can be seen to be physically reasonable in the extreme limits: If $m' \ll m$, one pictures the very light mass as vibrating back and forth between the heavy 'walls' provided by its neighbors. A heavy defect, $m' \gg m$, shoves aside its neighbors and its vibrations spread throughout the chain.

The problem of a small number of impurities in a noninteracting TBH can be treated analytically [3]. The procedure is sufficiently interesting and important to provide the initial steps here. In order to connect this discussion with the previous material, it is useful to recall an alternate approach to the solution of noninteracting TBHs.

We solved Eq. (2) by a rather sophisticated method, namely by doing a canonical transformation on the fermionic creation and annihilation operators which diagonalized $\hat{H}$. A less sophisticated solution is to construct the matrix for $\hat{H}$ using position occupation states as a basis. This is done in the usual way, by allowing $\hat{H}$ to act on each basis vector. Because $\hat{H}$ conserves particle number (fermion creation and destruction operators always appear as partners), its matrix consists of independent blocks corresponding to the particle number. For a linear chain of N sites with periodic boundary conditions, then

$$\begin{aligned}
\hat{H}\,|\,1\,0\,0\,0\,0\,0\cdots 0\,0\,\rangle &= -t\,|\,0\,1\,0\,0\,0\,0\cdots 0\,0\,\rangle - t\,|\,0\,0\,0\,0\,0\,0\cdots 0\,1\,\rangle \\
\hat{H}\,|\,0\,1\,0\,0\,0\,0\cdots 0\,0\,\rangle &= -t\,|\,1\,0\,0\,0\,0\,0\cdots 0\,0\,\rangle - t\,|\,0\,0\,1\,0\,0\,0\cdots 0\,0\,\rangle \\
\hat{H}\,|\,0\,1\,0\,0\,0\,0\cdots 0\,0\,\rangle &= -t\,|\,1\,0\,0\,0\,0\,0\cdots 0\,0\,\rangle - t\,|0\,0\,1\,0\,0\,0\cdots 0\,0\,\rangle \\
&\cdots \text{etc.}
\end{aligned} \tag{18}$$

The calculation of the single particle eigenstates ϕ and eigenenergies E, in the absence of an impurity, therefore corresponds to the linear algebra problem,

$$\sum_n L_{mn}\,\phi_n = 0 \qquad L_{mn} = E\,\delta_{mn} - t\,\delta_{m,n-1} - t\,\delta_{m,n+1}\ , \tag{19}$$

where L is the matrix of numbers which forms the single particle block of $\hat{H}$ in the occupation number basis.

The nontrivial solution of Eq. (19) requires the vanishing of the determinant $|\,L\,| = 0$. It is an easily proven that the k component of the nth eigenvector is $\phi_n = e^{ikn}$, and $E_k = -2\,t\,\cos k$, solve Eq. (19). The periodic boundary conditions discretize the allowed k values to $k = 2\pi n/N$ with $n = \{1, 2, 3, \ldots, N\}$. Notice that this solution is precisely the same as that arising from the transformation to momentum space operators, Eq. (5)!

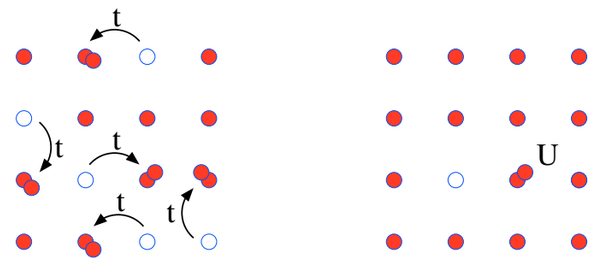


Fig. 3: *The qualitative physics of the Mott insulator is seen by considering a half-filled system (one particle per site). Left: When the on-site repulsion U between particles is weak, they are free to hop around the lattice. Empty, singly, and doubly occupied sites are all present, with only the average density equalling one particle per site. Right: On the other hand, when U is very large compared to t, it is energetically preferable for the particles to sit with exactly one fermion on each individual site.*

This second approach to the problem lends itself nicely to an attack on the behavior in the presence of randomness. One can write the problem as,

$$\sum_n L_{mn}\,\phi_n = \sum_k \delta L_{mk}\,\phi_k \qquad \rightarrow \qquad \left(\, I - G\,\delta L\right)\phi = 0\ , \tag{20}$$

where δL is the matrix which contains the local chemical potentials and $G = L^{-1}$. In the case of Eq. (16), δL has a single nonzero entry along its diagonal.
A solution to Eq. (20) is,

$$\phi_n = \sum_{lk} G_{nl}\,\delta L_{lk}\,\phi_k\ . \tag{21}$$

However, this a only 'formal' solution because the unknown variables ϕ_n appear on both sides of Eq. (21). However, note that the non-trivial solution of Eq. (21) requires $|\, I - G\,\delta L\,| = 0$. The important observation is that the sparsity or δL enormously simplifies the linear algebra problem. Instead of rank N, the matrix $I - G\,\delta\,L$ whose determinant must be computed has much lower rank. Furthermore, the solution of the eigenproblem of L is known, we have an explicit expression for the Green function, $G_{nl} = \sum_k e^{ik(n-l)}/E_k$. Amazingly, then, the problem of the modes in the presence of $n \ll N$ defects boils down to the diagonalization of an $n \times n$ matrix, whose elements involve the known defect potential δL and Green function G. Ref. [3] provides some explicit examples, and a beautiful graphical solution of several interesting cases.
Having discussed the situation when there is a single, or small number of, defects, it is natural to ask what happens when there are many impurities present, for example when there is a randomly chosen chemical potential on *every* site of the lattice. This is the problem of 'Anderson Localization' [4]. In one dimension, all the eigenstates become localized, for any amplitude of disorder. This is also true in two dimensions, although just barely [5]. In three dimensions, the eigenfunctions at the extremes of the spectrum (that is, those associated with the largest and

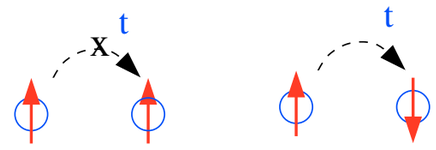


Fig. 4: *The Pauli Principle prevents fermions of like spin on adjacent sites from hopping (left), a process which is allowed if the fermions have opposite spin (right). In the case of antiparallel spin, the intermediate state created by the hop has a doubly occupied site, and hence a potential energy U. The resulting second order lowering of the energy relative to the parallel spin arrangement is proportional to $-t^2/U$.*

smallest eigenvalues) are localized, while the eigenfunctions near the center of the spectrum are extended. The energy which separates these two behaviors is referred to as the mobility edge.

In 3D one has the appearance of Anderson insulating behavior, and the possibility of associated metal-insulator transitions: If the chemical potential lies below the mobility edge, only localized eigenfunctions are occupied, and the system is an insulator. When μ crosses the mobility edge, extended states become occupied, and the system becomes a metal. It is important to emphasize that, in stark contrast to the band, SDW, and CDW insulators previously discussed, there is *no gap* in the spectrum. The compressibility κ is nonzero in the insulator, and a plot of density ρ as a function of chemical potential μ would show no marked signal at the transition from Anderson insulator to metal.

The final qualitative discussion concerns "Mott insulators," whose behavior arises from interactions, as opposed to gaps in the band structure or localization by disorder. Consider a single band Hubbard Hamiltonian, for example on a square lattice, at "half-filling" (one electron per site). The simple physical picture of a Mott insulator is that if the on-site repulsion U is very large, the energy cost for the double occupation which must occur in order for the electrons to move, overwhelms the kinetic energy and freezes the electrons in place. See Fig. 3.

Although in Fig. 3 the spin orientations of the fermions are not indicated, it is natural to ask if they have any preferred arrangement. There are several arguments which suggest AF order. The first treats the hopping term in the Hubbard Hamiltonian Eq. (13) perturbatively. Consider two adjacent sites, both singly occupied with fermions of parallel spin. The interaction energy is zero, and, because of the Pauli Principle, the matrix element of the kinetic energy in this state vanishes, so there is no shift in the energy. If the fermions have antiparallel spin, however, the kinetic energy operator connects to an intermediate state with one empty and one doubly occupied site, with energy U. Thus the energy of a pair of sites with antiparallel spin fermions is lowered by $\Delta E \sim -t^2/U$. See Fig. 4. There are other arguments suggesting AF dominates at half-filling, for example a calculation of the magnetic susceptibility of the Hubbard Hamiltonian within the random phase approximation. A very nice early discussion of these ideas, emphasizing several unique features of the square lattice dispersion, is contained in Ref. [2].

Figure 5 shows some quantum simulation results for the square lattice Hubbard Hamiltonian at $U = 4$. $\rho(\mu)$ develops a 'Slater-Mott' plateau at half-filling. (See below.) The figure uses a convention in which the interaction term is written in particle-hole symmetric form so that

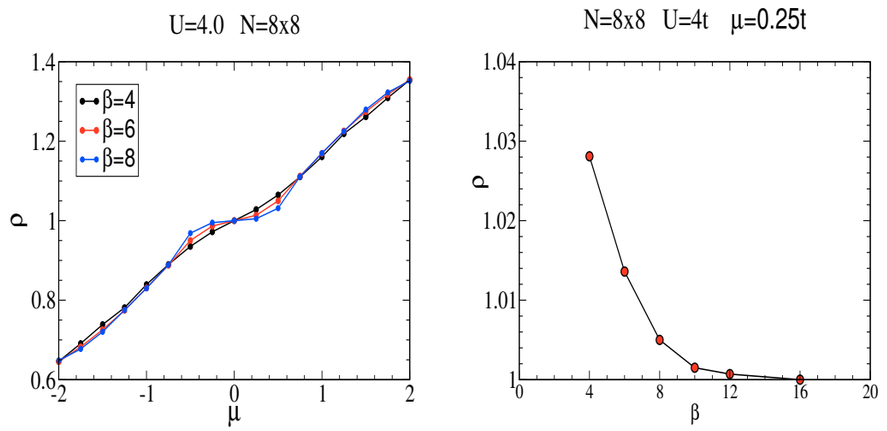


Fig. 5: *Left: The density ρ as a function of the chemical potential μ for the Hubbard Hamiltonian on a square lattice at $U = 4t$ and three different inverse temperatures $\beta = t/T = 4, 6, 8$. As β increases, Right: The density of fermions at a nonzero chemical potential, as a function of β shows that $\rho \to 1$.*

$\mu = 0$ corresponds to $\rho = 1$. The vanishing of the compressibility, $\kappa = 0$, at $\rho = 1$ is a truly remarkable change in behavior since, on the square lattice, the noninteracting system has a divergent density of states at half-filling: $\kappa = \infty$ at $U = 0$! The algorithm used in the figure is 'determinant QMC'. This approach treats the interactions between electrons exactly, on lattices of finite spatial extent (a few hundred up to about a thousand spatial sites), and thus provides a much more rigorous treatment than that provided by MFT. The reader is referred to Refs. [2, 6–8] for a discussion of DQMC and its application to magnetism in the 2D Hubbard Hamiltonian.

The review of these ideas emphasizes an important point: in many situations (especially on bipartite lattices) a 'Slater insulator,' which occurs at weak to intermediate U due to the opening of an AF gap, merges smoothly, as U increases, into the Mott insulator where the lack of transport predominantly arises from the high cost of double occupancy. There is no sharp boundary between these two types of insulator, but rather a gradual crossover. A very deep question indeed is whether for fermionic systems symmetry breaking such as AF order always accompanies the Mott insulator, or whether a featureless, translationally invariant Mott phase can occur, as for collections of bosonic particles [9].

5 Formal definitions

The proverbially alert reader will have noticed that the preceding discussion avoided what would seemingly be the most natural quantity to distinguish metals and insulators, namely the conductivity σ. This is because transport properties are a bit more subtle to deal with. We will now consider σ and develop an understanding, which unifies the preceding, more qualitative, discussion. An added bonus will be the fact that the superfluid density, the defining characteristic

of a superconductor, naturally arises. The discussion in this section very closely follows that of Ref. [1]. The derivation is a bit dense. The key 'practical results' are Eqs. (30) and (31) which allow for the determination of Drude weight D and superfluid density D_s from the current-current correlation function Λ_{xx}.

Consider the response of the current to presence of a vector potential $A_x(\mathbf{l})$. As shown in Ref. [10], this modifies the hopping term in the kinetic energy (suppressing the spin indices),

$$c^\dagger_{\mathbf{l}+x} c_{\mathbf{l}} + c^\dagger_{\mathbf{l}} c_{\mathbf{l}+x} \rightarrow e^{ieA_x(\mathbf{l})} c^\dagger_{\mathbf{l}+x} c_{\mathbf{l}} + e^{-ieA_x(\mathbf{l})} c^\dagger_{\mathbf{l}} c_{\mathbf{l}+x} \ . \tag{22}$$

This can be expanded in powers of A so that the kinetic energy K acquires an additional term which can be expressed in terms of the paramagnetic current density in the x direction $ej^p_x(\mathbf{l})$ and the kinetic energy density on bonds in the x direction, $k_x(\mathbf{l})$,

$$\begin{aligned} K_A &= K - \sum_{\mathbf{l}} \Big(e j^p_x(\mathbf{l}) A_x(\mathbf{l}) + \frac{e^2 k_x(\mathbf{l})}{2} A_x(\mathbf{l})^2 \Big) \\ j^p_x(\mathbf{l}) &= it \sum_\sigma \left(c^\dagger_{\mathbf{l}+x\,\sigma} c_{\mathbf{l}\sigma} - c^\dagger_{\mathbf{l}\sigma} c_{\mathbf{l}+x\,\sigma} \right) \\ k_x(\mathbf{l}) &= -t \sum_\sigma \left(c^\dagger_{\mathbf{l}+x\,\sigma} c_{\mathbf{l}\sigma} + c^\dagger_{\mathbf{l}\sigma} c_{\mathbf{l}+x\,\sigma} \right) \ . \end{aligned} \tag{23}$$

Differentiating Eq. (23) with respect to $A_x(\mathbf{l})$ yields the total current density, which includes both paramagnetic and diamagnetic contributions,

$$j_x(\mathbf{l}) = -\frac{\delta K_A}{\delta A_x(\mathbf{l})} = e\, j^p_x(\mathbf{l}) + e^2\, k_x(\mathbf{l})\, A_x(\mathbf{l}) \tag{24}$$

If one assumes a plane wave form for the vector potential,

$$A_x(\mathbf{l}, t) = \mathrm{Re}\big(A_x(\mathbf{q}, \omega)\, e^{i\mathbf{q}\cdot\mathbf{l} - i\omega t} \big) \ , \tag{25}$$

then the resulting current is,

$$\begin{aligned} \langle j_x(\mathbf{l}, t) \rangle &= \mathrm{Re}\big(\langle\, j_x(\mathbf{q}, \omega) \rangle\, e^{i\mathbf{q}\cdot\mathbf{l} - i\omega t} \big) \\ \langle\, j_x(\mathbf{q}, \omega) \rangle &= -e^2 \Big(\langle k_x \rangle - \Lambda_{xx}(\mathbf{q}, \omega) \Big) A_x(\mathbf{q}, \omega) \ . \end{aligned} \tag{26}$$

The real-frequency current-current correlation functions $\Lambda(\mathbf{q}, \omega)$ are related to those at Matsubara frequencies $i\omega_m = 2\pi m T$,

$$\Lambda_{xx}(\mathbf{q}, i\omega_m) = \frac{1}{N} \int_0^\beta d\tau e^{i\omega_m \tau} \langle j^p_x(\mathbf{q}, \tau) j^p_x(-\mathbf{q}, 0) \rangle \ , \tag{27}$$

by analytic continuation.

Equations (26), (27) and the calculations leading to them are simply somewhat more complex versions of the relations such as the one which expresses the magnetization induced by an applied Zeeman field, to the magnetization-magnetization correlation functions and thereby the magnetic susceptibility χ, or any of the other multitude of 'fluctuation-dissipation' relations which arise from linear response theory.

It remains to connect this rather abstract quantity to more physical objects like the superfluid density: For bosonic particles, the superfluid density can be measured in, for example, a torsional oscillator experiment. As T is decreased below the superfluid transition temperature, the moment of inertia of a liquid in a container abruptly decreases, because the liquid inside no longer couples to the walls of the container. As we discuss below, for fermionic particles the superfluid density determines the distance to which a magnetic field penetrates a superconductor.

One of the early fundamental advances in understanding superconductivity was London's observation that the Meissner effect follows if one assumes the current density is proportional to the vector potential,

$$j_x(q_y) = -\frac{1}{4\pi}\frac{1}{\lambda^2}A_x(q_y) \ . \tag{28}$$

That is, magnetic fields will be expelled from a superconductor at distances beyond the penetration depth λ,

$$\frac{1}{\lambda^2} = \frac{4\pi n_s e^2}{mc^2} \ , \tag{29}$$

which depends on the superfluid density n_s. A comparison of Eqs. (28), (29) with Eq. (26) provides a link between the superfluid weight $D_s = n_s/m$ and the current-current correlation function:

$$\frac{D_s}{\pi e^2} = -\langle -k_x \rangle - \Lambda_{xx}(q_x = 0, q_y \to 0, i\omega_m = 0) \ . \tag{30}$$

The usual relations between vector potential and electric field, $E_x = -\partial A_x/\partial t$, and between the conductivity and electric field, result in an analogous formula for the Drude weight, the delta function contribution $D\delta(\omega)$ to the conductivity,

$$\frac{D}{\pi e^2} = -\langle -k_x \rangle - \Lambda_{xx}(q_x = 0, q_y = 0, i\omega_m \to 0) \ . \tag{31}$$

Details of this connection are in Ref. [1].

The third limit, in which the longitudinal momentum is taken to zero, relates Λ to the kinetic energy,

$$\langle -k_x \rangle = \Lambda_{xx}(q_x \to 0, q_y = 0, i\omega_m = 0) \ . \tag{32}$$

Summarizing, the key results are the following: Depending on the limits in which the momenta and frequency are taken to zero, one can obtain superfluid density D_s and Drude weight D from the current-current correlation function.

The superfluid density D_s and the Drude weight D form a basis for distinguishing an insulator ($D = D_s = 0$), from a metal, ($D \neq 0, D_s = 0$), from a superconductor ($D_s \neq 0$). It is rather remarkable that these alternate limits of approaching zero momentum and frequency yield distinct results and profoundly different physical quantities, especially to physicists who are accustomed to not being overly worried about the subtleties of the order of operations.

We will introduce the simplified notation $\Lambda^{\rm L} \equiv \lim_{q_x \to 0} \ \Lambda_{xx}(q_x, q_y = 0; i\omega_n = 0)$ and $\Lambda^T \equiv \lim_{q_y \to 0} \ \Lambda_{xx}(q_x = 0, q_y; i\omega_n = 0)$ so that Eqs. (30) and (32) can be simply expressed as $D_s = \pi[-K_x - \Lambda^T]$ and $-K_x = \Lambda^L$ respectively.

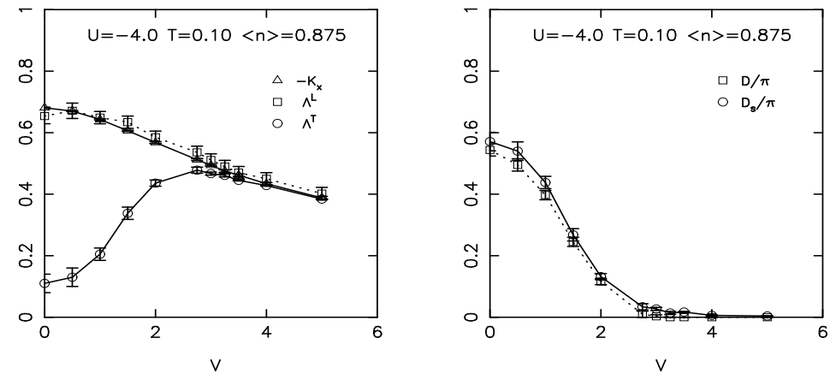


Fig. 6: *Left: Kinetic energy K_x, longitudinal Λ^L, and transverse Λ^T limits of current-current correlation function for the attractive Hubbard model with $U = -4\,t$ at temperature $T = 0.1$ and filling $\rho = 0.875$. The horizontal axis is the strength of random site energies $-V < \mu_i < +V$. The data indicate that $\Lambda^L = -K_x$ over all parameter ranges, as required by gauge invariance. Right: The superfluid density $D_s = \pi\big(-K_x - \Lambda^T\big)$ and Drude weight D. To within the accuracy of the numerics, $D = D_s$. See Eqs.* (30), (31)*, and also Fig. 7*

6 Applications of formal theory

Ref. [1] considered the simplest TBHs to check their formalism, namely the clean, single band attractive and repulsive Hubbard Hamiltonians on a square lattice. Here we present results [11] on a TBH which also includes disorder in the site energies (an additional term $\sum_i v_i(n_{i\uparrow} + n_{i\downarrow})$ in the Hubbard Hamiltonian), to illustrate how powerful and general Eqs. (30), (31) truly are. We use the same DQMC approach which generated the data shown in Fig. 5 (and which was used in [1]). We note, however, that the implementation of these criteria within DQMC requires the evaluation of imaginary time-dependent observables, as opposed to the algorithmically more simple equal time quantities like the energy, density, and magnetic, charge, and pairing structure factors. Such calculations slow down DQMC simulations quite significantly, especially at low temperatures and on large spatial lattices.

It is important to note that, while the presence of randomness breaks translation invariance for a *single* disorder realization, translation invariance is recovered after disorder averaging. Typically one finds calculations for 10-100 distinct instances of the local site potential $\{v_i\}$ are required in DQMC simulations such as those described here.

Results from [11] for Λ^T, Λ^L and $-K_x$ are plotted in Fig. 6(left) as a function of the strength V of randomness in the site energies $-V < v_i < +V$. The attractive interaction strength $U = -4$, temperature $T = 0.10$, and density $\rho = 0.875$. D and D_s are plotted in Fig. 6(right). They decrease monotonically with disorder. There is a critical value V_c beyond which $D = D_s = 0$ and the system becomes insulating. These results are consistent with a direct superconductor to insulator transition in 2D, without an extended intervening metallic phase.

Figure 7 provides some numerical details on the extrapolation in Matsubara frequencies which, following Eq. (31), is needed to capture D. Similar plots showing the momentum extrapolations

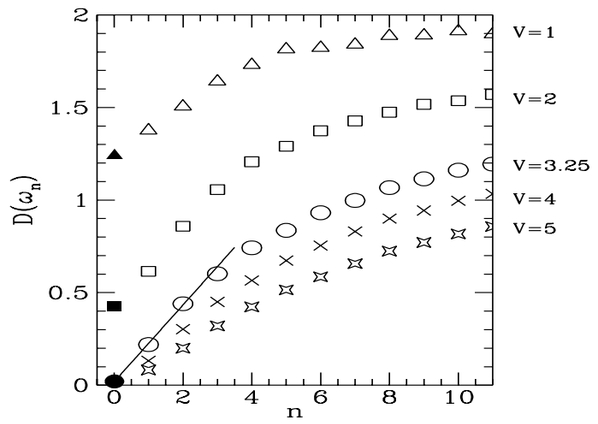


Fig. 7: *Illustration of the details of the extrapolation procedure to obtain the Drude weight D via Eq.* (31)*. The horizontal axis $n = \omega_n/(2\pi T)$. Parameters are as in Fig. 6. In the metallic phase, which occurs precisely at the critical point $V = V_c \sim 3.25$, the slope of $D(\omega_n)$ can be used to obtain σ_{dc}. See text.*

to obtain D_s and verify $\Lambda^L = -\langle K_x \rangle$ are not shown, but can be found in Ref. [1]. Note that in general the simulations are performed in a regime where there are up to several hundred Matsubara frequencies, but only 10-30 momenta in each direction, an order of magnitude less. Thus the momentum extrapolations needed for Λ^L and Λ^T are typically more challenging than those for D.

As argued in Ref. [1], the extrapolation in Fig. 7 can also be used to obtain the dc conductivity via

$$D(\omega_n) = \pi \sigma_{dc} |\omega_n| \ . \tag{33}$$

We will use this as a consistency check against alternate ways of quantifying the metal-insulator transition and obtaining σ_{dc}.

7 Conductivity and spectral function

This final section before the conclusions will focus on two further QMC approaches to the metal-insulator transition. The first technique, like those of Sec. 6, begins with the current-current correlation function, but has the advantage of avoiding analytic continuation and extrapolation to zero momentum or frequency. It is, however, approximate. The second method moves away from Λ_{xx} and instead considers the spectral function.

Consider the fluctuation-dissipation theorem

$$\Lambda_{xx}(\mathbf{q}, \tau) = \int_{-\infty}^{+\infty} \frac{d\omega}{\pi} \frac{\exp(-\omega\tau)}{1 - \exp(-\beta\omega)} \, \mathrm{Im}\, \Lambda_{xx}(\mathbf{q}, \omega) \ . \tag{34}$$

In principle one can invert this Laplace transform to get $\mathrm{Im}\, \Lambda_{xx}$, but this process is known to be very ill-conditioned [12]. We instead proceed as follows: If the temperature $T \ll \Omega$, the scale

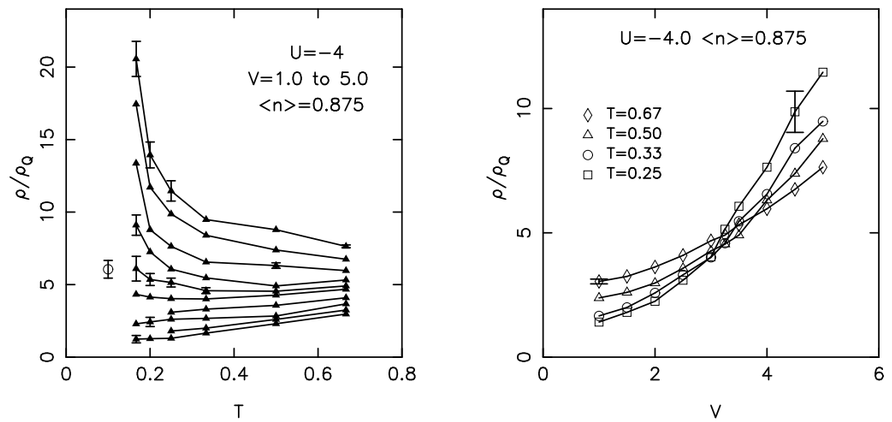


Fig. 8: *Left: The resistivity $\rho_{dc} = 1/\sigma_{dc}$ obtained from Eq.* (35) *as a function of temperature. The on-site attraction $U = -4\,t$ and the density $\rho = 0.875$. The curves are (top to bottom) for disorder strengths $V = 5.0, 4.5, 4.0, 3.5, 2.0, 2.5, 2.0, 1.5, 1.0$. For large V, ρ_{dc} increases as T is lowered, indicating insulating behavior. For small V, ρ_{dc} decreases as T is lowered, indicating metallic behavior. The open symbol at $T = 0.10$ is the value of ρ_{dc} inferred from the $V = V_c$ data in Fig. 7. See text. Right: The data are replotted to show ρ_{dc} as a function of V for curves of constant T. The crossing indicates the approximate position V_c of the metal-insulator transition. For these parameters, the superconducting transition temperature $T_c \lesssim 0.05t$, so no abrupt drop in ρ_{dc} occurs. The quantum of resistance $\rho_Q = h/(4e^2) = \pi/2$ in our units ($\hbar = e^2 = 1$).*

at which $\mathrm{Im}\,\Lambda$ deviates from its low frequency behavior $\mathrm{Im}\,\Lambda \sim \omega\sigma_{dc}$, it is useful to evaluate Eq. (34) at the largest possible imaginary time, $\tau = \beta/2$. By doing this, the factor $e^{-\omega\tau}$ cuts off all contributions to the integral for frequencies above Ω, allowing us to replace $\mathrm{Im}\,\Lambda$ by $\omega\sigma_{dc}$, and enabling an analytic evaluation of the integral. The result

$$\sigma_{\mathrm{dc}} = \frac{\beta^2}{\pi}\Lambda_{xx}(\mathbf{q} = 0, \tau = \beta/2)\ , \tag{35}$$

provides a very useful approximate formula for σ_{dc}, subject to the restrictions noted above.

The reasoning leading to Eq. (35) is dubious for non-random systems: for example, for a Fermi liquid, the scale $\Omega \simeq 1/\tau_{\mathrm{e-e}} \sim N(0)T^2$, so that it is impossible to satisfy $T \ll \Omega$ at low T. However, in the presence of strong disorder Ω it is set by V. Since Ω is T-independent, it is possible to lower the temperature sufficiently far in the DQMC simulation to make Eq. (35) applicable.

There is a quite nice consistency between the different methodologies to characterize the phases of the model, and even the quantitative values of the conductivity. For example, Fig. 6 shows an onset of nonzero D and D_s for V in the range $3 \lesssim V \lesssim 4$ as the disorder strength is decreased. These results are based on Eqs. (30), (31). Meanwhile, the crossings of the data for ρ_{dc} in Fig. 8 indicate $V_c \sim 3.5$. Here Eq. (35) was utilized. Analysis of D based on Eq. (31) yields $V_c \sim 3.25$ and, furthermore, via Eq. (33), gives a numerical value for σ_{dc} which agrees quite closely with Eq. (35). This sort of careful cross-checking of numerics is of course essential in

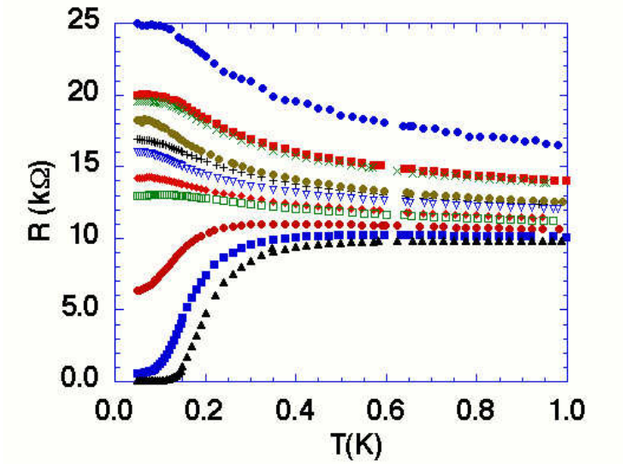


Fig. 9: *Superconducting-insulator transition in thin amorphous Bi films as a function of carrier density [17]. Note the qualitative similarity to Fig. 8. (As noted in the caption to Fig. 8, the less abrupt drop in the DQMC data is a consequence of the fact that $T > T_c$ over the temperature range shown.)*

any calculational approach, but is especially important in QMC studies of interacting fermions, where limitations of finite size and the sign problem are especially acute.

As a connection to real materials, we observe that the curves in Fig. 8 are remarkably similar to those found in the experimental literature on the two dimensional superconductor-insulator transition (SIT) in the presence of disorder [13]. In these studies, the SIT has been accessed in a wide variety of ways: by explicitly changing the degree of microscopic disorder (similar to the model studied here in which V is varied, by altering the film thickness, by applying a magnetic field, or by changing the carrier density. An example of the latter tuning method is given in Fig. 9. With DQMC, different ways of driving the SIT have also been explored with DQMC [11, 14–16].

One further method of distinguishing metals and insulators relies on the computation of the momentum-resolved spectral function $A(\mathbf{q}, \omega)$ and its sum, the density of states. The formalism is similar to that of Eq. (34), except involving the single-particle Green function $G(\mathbf{q}, \tau)$.

$$G(\mathbf{q}, \tau) = \int_{-\infty}^{+\infty} d\omega \frac{\exp(-\omega\tau)}{1 + \exp(-\beta\omega)} A(\mathbf{q}, \omega) \qquad N(\omega) = \sum_{\mathbf{q}} A(\mathbf{q}, \omega) \ . \tag{36}$$

Figure 10 shows what this diagnostic discloses concerning the square lattice Hubbard Hamiltonian at half-filling. One observes that $N(\omega = 0) \to 0$ as $T \to 0$ both for weak U, the 'Slater insulator' driven by SDW order, and at intermediate U where the crossover begins to Mott insulating behavior. The size of the insulating gap is roughly given by the temperature range over which $N(\omega = 0)$ is small. This is seen to increase with increasing U.

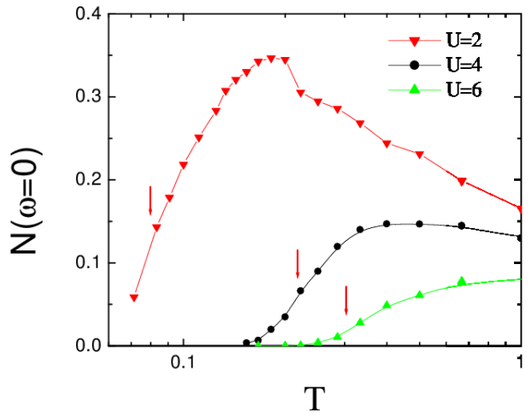


Fig. 10: *Density of states at the Fermi surface,* $N(\omega = 0)$ *of the half-filled Hubbard Hamiltonian as a function of temperature* T *for different values of the interaction strength* $U = 2, 4, 6$. *As* $T \to 0$, *the density of states vanishes in all cases. One concludes the Hubbard Hamiltonian on a square lattice with* $\rho = 1$ *is insulating over the entire range* $0 < U < \infty$.

8 Conclusions

We began this chapter with some simple qualitative pictures of metals and various types of insulators: (i) band insulators which arise when a TBH has several non-overlapping bands and the chemical potential lies between them; (ii) SDW and CDW insulators whose origin can be understood within a MFT treatment of interactions between the electrons or between electrons and phonons; (iii) Anderson insulators formed by disorder; and (iv) the most challenging situation, Mott insulators driven by strong interactions.

We then turned to a formal way of characterizing metals and insulators in terms of different limits of the current-current correlation function, and the implications for the conductivity and superfluid density. Our qualitative pictures of the distinction between metal and insulator in (i) and (ii) focussed on the spectrum of the Hamiltonian rather than the conductivity. The MFT treatment of the formal criteria showed the linkage between the two pictures.

The formal criteria have also been used in conjunction with QMC in the solution of the Hubbard Hamiltonian [1] to show that they indeed work when the interactions are treated more exactly than in MFT. We gave some illustrations of this approach when disorder and interactions are both present which serves as a specific model calculation for the superconducting to insulator phase transition [11], which is so well-explored experimentally [13]. Finally, we showed a few QMC results for the conductivity, spectral function, and density of states in determining insulating behavior.

It is worth noting two further approaches to the question of the metal-insulator transition which have also been widely used in QMC. The first is an analysis of the behavior of the electron self-energy at small Matsubara frequencies. For a illustration of this method, see [18]. The second is an analytic continuation of the imaginary time dependent spin, $\chi(\tau) = \langle M(\tau)M(0)\rangle$, and charge, $P(\tau) = \langle N(\tau)N(0)\rangle$, correlation functions. Here $M = \sum_i \left(n_{i_\uparrow} - n_{i\downarrow}\right)$ and $N =$

$\sum_i \left(n_{i\uparrow} + n_{i\downarrow}\right)$. The presence of 'spin and charge gaps' in the low frequency behavior of their Laplace transforms $\chi(\omega)$ and $P(\omega)$ can be used to infer the presence of insulating behavior associated with spin and charge order. See, for example, [19].

We finish by returning to the opening of this chapter, presenting the reader with a question: In our first encounter with the idea of conductors, one associates the resistance R in Ohm's law with some sort of scattering mechanism which provides for the loss of energy. Where is such dissipation in models like the clean Hubbard Hamiltonian?

References

[1] D.J. Scalapino, S.R. White, and S. Zhang, Phys. Rev. B **47**, 7995 (1993)

[2] J.E. Hirsch, Phys. Rev. B **31**, 4403 (1985)

[3] A.A. Maradudin: *Theoretical and Experimental Aspects of the Effects of Point Defects and Disorder on the Vibrations of Crystals*, Solid State Physics **18**, 273–420 (1966)

[4] P.W. Anderson, Phys. Rev. **109**, 1492 (1958)

[5] E. Abrahams, P.W. Anderson, D.C. Licciardello, and T.V. Ramakrishnan, Phys. Rev. Lett. **42**, 673 (1979)

[6] R. Blankenbecler, D.J. Scalapino, and R.L. Sugar, Phys. Rev. D **24**, 2278 (1981)

[7] J.E. Hirsch and S. Tang, Phys. Rev. Lett. **62**, 591 (1989)

[8] C.N. Varney, C.R. Lee, Z.J. Bai, S. Chiesa, M. Jarrell, and R.T. Scalettar, Phys. Rev. B **80**, 075116 (2009)

[9] M.P.A. Fisher, P.B. Weichman, G. Grinstein, and D.S. Fisher, Phys. Rev. B **40**, 546 (1989)

[10] R.E. Peierls, Z. Phys. **80**, 763 (1933)

[11] N. Trivedi, R.T. Scalettar, and M. Randeria, Phys. Rev. B **54**, 3756 (1996)

[12] J.E. Gubernatis, M. Jarrell, R.N. Silver, and D.S. Sivia, Phys. Rev. B **44**, 6011 (1991)

[13] D.B. Haviland, Y.Liu, and A.M. Goldman, Phys. Rev. Lett. **62**, 2180, (1989); A.F. Hebard and M.A. Paalanen, Phys. Rev. Lett. **65**, 927 (1990); J.M. Valles, R.C. Dynes, and J.P. Garno, Phys. Rev. Lett. **69**, 3567 (1992); and A. Yazdani and A. Kapitulnik, Phys. Rev. Lett. **74**, 3037 (1995)

[14] P.J.H. Denteneer, R.T. Scalettar, and N. Trivedi, Phys. Rev. Lett. **83**, 4610 (1999)

[15] P.J.H. Denteneer, R.T. Scalettar, and N. Trivedi, Phys. Rev. Lett. **87**, 146401 (2001)

[16] P.J.H. Denteneer and R.T. Scalettar, Phys. Rev. Lett. **90**, 246401 (2003)

[17] K.A. Parendo, K.H. Sarwa B. Tan, A. Bhattacharya, M. Eblen-Zayas, N.E. Staley, and A.M. Goldman, Phys. Rev. Lett. **94**, 197004 (2005)

[18] R.M. Noack and D.J. Scalapino, Phys. Rev. B **47**, 305 (1993)

[19] M. Vekic, J.W. Cannon, D.J. Scalapino, R.T. Scalettar, and R.L. Sugar, Phys. Rev. Lett. **74**, 2367 (1995)

3 The Insulating State of Matter: A Geometrical Theory

Raffaele Resta
Dipartimento di Fisica, Università di Trieste,
Strada Costiera 11, 34151 Trieste, Italy,
Consiglio Nazionale delle Ricerche (CNR)
Istituto Officina dei Materiali (IOM) DEMOCRITOS,
34136 Trieste, Italy

Contents

1 Introduction **2**

2 Linear response and conductivity **3**

3 Drude weight **6**

4 The Resta-Sorella approach **7**

5 The Souza-Wilkens-Martin sum rule (periodic boundary conditions) **8**

6 Many-body Chern number **10**

7 Bounded samples within open boundary conditions **10**

8 Independent electrons **12**

9 Geometry in the anomalous Hall effect **16**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

By definition, a macroscopically homogeneous material is insulating whenever its dc longitudinal conductivity vanishes, i.e., when the real symmetric part of the conductivity tensor $\sigma^{(+)}_{\alpha\beta}(\omega)$ goes to zero for $\omega \to 0$. Here and throughout Greek subscripts are Cartesian indices. For a d-dimensional system of volume L^d the conductance G equals $L^{d-2}\sigma$. When measured in klitzing^{-1} (symbol $R_{\rm K}^{-1}$) conductivity is dimensionless in $d = 2$, while it has the dimensions of an inverse length in $d = 3$. We remind that 1 $R_{\rm K}$ = $h/e^2 \simeq$ 25,813 ohm [1].
Longitudinal conductivity is an intensive material property whose most general form can be written as

$$\sigma^{(+)}_{\alpha\beta}(\omega) = D_{\alpha\beta}\left[\delta(\omega) + \frac{i}{\pi\omega}\right] + \sigma^{(\rm regular)}_{\alpha\beta}(\omega) = \sigma^{(\rm Drude)}_{\alpha\beta}(\omega) + \sigma^{(\rm regular)}_{\alpha\beta}(\omega), \tag{1}$$

where the constant $D_{\alpha\beta}$ goes under the name of Drude weight. The insulating behavior of a material implies that $D_{\alpha\beta} = 0$ and that the real symmetric part of $\sigma^{(\rm regular)}_{\alpha\beta}(\omega)$ goes to zero for $\omega \to 0$ at zero temperature.
Eqn. (1) will be expressed below using linear-response theory (Kubo formulas); it may include—at least in principle—disorder and correlation, but does not include any dissipative mechanisms. The conductivity obeys the f-sum rule

$$\int_0^\infty d\omega\ \mathrm{Re}\ \sigma_{\alpha\alpha}(\omega) = \frac{D_{\alpha\alpha}}{2} + \int_0^\infty d\omega\ \mathrm{Re}\ \sigma^{(\rm regular)}_{\alpha\alpha}(\omega) = \frac{\omega_{\rm p}^2}{8} = \frac{\pi e^2 n}{2m}, \tag{2}$$

where n is the electron density and $\omega_{\rm p}$ is the plasma frequency. For free electrons (a gas of noninteracting electrons in a flat potential) $\sigma^{(\rm regular)}_{\alpha\beta}(\omega)$ vanishes, while $D_{\alpha\beta}$ assumes the same value as in classical physics [2], i.e, $D_{\alpha\beta} = \pi e^2 (n/m)\, \delta_{\alpha\beta}$: this explains the extraordinary longevity of Drude theory, developed in the year 1900. Given eqn. (2), switching on the potential (one-body and two-body) has the effect of transferring some spectral weight from the Drude peak into the regular term.
Dissipation can be included phenomenologically in the Drude term by adopting a single-relaxation-time approximation, exactly as in the classical textbook case [2], i.e.,

$$\sigma^{(\rm Drude)}_{\alpha\beta}(\omega) = \frac{i}{\pi}\frac{D_{\alpha\beta}}{\omega + i/\tau}, \tag{3}$$

whose $\tau \to \infty$ limit coincides with first term in the expression (1).
In the special case of a band metal (i.e., a crystalline system of non interacting electrons) $\sigma^{(\rm regular)}_{\alpha\beta}(\omega)$ is a linear-response property, which accounts for interband transitions, and is nonvanishing only at frequencies higher than a finite threshold. Instead, $D_{\alpha\beta}$ is a ground-state property which accounts for the inertia of the many-electron system in the adiabatic limit, and provides an effective value of n/m, where the free-electron value is modified by the periodic potential. After an integration by parts, $D_{\alpha\beta}$ can be equivalently expressed as a Fermi-surface integral, and acquires then the meaning of an "intraband" term [3]. As said above, the free-electron Drude weight is an upper limit for the actual value of $D_{\alpha\beta}$.

In 1964 Kohn published the milestone paper "Theory of the insulating state" [5], according to which insulators and metals differ in their ground state. Even before the system is excited by any probe, a different organization of the electrons is present in the ground state and this is the key feature discriminating between insulators and metals. Kohn's theory remained little visited for many years until the late 1990s, when a breakthrough occurred in electronic structure theory: the modern theory of polarization (for historical presentations see, e.g., Refs. [6–8]).
The many-body version of polarization theory appeared in 1998 [9]; shortly afterwards—inspired by the fact that electrical polarization discriminates qualitatively between insulators and metals—Resta and Sorella [10] provided a definition of many-electron localization rather different from Kohn's, and derived by the theory of polarization. Their program was completed soon after by Souza et al. [11] (hereafter quoted as SWM), thus providing the foundations of the modern theory of the insulating state, deeply rooted in geometrical concepts. A couple of review papers appeared in 2002 [12] and in 2011 [13]. We are going to revisit the theory here. The present viewpoint differs somewhat from the previous one; some of the results given here are original and published for the first time.

2 Linear response and conductivity

To start with, we fix our conventions about Fourier transforms

$$f(\omega) = \int_{-\infty}^{\infty} dt\ \mathrm{e}^{i\omega t} f(t) \qquad f(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\omega\ \mathrm{e}^{-i\omega t} f(\omega); \tag{4}$$

different conventions can be found in the literature.
Suppose we have a general input signal $f_{\mathrm{input}}(t)$ and the corresponding output $f_{\mathrm{output}}(t)$, which is due to the response of a time-independent physical system. The most general linear response is given by a convolution

$$f_{\mathrm{output}}(t) = \int_{-\infty}^{\infty} dt'\ \chi(t-t') f_{\mathrm{input}}(t'), \tag{5}$$

where $\chi(t)$ is the generalized susceptibility. It is easily verified that $\chi(t)$ can equivalently be defined as the response an instantaneous δ-like "kick" at $t = 0$; causality implies that $\chi(t) = 0$ for $t < 0$. The convolution theorem yields

$$f_{\mathrm{output}}(\omega) = \chi(\omega)\ f_{\mathrm{input}}(\omega). \tag{6}$$

Within quantum mechanics at zero temperature, we define $\chi(t)$ by means of a perturbation in the Hamiltonian $\Delta\hat{H} = -\delta(t)\hat{A}$ (the "kick"), acting on the system in its ground state. The response is measured as the expectation value of another operator $\hat{B}$. Without loss of generality we simplify our notation by assuming that

$$\langle\Psi_0|\hat{A}|\Psi_0\rangle = 0, \quad \langle\Psi_0|\hat{B}|\Psi_0\rangle = 0. \tag{7}$$

Time-dependent perturbation theory leads to the Kubo formula for the generalized susceptibility, which we write in the ω domain adopting the compact notation due to Zubarev [14–16]

$$\chi(\omega) = -\langle\langle \hat{B} | \hat{A} \rangle\rangle_\omega ; \tag{8}$$

$$\langle\langle \hat{B} | \hat{A} \rangle\rangle_\omega = \frac{1}{\hbar} \lim_{\eta \to 0^+} \sum_{n \neq 0}{}' \left(\frac{\langle \Psi_0 | \hat{B} | \Psi_n \rangle \langle \Psi_n | \hat{A} | \Psi_0 \rangle}{\omega - \omega_{0n} + i\eta} - \frac{\langle \Psi_0 | \hat{A} | \Psi_n \rangle \langle \Psi_n | \hat{B} | \Psi_0 \rangle}{\omega + \omega_{0n} + i\eta} \right) , \tag{9}$$

where $\omega_{0n} = (E_n - E_0)/\hbar$. The positive infinitesimal η ensures causality, and we remind that

$$\lim_{\eta \to 0+} \frac{1}{\omega \pm i\eta} = \mathcal{P} \frac{1}{\omega} \mp i\pi\delta(\omega) , \tag{10}$$

where $\mathcal{P}$ indicates the principal part. We draw attention to the fact that the sign convention adopted in this chapter agrees with Zubarev [14, 15] and Chandler [17], but is opposite the one of McWeeny [16] and other textbooks.

We apply the general linear response theory by addressing an interacting N-electron system, whose most general Hamiltonian we write, in the Schrödinger representation and in Gaussian units, as

$$\hat{H}(\boldsymbol{\kappa}) = \frac{1}{2m} \sum_{i=1}^{N} \left| \mathbf{p}_i + \frac{e}{c} \mathbf{A}(\mathbf{r}_i) + \hbar \boldsymbol{\kappa} \right|^2 + \hat{V} ; \tag{11}$$

the potential $\hat{V}$ includes one-body (possibly disordered) and two-body (electron-electron) contributions. Equation (11) is exact in the nonrelativistic, infinite-nuclear-mass limit. The velocity in eqn. (11) is augmented with two terms: $\mathbf{A}(\mathbf{r})$ is a vector potential of electromagnetic origin, and $\boldsymbol{\kappa}$, having the dimensions of an inverse length, is called "flux" or "twist". Setting $\boldsymbol{\kappa} \neq 0$ amounts to a gauge transformation. The electrons are confined in a cubic box of volume L^d and the eigenstates $|\Psi_n(\boldsymbol{\kappa})\rangle$ are normalized to one in the hypercube of volume L^{Nd}; we will adopt the simplifying notation $|\Psi_n(\boldsymbol{\kappa} = 0)\rangle = |\Psi_n\rangle$.

Bulk properties of condensed matter are obtained from the thermodynamic limit: $N \to \infty$, $L \to \infty$, with N/L^d constant. Since the following formulas will comprise $\boldsymbol{\kappa}$-derivatives evaluated at $\boldsymbol{\kappa} = 0$, it is important to stress that the differentiation is performed first, and the thermodynamic limit afterwards.

Two kinds of boundary conditions can be adopted for the given Hamiltonian: either periodic (PBCs) or "open" (OBCs). We briefly address the latter case first: the cubic box confines the electrons in an infinite potential well, the eigenstates $|\Psi_n(\boldsymbol{\kappa})\rangle$ are square-integrable over $\mathbb{R}^{Nd}$, and the position operator $\hat{\mathbf{r}} = \sum_i \mathbf{r}_i$ is the ordinary multiplicative operator. Within OBCs the effect of the gauge is easily "gauged away": the ground-state energy is gauge-independent, while the ground state is $|\Psi_0(\boldsymbol{\kappa})\rangle = \mathrm{e}^{-i\boldsymbol{\kappa}\cdot\hat{\mathbf{r}}} |\Psi_0\rangle$.

We will come back below (Sec. 7) to OBCs. For the time being we adopt instead Born-von-Kàrmàn PBCs over each electron coordinate $\mathbf{r}_i$ independently, whose Cartesian components $r_{i,\alpha}$ are then equivalent to the angles $2\pi r_{i,\alpha}/L$. The potential $\hat{V}$ enjoys the same periodicity, which implies that the electric field averages to zero over the sample. As noticed by W. Kohn in 1964 [5], PBCs violate gauge invariance in the conventional sense: for instance, the ground state energy $E_0(\boldsymbol{\kappa})$ actually *depends* on $\boldsymbol{\kappa}$.

In order to address conductivity it is essential to adopt PBCs: there cannot be any steady state current within OBCs. Furthermore, since the multiplicative position $\hat{\mathbf{r}}$ is no longer a legitimate operator within PBCs [9], it is mandatory to adopt the vector-potential gauge for the macroscopic electric field $\boldsymbol{\mathcal{E}}$: the perturbation in the Hamiltonian is therefore an ω-dependent vector potential $\delta\mathbf{A}$, constant in space.
One key point is that the vector potential modifies the velocity operator. We stick to the symbol $\hat{\mathbf{v}}$ for the velocity in absence of the perturbation, and at $\boldsymbol{\kappa} = 0$: this may include a *ground-state* vector potential, but not the *perturbing* one, i.e.,

$$\hat{\mathbf{v}} = \frac{1}{m}\sum_{i=1}^{N}\left[\,\mathbf{p}_i + \frac{e}{c}\mathbf{A}(\mathbf{r}_i)\,\right]. \tag{12}$$

The current carried by a generic state $|\Psi\rangle$ after the perturbation is switched on is therefore

$$\mathbf{j} = -\frac{e}{L^d}\langle\Psi|\hat{\mathbf{v}}|\Psi\rangle - \frac{e^2N}{mcL^d}\delta\mathbf{A}. \tag{13}$$

Expansion of the Hamiltonian to first order in the perturbing vector potential $\delta\mathbf{A}$ yields

$$\Delta\hat{H} = \frac{e}{c}\,\delta\mathbf{A}\cdot\mathbf{v}. \tag{14}$$

If we set $\boldsymbol{\mathcal{E}}$ and $\delta\mathbf{A}$ along the β direction, the linearly induced current in the α direction is

$$j_\alpha = -\frac{e^2N}{mcL^d}\,\delta A\,\delta_{\alpha\beta} - \frac{e}{L^d}\langle\!\langle\,\hat{v}_\alpha\,|\,\frac{e}{c}\,\delta A\;\hat{v}_\beta\rangle\!\rangle_\omega = -\frac{e^2}{cL^d}\left(\frac{N}{m}\,\delta_{\alpha\beta} + \langle\!\langle\hat{v}_\alpha|\hat{v}_\beta\rangle\!\rangle_\omega\right)\delta A(\omega), \tag{15}$$

where we are restoring the ω dependence. The term in δA^2, being constant in space, has zero matrix elements; it is also second order in $\boldsymbol{\mathcal{E}}$.
In order to arrive at the conductivity we need to express $\delta A(\omega)$ in eqn. (15) in terms of $\mathcal{E}(\omega)$. In the time domain their relationship is $\mathcal{E} = -\frac{1}{c}\partial\delta A/\partial t$; a naive integration would yield $\delta A(\omega) = -ic\mathcal{E}(\omega)/\omega$, but this violates causality. The correct integration yields:

$$\delta A(\omega) = c\mathcal{E}(\omega)\left[\frac{1}{i\omega} - \pi\delta(\omega)\right]. \tag{16}$$

Therefore the current, as expressed directly in terms of the field intensity, is

$$j_\alpha(\omega) = \sigma_{\alpha\beta}(\omega)\,\mathcal{E}_\beta(\omega) = -\frac{e^2}{L^d}\left(\frac{N}{m}\delta_{\alpha\beta} + \langle\!\langle\hat{v}_\alpha|\hat{v}_\beta\rangle\!\rangle_\omega\right)\left[\frac{1}{i\omega} - \pi\delta(\omega)\right]\mathcal{E}_\beta(\omega). \tag{17}$$

We then write the Kubo formula as

$$\langle\!\langle v_\alpha|v_\beta\rangle\!\rangle_\omega = \frac{1}{\hbar}\lim_{\eta\to 0^+}{\sum_{n\neq 0}}'\left(\frac{\mathcal{R}_{n,\alpha\beta} + i\,\mathcal{I}_{n,\alpha\beta}}{\omega - \omega_{0n} + i\eta} - \frac{\mathcal{R}_{n,\alpha\beta} - i\,\mathcal{I}_{n,\alpha\beta}}{\omega + \omega_{0n} + i\eta}\right), \tag{18}$$

$$\mathcal{R}_{n,\alpha\beta} = \mathrm{Re}\,\langle\Psi_0|v_\alpha|\Psi_n\rangle\langle\Psi_n|v_\beta|\Psi_0\rangle, \quad \mathcal{I}_{n,\alpha\beta} = \mathrm{Im}\,\langle\Psi_0|v_\alpha|\Psi_n\rangle\langle\Psi_n|v_\beta|\Psi_0\rangle, \tag{19}$$

where $\mathcal{R}_{n,\alpha\beta}$ is symmetric and $\mathcal{I}_{n,\alpha\beta}$ antisymmetric. The longitudinal conductivity is the symmetric part $\sigma^{(+)}_{\alpha\beta}(\omega)$ of the tensor. Upon exploiting eqn. (10) we eventually get

$$D_{\alpha\beta} = \frac{\pi e^2}{L^d}\left(\frac{N}{m}\delta_{\alpha\beta} - \frac{2}{\hbar}{\sum_{n\neq 0}}'\frac{\mathcal{R}_{n,\alpha\beta}}{\omega_{0n}}\right), \tag{20}$$

$$\text{Re}\ \sigma^{(\text{regular})}_{\alpha\beta}(\omega) = \frac{\pi e^2}{\hbar L^d} {\sum_{n\neq 0}}' \frac{\mathcal{R}_{n,\alpha\beta}}{\omega_{0n}} [\,\delta(\omega-\omega_{0n}) - \delta(\omega+\omega_{0n})\,], \tag{21}$$

$$\text{Im}\ \sigma^{(\text{regular})}_{\alpha\beta}(\omega) = \frac{2e^2}{\hbar L^d} {\sum_{n\neq 0}}' \frac{\mathcal{R}_{n,\alpha\beta}}{\omega_{0n}} \frac{\omega}{\omega_{0n}^2-\omega^2}. \tag{22}$$

The two terms $\sigma^{(\text{Drude})}_{\alpha\beta}(\omega)$ and $\sigma^{(\text{regular})}_{\alpha\beta}(\omega)$ obey the Kramers-Kronig relationships *separately*; we also remind that only the longitudinal conductivity $\sigma^{(+)}_{\alpha\beta}(\omega)$ is addressed for the time being. The transverse conductivity $\sigma^{(-)}_{\alpha\beta}$ will be addressed in Secs. 6 and 9.

At any finite-size L the spectrum is discrete and the system is gapped, while in a metal the gap closes in the large-L limit. It is therefore necessary to regularize the singular sums in eqns. (20-22); this can be done in the following way [18]: One starts assuming a finite value of η in the Kubo formula, eqn. (9), with η much larger than the level spacing; then one takes the $L\to\infty$ limit first, and the $\eta\to 0^+$ limit afterwards.

The first term in the parenthesis in eqn. (20) yields the free-electron Drude weight, while the second term accounts for the (always negative) correction due to the one-body potential and to the electron-electron interaction. We have given here the Kubo formula for a many-body Hamiltonian; for independent electrons eqn. (18) is easily transformed into a double sum over occupied and unoccupied orbitals [3].

3 Drude weight

We have arrived at eqn. (20) by means of linear-response theory, while we have stressed above that $D_{\alpha\beta}$ must be regarded as a ground-state property, which measures the inertia of the many-electron system in the adiabatic limit. In order to show this, we follow W. Kohn, who in 1964 adopted the "twisted" Hamiltonian, eqn. (11). By expanding $E_0(\boldsymbol{\kappa})$ to second order one gets

$$E_0(\boldsymbol{\kappa}) \simeq \frac{N\hbar^2}{2m}\kappa^2 - \hbar\,\kappa_\alpha\kappa_\beta\ \text{Re}\ {\sum_{n\neq 0}}' \frac{\langle\Psi_0|\hat{v}_\alpha|\Psi_n\rangle\langle\Psi_n|\hat{v}_\beta|\Psi_0\rangle}{\omega_{0n}}; \tag{23}$$

the expansion is essentially the many-body analogue of the elementary $\mathbf{k}\cdot\mathbf{p}$ expansion for the band energy, leading to the effective mass.

By comparing eqn. (23) to (20) one immediately gets Kohn's result:

$$D_{\alpha\beta} = \frac{\pi e^2}{\hbar^2 L^d} \frac{\partial^2 E_0(\boldsymbol{\kappa})}{\partial\kappa_\alpha\partial\kappa_\beta}\bigg|_{\boldsymbol{\kappa}=0}. \tag{24}$$

We remind that it is crucial to set $\boldsymbol{\kappa}=0$ in the derivative *before* the thermodynamic limit is taken: this ensures that we are following the ground state adiabatically [19]. In insulators the second derivative is zero: this can be proved in various ways.

In the simple case of a band metal eqn. (24) becomes the Brillouin-zone (BZ) integral [3]:

$$D_{\alpha\beta} = \pi e^2 \sum_j \int_{\text{BZ}} [d\,\text{k}]\ \theta(\mu-\varepsilon_{j\mathbf{k}})\, m^{-1}_{j,\alpha\beta}(\mathbf{k}), \tag{25}$$

where $[dk] = d\mathbf{k}/(2\pi)^d$, μ is the Fermi level, $\varepsilon_{j\mathbf{k}}$ are band energies, and the effective inverse mass tensor of band j is

$$m^{-1}_{j,\alpha\beta}(\mathbf{k}) = \frac{1}{\hbar^2}\frac{\partial^2 \varepsilon_{j\mathbf{k}}}{\partial k_\alpha \partial k_\beta}. \tag{26}$$

4 The Resta-Sorella approach

We consider a special value $\boldsymbol{\kappa}_0 = \frac{2\pi}{L}\mathrm{e}_\alpha$, where e_α is the unit vector in any Cartesian direction. For this special $\boldsymbol{\kappa}_0$ the effect of the gauge is easily gauged away; in fact the state vector

$$|\tilde{\Psi}_0(\boldsymbol{\kappa}_0)\rangle = \mathrm{e}^{-i\boldsymbol{\kappa}_0\cdot\hat{\mathbf{r}}}|\Psi_0\rangle \tag{27}$$

obeys PBCs, and is an eigenstate of $\hat{H}(\boldsymbol{\kappa}_0)$ with eigenvalue E_0, similarly to the OBCs case. Now the issue is whether $|\tilde{\Psi}_0(\boldsymbol{\kappa}_0)\rangle$ coincides or not with the genuine $|\Psi_0(\boldsymbol{\kappa}_0)\rangle$, obtained, as said above, by following the ground state adiabatically while $\boldsymbol{\kappa}$ is switched on continuously. Eqn. (24) shows that whenever $D \neq 0$ the state $|\Psi_0(\boldsymbol{\kappa}_0)\rangle$ has an energy higher than E_0: it is therefore an excited eigenstate of $\hat{H}(\boldsymbol{\kappa}_0)$, orthogonal to $|\tilde{\Psi}_0(\boldsymbol{\kappa}_0)\rangle$. If instead $D = 0$, then the state $|\Psi_0(\boldsymbol{\kappa}_0)\rangle$ coincides—apart for a phase factor—with $|\tilde{\Psi}_0(\boldsymbol{\kappa}_0)\rangle$ (we are assuming a nondegenerate ground state):

$$\langle\tilde{\Psi}_0(\boldsymbol{\kappa}_0)|\Psi_0(\boldsymbol{\kappa}_0)\rangle \;=\; \langle\Psi_0|\mathrm{e}^{i\boldsymbol{\kappa}_0\cdot\hat{\mathbf{r}}}|\Psi_0(\boldsymbol{\kappa}_0)\rangle = 0\,, \qquad D \neq 0, \tag{28}$$

$$\langle\tilde{\Psi}_0(\boldsymbol{\kappa}_0)|\Psi_0(\boldsymbol{\kappa}_0)\rangle \;=\; \langle\Psi_0|\mathrm{e}^{i\boldsymbol{\kappa}_0\cdot\hat{\mathbf{r}}}|\Psi_0(\boldsymbol{\kappa}_0)\rangle = \mathrm{e}^{i\gamma}, \quad D = 0. \tag{29}$$

We notice, en passant, that γ is the single-point Berry phase determining the polarization [9]; we are not discussing the issue here.

Replacing now $|\Psi_0(\boldsymbol{\kappa}_0)\rangle$ with $|\Psi_0\rangle$ we are approximating eqns. (28) and (29) to order $1/L$, i.e.,

$$|\mathfrak{z}_N| \;=\; |\,\langle\Psi_0|\mathrm{e}^{i\boldsymbol{\kappa}_0\cdot\hat{\mathbf{r}}}|\Psi_0\rangle\,| = \quad \mathcal{O}(1/L), \quad D \neq 0, \tag{30}$$

$$|\mathfrak{z}_N| \;=\; |\,\langle\Psi_0|\mathrm{e}^{i\boldsymbol{\kappa}_0\cdot\hat{\mathbf{r}}}|\Psi_0\rangle\,| = 1 - \mathcal{O}(1/L), \quad D = 0. \tag{31}$$

The Resta-Sorella [10] localization length is defined for an isotropic system in dimension d as

$$\lambda^2 = -\frac{1}{4\pi^2 n^{2/d}} \lim_{N\to\infty} N^{2/d-1} \log|\mathfrak{z}_N|^2 = -\frac{1}{4\pi^2} \lim_{N\to\infty} \frac{L^2}{N} \log|\mathfrak{z}_N|^2, \tag{32}$$

where $n = N/L^d$ is the density. Owing to eqns. (30) and (31) the localization length diverges when $D \neq 0$ and converges to a finite limit otherwise.

A very successful application of this theory concerns the Mott transition in 1-dimensional hydrogen chains within PBCs [20, 21]. We reproduce here Fig. 1 from Ref. [20] by Stella et al. who have performed variational quantum Monte Carlo studies, up to 66 atoms. The crossover between the weakly correlated (band) metallic regime—at small a—and the strongly correlated (Mott) insulating regime—at large a—is clearly visible in both panels of Fig. 1, which indicate the transition at $a \simeq 3.5$ bohr. The bottom panel shows that the modulus of the matrix element in eqns. (30) and (31). Top panel: λ/a. Bottom panel: the modulus of the matrix element

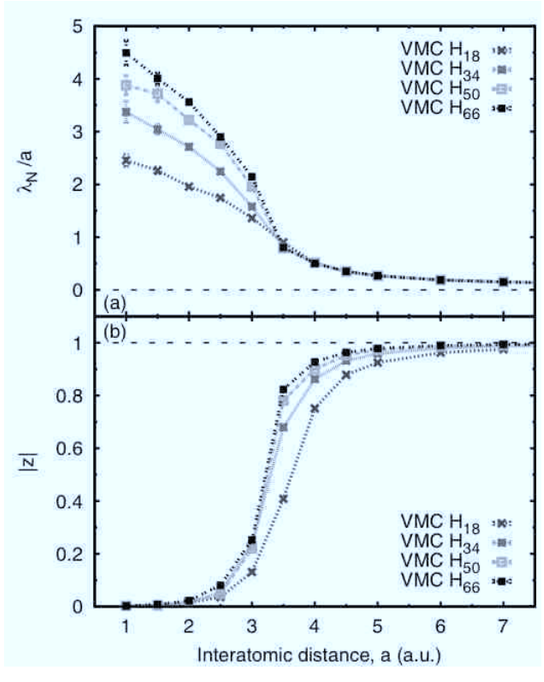


Fig. 1: *Results for chains of H atoms of different lengths as a function of the interatomic distance a, after Ref. [20]. Top panel: λ/a. Bottom panel: the modulus of the matrix element in eqns. (30) and (31).*

switches from zero to one in a narrow a region, the transition becoming sharper with increasing size. The top panel perspicuously shows that in the Mott-insulating regime λ is size-insensitive, while in the metallic regime it diverges with size. Unfortunately, the authors have chosen to plot λ/a instead of λ itself. Therefore the λ value in the large a limit cannot be verified: we expect that it goes to the isolated-atom limit, i.e., $\lambda = 1$ bohr.

5 The Souza-Wilkens-Martin sum rule (periodic boundary conditions)

The modern theory of the insulating state is also rooted in a sum rule, introduced in 2000 by SWM [11]. They define the insulating/metallic character of a homogenous material via the frequency integral

$$I_{\mathrm{SWM}} = \int_0^\infty \frac{d\omega}{\omega} \,\mathrm{Re}\, \sigma_{\alpha\alpha}(\omega); \tag{33}$$

for the sake of simplicity we address isotropic materials only. I_{SWM} converges in all insulators and diverges in all metals. In fact the integral converges at the upper limit—compare to eqn. (2)—but it diverges at the lower one whenever $D_{\alpha\alpha} \neq 0$ and also whenever Re $\sigma^{(\mathrm{regular})}_{\alpha\alpha}(0)$ is finite. The SWM integral has instead a finite value when the system has either a spectral gap or a mobility gap. We evaluate I_{SWM} using the *regular part* only of longitudinal conductivity,

eqn. (22):

$$I_{\rm SWM} = \frac{\pi e^2}{\hbar L^d} \sum_{n\neq 0}{}' \frac{|\langle \Psi_0 | \hat{v_\alpha} | \Psi_0 \rangle|^2}{\omega_{0n}^2}. \tag{34}$$

The SWM integral is related by a sum rule to the quantum metric, defined according to Provost and Vallee [22], and where the relevant parameter is the twist $\boldsymbol{\kappa}$. The metric-curvature tensor at $\boldsymbol{\kappa} = 0$ is

$$\mathcal{F}_{\alpha\beta} = \frac{1}{N} \big(\langle \partial_{\kappa_\alpha} \Psi_0 | \partial_{\kappa_\beta} \Psi_0 \rangle - \langle \partial_{\kappa_\alpha} \Psi_0 | \Psi_0 \rangle \langle \Psi_0 | \partial_{\kappa_\beta} \Psi_0 \rangle \big), \tag{35}$$

where we have divided by N in order to get an intensive quantity. This tensor is real symmetric in time-reversal invariant system, and may be endowed with an antisymmetric imaginary part if time-reversal invariance is lacking. The latter feature is discussed in the next Section.
The metric tensor at $\boldsymbol{\kappa} = 0$ is the real symmetric part of $\mathcal{F}_{\alpha\beta}$:

$$g_{\alpha\beta} = \frac{1}{N} \big(\text{Re } \langle \partial_{\kappa_\alpha} \Psi_0 | \partial_{\kappa_\beta} \Psi_0 \rangle - \langle \partial_{\kappa_\alpha} \Psi_0 | \Psi_0 \rangle \langle \Psi_0 | \partial_{\kappa_\beta} \Psi_0 \rangle \big); \tag{36}$$

since $g_{\alpha\beta}$ is gauge-invariant, we are going to evaluate it in the parallel-transport gauge, where

$$|\partial_{\kappa_\alpha} \Psi_0 \rangle = \hbar \sum_{n\neq 0}{}' |\Psi_n\rangle \frac{\langle \Psi_n | \hat{v}_\alpha | \Psi_0 \rangle}{E_0 - E_n} = -\sum_{n\neq 0}{}' |\Psi_n\rangle \frac{\langle \Psi_n | \hat{v}_\alpha | \Psi_0 \rangle}{\omega_{0n}}, \tag{37}$$

$$\mathcal{F}_{\alpha\beta} = \frac{1}{N} \sum_{n\neq 0}{}' \frac{\langle \Psi_0 | \hat{v}_\alpha | \Psi_n \rangle \langle \Psi_n | \hat{v}_\beta | \Psi_0 \rangle}{\omega_{0n}^2} = \frac{1}{N} \sum_{n\neq 0}{}' \frac{\mathcal{R}_{n,\alpha\beta} + i\mathcal{I}_{n,\alpha\beta}}{\omega_{0n}^2}. \tag{38}$$

From eqns. (34) and (38) we thus get

$$g_{\alpha\alpha} = \frac{\hbar}{\pi e^2 n} \int_0^\infty \frac{d\omega}{\omega} \text{ Re } \sigma_{\alpha\alpha}^{(\rm regular)}(\omega), \tag{39}$$

where $n = N/L^d$. We observe that in eqn. (39) the l.h.s. is a ground-state property, while the r.h.s. concerns the excitations of the system.
In the insulating case $\sigma^{(\rm regular)}$ coincides with the full conductivity; if $\varepsilon_{\rm g}$ is either the spectral gap or the mobility gap, the SWM sum rule reads

$$g_{\alpha\alpha} = \frac{\hbar}{\pi e^2 n} \int_{\varepsilon_{\rm g}/\hbar}^\infty \frac{d\omega}{\omega} \text{ Re } \sigma_{\alpha\alpha}(\omega). \tag{40}$$

The f-sum rule leads to the inequality:

$$g_{\alpha\alpha} < \frac{\hbar^2}{\pi e^2 n \, \varepsilon_{\rm g}} \int_{\varepsilon_{\rm g}/\hbar}^\infty d\omega \text{ Re } \sigma_{\alpha\alpha}(\omega) = \frac{\hbar^2}{2m\varepsilon_{\rm g}}. \tag{41}$$

From the above it becomes clear that the PBCs metric, when defined via eqn. (36), does *not* discriminate between insulators and metals: in the latter case it misses the (diverging) Drude contribution to eqn. (33). For instance, eqn. (39) vanishes for the paradigmatic metal: the free electron gas.
The author has recently shown how to remove this drawback of the PBCs metric, upon defining it in a somewhat more general way [23]. The novel metric coincides with the established one in the insulating case, but diverges in metals. The theory is incomplete, in that it only addresses independent electrons (in both the crystalline and noncrystalline cases). This recent development is reviewed in Sec. 8.

6 Many-body Chern number

The celebrated TKNN paper (Thouless, Kohmoto, Nightingale, and den Nijs) [24] explains the integer quantum Hall effect as the manifestation of a topological invariant of the electronic ground state in $2d$, the integer $C_1 \in \mathbb{Z}$, called Chern number of the first class. The choice of the sign of C_1 is not uniform across the literature: the one adopted in this Chapter is consistent with most of the recent papers.
Later Niu, Thouless, and Wu [25] addressed the fractional quantum Hall effect, where the many-body wavefunction is known to be strongly correlated. They provided a many-body definition of C_1 which, in the notations of the present work, reads:

$$C_1 = \frac{i}{2\pi} \int_0^{2\pi/L} d\kappa_x \int_0^{2\pi/L} d\kappa_y \, \big(\langle \partial_{\kappa_x} \Psi_0(\boldsymbol{\kappa}) | \partial_{\kappa_y} \Psi_0(\boldsymbol{\kappa}) \rangle - \langle \partial_{\kappa_y} \Psi_0(\boldsymbol{\kappa}) | \partial_{\kappa_x} \Psi_0(\boldsymbol{\kappa}) \rangle \big). \tag{42}$$

Since the $L \to \infty$ is implicit in the definition, we observed in Ref. [26] that the mean-value theorem yields

$$C_1 = \frac{i}{2\pi} \left(\frac{2\pi}{L} \right)^2 \big(\langle \partial_{\kappa_x} \Psi_0 | \partial_{\kappa_y} \Psi_0 \rangle - \langle \partial_{\kappa_y} \Psi_0 | \partial_{\kappa_x} \Psi_0 \rangle \big). \tag{43}$$

This is clearly proportional to the imaginary part of the metric-curvature tensor, as defined in eqn. (35):

$$C_1 = -4\pi n \, \mathrm{Im} \, \mathcal{F}_{xy} \,, \tag{44}$$

where $n = N/L^2$ is the $2d$ density. A minor detail is worth mentioning: the ground-state wavefunction is a singlet state in the previous Section, while it is instead spin-polarized in the quantum-Hall regime.
The main result by Niu, Thouless, and Wu is the expression of the quantized Hall conductivity in terms of the many-body Chern number C_1. From eqns. (17) and (18) one gets

$$\mathrm{Re} \, \sigma_{xy}^{(-)}(\omega) = -\frac{e^2}{\omega L^2} \mathrm{Im} \, \langle\!\langle v_x | v_y \rangle\!\rangle_\omega = -\frac{2e^2}{\hbar L^2} {\sum_{n\neq 0}}' \frac{\mathcal{I}_{n,xy}}{\omega^2 - \omega_{0n}^2}, \tag{45}$$

and from eqns. (38) and (44) the final result is

$$\mathrm{Re} \, \sigma_{xy}^{(-)}(0) = \frac{2e^2 n}{\hbar} \, \mathrm{Im} \, \mathcal{F}_{xy} = -\frac{e^2}{h} C_1. \tag{46}$$

7 Bounded samples within open boundary conditions

At variance with the PBCs results presented in Sec. 5, the OBCs metric *does* carry the information to discriminate between insulators and metals. As said above, within OBCs the twist is easily gauged away and one has $|\Psi_0(\boldsymbol{\kappa})\rangle = \mathrm{e}^{-i\boldsymbol{\kappa}\cdot\hat{\mathbf{r}}} |\Psi_0\rangle$, where $\hat{\mathbf{r}} = \sum_i \mathbf{r}_i$ is the ordinary position operator, well defined within OBCs.
It is expedient to adopt a $\boldsymbol{\kappa}$-dependent phase factor and write instead:

$$|\Psi_0(\boldsymbol{\kappa})\rangle = \mathrm{e}^{-i\boldsymbol{\kappa}\cdot(\hat{\mathbf{r}}-\mathbf{d})} |\Psi_0\rangle, \quad \mathbf{d} = \langle \Psi_0 | \hat{r} | \Psi_0 \rangle. \tag{47}$$

The gauge-invariant metric, eqn. (36), takes then the form

$$\tilde{g}_{\alpha\beta} = \frac{1}{N}\big(\langle\Psi_0|\hat{r}_\alpha\hat{r}_\beta|\Psi_0\rangle - \langle\Psi_0|\hat{r}_\alpha|\Psi_0\rangle\langle\Psi_0|\hat{r}_\beta\Psi_0\rangle \big), \tag{48}$$

where the different symbol emphasizes the different boundary conditions adopted for $|\Psi_0\rangle$. eqn. (48) clearly shows that $\tilde{g}_{\alpha\beta}$ is the second cumulant moment of the position, or equivalently the ground-state quantum fluctuation of polarization. The basic tenet of the modern theory of the insulating state is that the OBCs metric, eqn. (48), in the large-N limit diverges in all metals and converges in all insulators.

We are going to recast $\tilde{g}_{\alpha\beta}$ in terms of one-body and two body densities, defined as

$$n(\mathbf{r}_1) = N \sum_{\sigma_1} \int d\mathbf{x}_2 d\mathbf{x}_3 \dots d\mathbf{x}_N |\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N)|^2, \tag{49}$$

$$n^{(2)}(\mathbf{r}_1, \mathbf{r}_2) = N(N-1) \sum_{\sigma_1\sigma_2} \int d\mathbf{x}_3 \dots d\mathbf{x}_N |\Psi(\mathbf{x}_1, \mathbf{x}_2, \dots \mathbf{x}_N)|^2, \tag{50}$$

where $\mathbf{x}_i \equiv (\mathbf{r}_i, \sigma_i)$ are the space and spin coordinates of the i-the electron, and a singlet ground state is assumed. Straightforward manipulations lead to the equivalent form:

$$\tilde{g}_{\alpha\beta} = \frac{1}{2N} \int d\mathbf{r} d\mathbf{r}' \; (\mathbf{r} - \mathbf{r}')_\alpha (\mathbf{r} - \mathbf{r}')_\beta [\, n(\mathbf{r}) n(\mathbf{r}') - n^{(2)}(\mathbf{r}, \mathbf{r}') \,], \tag{51}$$

showing that $\tilde{g}_{\alpha\beta}$ is the second moment of the exchange-correlation hole, averaged over the sample.

We have not justified yet why the OBCs metric discriminates between insulators and metals. In a bounded sample there cannot be a steady-state current, nonetheless an oscillating field induces charge sloshing and an oscillating macroscopic current. Therefore at $\omega \neq 0$ a linear relationship of the kind $j_\alpha(\omega) = \tilde{\sigma}_{\alpha\beta}(\omega) E_\beta(\omega)$ holds. The *definition* of the insulating state, making reference to large bounded samples, is that even $\tilde{\sigma}_{\alpha\beta}(\omega)$ vanishes in the $\omega \to 0$ limit. The order of limits is crucial: first $N \to \infty$, and then $\omega \to 0$. The SWM integral bypasses this problem of limits: the insulating state requires that

$$\tilde{I}_{\rm SWM} = \int_0^\infty \frac{d\omega}{\omega} \; \text{Re}\; \tilde{\sigma}_{\alpha\alpha}(\omega), \tag{52}$$

stays finite in the large-N limit. We stress that $\tilde{\sigma}_{\alpha\beta}(\omega)$ differs from the genuine longitudinal conductivity $\sigma^{(+)}_{\alpha\beta}(\omega)$ in two respects: it lacks the Drude peak, and it includes contributions from the sample boundary. The latter feature enters the Kubo formula by means of the matrix elements.

We are going to relate $\tilde{I}_{\rm SWM}$ to the OBCs metric $\tilde{g}_{\alpha\beta}$. To this aim we start converting $\tilde{g}_{\alpha\beta}$ into a sum-over-states form. Using again eqn. (37), we get an expression identical in form to eqn. (38), i.e.,

$$\tilde{g}_{\alpha\beta} = \frac{1}{N} {\sum_{n\neq 0}}' \frac{\langle\Psi_0|\hat{v}_\alpha|\Psi_n\rangle\langle\Psi_n|\hat{v}_\beta|\Psi_0\rangle}{\omega_{0n}^2}; \tag{53}$$

the key point is that the velocity matrix elements therein are *very different*, owing to the different boundary conditions. Addressing once more the extreme case of the free electron gas, all matrix elements in eqn. (38) vanish (by an obvious selection rule); they don't vanish for a bounded sample within OBCs, as in eqn. (53).

Within OBCs we may safely adopt the scalar potential gauge, where $\tilde{\sigma}$ has the compact expression

$$\tilde{\sigma}_{\alpha\beta}(\omega) = -\frac{e^2}{L^3}\langle\langle \hat{v}_\alpha | \hat{r}_\beta \rangle\rangle_\omega \tag{54}$$

in Zubarev's notations. The matrix elements of $\hat{\mathbf{r}}$ are converted into the matrix elements of $\hat{v}$ by means of the commutator $[\hat{H}, \hat{\mathbf{r}}]$, i.e., $\langle\Psi_0|\hat{\mathbf{r}}|\Psi_n\rangle = i\langle\Psi_0|\hat{\mathbf{v}}|\Psi_n\rangle/\omega_{0n}$, to obtain

$$\text{Re}\ \tilde{\sigma}_{\alpha\alpha}(\omega) = \frac{e^2\pi}{\hbar L^3}\sum_{n\neq 0}{}' \frac{|\langle\Psi_0|\hat{v}_\alpha|\Psi_n\rangle|^2}{\omega_{0n}}[\ \delta(\omega-\omega_{0n}) + \delta(\omega-\omega_{0n})\ ], \tag{55}$$

leading to the OBCs version of the SWM sum rule

$$\tilde{g}_{\alpha\alpha} = \frac{\hbar}{\pi e^2 n}\int_0^\infty \frac{d\omega}{\omega}\ \text{Re}\ \tilde{\sigma}_{\alpha\alpha}(\omega). \tag{56}$$

Once more, eqns. (55) and (56) are identical in form to their PBCs counterpart, eqns. (21) and (39), but their physical content—as well as their defining quantities—are very different in the metallic case. For instance, $\tilde{\sigma}_{\alpha\beta}(\omega)$ by itself obeys the f-sum rule, while $\sigma^{(\text{regular})}_{\alpha\beta}(\omega)$ does not: see eqn. (2).

In the insulating case, instead, the PBCs conductivity $\sigma(\omega)$ coincides with the OBCs one $\tilde{\sigma}(\omega)$. It follows that the (finite) metric $\tilde{g}_{\alpha\alpha}$ coincides with $g_{\alpha\alpha}$ and obeys the SWM sum rule in the form of eqn. (40). This can be proved in various ways; the basic feature is that the macroscopic polarization $P_\alpha(\omega)$ linearly induced by an oscillating field stays finite for $\omega \to 0$ in insulating materials, and can therefore be evaluated using either OBCs or PBCs, in any gauge.

Finally, we observe that the l.h.s. of eqn. (56) is a ground-state quantum fluctuation, while the r.h.s. is a property of the system *excitations*. Eqn. (56) belongs then to the general class of fluctuation-dissipation theorems.

8 Independent electrons

Owing to eqn. (51), in the noninteracting case the OBCs metric $\tilde{g}_{\alpha\beta}$ is expressed in terms of the one-body density matrix as $\rho(\mathbf{r},\mathbf{r}') = 2\langle\mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle$ as

$$\tilde{g}_{\alpha\beta} = \frac{1}{N}\int d\mathbf{r}d\mathbf{r}'\ (\mathbf{r}-\mathbf{r}')_\alpha(\mathbf{r}-\mathbf{r}')_\beta\,|\langle\mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle|^2. \tag{57}$$

As said above, the convergence/divergence of $\tilde{g}_{\alpha\beta}$ in the large-N limit discriminates between electrons and metals. For instance, the well known $\mathcal{P}$ expression for the free-electron gas [27], when inserted in eqn. (57), yields a diverging $\tilde{g}_{\alpha\beta}$ in $d = 1$, 2, and 3. This is what is expected in a metal, and is in sharp contrast with the OBCs metric $g_{\alpha\beta}$ which—if defined as in eqn. (36)—vanishes. The difference is to be ascribed to the different order of limits. In this Section we are going to provide a more general definition of the OBCs metric $g_{\alpha\beta}$, which coincides with eqn. (36) in the insulating case, but has the virtue of diverging in the metallic case.

Crystalline systems

Besides the electron-gas case, where $\mathcal{P}$ is known analytically, simulations for model noninteracting systems within OBCs have indeed demonstrated the large-N divergence of the OBCs metric $\tilde{g}$, eqn. (58), in the metallic case [28,29]. Simulations and heuristic arguments altogether suggest that the metallic divergence of $\tilde{g}$ is of order of the linear dimension L of the system in $d = 1$, 2, or 3.

The PBCs metric $g_{\alpha\beta}$, as defined so far, *does not* diverge in the metallic case and requires therefore a somewhat different definition in order to acquire the same desirable feature. The novel definition provided here follows Ref. [23].

In the crystalline case the PBCs ground-state projector $\mathcal{P}$ is

$$\mathcal{P} = V_{\text{cell}} \sum_j \int_{\text{BZ}} [d\,\text{k}]\ \theta(\mu - \varepsilon_{j\mathbf{k}}) |\psi_{j\mathbf{k}}\rangle\langle\psi_{j\mathbf{k}}|, \tag{58}$$

where BZ is the Brillouin-zone, $|\psi_{j\mathbf{k}}\rangle = \text{e}^{i\mathbf{k}\cdot\mathbf{r}}|u_{j\mathbf{k}}\rangle$ are the Bloch states (normalized to one over the unit cell of volume V_{cell}), $\varepsilon_{j\mathbf{k}}$ are the band energies, μ is the Fermi level, the integration is over $[d\,\text{k}] = d\mathbf{k}/(2\pi)^d$, and d is the dimension. We recast eqn. (58) in terms of Bloch projectors $\mathcal{P}_{\mathbf{k}}$ as

$$\langle\mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle = V_{\text{cell}} \int_{\text{BZ}} [d\,\text{k}]\ \text{e}^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} \langle\mathbf{r}|\mathcal{P}_{\mathbf{k}}|\mathbf{r}'\rangle, \qquad \mathcal{P}_{\mathbf{k}} = \sum_j \theta(\mu - \varepsilon_{j\mathbf{k}}) |u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}|, \tag{59}$$

and we choose a gauge which makes $|u_{j\mathbf{k}}\rangle$ smooth on the whole BZ: this is always possible, even in topologically nontrivial materials. The Bloch projectors $\mathcal{P}_{\mathbf{k}}$ are gauge-invariant in the generalized Marzari-Vanderbilt sense [30, 31], i.e., they are invariant for any unitary transformation of the occupied $|u_{j\mathbf{k}}\rangle$ at the given $\mathbf{k}$.

The BZ integrand is smooth in insulators, and only piecewise continuous in metals. In the latter case, the sharpness of the Fermi surface is responsible for the power-law decay of $\langle\mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle$ for $|\mathbf{r} - \mathbf{r}'| \to \infty$; the decay is instead quasi-exponential (i.e. exponential times a power) in insulators [32].

The ground-state projector is lattice-periodic, i.e.,

$$\langle\mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle = \langle\mathbf{r} + \mathbf{R}|\mathcal{P}|\mathbf{r}' + \mathbf{R}\rangle, \tag{60}$$

where $\mathbf{R}$ is a lattice translation. Therefore in the large-N limit the crystalline form of eqn. (57) is

$$\tilde{g}_{\alpha\beta} = \frac{1}{N_{\text{c}}} \int_{\text{cell}} d\mathbf{r} \int_{\text{all space}} d\mathbf{r}'\ (\mathbf{r} - \mathbf{r}')_\alpha (\mathbf{r} - \mathbf{r}')_\beta\, |\langle\mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle|^2, \tag{61}$$

where N_{c} is the number of electrons per crystal cell.

Next we are going to address the PBCs metric $g_{\alpha\beta}$, starting with the insulating case, where it coincides with the OBCs metric $\tilde{g}_{\alpha\beta}$ and obeys the SWM sum rule in the form of eqn. (40). The number of occupied bands is $N_{\text{c}}/2$, independent of $\mathbf{k}$. A well known result, first shown in Ref. [33], is

$$g_{\alpha\beta} = \frac{2}{n} \int_{\text{BZ}} [d\,\text{k}]\ \text{Re}\ \mathcal{F}_{\alpha\beta}(\mathbf{k}), \tag{62}$$

where n is the electron density and $\mathcal{F}_{\alpha\beta}(\mathbf{k})$ is the $\mathbf{k}$-dependent metric-curvature tensor [22, 30, 31, 33]:

$$\mathcal{F}_{\alpha\beta}(\mathbf{k}) = \sum_{j=1}^{N_c/2} \langle \partial_{k_\alpha} u_{j\mathbf{k}} | \partial_{k_\beta} u_{j\mathbf{k}} \rangle - \sum_{j,j'=1}^{N_c/2} \langle \partial_{k_\alpha} u_{j\mathbf{k}} | u_{j'\mathbf{k}} \rangle \langle u_{j'\mathbf{k}} | \partial_{k_\beta} u_{j\mathbf{k}} \rangle. \tag{63}$$

The BZ integral in the r.h.s. of eqn. (62) made its first appearance in the Marzari-Vanderbilt theory of maximally localized Wannier function, where it provides the gauge-invariant term in the quadratic spread, universally indicated as $\Omega_{\rm I}$ in the literature. The relationship is [30, 31]:

$$\Omega_{\rm I} = V_{\rm cell} \int_{\rm BZ} [d\,{\rm k}] \ \sum_\alpha \mathcal{F}_{\alpha\alpha}(\mathbf{k}) = \frac{N_c}{2} \sum_\alpha g_{\alpha\alpha}. \tag{64}$$

Notice that the original definition of $\Omega_{\rm I}$ is not intensive.

Following Ref. [23] the PBCs metric of a band insulator, eqn. (62), can be recast in a compact trace form, which has the virtue of showing gauge invariance explicitly. A tedious calculation shows that

$$\mathcal{F}_{\alpha\beta}(\mathbf{k}) = {\rm Tr}\, \{ \mathcal{P}_\mathbf{k} (\partial_{k_\alpha} \mathcal{P}_\mathbf{k}) (\partial_{k_\beta} \mathcal{P}_\mathbf{k}) \}. \tag{65}$$

We may extend the definition of eqns. (62) and (65) to the metallic case as well, noticing that the $\mathbf{k}$-derivative of the Bloch projector acquires a singular δ-like term at the Fermi level:

$$\partial_{k_\alpha} \mathcal{P}_\mathbf{k} = -\sum_j \delta(\mu - \varepsilon_{j\mathbf{k}}) \partial_{k_\alpha} \varepsilon_{j\mathbf{k}} \ |u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}| + \sum_j \theta(\mu - \varepsilon_{j\mathbf{k}}) (|u_{j\mathbf{k}}\rangle\langle \partial_{k_\alpha} u_{j\mathbf{k}}| + |\partial_{k_\alpha} u_{j\mathbf{k}}\rangle\langle u_{j\mathbf{k}}|). \tag{66}$$

In the insulating case the singularity vanishes, and we thus retrieve the previous result, while the squared δ, when inserted into eqn. (62), provides the sought for divergence.

The second term in eqn. (66) is smooth in insulators; instead it is only piecewise continuous—and therefore integrable—in metals. If eqn. (62) is evaluated using this term *only*, we retrieve the nondivergent SWM sum rule in the PBCs form of eqn. (39).

Noncrystalline systems

The OBCs metric, eqn. (58), has been implemented to study the metal-insulator (Anderson) transition in disordered systems. It is well known that in $1d$ any amount of (uncorrelated) disorder yields an insulating ground state. OBCs simulations over a lattice model in $1d$ have shown that the system has no spectral gap but eqn. (58) converges nonetheless to a finite value in the large-N limit [28].

In $3d$ matters are different: a genuine metal-insulator transition may occur. The integral in eqn. (57) converges whenever $\langle \mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle$ is exponential in $|\mathbf{r}-\mathbf{r}'|$ (as in crystalline insulators), as well as when $|\langle \mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle|^2$ decays as $|\mathbf{r}-\mathbf{r}'|^{-a}$, with $a > 5$. A detailed study of the Anderson transition on a paradigmatic lattice model, based on eqn. (57), has recently appeared [34]. This confirms that the OBCs metric is an alternative tool with respect to the ones currently adopted in the literature. The standard computational methods to address the Anderson transition are often peculiar to lattice models (recursive methods and the like) [35], while our approach has a

general ab-initio formulation and could in principle be applied to realistic disordered materials by standard electronic structure methods.

A novel approach to disordered and macroscopically inhomogeneous systems (such as heterojunctions) has been proposed in Ref. [23]. We recast here the PBCs metric in trace form, i.e.,

$$g_{\alpha\beta} = \frac{2}{n} \,\mathrm{Re} \int_{\mathrm{BZ}} [d\,\mathrm{k}]\ \mathrm{Tr}\,\{\mathcal{P}_{\mathbf{k}}(\partial_{k_\alpha}\mathcal{P}_{\mathbf{k}})(\partial_{k_\beta}\mathcal{P}_{\mathbf{k}})\}. \tag{67}$$

We are going to show that eqn. (67) can be equivalently cast in a form where any explicit reference to lattice periodicity disappears.

In order to arrive at such a transformation, we start noticing that the integrand in eqn. (58) is periodical over the reciprocal lattice, and therefore the BZ integral of its $\mathbf{k}$-gradient vanishes:

$$i(\mathbf{r}-\mathbf{r}')\langle\mathbf{r}|\mathcal{P}|\mathbf{r}'\rangle + V_{\mathrm{cell}} \int_{\mathrm{BZ}} [d\,\mathrm{k}]\ \mathrm{e}^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')}\langle\mathbf{r}|\partial_{\mathbf{k}}\mathcal{P}_{\mathbf{k}}|\mathbf{r}'\rangle = 0; \tag{68}$$

we remind that eqn. (68) is a well behaved expression only in insulators. The first term therein is i times $[\mathbf{r},\mathcal{P}]$: a lattice periodic operator (unlike $\mathbf{r}$ itself). The trace of eqn. (68) can therefore be cast as

$$\int_{\mathrm{BZ}} [d\,\mathrm{k}]\ \mathrm{Tr}\,\{\partial_{k_\alpha}\mathcal{P}_{\mathbf{k}}\} = -\frac{i}{V_{\mathrm{cell}}} \int_{\mathrm{cell}} d\mathbf{r}\ \langle\mathbf{r}|\,[r_\alpha,\mathcal{P}]\,|\mathbf{r}\rangle = -i\mathrm{Tr}_V\,\{[r_\alpha,\mathcal{P}]\}, \tag{69}$$

where Tr_V indicates the trace per unit volume in the Schrödinger representation.

Using similar arguments it is not difficult to prove that, for an unbounded sample within PBCs,

$$g_{\alpha\beta} = \frac{2}{n} \,\mathrm{Re} \int_{\mathrm{BZ}} [d\,\mathrm{k}]\ \mathrm{Tr}\,\{\mathcal{P}_{\mathbf{k}}(\partial_{k_\alpha}\mathcal{P}_{\mathbf{k}})(\partial_{k_\beta}\mathcal{P}_{\mathbf{k}})\} = -\frac{2}{n}\,\mathrm{Re}\,\mathrm{Tr}_V\,\{\mathcal{P}\,[r_\alpha,\mathcal{P}]\,[r_\beta,\mathcal{P}]\}. \tag{70}$$

The second expression on the r.h.s. has two outstanding virtues: (i) it is expressed directly in the Schrödinger representation, making no reference to reciprocal space, and (ii) it can be adopted as such for supercells of arbitrarily large size, thus extending the concept of PBCs metric to noncrystalline systems, such as alloys and liquids. We have not proved yet that such form can be adopted as it stands even for bounded samples within OBCs.

If we evaluate the trace per unit volume over the whole sample of volume V, eqn. (70) yields

$$g_{\alpha\beta} = -\frac{2}{N}\mathrm{Re}\ \mathrm{Tr}\,\{\mathcal{P}\,[r_\alpha,\mathcal{P}]\,[r_\beta,\mathcal{P}]\} = \frac{2}{N}\mathrm{Tr}\,\{\mathcal{P}r_\alpha r_\beta\mathcal{P}\} - \frac{2}{N}\mathrm{Tr}\,\{\mathcal{P}\alpha\mathcal{P}r_\beta\mathcal{P}\}. \tag{71}$$

We have stated above that in insulators the PBCs metric $g_{\alpha\beta}$ is finite and coincides with the OBCs metric $\tilde{g}_{\alpha\beta}$: a simple calculation confirms that eqn. (71) is indeed identical to eqn. (57). Our novel approach reconciles the PBCs metric with the OBCSs one: both metrics yield the same message even in the metallic case. Looking at eqn. (70), the first expression on the r.h.s. diverges because of the sharpness of the Fermi surface embedded in eqn. (66), while the divergence of the second expression has been discussed in Sec. 7.

The next issue is whether one may adopt eqn. (70) *locally*, in order to address inhomogeneous systems: preliminary results indicate that the answer is affirmative [36]. For an isotropic system the local marker for the insulating state is the real function

$$\mathcal{L}_{\alpha\alpha}(\mathbf{r}) \propto \mathrm{Re}\ \langle\mathbf{r}|\,\mathcal{P}\,[r_\alpha,\mathcal{P}]\,[r_\alpha,\mathcal{P}]\,|\mathbf{r}\rangle\ : \tag{72}$$

when averaged locally in a homogenous region of the sample it detects the insulating vs. metallic character of that region. For instance in a metal/insulator heterojunction it diverges on the metallic side and converges to a finite value on the insulating side. It therefore provides a marker complementary to the (commonly used) local density of states; at variance with it, $\mathcal{L}_{\alpha\alpha}(\mathbf{r})$ probes locally the organization of the electrons in the *ground state*, the main property that—according to the outstanding Kohn's message [5]—discriminates insulators from metals.

9 Geometry in the anomalous Hall effect

On the theory side, the anomalous Hall effect (in both insulators and metals) is closely related to the theory of the insulating state. In this section we remain at the independent-particle level. Furthermore, in agreement with the literature on the topic, we adopt a spinless-electron formalism: factors of two here will differ from the previous sections. The metric-curvature tensor for a band insulator, eqn. (63), is rewritten as

$$\mathcal{F}_{\alpha\beta}(\mathbf{k}) = \sum_{j=1}^{N_c} \langle \partial_{k_\alpha} u_{j\mathbf{k}} | \partial_{k_\beta} u_{j\mathbf{k}} \rangle - \sum_{j,j'=1}^{N_c} \langle \partial_{k_\alpha} u_{j\mathbf{k}} | u_{j'\mathbf{k}} \rangle \langle u_{j'\mathbf{k}} | \partial_{k_\beta} u_{j\mathbf{k}} \rangle . \tag{73}$$

So far, we have addressed the real symmetric part of $\mathcal{F}_{\alpha\beta}(\mathbf{k})$, i.e., the $\mathbf{k}$-space metric first introduced by Marzari and Vanderbilt in the theory of maximally localized Wannier functions [30, 31]. The imaginary antisymmetric part (times -2) is the Berry curvature of the occupied manifold:

$$\Omega_{\alpha\beta}(\mathbf{k}) = -2\,\mathrm{Im}\,\mathcal{F}_{\alpha\beta}(\mathbf{k}) = i \sum_{j=1}^{N_c} \big(\langle \partial_{k_\alpha} u_{j\mathbf{k}} | \partial_{k_\beta} u_{j\mathbf{k}} \rangle - \langle \partial_{k_\beta} u_{j\mathbf{k}} | \partial_{k_\alpha} u_{j\mathbf{k}} \rangle \big) . \tag{74}$$

From eqn. (65) we equivalently get

$$\Omega_{\alpha\beta}(\mathbf{k}) = i\,\mathrm{Tr}\,\{\mathcal{P}_{\mathbf{k}}[\,\partial_{k_\alpha}\mathcal{P}_{\mathbf{k}}, \partial_{k_\beta}\mathcal{P}_{\mathbf{k}}\,]\}. \tag{75}$$

While this form was used in the past for the insulating case only, we stress that it holds for the metallic case as well: in fact, the singular term in eqn. (66) disappears after antisymmetrization. The key difference is that the Berry curvature of the occupied manifold is smooth in insulators and only piecewise continuous in metals: its BZ integral is well defined and finite in both cases. The anomalous Hall conductivity (AHC) is by definition the Hall conductivity in zero magnetic field; it can be nonvanishing only if the Hamiltonian lacks time-reversal symmetry. When expressed in klitzing^{-1} it is dimensionless for $d = 2$, while it has the dimensions of an inverse length for $d = 3$. The known expression for the $\omega = 0$ AHC in both metals and insulators is

$$\sigma^{(-)}_{\alpha\beta}(0) = \frac{4\pi e^2}{h} \int_{\mathrm{BZ}} [d\,\mathrm{k}]\ \ \mathrm{Im}\,\mathcal{F}_{\alpha\beta}(\mathbf{k}) = -\frac{e^2}{h}\frac{1}{2\pi}\int_{\mathrm{BZ}} d\mathbf{k}\ \Omega_{\alpha\beta}(\mathbf{k}), \tag{76}$$

and this expressions holds for both $d = 2$ and $d = 3$; notice the two equivalent forms, where the integral is either in $[d\,\mathrm{k}] = d\mathbf{k}/(2\pi)^d$ or in $d\mathbf{k}$.

We address the insulating case first: the AHC is quantized, and in natural units it equals minus the Chern invariant C_γ, usually defined as

$$C_\gamma = \frac{1}{4\pi}\varepsilon_{\gamma\alpha\beta}\int_{\mathrm{BZ}} d\mathbf{k}\ \Omega_{\alpha\beta}(\mathbf{k}). \tag{77}$$

In $2d$ the Chern invariant is a dimensionless integer $\in \mathbb{Z}$. The definition of eqn. (77) coincides indeed with the Chern number C_1, as defined in Sec. 6 for a many-body wavefunction [26].
In the metallic case eqn. (76) is nonquantized: the difference owes to the fact that the Berry curvature of the occupied manifold $\Omega_{\alpha\beta}(\mathbf{k})$ is smooth in insulators and only piecewise continuous in metals.
In the metallic case eqn. (76) yields only the intrinsic (or geometric) contribution to the AHC; extrinsic contributions, known as "skew scattering" and "side jump" must be added [37]. We stress that, instead, extrinsic contributions have no effect in insulators, owing to the robustness of topological observables.
The same transformation as in eqn. (70) can be carried over for the antisymmetric imaginary part of $\mathcal{F}_{\alpha\beta}(\mathbf{k})$, leading to

$$\sigma^{(-)}_{\alpha\beta}(0) = -\frac{4\pi e^2}{h}\mathrm{Im\ Tr}_V\ \{\mathcal{P}[r_\alpha,\mathcal{P}]\,[r_\beta,\mathcal{P}]\}, \tag{78}$$

where we address a possibly disordered sample, although still unbounded within PBCs. If we try to proceed analogously to what we did for the real symmetric part, by adopting eqn. (78) even for a bounded sample within OBCs and evaluating the trace over the whole sample, we get a vanishing result: the tensor entering eqns. (70) and (78) is obviously real symmetric. This stems from the fact that even the original definition of $\tilde{g}_{\alpha\beta}$, eqn. (48), is *not* endowed with an antisymmetric term.
The solution of the paradox was found in Ref. [38]. The real function

$$\mathfrak{C}(\mathbf{r}) = 4\pi\ \mathrm{Im}\ \langle\mathbf{r}|\,\mathcal{P}[r_\alpha,\mathcal{P}]\,[r_\beta,\mathcal{P}]\,|\mathbf{r}\rangle \tag{79}$$

carries indeed the information which allows evaluating the AHC *locally*; but its average has to be evaluated using an inner region of the bounded sample and *not* the whole sample. The boundary provides a compensating contribution. When the bounded sample is a crystallite, one may integrate $\mathfrak{C}(\mathbf{r})$ over the central cell; this integral, divided by the cell volume (area in $2d$), provides the AHC value in the large-sample limit.
In the insulating case the function $\mathfrak{C}(\mathbf{r})$ samples the topological nature of the ground state *locally*: it has therefore been dubbed "topological marker" [38]. Simulations on a paradigmatic lattice model in $2d$ for bounded samples (crystalline and disordered) and for heterojunctions have shown that $\mathfrak{C}(\mathbf{r})$ samples indeed the local Chern number (equal to minus the Hall conductivity in natural units).
The metallic case differs from the insulating one in two important respects: (i) the macroscopic current flows across the whole sample, while it only flows at the boundaries in topological insulators; (ii) the ground-state projector entering eqn. (79) is power-law in $|\mathbf{r}-\mathbf{r}'|$, while it

is quasi-exponential in insulators (including topological insulators). Despite these differences, simulations reported in Ref. [39] demonstrate that $\mathfrak{C}(\mathbf{r})$, eqn. (79), provides in the metallic case a "geometrical marker", which allows to evaluate the geometrical contribution to the AHC locally in both homogenous and inhomogeneous samples.
The homogeneous case of a "dirty" metal deserves a comment. The trace per unit volume of $\mathfrak{C}(\mathbf{r})$ clearly includes some geometrical effects due to the impurities. It is argued that the AHC evaluated in this way may yield the sum of the intrinsic and side-jump contributions to the AHC, while instead it may not include the skew scattering [37, 40].
Ref. [39] also provides a convergence study. Therein, a metallic crystallite is addressed vs. an insulating one, and the AHC of the material is evaluated, as said above, by averaging the respective $\mathfrak{C}(\mathbf{r})$ over the central cell. The convergence to the bulk value is—as expected—exponential in the insulating case. In the metallic case the convergence is instead of the order L^{-3}, where L is the linear dimension of the sample. While the actual simulations are in $2d$, it is conjectured that the convergence is of order L^{-3} in any dimension, in analogy with what happens to the large-sample metallic divergence of the metric (of order L in any d).

Acknowledgments

During several stays at the Donostia International Physics Center in San Sebastian (Spain) over the years, I have discussed thoroughly the topics in the present Review with Ivo Souza; his invaluable contribution is gratefully acknowledged. The hospitality by the Center is acknowledged as well. This work was supported by the ONR Grant No. N00014-12-1-1041.

References

[1] `http://physics.nist.gov/cuu/Constants/`

[2] N.W. Ashcroft and N.D. Mermin, *Solid State Physics* (Saunders, Philadelphia, 1976) Ch. 1

[3] P.B. Allen, in: [4], p. 165

[4] S.G. Louie and M.L. Cohen (eds.): *Conceptual foundations of materials: A standard model for ground- and excited-state properties* (Elsevier, 2006)

[5] W. Kohn, Phys. Rev. **133**, A171 (1964)

[6] D. Vanderbilt and R. Resta, in [4], p. 139

[7] R. Resta and D. Vanderbilt, in Ch.H. Ahn, K.M. Rabe, and J.-M. Triscone (eds.): *Physics of Ferroelectrics: a Modern Perspective*, Topics in Applied Physics Vol. **105** (Springer, 2007), p. 31

[8] R. Resta, J. Phys.: Condens. Matter **22** 123201 (2010)

[9] R. Resta, Phys. Rev. Lett. **80**, 1800 (1998)

[10] R. Resta and S. Sorella, Phys. Rev. Lett. **82**, 370 (1999)

[11] I. Souza, T. Wilkens, and R.M. Martin, Phys. Rev. B **62**, 1666 (2000)

[12] R. Resta, J. Phys.: Condens. Matter **14**, R625 (2002)

[13] R. Resta, Eur. Phys. J. B **79**, 121 (2011)

[14] D.N. Zubarev, *Non-Equilibrium Statistical Mechanics* (Consultants Bureau, New York, 1974)

[15] D.N. Zubarev, Soviet Phys. Ushpekhi **3**, 320 (1960)

[16] R. McWeeny, *Methods of Molecular Quantum Mechanics*, Second Edition (Academic, London, 1992)

[17] D. Chandler, *Introduction to Modern Statistical Mechanics* (Oxford University Press, 1987)

[18] E. Akkermans, J. Math. Phys. **38**, 1781 (1997)

[19] D.J. Scalapino, S.R. White, and S.C. Zhang, Phys. Rev. **47**, 7995 (1993)

[20] L. Stella, C. Attaccalite, S. Sorella, and A. Rubio, Phys. Rev. B **84**, 245117 (2011)

[21] M. El Khatib *et al.*, J. Chem. Phys. **142**, 094113 (2015)

[22] J.P. Provost and G. Vallee, Commun. Math Phys. **76**, 289 (1980)

[23] R. Resta, arXiv:1703.00712

[24] D.J. Thouless, M. Kohmoto, M.P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982)

[25] Q. Niu, D.J. Thouless, and Y.S. Wu, Phys. Rev. B **31**, 3372 (1985)

[26] R. Resta, Phys. Rev. Lett. **95**, 196805 (2005)

[27] G.F. Giuliani and G. Vignale, *Quantum Theory of the Electron Liquid* (Cambridge University Press, Cambridge, 2005)

[28] G.L. Bendazzoli, S. Evangelisti, A. Monari, and R. Resta, J. Chem. Phys. **133**, 064703 (2010)

[29] G.L. Bendazzoli, S. Evangelisti, and A. Monari, Int. J. Quantum Chem. **112**, 653 (2012)

[30] N. Marzari and D. Vanderbilt, Phys. Rev. B **56**, 12847 (1997)

[31] N. Marzari, A.A. Mostofi, J.R. Yates, I. Souza, and D. Vanderbilt, Rev. Mod. Phys. **84**, 1419 (2012)

[32] L. He and D. Vanderbilt, Phys. Rev. Lett. **86**, 5341 (2001)

[33] C. Sgiarovello, M. Peressi, and R. Resta, Phys. Rev. **64**, 115202 (2001)

[34] T. Olsen, R. Resta, and I. Souza, Phys. Rev. B **95**, 045109 (2017)

[35] B. Kramer and A. MacKinnon, Rep. Prog. Phys. **56**, 1469 (1993)

[36] A. Marrazzo and R. Resta, unpublished

[37] N. Nagaosa, J. Sinova, S. Onoda, A.H. MacDonald, and N.P. Ong, Rev. Mod. Phys. **82**, 1539 (2010)

[38] R. Bianco and R. Resta, Phys. Rev. B **84**, 241106(R) (2011)

[39] A. Marrazzo and R. Resta, Phys. Rev. B **95**, 121114(R) (2017)

[40] R. Bianco, R. Resta, and I. Souza, Phys. Rev. B **90**, 125153 (2014)

4 Exchange Mechanisms

Erik Koch
Institute for Advanced Simulation
Forschungszentrum Jülich

Contents

1 Introduction **2**

2 Coulomb exchange **4**

3 Kinetic exchange **7**
3.1 A toy model 7
3.2 Direct exchange 8
3.3 Second quantization for pedestrians 10
3.4 Mean-field treatment 12
3.5 Superexchange 13
3.6 Ferromagnetic superexchange 16

4 Double exchange **18**

5 Orbital-ordering **21**

6 Extended systems **23**
6.1 Hubbard model 23
6.2 Mott transition 23
6.3 Heisenberg model 25

7 Conclusion **27**

A Atomic units **28**

B Downfolding **29**

C Pauli matrices **30**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

One of the profound *Surprises in Theoretical Physics* [1] is that magnetism is an inherently quantum mechanical effect. Classically, magnetic moments originate from electric currents: A current density $\vec{j}(\vec{r})$ generates a magnetic moment

$$\vec{\mu} = \frac{1}{2}\int \vec{r}\times\vec{j}\,d^3r\,. \tag{1}$$

These moments interact via the dipole-dipole interaction. The magnetostatic interaction energy between two dipoles at a distance R, $\hat{R}$ being the unit-vector from the position of the first to that of the second dipole,

$$\Delta E = \frac{\mu_0}{4\pi}\,\frac{\vec{\mu}_1\cdot\vec{\mu}_2 - 3(\hat{R}\cdot\vec{\mu}_1)(\hat{R}\cdot\vec{\mu_2})}{R^3} = \frac{\vec{\mu}_1\cdot\vec{\mu}_2 - 3(\hat{R}\cdot\vec{\mu}_1)(\hat{R}\cdot\vec{\mu_2})}{4\pi\varepsilon_0 c^2\,R^3} \tag{2}$$

depends on their distance and relative orientation. This can, however, not be the origin of the magnetism found in actual materials: In a classical system charges cannot flow in thermodynamic equilibrium, the celebrated Bohr-van Leeuwen theorem, and hence there are no magnetic moments to begin with [2].

In quantum mechanics, however, non-vanishing charge currents in the ground state are not uncommon: An electron in state $\Psi(\vec{r})$ corresponds to a current density

$$\vec{j}(\vec{r}) = -\frac{e\hbar}{2im_e}\left(\overline{\Psi(\vec{r})}\,\nabla\Psi(\vec{r}) - \Psi(\vec{r})\,\nabla\overline{\Psi(\vec{r})}\right) \tag{3}$$

which, for a complex wave function $\Psi(\vec{r})$, is usually non-vanishing. According to (1) it produces a magnetic moment proportional to the expectation value of the angular momentum

$$\vec{\mu}_L = -\frac{e\hbar}{2m_e}\,\langle\vec{L}\,\rangle = -\mu_B\,\langle\vec{L}\,\rangle\,. \tag{4}$$

The constant of proportionality is the Bohr magneton μ_B. In particular, an atomic orbital $|n,l,m\rangle$ has a magnetic moment proportional to its magnetic quantum number $\vec{\mu} = -\mu_B\,m\,\hat{z}$. Also the electron spin $\vec{S}$ carries a magnetic moment

$$\vec{\mu}_S = -g_e\mu_B\,\langle\vec{S}\,\rangle\,. \tag{5}$$

The constant of proportionality between spin and magnetic moment differs from that between orbital momentum and moment by the gyromagnetic ratio g_0. Dirac theory gives $g_e = 2$, which is changed to $g_e \approx 2.0023\ldots$ by QED corrections.

Atomic moments are thus of the order of μ_B. For two such moments at a distance of 1 Å the magnetostatic energy (2) is of the order of 0.05 meV, corresponding to a temperature of less than 1 K. Therefore, magnetic ordering which, e.g., in magnetite (Fe_3O_4), persists till about 860 K, must originate from an interaction other than the magnetostatic interaction of dipoles. Indeed, it is the interplay of electronic properties which are apparently unrelated to magnetism, the Pauli principle in combination with the Coulomb repulsion (Coulomb exchange) as well

as the hopping of electrons (kinetic exchange) that leads to an effective coupling between the magnetic moments in a solid.

The basic mechanisms of the exchange coupling are quite simple: Since many-body wave functions must change sign under the permutation of Fermions, electrons of the same spin cannot be at the same position. Electrons of like spin thus tend to avoid each other, i.e., the probability of finding them close to each other tends to be lower than for electrons of opposite spin (exchange hole). In that sense the Coulomb energy between two electrons depends on their relative spins. By this argument, aligning electron spins tends be energetically favorable. This *Coulomb exchange* is the basis of Hund's first rule. When more than one atom is involved, electrons can hop from one site to its neighbor. This kinetic term is, again, modified by the Pauli principle, as the hopping to an orbital on the neighboring atom will only be possible, if there is not already an electron of the same spin occupying that orbital and by the Coulomb repulsion among the electrons. This is the idea of *kinetic exchange*. When Coulomb exchange and kinetic terms work together we speak of *double exchange*. In that case the electron-hopping serves to mediate the spin-correlation created on an atom to its neighbors.

Exchange mechanisms are idealizations of characteristic situations found in real materials. As such they are merely approximations, but they afford a simplification of the complicated realistic description, which provides a good basis for thinking about the relevant effects in a real material. We will start by discussing the effect of Coulomb exchange matrix elements (Sec. 2). To keep things simple, we will discuss a two-orbital model and only mention atomic multiplets and Hund's rule. Next we turn to exchange mechanisms involving also hopping (Sec. 3). We start by looking at the a simple two-site model with two electrons. Focussing on the limit of strong electronic correlations (Coulomb repulsion dominating electron hopping), we introduce the method of downfolding to derive an effective Hamiltonian in which an explicit coupling of the electron spins appears. While conceptually simple, this direct exchange mechanism is rarely found in real materials. There hopping between correlated orbitals is usually mediated by a weakly correlated orbital. This is the superexchange mechanism. The derivation is very similar to that of kinetic exchange. However, the number of states involved, makes explicit book-keeping tedious. To simplify our work, we introduce second quantization as a simple notation of many-electron states. This also enables us to easily discuss double exchange, which combines direct exchange on an atom with coupling to the neighbors via electron hopping. Examples are the superexchange between transition metal atoms bridged by an oxygen at a right angle, which arises from the Coulomb exchange on the oxygen, as well as the exchange in mixed-valence compounds (Sec. 4). The competition between kinetic and double exchange is described by the Goodenough-Kanamori rules. Finally we show that exchange is not restricted to coupling spins, but can also produce interactions between orbital occupations (Sec. 5).

How exchange gives rise to an effective coupling of momenta is most easily shown for single- or two-site models. To see how these results carry over to solids, we consider the case of direct exchange (Sec. 6). Starting from the Hubbard model we show how taking the limit of strong correlations leads to the t-J-model, which, for half-filling, simplifies to the Heisenberg model.

2 Coulomb exchange

The Coulomb repulsion between electrons,

$$H_U = \sum_{i<j} \frac{1}{|\vec{r}_i - \vec{r}_j|}\,, \tag{6}$$

is manifestly spin-independent. Nevertheless, because of the antisymmetry of the many-electron wave function, the eigenenergies of H_U depend on spin. This is the basis of the multiplet structure in atoms and of Hund's first two rules.

To understand the mechanism of this Coulomb exchange we consider a simple two-electron model. In the spirit of tight-binding, we assume that we have solved the two-electron Hamiltonian H_0, replacing the interaction term H_U, e.g., as a self-consistent potential $\sum_i U(\vec{r}_i)$, obtaining an orthonormal set of one-electron eigenstates $\varphi_\alpha(\vec{r})$ with eigenvalues ε_α. We now ask for the effect of re-introducing the interaction $H_U - \sum_i U(\vec{r}_i)$. The largest effect we will find for states that are degenerate.

Let us consider two orbitals $\alpha = a,\, b$. Then the two-electron Slater determinants with spins σ and σ'

$$\begin{aligned}\Psi_{a,\sigma;\,b\sigma'}(\vec{r}_1,s_1;\,\vec{r}_2,s_2) &= \frac{1}{\sqrt{2}}\begin{vmatrix}\varphi_a(\vec{r}_1)\,\sigma(s_1) & \varphi_a(\vec{r}_2)\,\sigma(s_2)\\ \varphi_b(\vec{r}_1)\,\sigma'(s_1) & \varphi_b(\vec{r}_2)\,\sigma'(s_2)\end{vmatrix} \\ &= \frac{1}{\sqrt{2}}\Big(\varphi_a(\vec{r}_1)\,\varphi_b(\vec{r}_2)\;\sigma(s_1)\,\sigma'(s_2) - \varphi_b(\vec{r}_1)\,\varphi_a(\vec{r}_2)\;\sigma'(s_1)\,\sigma(s_2)\Big)\end{aligned} \tag{7}$$

are degenerate eigenstates of H_0 with eigenvalue $\varepsilon_a + \varepsilon_b$, independent of the spin orientations. To see how this degeneracy is lifted, we calculate the matrix elements of H_U in the basis of the Slater determinants $\Psi_{a,\sigma;\,b\sigma'}$.

When both electrons have the same spin ($\sigma = \sigma'$), we can factor out the spin functions

$$\Psi_{a,\sigma;\,b\sigma} = \frac{1}{\sqrt{2}}\Big(\varphi_a(\vec{r}_1)\varphi_b(\vec{r}_2) - \varphi_b(\vec{r}_1)\varphi_a(\vec{r}_2)\Big)\;\sigma(s_1)\sigma(s_2) \tag{8}$$

and obtain

$$\left\langle \Psi_{a,\sigma;\,b,\sigma}\left|\frac{1}{|\vec{r}_1-\vec{r}_2|}\right|\Psi_{a,\sigma;\,b,\sigma}\right\rangle = \frac{1}{2}\left(U_{ab} - J_{ab} - J_{ba} + U_{ba}\right) = U_{ab} - J_{ab} \tag{9}$$

where the direct terms are the Coulomb integral

$$U_{ab} = \int d^3r_1 \int d^3r_2\, \frac{|\varphi_a(\vec{r}_1)|^2\;|\varphi_b(\vec{r}_2)|^2}{|\vec{r}_1-\vec{r}_2|} \tag{10}$$

while the cross terms give the exchange integral

$$J_{ab} = \int d^3r_1 \int d^3r_2\, \frac{\overline{\varphi_a(\vec{r}_1)}\;\varphi_b(\vec{r}_1)\;\overline{\varphi_b(\vec{r}_2)}\;\varphi_a(\vec{r}_2)}{|\vec{r}_1-\vec{r}_2|}\,. \tag{11}$$

For the states where the electrons have opposite spin ($\sigma' = -\sigma$)

$$\left\langle \Psi_{a,\sigma;\,b,-\sigma}\left|\frac{1}{|\vec{r}_1-\vec{r}_2|}\right|\Psi_{a,\sigma;\,b,-\sigma}\right\rangle = U_{ab} \tag{12}$$

the diagonal matrix element has no exchange contribution, as the overlap of the spin functions for the cross terms vanish. There are however off-diagonal matrix elements

$$\left\langle \Psi_{a\uparrow;b\downarrow} \left| \frac{1}{|\vec{r}_1 - \vec{r}_2|} \right| \Psi_{a\downarrow;b\uparrow} \right\rangle = -J_{ab} \; . \tag{13}$$

Since H_U does not change the spins, these are the only non-zero matrix elements. In the basis of the states $\Psi_{\uparrow\uparrow}$, $\Psi_{\uparrow\downarrow}$, $\Psi_{\downarrow\uparrow}$ and $\Psi_{\downarrow\downarrow}$ the Coulomb term is thus given by

$$H_U = \begin{pmatrix} U_{ab} - J_{ab} & 0 & 0 & 0 \\ 0 & U_{ab} & -J_{ab} & 0 \\ 0 & -J_{ab} & U_{ab} & 0 \\ 0 & 0 & 0 & U_{ab} - J_{ab} \end{pmatrix} . \tag{14}$$

The triplet states $\Psi_{\uparrow\uparrow}$ and $\Psi_{\downarrow\downarrow}$ are obviously eigenstates of H_U with eigenenergy

$$\Delta\varepsilon_{\text{triplet}} = U_{ab} - J_{ab} \; . \tag{15}$$

Diagonalizing the 2×2 submatrix, we obtain the third triplet state $(\Psi_{\uparrow\downarrow} + \Psi_{\downarrow\uparrow})/\sqrt{2}$ and the singlet state $(\Psi_{\uparrow\downarrow} - \Psi_{\downarrow\uparrow})/\sqrt{2}$

$$\frac{1}{\sqrt{2}} \left(\Psi_{\uparrow\downarrow} - \Psi_{\downarrow\uparrow} \right) = \frac{1}{\sqrt{2}} \Big(\varphi_a(\vec{r}_1)\varphi_b(\vec{r}_2) + \varphi_b(\vec{r}_1)\varphi_a(\vec{r}_2) \Big) \; \frac{1}{\sqrt{2}} \Big(|\downarrow\uparrow\rangle - |\uparrow\downarrow\rangle \Big) \tag{16}$$

with energy

$$\Delta\varepsilon_{\text{singlet}} = U_{ab} + J_{ab} \; . \tag{17}$$

To see whether the triplet or the singlet is lower in energy, we need to know the sign of the exchange matrix element. While the Coulomb integral U_{ab}, having a positive integrand, is obviously positive, it is less obvious that also $J_{ab} > 0$. Introducing $\Phi(\vec{r}) = \overline{\varphi_a(\vec{r})}\varphi_b(\vec{r})$ and Fourier transforming to $\Phi(\vec{k}) = \int d^3k \, \Phi(\vec{r}) \, e^{-i\vec{k}\cdot\vec{r}}$ we obtain [3,4]:

$$J_{ab} = \int d^3r_1 \, \overline{\Phi(\vec{r}_1)} \underbrace{\int d^3r_2 \, \frac{1}{|\vec{r}_1 - \vec{r}_2|} \, \Phi(\vec{r}_2)}_{=(2\pi)^{-3} \int dk \, \Phi(k) e^{ikr_1} 4\pi/k^2} \tag{18}$$

$$= \frac{1}{(2\pi)^3} \int d^3k \underbrace{\int d^3r_1 \, e^{i\vec{k}\cdot\vec{r}_1} \overline{\Phi(\vec{r}_1)}}_{=\overline{\Phi(k)}} \Phi(\vec{k}) \, \frac{4\pi}{k^2} \tag{19}$$

$$= \frac{1}{(2\pi)^3} \int d^3k \, |\Phi(\vec{k})|^2 \, \frac{4\pi}{k^2} > 0 \tag{20}$$

Thus the triplet states are below the singlet state by an energy $2J_{ab}$. If the φ_α are degenerate atomic orbitals, this is an example of Hund's first rule: For an atomic shell, the lowest state will have maximum spin.

Since H_U only contains interactions within the system of electrons, it commutes with the total orbital momentum $[H_U, \vec{L}_{\text{tot}}] = 0$. Obviously it also commutes with the total spin $\vec{S}_{\text{tot}}$. The eigenstates of $H_0 + H_U$ can thus be classified by their quantum numbers L and S. These

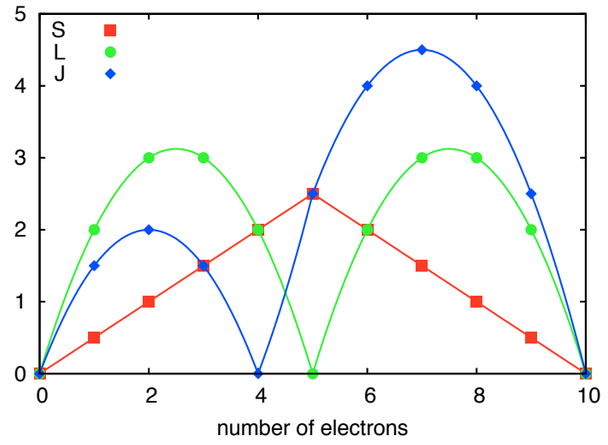


Fig. 1: *Angular momenta of the Hund's rules ground state* $^{2S+1}L_J$ *for* d*-shells.*

terms are written as ^{2S+1}L. For p- and d-shells they are listed in table 1. Hund's rules give the multiplet term with the lowest energy: For a given shell, this lowest state has the largest possible spin (Hund's first rule). If there are several terms of maximum multiplicity, the one with lowest energy has the largest total orbital momentum (Hund's second rule). There is a third Hund's rule, which, however, is not related with the electron-electron repulsion but with spin-orbit coupling: Within L-S coupling H_{SO} splits the atomic orbitals into eigenstates of the total angular momentum $\vec{J} = \vec{L}_{\text{tot}} + \vec{S}_{\text{tot}}$. The multiplets ^{2S+1}L thus split into $^{2S+1}L_J$. The term with the lowest energy is the one with smallest J if the shell is less than half-filled and largest J if it is more than half-filled (Hund's third rule). These rules are illustrated for d-shells in Fig. 2. A more detailed discussion of multiplet effects and the Coulomb interaction in atomic-like systems can be found in [5, 6], calculations of multiplets including spin-orbit coupling can be performed online, at `http://www.cond-mat.de/sims/multiplet`.

| | | |

3 Kinetic exchange

When electron-hopping plays the main role in the exchange mechanism, we speak of kinetic exchange. In contrast to Coulomb exchange the resulting interactions are usually antiferromagnetic, i.e., they prefer antiparallel spins. The physical principle of kinetic exchange can be understood in a simple two-site system. We discuss this problem in some detail and introduce two key concepts along the way: downfolding and second quantization. As we will see in the subsequent sections, realistic exchange mechanisms are natural generalizations of this simple mechanism [7–9].

3.1 A toy model

As a toy model, we consider the minimal model of an H_2 molecule. We restrict ourselves to two (orthonormal) orbitals, φ_1 and φ_2, separated by some distance. If we add an electron to the system, that electron will be able to move between the two orbitals, with a matrix element $-t$. Because we allow the electron to only occupy two orbitals, the Hamiltonian is a 2×2 matrix

$$H = \begin{pmatrix} 0 & -t \\ -t & 0 \end{pmatrix} . \tag{21}$$

This tight-binding Hamiltonian is easily diagonalized giving the linear combinations

$$\varphi_\pm = \frac{1}{\sqrt{2}} \Big(\varphi_1 \pm \varphi_2 \Big) \tag{22}$$

as eigenstates with eigenenergies $\varepsilon_\pm = \mp t$. We have written the hopping matrix element as $-t$, so that for $t > 0$ the state without a node, φ_+, is the ground state.
Pictorially we can write the basis states by specifying which orbital the electron occupies. For a spin-up electron we then write

$$\varphi_1 = |\uparrow , \cdot \rangle \qquad \text{and} \qquad \varphi_2 = | \cdot , \uparrow \rangle \tag{23}$$

where we now represent the basis states by where the electron is located.
If there are two electrons in the system, i.e., one electron per orbital, we can again use basis states which just specify, which orbitals the electrons occupy. For two electrons of opposite spin we then find two states where the electrons are in different orbitals

$$|\uparrow , \downarrow\rangle \qquad |\downarrow , \uparrow\rangle \qquad \text{“covalent states”}$$

and two states where the electrons are in the same orbital

$$|\updownarrow , \cdot \rangle \qquad | \cdot , \updownarrow\rangle \qquad \text{“ionic states”.}$$

In this basis the Hamiltonian matrix for our simple model of the H_2 molecule has the form

$$H = \begin{pmatrix} 0 & 0 & -t & -t \\ 0 & 0 & +t & +t \\ -t & +t & U & 0 \\ -t & +t & 0 & U \end{pmatrix} \qquad \begin{matrix} |\uparrow , \downarrow\rangle \\ |\downarrow , \uparrow\rangle \\ |\updownarrow , \cdot \rangle \\ | \cdot , \updownarrow\rangle \end{matrix} \tag{24}$$

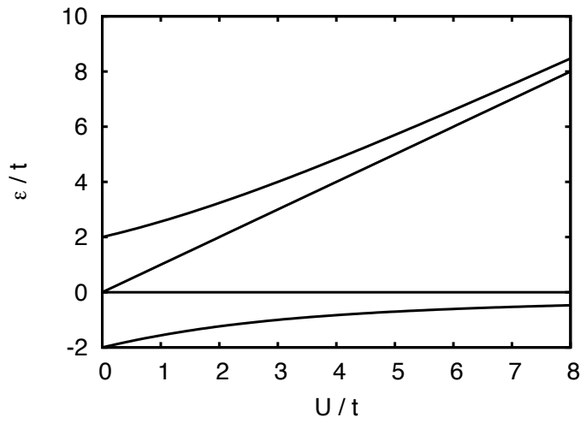


Fig. 2: *Spectrum of the two-site Hubbard model as a function of U. For large U there are two levels with energy close to zero. Their energy difference corresponds to the exchange energy. The remaining two states with ionic character have an energy roughly proportional to U.*

As before, moving an electron to a neighboring orbital gives a matrix element $-t$, with an additional sign when the order of the electrons is changed (Fermi statistics!). For the ionic states, where both electrons are in the same orbital, we have the Coulomb matrix element U. Coulomb matrix elements involving electrons on different sites are, for reasonably large distance between the sites, negligible. So there is no Coulomb exchange, just the local Coulomb repulsion in our model. Diagonalizing H we find the energy spectrum and the corresponding eigenstates:

$$\varepsilon_\pm = \frac{U}{2} \pm \frac{\sqrt{U^2+16\,t^2}}{2} \quad , \quad \Psi_\pm = \frac{\Big(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle - \frac{\varepsilon_\pm}{2t}\big[\,|\!\updownarrow,\cdot\,\rangle + |\,\cdot\,,\!\updownarrow\rangle\,\big]\Big)}{\sqrt{2+\varepsilon_\pm^2/(2t^2)}}$$

$$\varepsilon_{\mathrm{cov}} = 0 \quad , \quad \Psi_{\mathrm{cov}} = \frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle\Big)$$

$$\varepsilon_{\mathrm{ion}} = U \quad , \quad \Psi_{\mathrm{ion}} = \frac{1}{\sqrt{2}}\Big(|\!\updownarrow,\cdot\,\rangle - |\,\cdot\,,\!\updownarrow\rangle\Big)$$

The eigenenergies as a function of U are shown in figure 2.

3.2 Direct exchange

Again, we have found that the energy of two-electron states depends on the relative spin of the electrons. To understand this more clearly we analyze the limit when U is much larger than t. From Fig. 2 we see that there are two states with energies that increase with U. They are the states Ψ_{ion} and Ψ_+ that have considerable contributions of the ionic states. Then there are two states whose energy is close to zero. They are the states that have mainly covalent character. To find the energy and the character of these levels in the limit $U \to \infty$ we can just expand $\varepsilon_- \to -4t^2/U$ and $\varepsilon_+ \to U + 4t^2/U$. We thus see that while the purely covalent state, the spin-triplet state Ψ_{cov}, is independent of U, Ψ_- has a slightly lower energy due to some small

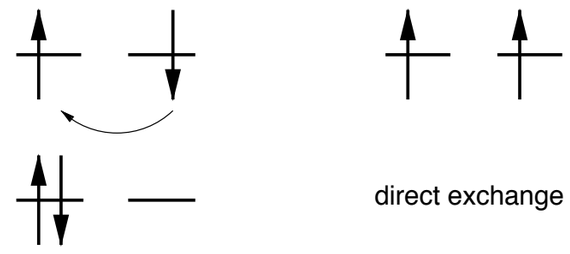


Fig. 3: *Simple picture of direct exchange: The antiparallel alignment of the spins (left) is favored, since it allows the electrons to hop to the neighboring site. For parallel spins (right) hopping is suppressed by the Pauli principle.*

admixture of the ionic states. In the limit $U \to \infty$ it becomes the maximally entangled state $(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle)/\sqrt{2}$. We see that for large U, Ψ_- cannot be expressed, even approximately, as a Slater determinant, see also Sec. 3.4. This is the reason why strongly correlated systems are so difficult to describe.

An instructive method to analyze the large-U limit, which can readily be generalized to more complex situations, where we can no longer diagonalize the full Hamiltonian, is the downfolding technique. The mathematical background is explained in the appendix. The idea of downfolding is to partition the Hilbert space into parts that are of interest, here the low-energy covalent type states, and states that should be projected out, here the high-energy ionic states. With this partitioning we can view the Hamiltonian matrix (24) as built of 2×2 submatrices. Calculating the inverse on the space of covalent states (see Eqn. (85) in the appendix) we find an *effective Hamiltonian* which now operates on the covalent states only:

$$H_{\text{eff}}(\varepsilon) = \begin{pmatrix} -t & -t \\ +t & +t \end{pmatrix} \begin{pmatrix} \varepsilon - U & 0 \\ 0 & \varepsilon - U \end{pmatrix}^{-1} \begin{pmatrix} -t & +t \\ -t & +t \end{pmatrix} \approx -\frac{2t^2}{U} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \qquad (25)$$

In the last step we have made an approximation by setting ε to zero, which is roughly the energy of the states with covalent character.

The process of eliminating the ionic states thus gives rise to an effective interaction between the covalent states, which was not present in the original Hamiltonian (24). Diagonalizing the effective Hamiltonian, we find

$$\varepsilon_s = -\frac{4t^2}{U} \quad , \qquad \Psi_s = \frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle - |\downarrow,\uparrow\rangle\Big)$$
$$\varepsilon_t = 0 \qquad , \qquad \Psi_t = \frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle + |\downarrow,\uparrow\rangle\Big)$$

These states correspond to the singlet and triplet states in the hydrogen molecule. Here the singlet-triplet splitting is $2J_{\text{direct}} = -4t^2/U$. The other states in the triplet are those with two electrons of parallel spin: $|\uparrow,\uparrow\rangle$ and $|\downarrow,\downarrow\rangle$. They, of course, also have energy zero, as hopping is impossible due to the Pauli principle.

To understand the nature of the effective interaction in the low-energy Hamiltonian we observe that the off-diagonal matrix elements in (25) correspond to flipping the spin of both electrons ("exchange"). Remembering that

$$\vec{S}_1 \cdot \vec{S}_2 = S_1^z S_2^z + \frac{1}{2}\Big(S_1^+ S_2^- + S_1^- S_2^+\Big) \tag{26}$$

we see that the effective interaction will contain a spin-spin coupling term.

3.3 Second quantization for pedestrians

A systematic way for obtaining the form of the effective interaction is by using second quantization, which will also help us simplify our notation. For a mathematically rigorous introduction see, e.g., [10]. In second quantization we use operators to specify in which orbital an electron is located. As an example, $c^\dagger_{1,\uparrow}$ puts a spin-up electron in orbital φ_1. Denoting the system with no electrons by $|0\rangle$, the basis states that we have considered so far are written as

$$\begin{aligned}
|\uparrow,\cdot\rangle &= c^\dagger_{1\uparrow}|0\rangle \\
|\cdot,\uparrow\rangle &= c^\dagger_{2\uparrow}|0\rangle
\end{aligned}$$

for the single-electron states, and

$$\begin{aligned}
|\uparrow,\downarrow\rangle &= c^\dagger_{2\downarrow}c^\dagger_{1\uparrow}|0\rangle \\
|\downarrow,\uparrow\rangle &= c^\dagger_{2\uparrow}c^\dagger_{1\downarrow}|0\rangle \\
|\updownarrow,\cdot\rangle &= c^\dagger_{1\downarrow}c^\dagger_{1\uparrow}|0\rangle \\
|\cdot,\updownarrow\rangle &= c^\dagger_{2\downarrow}c^\dagger_{2\uparrow}|0\rangle
\end{aligned} \tag{27}$$

for the two-electron states. In order to describe the hopping of an electron from one orbital to another, we introduce operators that annihilate an electron. For example $c_{1\uparrow}$ removes a spin-up electron from orbital φ_1. The hopping of an up electron from φ_1 to φ_2 is thus described by the operator $c^\dagger_{2\uparrow}c_{1\uparrow}$ that first takes an electron out of orbital 1 and then creates one in orbital 2. The Hamiltonian for a spin-up electron hopping between two orbitals can thus be written as

$$H = -t\Big(c^\dagger_{1\uparrow}c_{2\uparrow} + c^\dagger_{2\uparrow}c_{1\uparrow}\Big)\,. \tag{28}$$

Calculating the matrix elements with the single-electron basis states, we recover the matrix (21). For the calculation we need to know that the operators that describe the electrons *anticommute*. This reflects the fact that a many-electron wave function changes sign when two electrons are exchanged. Using the notation $\{a,b\} = ab + ba$ we have

$$\Big\{c_{i\sigma},\, c_{j\sigma'}\Big\} = 0 \qquad \Big\{c^\dagger_{i\sigma},\, c^\dagger_{j\sigma'}\Big\} = 0 \qquad \Big\{c_{i\sigma},\, c^\dagger_{j\sigma'}\Big\} = \delta_{i,j}\delta_{\sigma,\sigma'}$$

Moreover, trying to annihilate an electron in a state where there is no electron, results in zero: $c_{i\sigma}|0\rangle = 0$. Finally, as the notation implies, $c^\dagger_{i\sigma}$ is the adjoint of $c_{i\sigma}$ and $\langle 0|0\rangle = 1$.

To describe the Coulomb repulsion between two electrons in the same orbital we use that $n_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$ returns 0 when operating on a basis state with no spin-σ electron in orbital φ_i, and has eigenvalue 1 for a basis state with a spin-σ electron in orbital φ_i. It is thus called the occupation-number operator. The Coulomb repulsion in orbital φ_1 is then described by the operator $U n_{1\uparrow} n_{1\downarrow}$, which is non-zero only when there is a spin-up *and* a spin-down electron in φ_1. The Hamiltonian for our two-orbital model, where both up- and down-spin electrons can hop, and including the Coulomb repulsion for two electrons in the same orbital, is thus given by

$$\begin{aligned} H &= -t\left(c^\dagger_{1\uparrow}c_{2\uparrow} + c^\dagger_{2\uparrow}c_{1\uparrow} + c^\dagger_{1\downarrow}c_{2\downarrow} + c^\dagger_{2\downarrow}c_{1\downarrow}\right) + U\left(n_{1\uparrow}n_{1\downarrow} + n_{2\uparrow}n_{2\downarrow}\right) \\ &= -t\sum_{i,j,\sigma} c^\dagger_{j\sigma}c_{i\sigma} + U\sum_i n_{i\uparrow}n_{i\downarrow}\,. \end{aligned} \tag{29}$$

You should convince yourself that when you calculate the matrix elements for the two-electron states, you recover the matrix (24). The great advantage of writing the Hamiltonian in second-quantized form is that it is valid for any number of electrons, while the matrix form is restricted to a particular number of electrons.

Coming back to the effective Hamiltonian (25), we can rewrite H_{eff} in second quantized form:

$$\begin{aligned} H_{\text{eff}} &= -\frac{2t^2}{U}\left(c^\dagger_{2\uparrow}c^\dagger_{1\downarrow}c_{1\downarrow}c_{2\uparrow} - c^\dagger_{2\downarrow}c^\dagger_{1\uparrow}c_{1\downarrow}c_{2\uparrow} - c^\dagger_{2\uparrow}c^\dagger_{1\downarrow}c_{1\uparrow}c_{2\downarrow} + c^\dagger_{2\downarrow}c^\dagger_{1\uparrow}c_{1\uparrow}c_{2\downarrow}\right) \\ &= -\frac{2t^2}{U}\left(c^\dagger_{1\downarrow}c_{1\downarrow}c^\dagger_{2\uparrow}c_{2\uparrow} - c^\dagger_{1\uparrow}c_{1\downarrow}c^\dagger_{2\downarrow}c_{2\uparrow} - c^\dagger_{1\downarrow}c_{1\uparrow}c^\dagger_{2\uparrow}c_{2\downarrow} + c^\dagger_{1\uparrow}c_{1\uparrow}c^\dagger_{2\downarrow}c_{2\downarrow}\right) \end{aligned} \tag{30}$$

Looking at equation (90) in the appendix we see that the spin operators are given in second quantization by

$$S^x_i = \frac{1}{2}\left(c^\dagger_{i\uparrow}c_{i\downarrow} + c^\dagger_{i\downarrow}c_{i\uparrow}\right) \qquad S^y_i = -\frac{i}{2}\left(c^\dagger_{i\uparrow}c_{i\downarrow} - c^\dagger_{i\downarrow}c_{i\uparrow}\right) \qquad S^z_i = \frac{1}{2}\left(n_{i\uparrow} - n_{i\downarrow}\right). \tag{31}$$

From this we find (after some calculation) that the effective Hamiltonian can be written in terms of the spin operators

$$H_{\text{eff}} = \frac{4t^2}{U}\left(\vec{S}_1 \cdot \vec{S}_2 - \frac{n_1\, n_2}{4}\right). \tag{32}$$

To conclude, we again find that the completely spin-independent Hamiltonian (29), in the limit of large U, gives rise to a spin-spin interaction. Since the exchange coupling $J = 4t^2/U$ is positive, states with antiparallel spins have lower energy. Thus direct exchange leads to *antiferromagnetism*.

It is important to realize that the singlet-triplet splitting for the effective Hamiltonian really arises from the admixture of ionic states into the singlet. By downfolding we eliminate the high-energy ionic states, i.e., charge fluctuations, from our Hilbert space. The eliminated states then give rise to an effective spin-spin interaction on the new reduced low-energy Hilbert space. We must therefore keep in mind that, when working with the effective Hamiltonian (32), we are considering slightly different states than when working with the original Hamiltonian (29).

3.4 Mean-field treatment

To conclude our discussion of the simplest kinetic exchange mechanism, it is instructive to consider the results of a mean-field treatment. For the two-electron Hamiltonian (24) it is straightforward to find the Hartree-Fock solution by directly minimizing the energy expectation value for a two-electron Slater determinant. The most general ansatz is a Slater determinant constructed from an orbital $\varphi(\theta_\uparrow) = \sin(\theta_\uparrow)\,\varphi_1 + \cos(\theta_\uparrow)\,\varphi_2$ for the spin-up, and $\varphi(\theta_\downarrow) = \sin(\theta_\downarrow)\,\varphi_1 + \cos(\theta_\downarrow)\,\varphi_2$ for the spin-down electron:

$$|\Psi(\theta_\uparrow,\theta_\downarrow)\rangle = \Big(\sin(\theta_\downarrow)\,c^\dagger_{1\downarrow} + \cos(\theta_\downarrow)\,c^\dagger_{2\downarrow}\Big)\;\Big(\sin(\theta_\uparrow)\,c^\dagger_{1\uparrow} + \cos(\theta_\uparrow)\,c^\dagger_{2\uparrow}\Big)\;|0\rangle\;. \tag{33}$$

Translating the second quantized states via (27) into the basis used for writing the Hamiltonian matrix (24), we find the expectation value

$$\begin{aligned}\langle\Psi(\theta_\uparrow,\theta_\downarrow)|H|\Psi(\theta_\uparrow,\theta_\downarrow)\rangle \;=\; & -2t\,(\sin\theta_\uparrow\sin\theta_\downarrow + \cos\theta_\uparrow\cos\theta_\downarrow)\,(\cos\theta_\uparrow\sin\theta_\downarrow + \sin\theta_\uparrow\cos\theta_\downarrow)\\ & +U\left(\sin^2\theta_\uparrow\sin^2\theta_\downarrow + \cos^2\theta_\uparrow\cos^2\theta_\downarrow\right)\;.\end{aligned} \tag{34}$$

If the Slater determinant respects the mirror symmetry of the H_2 molecule, it follows that the Hartree-Fock orbitals for both spins are the bonding state φ_+ ($\theta = \pi/4$). This is the *restricted Hartree-Fock* solution. The corresponding energy is $E(\pi/4,\pi/4) = -2t + U/2$. The excited states are obtained by replacing occupied orbitals φ_+ with φ_-. Altogether we obtain the restricted Hartree-Fock spectrum

$$\begin{aligned} E(\;\pi/4,\;\pi/4) &= -2t + U/2\\ E(\;\pi/4,-\pi/4) &= \;\;\;U/2\\ E(-\pi/4,\;\pi/4) &= \;\;\;U/2\\ E(-\pi/4,-\pi/4) &= \;\;2t + U/2\end{aligned} \tag{35}$$

Comparing to the energy for a state with both electrons of the same spin ($E = 0$), we see that there is no spin-triplet, i.e., Hartree-Fock breaks the spin symmetry. The states (35) are *spin-contaminated* [11]. Even worse, the Hartree-Fock ground state, and consequently all the states, are independent of U. The weight of the ionic states is always $1/2$, leading to an increase of the energy with $U/2$.

To avoid this, we can allow the Hartree-Fock solution to break the symmetry of the molecule (*unrestricted Hartree-Fock*), putting, e.g., more of the up-spin electron in the orbital on site 1 and more of the down-spin electron in orbital 2. For $U < 2t$ this does not lead to a state of lower energy. For larger U there is a symmetry-broken ground state

$$\Psi_{UHF} = \Psi(\theta,\pi/2-\theta)\;\;\text{with}\quad \theta(U) = \frac{\pi}{4} \pm \frac{1}{2}\arccos\left(\frac{2t}{U}\right)\;. \tag{36}$$

Its energy is $E_{UHF} = -2t^2/U$. This looks similar to the singlet energy ε_s, however, with a different prefactor. Still there is no triplet state (spin contamination) and, for $U \to \infty$, the overlap with the true singlet ground state goes to $|\langle\Psi_{UHF}|\Psi_-\rangle|^2 = 1/2$. In an extended system the breaking of the symmetry implies long-range order.

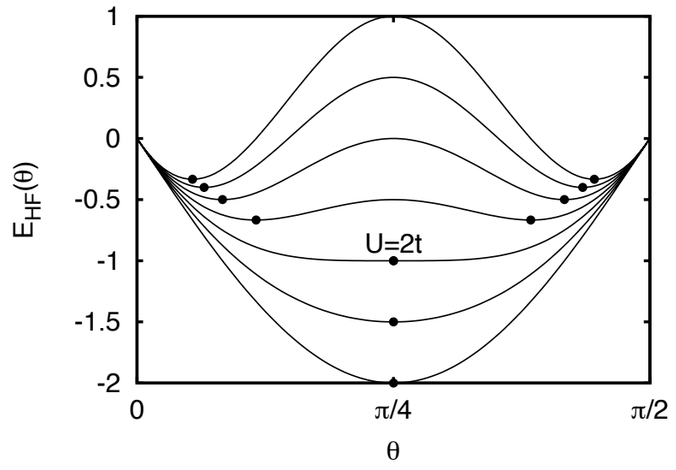


Fig. 4: *Energy expectation value for a Slater determinant $\Psi(\theta, \pi/2-\theta)$ for $U=0,\ t,\ 2t,\ \ldots,\ 6t$. When $U \leq 2t$ the minimum is at $\theta = \pi/4$. This is the Hartree-Fock solution with the bonding orbitals φ_+ occupied. For $U \geq 2t$, $\theta = \pi/4$ is still an extremal point (restricted Hartree-Fock solution), but an energy minimum is only attained when the symmetry is broken (unrestricted Hartree-Fock solution).*

3.5 Superexchange

For the direct exchange mechanism discussed above, it is crucial that there is hopping between the orbitals. These orbitals are typically localized d-orbitals of transition-metals. However, direct exchange cannot explain the antiferromagnetism of most transition-metal compounds: Since the d-orbitals are so localized, hopping can only occur between orbitals on different atoms that are very close to each other. But most antiferromagnetic insulators are transition-metal *oxides*, so that the transition-metal cations are separated by large oxygen anions. In such a situation, shown in figure 5, direct hopping between the d-orbitals is very unlikely. The concept of direct exchange can, however, be extended to these cases by taking into account hopping via the intermediate p-orbital. This mechanism is called superexchange.

To understand superexchange, we consider two d-orbitals with an oxygen p-orbital in-between. We introduce the operator $c^\dagger_{i\sigma}$, which creates a spin-σ electron in the d-orbital at site i, where $i = 1$ denotes the d-orbital on the left and $i = 2$ the one on the right (see Fig. 5). Likewise $c^\dagger_{p\sigma}$ creates an electron in the p-orbital. The energy of an electron in a d- or p-orbital is ε_d and ε_p, respectively. The Coulomb repulsion between two electrons in a d-orbital is U_d, while we neglect the repulsion between electrons in the p-orbital. Finally, $-t_{pd}$ is the hopping between p and d orbitals. The Hamiltonian for the system of figure 5 is then given by

$$H = \sum_\sigma \left(\varepsilon_d \sum_i n_{i\sigma} + \varepsilon_p \, n_{p\sigma} - t_{pd} \sum_i \left(c^\dagger_{i\sigma} c_{p\sigma} + c^\dagger_{p\sigma} c_{i\sigma} \right) \right) + U_d \sum_i n_{i\uparrow} n_{i\downarrow} \, . \tag{37}$$

In the absence of hopping, the ground state will have singly occupied d-orbitals, corresponding

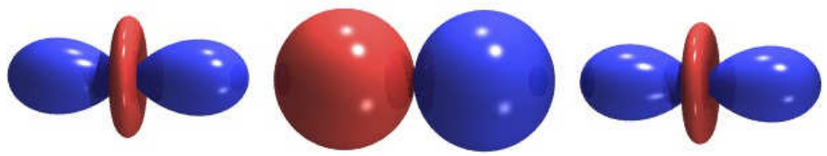


Fig. 5: *In superexchange an oxygen p-orbital mediates the exchange interaction between two transition-metal d-orbitals.*

to a positively charged transition-metal ion, and a doubly occupied p-orbital, corresponding to an O^{2-} ion. To study a possible coupling between the spins on the d-orbitals, we first look at the case where both d-spins point upwards (see the far right of Fig. 6). The Hamiltonian matrix in the corresponding Hilbert space is then given by

$$H = \left(\begin{array}{c|cc} 0 & t_{pd} & t_{pd} \\ \hline t_{pd} & U_d + \Delta_{pd} & 0 \\ t_{pd} & 0 & U_d + \Delta_{pd} \end{array}\right) \begin{array}{l} c^\dagger_{2\uparrow} c^\dagger_{p\downarrow} c^\dagger_{p\uparrow} c^\dagger_{1\uparrow} |0\rangle \\ c^\dagger_{2\uparrow} c^\dagger_{p\uparrow} c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} |0\rangle \\ c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} c^\dagger_{p\uparrow} c^\dagger_{1\uparrow} |0\rangle \end{array} \tag{38}$$

where we have chosen $2(\varepsilon_p + \varepsilon_d)$ as the zero of our energy scale and defined $\Delta_{pd} = \varepsilon_d - \varepsilon_p$. The basis states of the Hilbert space are given on the right and the lines indicate the partitioning of the Hilbert space for downfolding. The effective Hamiltonian for parallel spins on d-orbitals is then

$$H_{\text{eff}} = (t_{pd}, t_{pd}) \begin{pmatrix} \varepsilon - (U_d + \Delta_{pd}) & 0 \\ 0 & \varepsilon - (U_d + \Delta_{pd}) \end{pmatrix} \begin{pmatrix} t_{pd} \\ t_{pd} \end{pmatrix} \approx -\frac{2t_{pd}^2}{U_d + \Delta_{pd}} \tag{39}$$

where in the last step we have set ε to zero.

For antiparallel spins the Hilbert space is nine-dimensional. We sort the basis states into groups that are connected by the hopping of one electron. Starting from the two states with singly occupied d-orbitals, the second group has one of the p-electrons transferred to a d-orbital, leading to one doubly occupied d, while the last group has a second electron hopped, leading to either an empty p- or an empty d-orbital. The corresponding Hamiltonian matrix is

$$\left(\begin{array}{cc|cccc|ccc} 0 & 0 & +t_{pd} & +t_{pd} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & +t_{pd} & +t_{pd} & 0 & 0 & 0 \\ \hline +t_{pd} & 0 & U_d + \Delta_{pd} & 0 & 0 & 0 & -t_{pd} & 0 & -t_{pd} \\ +t_{pd} & 0 & 0 & U_d + \Delta_{pd} & 0 & 0 & 0 & -t_{pd} & -t_{pd} \\ 0 & +t_{pd} & 0 & 0 & U_d + \Delta_{pd} & 0 & +t_{pd} & 0 & +t_{pd} \\ 0 & +t_{pd} & 0 & 0 & 0 & U_d + \Delta_{pd} & 0 & +t_{pd} & +t_{pd} \\ \hline 0 & 0 & -t_{pd} & 0 & +t_{pd} & 0 & U_d & 0 & 0 \\ 0 & 0 & 0 & -t_{pd} & 0 & +t_{pd} & 0 & U_d & 0 \\ 0 & 0 & -t_{pd} & -t_{pd} & +t_{pd} & +t_{pd} & 0 & 0 & 2(U_d + \Delta_{pd}) \end{array}\right) \begin{array}{l} c^\dagger_{2\downarrow} c^\dagger_{p\downarrow} c^\dagger_{p\uparrow} c^\dagger_{1\uparrow} |0\rangle \\ c^\dagger_{2\uparrow} c^\dagger_{p\downarrow} c^\dagger_{p\uparrow} c^\dagger_{1\downarrow} |0\rangle \\ c^\dagger_{2\downarrow} c^\dagger_{p\uparrow} c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} |0\rangle \\ c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} c^\dagger_{p\downarrow} c^\dagger_{1\uparrow} |0\rangle \\ c^\dagger_{2\uparrow} c^\dagger_{p\downarrow} c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} |0\rangle \\ c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} c^\dagger_{p\uparrow} c^\dagger_{1\downarrow} |0\rangle \\ c^\dagger_{p\downarrow} c^\dagger_{p\uparrow} c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} |0\rangle \\ c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} c^\dagger_{p\downarrow} c^\dagger_{p\uparrow} |0\rangle \\ c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} |0\rangle \end{array}$$

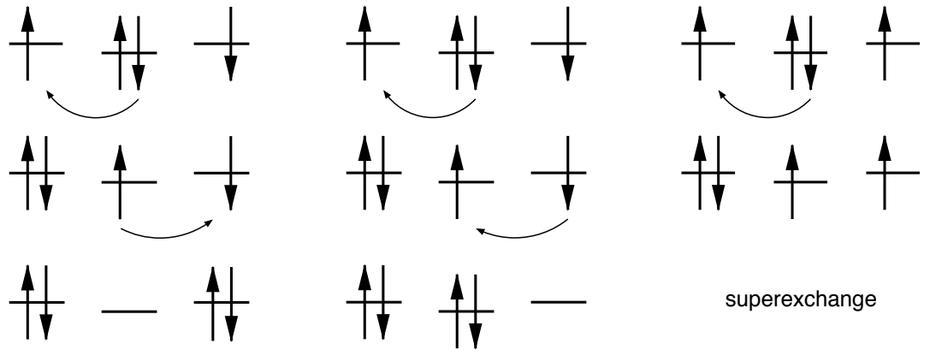


Fig. 6: *Simple picture of superexchange. Here the orbital on the central site is different from the orbitals on the sides. Typically, in the center there is an oxygen p-orbital coupling two d-orbitals. This situation is illustrated in Fig. 5. For antiparallel spins on the d-orbitals there are two ways that two consecutive hopping processes are possible. For parallel spins the Pauli principle suppresses the second hopping process.*

Downfolding the high energy states with at least one doubly occupied d-orbital, setting $\varepsilon = 0$ and expanding in $1/U_d$ (remembering $(A + \Delta)^{-1} \approx A^{-1}(1 - \Delta A^{-1})$), which is equivalent to second-order perturbation theory, leads to

$$\begin{aligned} H_{\text{eff}} &= H_{00} + T_{01}\left(\varepsilon - \left(H_{11} + T_{12}\left(\varepsilon - H_{22}\right)^{-1} T_{21}\right)\right)^{-1} T_{10} \\ &\approx H_{00} - T_{01}H_{11}^{-1}T_{10} - T_{01}H_{11}^{-1}T_{12}H_{22}^{-1}T_{21}H_{11}^{-1}T_{10} \qquad (40) \\ &= -\frac{2t_{pd}^2}{U_d + \Delta_{pd}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \frac{2t_{pd}^4}{(U_d + \Delta_{pd})^2}\left(\frac{1}{U_d} + \frac{1}{U_d + \Delta_{pd}}\right)\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \qquad (41) \end{aligned}$$

The first term is the same as for parallel spins (39). The additional term is of the same type as that found for the direct exchange mechanism. Again, it can be written in terms of spin operators. In the present case they are the spin operators for the d-orbitals, while the p-orbital does no longer appear in the spin Hamiltonian. The spin coupling is now given by

$$J = \frac{4t_{pd}^4}{(U_d + \Delta_{pd})^2}\left(\frac{1}{U_d} + \frac{1}{U_d + \Delta_{pd}}\right), \qquad (42)$$

which reflects that the superexchange mechanism involves four hopping processes (see Fig. 6), while direct exchange only involves two hoppings (see Fig. 3). The hopping process involving only a single doubly occupied d-orbital (middle of Fig. 6) is a generalization of the simple direct exchange with an effective hopping $t_{\text{eff}} = t_{pd}^2/(U_d + \Delta_{pd})$ between the d-orbitals and gives the first term, $4t_{\text{eff}}^2/U_d$, in (42), while the hopping process involving two occupied d-orbitals (left in Fig. 6) gives the second term $4t_{pd}^4/(U_d + \Delta_{pd})^3$.

3.6 Ferromagnetic superexchange

In the discussion of superexchange we have, so far, assumed that the oxygen ion lies between the two d-orbitals. This 180° geometry is shown on the left of Fig. 7. The situation is quite different, when the oxygen forms a 90° bridge between the two d-orbitals, see the right of Fig. 7. By symmetry, there is only hopping between the d- and the p-orbital that point towards each other (cf. the Slater-Koster integrals). As there is also no hopping between the p-orbitals on the same site, the Hamiltonian for the system separates into two parts, one involving only the d orbital on site 1 and the p_x orbital and the other only involving d on site 2 and p_y, e.g.:

$$H_1 = \begin{pmatrix} 0 & +t_{pd} \\ +t_{pd} & U_d + \Delta_{pd} \end{pmatrix} \quad \begin{matrix} c^\dagger_{x\downarrow} c^\dagger_{x\uparrow} c^\dagger_{1\downarrow} |0\rangle \\ c^\dagger_{x\downarrow} c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} |0\rangle \end{matrix} \tag{43}$$

Since it is not possible for an electron on site 1 to reach site 2, none of the superexchange processes discussed above are operational. Nevertheless, the energy for the system depends on the relative orientation of the electron spins in the two d-orbitals. To see this, we have to remember that Coulomb exchange prefers a triplet for two electrons in different orbitals on the same site (Hund's first rule). Including J_{xy} on the oxygen (but neglecting U_p for simplicity), we get, for the triplet state with two up-electrons, the Hamiltonian (note that there is no Hund's rule term for the states with three electrons, i.e. one hole, on the two oxygen orbitals p_x and p_y)

$$\left(\begin{array}{c|ccc} 0 & t_{pd} & t_{pd} & 0 \\ \hline t_{pd} & U_d + \Delta_{pd} & 0 & t_{pd} \\ t_{pd} & 0 & U_d + \Delta_{pd} & t_{pd} \\ \hline 0 & t_{pd} & t_{pd} & 2(U_d + \Delta_{pd}) - J_{xy} \end{array}\right) \quad \begin{matrix} c^\dagger_{1\uparrow} c^\dagger_{x\downarrow} c^\dagger_{x\uparrow} c^\dagger_{y\downarrow} c^\dagger_{y\uparrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} c^\dagger_{x\uparrow} c^\dagger_{y\downarrow} c^\dagger_{y\uparrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\uparrow} c^\dagger_{x\downarrow} c^\dagger_{x\uparrow} c^\dagger_{y\uparrow} c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} c^\dagger_{x\uparrow} c^\dagger_{y\uparrow} c^\dagger_{2\downarrow} c^\dagger_{2\downarrow} |0\rangle \end{matrix} . \tag{44}$$

The first state has the two up-electrons on the d-orbitals. The second group of states has one d-orbital doubly occupied, while the last state has both d doubly occupied, i.e., two electrons on the two p-orbitals – the situation discussed in Sec. 2. Calculating the effective Hamiltonian as in (40) gives the energy of the triplet state

$$H_{\text{eff}} = -\frac{2t_{pd}^2}{U_d + \Delta_{pd}} - \frac{4t_{pd}^4}{(U_d + \Delta_{pd})^2} \frac{1}{2(U_d + \Delta_{pd}) - J_{xy}} . \tag{45}$$

Starting from singly occupied d orbitals with opposite spin, we obtain

$$\left(\begin{array}{cc|cccc|cc} 0 & 0 & t_{pd} & 0 & t_{pd} & 0 & 0 & 0 \\ 0 & 0 & 0 & t_{pd} & 0 & t_{pd} & 0 & 0 \\ \hline t_{pd} & 0 & U_d + \Delta_{pd} & 0 & 0 & 0 & t_{pd} & 0 \\ 0 & t_{pd} & 0 & U_d + \Delta_{pd} & 0 & 0 & 0 & t_{pd} \\ t_{pd} & 0 & 0 & 0 & U_d + \Delta_{pd} & 0 & t_{pd} & 0 \\ 0 & t_{pd} & 0 & 0 & 0 & U_d + \Delta_{pd} & 0 & t_{pd} \\ \hline 0 & 0 & t_{pd} & 0 & t_{pd} & 0 & 2(U_d + \Delta_{pd}) & -J_{xy} \\ 0 & 0 & 0 & t_{pd} & 0 & t_{pd} & -J_{xy} & 2(U_d + \Delta_{pd}) \end{array}\right) \quad \begin{matrix} c^\dagger_{1\uparrow} c^\dagger_{x\downarrow} c^\dagger_{x\uparrow} c^\dagger_{y\downarrow} c^\dagger_{y\uparrow} c^\dagger_{2\downarrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{x\downarrow} c^\dagger_{x\uparrow} c^\dagger_{y\downarrow} c^\dagger_{y\uparrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} c^\dagger_{x\uparrow} c^\dagger_{y\downarrow} c^\dagger_{y\uparrow} c^\dagger_{2\downarrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} c^\dagger_{x\downarrow} c^\dagger_{y\downarrow} c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\uparrow} c^\dagger_{x\downarrow} c^\dagger_{x\uparrow} c^\dagger_{y\uparrow} c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{x\downarrow} c^\dagger_{x\uparrow} c^\dagger_{y\uparrow} c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} c^\dagger_{x\uparrow} c^\dagger_{y\downarrow} c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} |0\rangle \\ c^\dagger_{1\downarrow} c^\dagger_{1\uparrow} c^\dagger_{x\downarrow} c^\dagger_{y\uparrow} c^\dagger_{2\downarrow} c^\dagger_{2\uparrow} |0\rangle \end{matrix}$$

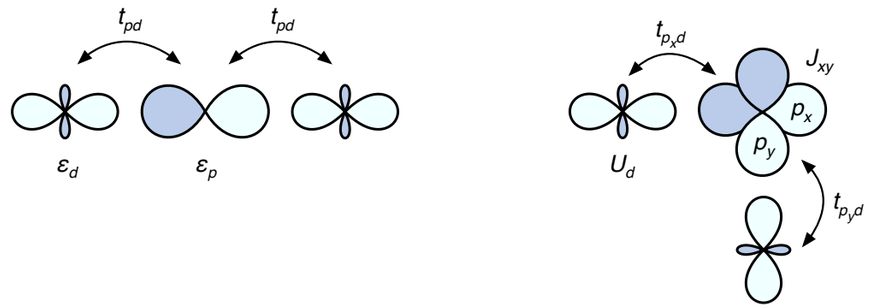


Fig. 7: *Dependence of superexchange on geometry: When the d-orbitals interact via an oxygen in-between (the 180° geometry shown on the left), both d-orbitals couple to the same p-orbital, while the hopping to the two other p-orbitals vanishes by symmetry. The result is antiferromagnetic superexchange. When the angle of the M-O-M group is 90° (right), the d-orbitals couple to orthogonal p-orbitals, making it impossible for an electron on one d-orbital to reach the d-orbital on the other site. In this case, superexchange is mediated via the Coulomb exchange on the connecting oxygen.*

giving the effective Hamiltonian

$$H_{\text{eff}} = -\frac{2t_{pd}^2}{U_d+\Delta_{pd}}\begin{pmatrix}1&0\\0&1\end{pmatrix} - \frac{4t_{pd}^4}{(U_d+\Delta_{pd})^2}\frac{1}{4(U_d+\Delta_{pd})^2-J_{xy}^2}\begin{pmatrix}2(U_d+\Delta_{pd}) & +J_{xy}\\ +J_{xy} & 2(U_d+\Delta_{pd})\end{pmatrix}.$$

Rearranging the matrices, we can bring this to the canonical form

$$\begin{aligned} H_{\text{eff}} = &-\left(\frac{2t_{pd}^2}{U_d+\Delta_{pd}} + \frac{4t_{pd}^4}{(U_d+\Delta_{pd})^2}\frac{1}{2(U_d+\Delta_{pd})-J_{xy}}\right) \\ &+ \frac{4t_{pd}^4}{(U_d+\Delta_{pd})^2}\frac{J_{xy}}{4(U_d+\Delta_{pd})^2-J_{xy}^2}\begin{pmatrix}1&-1\\-1&1\end{pmatrix}. \end{aligned} \tag{46}$$

The first term is just the energy of the triplet state (45). The second gives the difference in energy to the singlet. Despite the fact that the electrons cannot be transferred between the d orbitals we thus get a singlet-triplet splitting. This coupling of the spins originates from the states with both d-orbitals doubly occupied: the two remaining electrons, one each on the p_x- and p_y-orbital, respectively, form a triplet of energy $2J_{xy}$ lower than that of the singlet (see Eqn. (15)). When the electrons hop back from the d-orbital, the entanglement of the spins is transferred to the remaining electron on the d. Originating from the Coulomb exchange on the oxygen, the exchange coupling is ferromagnetic

$$J = -\frac{4t_{pd}^4}{(U_d+\Delta_{pd})^2}\frac{2J_{xy}}{4(U_d+\Delta_{pd})^2-J_{xy}^2}. \tag{47}$$

It tends to be significantly weaker than the antiferromagnetic 180° superexchange coupling (42). When the angle of the M-O-M group is larger than 90°, hopping to both p-orbitals becomes possible according to the Slater-Koster rules and the antiferromagnetic superexchange processes of Fig. 6 start to compete with the ferromagnetic superexchange mediated by the Coulomb exchange on the oxygen. This is one basis of the Goodenough-Kanamori rules [7, 12].

4 Double exchange

Double exchange takes its name from the fact that it results from a combination of Coulomb- and kinetic-exchange. In that sense the 90° superexchange mechanism discussed above is a double exchange mechanism. More commonly, double exchange is encountered in mixed-valence compounds. So far we have considered systems with an integer number of electrons per site. When correlations are strong the lowest energy state will essentially have the same number of electrons on every site and hopping will be strongly suppressed by the Coulomb repulsion energy U as we have seen for the simple two-site model of kinetic exchange. In a mixed valence system the number of electrons per site is non-integer, so even for large U some site will have more electrons than others. Thus electrons can hop between such sites without incurring a cost U. Hence these compounds are usually metallic.

As a simple example we consider two sites with two orbitals of the type discussed in Sec. 2. We assume that each site has one electron in orbital a, and that there is only a single electron in the b-orbitals. This electron can hop between the sites via a hopping matrix element t_{bb}. The situation is illustrated in Fig. 8.

When all three spins are up, $S^z_{\text{tot}} = 3/2$, we have a simple 2×2 Hamiltonian, taking U_{ab} as our zero of energy

$$H = \begin{pmatrix} -J_{ab} & -t_{bb} \\ -t_{bb} & -J_{ab} \end{pmatrix} . \tag{48}$$

The eigenstates are the bonding/antibonding linear combinations of the Hund's rule triplets. Their dispersion is $\pm t$:

$$\varepsilon_\pm = -J_{ab} \pm t_{bb} . \tag{49}$$

We see that the hopping couples the two sites into a state with the electrons in the a-orbital in a triplet state:

$$\Psi_\pm = \frac{1}{\sqrt{2}}\Big(|\uparrow,\uparrow\rangle_1 |\cdot,\uparrow\rangle_2 \pm |\cdot,\uparrow\rangle_1 |\uparrow,\uparrow\rangle_2\Big) = \frac{1}{\sqrt{2}}\Big(|\uparrow,\cdot\rangle_b \pm |\cdot,\uparrow\rangle_b\Big) |\uparrow,\uparrow\rangle_a . \tag{50}$$

In the language of quantum information, the hopping electron teleports the local Hund's rule triplet to the a-orbitals.

To obtain the Hamiltonian for the $S^z_{\text{tot}} = 1/2$ states, we arrange the basis states in the order they are connected by matrix elements, see Fig. 8. We obtain the tridiagonal Hamiltonian

$$H = \begin{pmatrix} -J_{ab} & -t_{bb} & 0 & 0 & 0 & 0 \\ -t_{bb} & 0 & -J_{ab} & 0 & 0 & 0 \\ 0 & -J_{ab} & 0 & -t_{bb} & 0 & 0 \\ 0 & 0 & -t_{bb} & 0 & -J_{ab} & 0 \\ 0 & 0 & 0 & -J_{ab} & 0 & -t_{bb} \\ 0 & 0 & 0 & 0 & -t_{bb} & -J_{ab} \end{pmatrix} \tag{51}$$

The ground-state is the equally weighted linear combination of all basis states. It has energy $\varepsilon = -J_{ab} - t_{bb}$ and belongs to the sector with $S_{\text{tot}} = 3/2$. Again, the hopping electron teleports

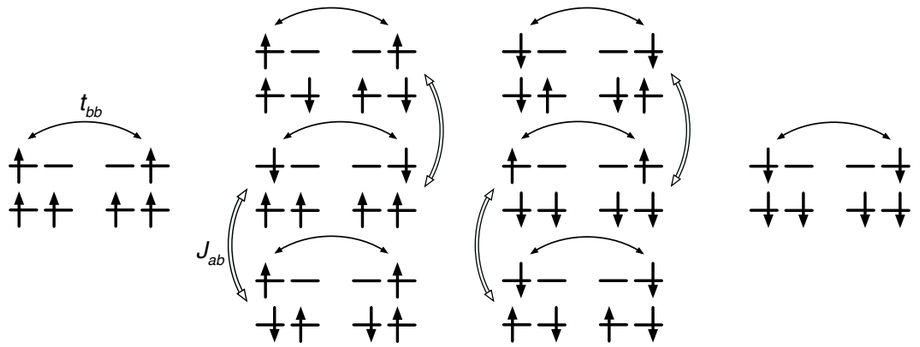


Fig. 8: *Matrix elements entering the double-exchange Hamiltonian. Hopping matrix elements t_{bb} are indicated as double arrows, Coulomb-exchange matrix elements J_{ab} as double lines. Note that the right half of the states are obtained from the left by flipping all spins.*

the triplets from the sites into a triplet state of the spins in the a-orbitals:

$$\frac{1}{\sqrt{6}}\Big(|\uparrow,\uparrow\rangle_1|\cdot,\downarrow\rangle_2+|\cdot,\uparrow\rangle_1|\uparrow,\downarrow\rangle_2+|\cdot,\uparrow\rangle_1|\downarrow,\uparrow\rangle_2+|\downarrow,\uparrow\rangle_1|\cdot,\uparrow\rangle_2+|\uparrow,\downarrow\rangle_1|\cdot,\uparrow\rangle_2+|\cdot,\downarrow\rangle_1|\uparrow,\uparrow\rangle_2\Big)$$
$$= \frac{1}{\sqrt{2}}\Big(|\uparrow,\cdot\rangle_b+|\cdot,\uparrow\rangle_b\Big)\,\frac{1}{\sqrt{2}}\Big(|\uparrow,\downarrow\rangle_a+|\downarrow,\uparrow\rangle_a\Big)+\frac{1}{\sqrt{2}}\Big(|\downarrow,\cdot\rangle_b+|\cdot,\downarrow\rangle_b\Big)\,|\uparrow,\uparrow\rangle_a$$

As in the $S^z_{\rm tot}=3/2$-sector, there is a corresponding eigenstate of energy $\varepsilon=-J_{ab}+t_{bb}$ with the b-electron antibonding. Again, we find that the triplet state is centered at $-J_{ab}$ with dispersion $\pm t_{bb}$. Thus the hopping electron in orbital b tends to align the spins in orbital a.
While the total spin is conserved, this is not true for the spin on site i, $\vec{S}_{i,a}+\vec{S}_{i,b}$ or for the spin in the a-orbitals $\vec{S}_{1a}+\vec{S}_{2a}$. Consequently the hopping mixes the Hund's rule singlets and triplets and therefore does not produce a singlet state of the a electrons. Instead, for $t_{bb}\ll J_{ab}$, we find in first order perturbation theory

$$\begin{array}{ll}
-J_{ab}-t_{bb} & \big(\ 1,\ \ 1,\ \ 1,\ \ 1,\ \ 1,\ \ 1\big)^T/\sqrt{6}\\
-J_{ab}-t_{bb}/2 & \big(\ 2,\ \ 1,\ \ 1,-1,-1,-2\big)^T/\sqrt{12}\\
-J_{ab}+t_{bb}/2 & \big(\ 2,-1,-1,-1,-1,\ \ 2\big)^T/\sqrt{12}\\
-J_{ab}+t_{bb} & \big(\ 1,-1,-1,\ \ 1,\ \ 1,-1\big)^T/\sqrt{6}\\
\\
+J_{ab}-t_{bb}/2 & \big(\ 0,\ \ 1,-1,-1,\ \ 1,\ \ 0\big)^T/2\\
+J_{ab}+t_{bb}/2 & \big(\ 0,\ \ 1,-1,\ \ 1,-1,\ \ 0\big)^T/2
\end{array}\qquad (52)$$

While the triplet states, $S_{\rm tot}=3/2$, are centered around $-J_{ab}$ with dispersion $\pm t_{bb}$, states with singlet character are centered at the same energy, but have smaller dispersion, $\pm t_{bb}/2$.
We can look at the situation from a different perspective, focusing on the effect of the spins in the a-orbitals on the hopping electron. This is another source of Goodenough-Kanamori rules [12]. We choose the quantization-axis on site 2 rotated relative to that on site 1 by an angle ϑ. Taking the original quantization axis as $\hat{z}$ and the direction from site 1 to site 2 as $\hat{x}$,

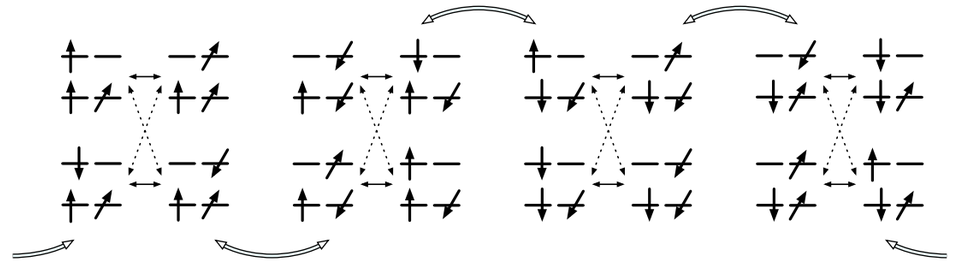


Fig. 9: *With quantization axes tilted between the sites, all states couple. Matrix elements are indicated by arrows: Hopping only couples sites with the same occupation of the a-orbitals. Full lines stand for $t_{bb}\cos(\vartheta/2)$, dotted lines for matrix elements proportional to $t_{bb}\sin(\vartheta/2)$. These states are coupled by off-diagonal Coulomb exchange matrix elements J_{ab}, shown as double lines.*

the rotation in spin space is given by $exp(-i\sigma_y\,\vartheta/2)$ (see appendix C). Introducing operators $d_{2b\sigma}$ in the rotated basis, we have, in terms of the original operators,

$$d_{2b\uparrow} = \cos(\vartheta/2)\,c_{2b\uparrow} - \sin(\vartheta/2)\,c_{2b\downarrow} \tag{53}$$

$$d_{2b\downarrow} = \sin(\vartheta/2)\,c_{2b\uparrow} + \cos(\vartheta/2)\,c_{2b\downarrow} \tag{54}$$

so the hopping becomes

$$-t_{bb}\,c^\dagger_{2b\uparrow}c_{1b\uparrow} = -t_{bb}\left(+\cos(\vartheta/2)\,d^\dagger_{2b\uparrow} + \sin(\vartheta/2)\,d^\dagger_{2b\downarrow}\right)c_{1b\uparrow} \tag{55}$$

$$-t_{bb}\,c^\dagger_{2b\downarrow}c_{1b\downarrow} = -t_{bb}\left(-\sin(\vartheta/2)\,d^\dagger_{2b\uparrow} + \cos(\vartheta/2)\,d^\dagger_{2b\downarrow}\right)c_{1b\downarrow}\,. \tag{56}$$

Obviously, such a change of basis does not change the spectrum of the resulting Hamiltonian. We do get a new situation, however, when we assume that the spin on orbital a is *fixed*. This is, e.g., a good approximation when the spin in the a-orbital arises actually not a from a single electron, but from many electrons coupled by Hund's rule, e.g., in a half-filled t_{2g}-level, like in the manganites. Then there are no off-diagonal exchange terms (double lines in Fig. 9) and the Hamiltonian splits into 4×4 blocks with only hopping (solid and dotted lines in Fig. 9) and on-site Coulomb exchange J_{ab}. The Hamiltonian then becomes

$$H = \left(\begin{array}{cc|cc} -J_{ab} & +t_{bb}\,\cos(\vartheta/2) & +t_{bb}\,\sin(\vartheta/2) & 0 \\ +t_{bb}\,\cos(\vartheta/2) & -J_{ab} & 0 & -t_{bb}\,\sin(\vartheta/2) \\ \hline +t_{bb}\,\sin(\vartheta/2) & 0 & 0 & +t_{bb}\,\cos(\vartheta/2) \\ 0 & -t_{bb}\,\sin(\vartheta/2) & +t_{bb}\,\cos(\vartheta/2) & 0 \end{array}\right), \tag{57}$$

where the a-spin simply produces a Zeeman splitting of orbital b, proportional to the exchange coupling J_{ab}. In the limit $t_{bb} \ll J_{ab}$ we can neglect the states with misaligned spins and obtain

$$\varepsilon = -J_{ab} \pm t_{bb}\,\cos(\vartheta/2)\,, \tag{58}$$

i.e., for parallel spins, $\vartheta = 0$, the gain in kinetic energy is maximized, giving the ground-state energy of the full Hamiltonian, while for anti-parallel spins, $\vartheta = \pi$ the dispersion vanishes.

5 Orbital-ordering

Exchange mechanisms are not restricted to the coupling of spins. As pointed out by Kugel and Khomskii [13], also orbital occupations can interact. Such a coupling leads, besides an ordering of the spins, to an ordering of the orbitals.

To understand the mechanism of orbital-ordering, we consider an e_g-molecule, i.e., two sites with two orbitals a and b, as discussed in Sec. 2. The Hamiltonian on the sites is thus given by (14). In addition, the two sites are coupled by hopping matrix elements t_{aa} and t_{bb}, i.e., hopping does not change the type, a or b, of the occupied orbital. We now consider the case of one electron in orbital a and the other in orbital b.

First, we consider the situation when both electrons have the same spin, e.g., spin-up. The basis states are shown in Fig. 10. Setting up the Hamiltonian is analogous to setting up (24)

$$H^{\uparrow\uparrow} = \left(\begin{array}{cc|cc} 0 & 0 & -t_{bb} & -t_{aa} \\ 0 & 0 & +t_{aa} & +t_{bb} \\ \hline -t_{bb} & +t_{aa} & U_{ab}-J_{ab} & 0 \\ -t_{aa} & +t_{bb} & 0 & U_{ab}-J_{ab} \end{array}\right). \tag{59}$$

Downfolding to the states without doubly occupied sites, we obtain

$$H^{\uparrow\uparrow}_{\text{eff}} \approx -\frac{1}{U_{ab}-J_{ab}} \begin{pmatrix} t_{aa}^2+t_{bb}^2 & -2t_{aa}t_{bb} \\ -2t_{aa}t_{bb} & t_{aa}^2+t_{bb}^2 \end{pmatrix} = -\frac{(t_{aa}-t_{bb})^2}{U_{ab}-J_{ab}} - \frac{2t_{aa}t_{bb}}{U_{ab}-J_{ab}} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \tag{60}$$

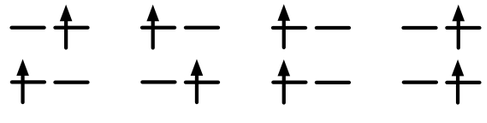
Thus we find that there is an interaction between the states with exchanged orbital-occupation, i.e., an orbital-exchange. For the present case of ferromagnetically aligned spins, the direct orbital exchange coupling favors the orbital singlet, when the hopping matrix elements are of the same sign. In analogy with the situation in kinetic exchange, this is called antiferro orbital exchange. To make the relation with kinetic exchange even more explicit, we can introduce, in analogy to (31), pseudo-spin operators $\vec{T}_{i\sigma}$

$$T^x_{i\sigma} = \frac{1}{2}\left(c^\dagger_{ai\sigma}c_{bi\sigma} + c^\dagger_{bi\sigma}c_{ai\sigma}\right),\ T^y_{i\sigma} = -\frac{i}{2}\left(c^\dagger_{ai\sigma}c_{bi\sigma} - c^\dagger_{bi\sigma}c_{ai\sigma}\right),\ T^z_{i\sigma} = \frac{1}{2}\left(n_{ai\sigma} - n_{bi\sigma}\right) \tag{61}$$

so that we can write

$$H^{\uparrow\uparrow}_{\text{eff}} = -\frac{(t_{aa}-t_{bb})^2}{U_{ab}-J_{ab}} + \frac{4t_{aa}t_{bb}}{U_{ab}-J_{ab}}\left(\vec{T}_{1\uparrow}\cdot\vec{T}_{2\uparrow} - \frac{1}{4}\right). \tag{62}$$

Fig. 10: *Basis states for an up-electron in orbital a and another up-electron in orbital b. Note that the states are ordered as in Eqn. (24).*



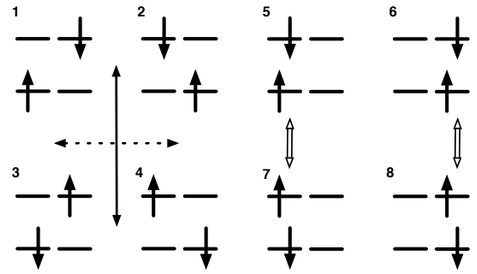


Fig. 11: *Basis states for electrons of opposite spin. The numbering used for the matrix (63) is indicated. Spin exchange is indicated by the full, orbital exchange by the dotted arrow. The states with both electrons on the same site are coupled via Coulomb exchange (double arrows).*

When the two electrons have opposite spin, we can study the interplay of spin- and orbital-exchange. The basis states are shown in Fig. 11. We expect orbital exchange to operate between the first two states in each row and spin exchange between the states between the rows. The Hamiltonian is

$$H^{\uparrow\downarrow} = \left(\begin{array}{cccc|cccc} 0 & 0 & 0 & 0 & -t_{bb} & -t_{aa} & 0 & 0 \\ 0 & 0 & 0 & 0 & +t_{aa} & +t_{bb} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -t_{bb} & -t_{aa} \\ 0 & 0 & 0 & 0 & 0 & 0 & +t_{aa} & +t_{bb} \\ \hline -t_{bb} & +t_{aa} & 0 & 0 & U_{ab} & 0 & -J_{ab} & 0 \\ -t_{aa} & +t_{bb} & 0 & 0 & 0 & U_{ab} & 0 & -J_{ab} \\ 0 & 0 & -t_{bb} & +t_{aa} & -J_{ab} & 0 & U_{ab} & 0 \\ 0 & 0 & -t_{aa} & +t_{bb} & 0 & -J_{ab} & 0 & U_{ab} \end{array}\right) \tag{63}$$

from which we obtain

$$\begin{aligned} H^{\uparrow\downarrow}_{\text{eff}} &\approx -\frac{1}{U_{ab}^2 - J_{ab}^2} \begin{pmatrix} (t_{aa}^2+t_{bb}^2)U_{ab} & -2t_{aa}t_{bb}\,U_{ab} & (t_{aa}^2+t_{bb}^2)J_{ab} & -2t_{aa}t_{bb}\,J_{ab} \\ -2t_{aa}t_{bb}\,U_{ab} & (t_{aa}^2+t_{bb}^2)U_{ab} & -2t_{aa}t_{bb}\,J_{ab} & (t_{aa}^2+t_{bb}^2)J_{ab} \\ (t_{aa}^2+t_{bb}^2)J_{ab} & -2t_{aa}t_{bb}\,J_{ab} & (t_{aa}^2+t_{bb}^2)U_{ab} & -2t_{aa}t_{bb}\,U_{ab} \\ -2t_{aa}t_{bb}\,J_{ab} & (t_{aa}^2+t_{bb}^2)J_{ab} & -2t_{aa}t_{bb}\,U_{ab} & (t_{aa}^2+t_{bb}^2)J_{ab} \end{pmatrix} \\ &= -\frac{1}{U_{ab}^2 - J_{ab}^2} \begin{pmatrix} U_{ab} & J_{ab} \\ J_{ab} & U_{ab} \end{pmatrix} \otimes \begin{pmatrix} t_{aa}^2+t_{bb}^2 & -2t_{aa}t_{bb} \\ -2t_{aa}t_{bb} & t_{aa}^2+t_{bb}^2 \end{pmatrix} \\ &= -\frac{1}{U_{ab}^2 - J_{ab}^2} \left[U_{ab} + J_{ab} - J_{ab} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right] \otimes \left[(t_{aa}-t_{bb})^2 + 2t_{aa}t_{bb} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}\right]. \end{aligned} \tag{64}$$

I.e., we get a simultaneous coupling of the spin- and orbital degrees of freedom. The first term describes the coupling of the spins, which is antiferromagnetic, while the coupling of the orbitals is, for hopping matrix elements of the same sign, ferro, i.e., orbital triplet. In terms of the spin and pseudo-spin operators we can write, with $\vec{T}_i = \sum_\sigma \vec{T}_{i\sigma}$ and $\vec{S}_i = \sum_{\alpha\in\{a,b\}} \vec{S}_{\alpha,i}$

$$H^{\uparrow\downarrow}_{\text{eff}} = -\frac{1}{U_{ab}^2 - J_{ab}^2} \left[(U_{ab}+J_{ab}) + 2J_{ab}\left(\vec{S}_1\cdot\vec{S}_2 - \frac{1}{4}\right)\right] \left[(t_{aa}-t_{bb})^2 - 4t_{aa}t_{bb}\left(\vec{T}_1\cdot\vec{T}_2 - \frac{1}{4}\right)\right].$$

There will be additional terms when we allow states with both electrons in the same orbital.

6 Extended systems

6.1 Hubbard model

We now turn to *extended* systems. For this we consider the Hubbard model [14] on an infinite lattice. Note that now the Hilbert space is infinitely dimensional, so we can no longer write down the Hamiltonian in its matrix form but have to rely on the second quantized form (29)

$$H = -t \sum_{i,j,\sigma} c^\dagger_{j\sigma} c_{i\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow} \,. \tag{65}$$

As in our toy model we still assume that each atom has only a single relevant orbital. There are links between the neighboring atoms with matrix elements t, which can be intuitively interpreted as hopping from site to site. In the absence of other terms the hopping gives rise to a band. A second energy scale is given by the Coulomb repulsion U between two electrons on the same atom. If this on-site Coulomb repulsion is comparable to or even larger than the band width, the electrons can no longer be considered independent; since the double occupation of an atom is energetically very costly, the movement of an electron will be hindered by the Coulomb repulsion. One says that the electrons move in a *correlated* way. We should note that also the Pauli principle hinders the movement of an electron. This effect can, however, be efficiently described by constructing a Slater determinant of independent-electron wave functions. Correlations, on the other hand, are notoriously difficult to describe since no simple wave functions for such systems are available. In the case of strong correlations, i.e., for $U \gg t$, we will treat the hopping as a perturbation. This is called the *atomic limit*, since the sites behave as almost independent atoms. Thus it is most appropriate to describe strongly correlated electrons in a local picture, i.e., in terms of electron configurations, which are the states that diagonalize the Coulomb term.

6.2 Mott transition

The physics described by the Hubbard model is the interplay between kinetic energy and Coulomb repulsion. Writing the Hubbard-Hamiltonian either in real or in k-space

$$\begin{aligned} H = \;& -t \sum_{i,j,\sigma} c^\dagger_{j\sigma} c_{i\sigma} && + U \sum_i n_{i\uparrow} n_{i\downarrow} \\ = \;& \sum_{k\sigma} \varepsilon_k \, c^\dagger_{k\sigma} c_{k\sigma} && + \frac{U}{M} \sum_{k,k',q} c^\dagger_{k\uparrow} c_{k-q\uparrow} c^\dagger_{k'\downarrow} c_{k'+q\downarrow} \,, \end{aligned}$$

where M is the number of lattice sites, we see that there are obviously two limiting cases: There is the non-interacting- or band-limit, when $t \gg U$. In that case, only the hopping term survives, i.e., there are no interactions, and the Hamiltonian can be solved easily in k-space. The energy levels then form a band and the system is metallic, except when the band is completely filled. In the opposite case, the atomic limit, the interaction term dominates. In that limit, to minimize the Coulomb energy, the electrons will be distributed over the lattice sites as uniformly

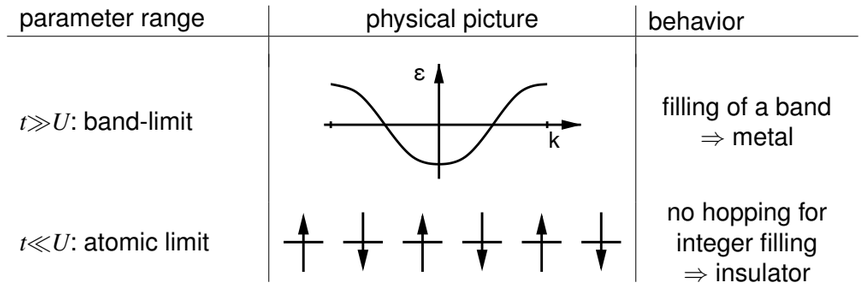


Fig. 12: *Metal-insulator transition for half-filling, i.e., one electron per site.*

as possible. For a non-degenerate, half-filled system this means, that every site carries exactly one electron, and hopping is suppressed, because it would create a doubly occupied site, which would increase the energy by $U \gg t$. Thus in the atomic limit the half-filled system will be an insulator. Clearly, in-between these two limiting cases there must be, at some value U_c, the so-called *critical* U, a transition from a metallic to an insulating state – the Mott transition [15]. Usually this transition is expected when U becomes of the order of the (non-interacting) band width W.

As the criterion for determining the metal-insulator transition we can use the opening of the gap for charge-carrying single-electron excitations

$$E_g = E(N+1) - 2E(N) + E(N-1)\,, \tag{66}$$

where $E(N)$ denotes the total energy of a cluster of M atoms with N electrons. For the half-filled system we have $N = M$. It is instructive to again consider the two limiting cases. In the non-interacting limit the total energy is given by the sum over the eigenvalues of the hopping Hamiltonian $\sum_{n:\mathrm{occ}} \varepsilon_n$. Thus, in the non-interacting limit $E_g^{\mathrm{band}} = \varepsilon_{N+1} - \varepsilon_N$, which, for a partly filled band, will vanish in the limit of infinite system size. On the other hand, in the atomic limit, the Coulomb energy for a single site with n electrons is $Un(n-1)/2$. Thus, for half-filling of we have

$$E_g^{\mathrm{atml}} = U\,, \tag{67}$$

i.e., the insulating state in the atomic limit is characterized by a finite gap.

For an infinite system the gap E_g can be rewritten in terms of the chemical potential. In the thermodynamic limit ($M \to \infty$ with N/M constant) we have to distinguish two types: the energy needed to add an electron to the system (electron affinity)

$$\mu^+ = \lim(E(N+1) - E(N)) = \left.\frac{d\varepsilon(n)}{dn}\right|_{n\searrow 1}\,, \tag{68}$$

and the energy required to extract an electron from the system (ionization energy)

$$\mu^- = \lim(E(N) - E(N-1)) = \left.\frac{d\varepsilon(n)}{dn}\right|_{n\nearrow 1}\,. \tag{69}$$

The gap is then given by the discontinuity in the left- and right-derivative of the energy per site $\varepsilon(n) = \lim E(N)/M$: $E_g = \mu^+ - \mu^-$.

6.3 Heisenberg model

We now consider the Hubbard model in the limit of large U. This is the generalization of the discussion of direct kinetic exchange in Sec. 3.2 to an extended system. For large U we work with the electron configurations, in which the interaction term is diagonal. Configurations with doubly occupied sites will have energies of the order of U or larger, so these are the configurations that we would like to project out. For downfolding we thus partition the configuration basis, and hence the Hilbert space, into the set of low-energy states which have no doubly occupied sites

$$S = \left\{ |n_{1\uparrow}, n_{1\downarrow}, n_{2\uparrow}, n_{2\downarrow}, \ldots\rangle \,\middle|\, \forall i : \; n_{i\uparrow} + n_{i\downarrow} \leq 1 \right\} \tag{70}$$

and the set of high-energy states with one or more doubly occupied sites

$$D = \left\{ |n_{1\uparrow}, n_{1\downarrow}, n_{2\uparrow}, n_{2\downarrow}, \ldots\rangle \,\middle|\, \exists i : \; n_{i\uparrow} + n_{i\downarrow} = 2 \right\}. \tag{71}$$

The hopping term T, which for large U is a perturbation to the interaction term I, couples the subspaces by hopping an electron into or out of a doubly occupied site. In addition it lifts the degeneracies within the subspaces. Hence the Hamiltonian can be partitioned as (note that $I \equiv 0$ on subspace S)

$$\hat{H} = \begin{pmatrix} P_S\, T\, P_S & P_S\, T\, P_D \\ P_D\, T\, P_S & P_D\, (T + I)\, P_D \end{pmatrix}, \tag{72}$$

Since we are dealing with an extended system, the subspaces are infinite, so we cannot write the Hamiltonian on the subspaces as matrices. Instead we restrict the operators to the appropriate subspace by using projection operators, P_S projecting on the low-energy configurations S, P_D projecting on D. Just like in 3.2 we can then write down an effective Hamiltonian operating on the low-energy configurations only:

$$H_{\text{eff}} = P_S\, T\, P_S + P_S\, T\, P_D \; \left[P_D\, (\varepsilon - (I + T))\; P_D\right]^{-1} P_D\, T\, P_S \;, \tag{73}$$

Unlike in the derivation of direct exchange, for the extended system we have no way of calculating the inverse in the second term explicitly. We can, however, expand in powers of t/U. This is Kato's method for perturbation theory (see, e.g., section 16.3 of [16]). Essentially we only need to consider configurations with a *single* double-occupancy – these correspond to the states of lowest energy in D. On this subspace the interaction term is diagonal with eigenvalue U and can thus be easily inverted. We then obtain the Hamiltonian

$$H_{t-J} = P_S \left[T - \frac{t^2}{U} \sum_{\langle ij\rangle\langle jk\rangle\sigma\sigma'} c^\dagger_{k\sigma'} c_{j\sigma'}\, n_{j\uparrow} n_{j\downarrow}\, c^\dagger_{j\sigma} c_{i\sigma} \right] P_S \;, \tag{74}$$

which is called the t-J Hamiltonian. The first term describes the hopping, constrained to configurations with no doubly occupied sites. Thus it essentially describes the hopping of empty sites (holes). To understand what the second term does, we observe that, because of the operators $n_{j\uparrow} n_{j\downarrow}$, there are only contributions for states with a singly occupied site j: $n_{j\sigma} = 0$ and $n_{j,-\sigma} = 1$. After applying the second term, site j will again be singly occupied with $n_{j\sigma'} = 0$

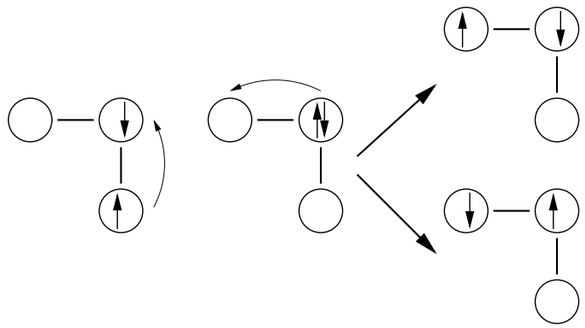


Fig. 13: *Processes contained in the three-site term T': indirect hopping processes to a second-nearest neighbor site with an intermediate (virtual) doubly occupied state. In the first process the two hopping processes are performed by the same electron, in the second process each electron hops once and thus the spin on the intermediate site is flipped.*

and $n_{j,-\sigma'} = 1$. Hence, for $\sigma \neq \sigma'$ the spin on site j will be flipped. Moreover, we distinguish the contributions where only two different sites are involved ($k = i$) from the three-site terms ($k \neq i$). The terms for $k = i$ are just the ones we already know from the kinetic exchange mechanism. The three-site terms describe a second-nearest neighbor hopping of an electron from site i to site k via a singly occupied intermediate site j. For $\sigma = \sigma'$ the spin of the hopping electron is opposite to that on the intermediate site. For $\sigma \neq \sigma'$ the spin of the intermediate site is flipped – as is that of the hopping electron. This is shown in Fig. 13. The t-J Hamiltonian is

$$H_{t-J} = P_S \, [T + H_H + T'] \, P_S \tag{75}$$

with

$$T = -t \sum_{\langle ij \rangle, \sigma} c^\dagger_{j\sigma} c_{i\sigma} \tag{76}$$

$$H_H = \frac{4t^2}{U} \sum_{\langle ij \rangle} \left(\vec{S}_j \cdot \vec{S}_i - \frac{n_i n_j}{4} \right) \tag{77}$$

$$T' = -\frac{t^2}{U} \sum_{\substack{\langle ij \rangle \langle jk \rangle \\ i \neq k}} \sum_{\sigma} \left(c^\dagger_{k\sigma} (1 - n_{j\sigma}) c_{i\sigma} - c^\dagger_{k,-\sigma} c^\dagger_{j\sigma} c_{j,-\sigma} c_{i\sigma} \right) \, n_{j,-\sigma} \tag{78}$$

In the case of half-filling, when $n_i = 1$, all hopping processes are suppressed, i.e., the projection P_S annihilates T and T'. Thus for a Mott insulator the t-J model reduces to the spin $1/2$ Heisenberg model

$$H_H = J \sum_{\langle ij \rangle} \vec{S}_j \cdot \vec{S}_i + \text{const.} \tag{79}$$

with the exchange coupling $J = 4t^2/U$ given by the direct kinetic exchange mechanism. We again stress that the spin-spin interaction is a result of projecting out the states with double occupancies.

7 Conclusion

We have seen that magnetic interactions in matter arise from the interplay of the Pauli principle and Coulomb interaction, kinetic energy, or both. The resulting effective couplings between magnetic moments are thus not fundamental interactions and, usually, take quite complex forms. However, in limiting cases they can become quite simple and transparent. These scenarios are called exchange mechanisms, of which we have discussed here a small selection. They give an idea of what magnetic interactions can be expected in real materials. Thus, despite their simplicity, exchange mechanisms provide vital guides for understanding the physics of complex ordering phenomena, of spins and orbital-occupations, from simple concepts.

Appendices

A Atomic units

Practical electronic structure calculations are usually done in atomic units, a.u. for short. While the idea behind the atomic units is remarkably simple, in practice there is often some confusion when trying to convert to SI units. We therefore give a brief explanation.
The motivation for introducing atomic units is to simplify the equations. For example, in SI units the Hamiltonian of a hydrogen atom is

$$H = -\frac{\hbar^2}{2m_e}\nabla^2 - \frac{e^2}{4\pi\epsilon_0\, r}\ . \tag{80}$$

When we implement such an equation in a computer program, we need to enter the numerical values of all the fundamental constants. We can avoid this by inventing a system of units in which the *numerical values* of the electron mass m_e, the elementary charge e, the Planck-constant $\hbar$, and the dielectric constant $4\pi\epsilon_0$ are all equal to one. In these units the above equation can be programmed as

$$H = -\frac{1}{2}\nabla^2 - \frac{1}{r}\ . \tag{81}$$

This immediately tells us: 1 a.u. mass $= m_e$ and 1 a.u. charge $= e$. To complete the set of basis units we still need the atomic unit of length, which we call a_0, and of time, t_0. To find the values of a_0 and t_0 we write $\hbar$ and $4\pi\epsilon_0$ (using simple dimensional analysis) in atomic units: $\hbar = 1\, m_e a_0^2/t_0$ and $4\pi\epsilon_0 = 1\, t_0^2 e^2/(m_e a_0^3)$. Solving this system of equations, we find

1 a.u. length	$= a_0 = 4\pi\epsilon_0\hbar^2/m_e e^2$	$\approx 5.2918\cdot 10^{-11}$	m
1 a.u. mass	$= m_e =$	$\approx 9.1095\cdot 10^{-31}$	kg
1 a.u. time	$= t_0 = (4\pi\epsilon_0)^2\hbar^3/m_e e^4$	$\approx 2.4189\cdot 10^{-17}$	s
1 a.u. charge	$= e =$	$\approx 1.6022\cdot 10^{-19}$	C

The atomic unit of length, a_0, is the Bohr radius. As the dimension of energy is mass times length squared divided by time squared, its atomic unit is $m_e a_0^2/t_0^2 = m_e e^4/(4\pi\epsilon_0)^2\hbar^2$. Because of its importance the atomic unit of energy has a name, the Hartree. One Hartree is minus twice the ground-state energy of the hydrogen atom, about 27.211 eV.
It would be tempting to try to set the numerical value of all fundamental constants to unity. But this must obviously fail, as the system of equations to solve becomes overdetermined when we try to prescribe the numerical values of constants that are not linearly independent in the space of basis units. Thus, we cannot, e.g., choose also the speed of light to have value one, as would be practical for relativistic calculations. Instead, in atomic units it is given by $c\, t_0/a_0 = 4\pi\varepsilon_0\hbar c/e^2 = 1/\alpha$, where α is the fine structure constant. Thus $c = \alpha^{-1}$ a.u. ≈ 137 a.u. The Bohr magneton is $\mu_B = 1/2$ a.u. The Boltzmann constant k_B, on the other hand, is independent of the previous constants. Setting its value to one fixes the unit of temperature to 1 a.u. temperature $= m_e e^4/(4\pi\epsilon_0)^2\hbar^2 k_B = Ha/k_B \approx 3.158\cdot 10^5$ K.

B Downfolding

To integrate-out high-energy degrees of freedom, we partition the Hilbert space of the full system into states of interest (low-energy states) and 'other' states, which will be integrated out. The Hamiltonian is then written in blocks

$$H = \begin{pmatrix} H_{00} & T_{01} \\ T_{10} & H_{11} \end{pmatrix}, \tag{82}$$

where H_{00} is the Hamiltonian restricted to the states of interest (reduced Hilbert space), H_{11} the Hamiltonian for the 'other' states, and the T matrices describe transitions between the two subspaces. The resolvent is partitioned likewise

$$G(\omega) = (\omega - H)^{-1} = \begin{pmatrix} \omega - H_{00} & -T_{01} \\ -T_{10} & \omega - H_{11} \end{pmatrix}^{-1}. \tag{83}$$

Its elements are easily determined by solving the system of two linear matrix equations

$$\begin{pmatrix} \omega - H_{00} & -T_{01} \\ -T_{10} & \omega - H_{11} \end{pmatrix} \begin{pmatrix} G_{00} & G_{01} \\ G_{10} & G_{11} \end{pmatrix} = \begin{pmatrix} \mathbb{1} & \mathbb{0} \\ \mathbb{0} & \mathbb{1} \end{pmatrix}, \tag{84}$$

keeping track of the order of the sub-matrix products. The resolvent on the reduced Hilbert space is thus given by

$$G_{00}(\omega) = \left(\omega - \underbrace{[H_{00} + T_{01}(\omega - H_{11})^{-1} T_{10}]}_{=H_{\text{eff}}(\omega)} \right)^{-1}. \tag{85}$$

This expression looks just like the resolvent for a Hamiltonian H_{eff} on the reduced Hilbert space. This effective Hamiltonian describes the physics of the full system, but operates only on the small reduced Hilbert space: For an eigenvector $H|\Psi\rangle = E|\Psi\rangle$ on the full Hilbert space

$$H|\Psi\rangle = \begin{pmatrix} H_{00} & T_{01} \\ T_{10} & H_{11} \end{pmatrix} \begin{pmatrix} |\Psi_0\rangle \\ |\Psi_1\rangle \end{pmatrix} = E \begin{pmatrix} |\Psi_0\rangle \\ |\Psi_1\rangle \end{pmatrix} \tag{86}$$

its projection $|\Psi_0\rangle$ onto the reduced Hilbert space is an eigenstate of $H_{\text{eff}}(E)$. On the other hand, we can construct the full eigenstate from a solution $H_{\text{eff}}(E)|\Psi_0\rangle = E|\Psi_0\rangle$ on the reduced Hilbert space by upfolding $|\Psi\rangle \propto (\mathbb{1} + (E - H_{11})^{-1} T_{10})|\Psi_0\rangle$.

Of course, this drastic simplification comes at a price: the effective Hamiltonian is energy dependent. If the hopping matrix elements in T_{01} are small, and/or if the states in the part of the Hilbert space that has been integrated out are energetically well-separated from the states that are explicitly considered, this energy dependence can, to a good approximation, be neglected. We can then replace ω by some characteristic energy ε_0 for the states in the reduced Hilbert space to obtain an energy-independent Hamiltonian

$$H_{\text{eff}}(\omega) = H_{00} + T_{01}(\omega - H_{11})^{-1} T_{10} \approx H_{00} + T_{01}(\varepsilon_0 - H_{11})^{-1} T_{10} = H_{\text{eff}}(\varepsilon_0) \tag{87}$$

that gives a good description of the electrons in the reduced Hilbert space, i.e., the states with an energy close to ε_0. Expanding $(\omega - H_{11})^{-1}$ about ε_0, we can systematically improve the approximation (linear and higher-order methods).

C Pauli matrices

Here we collect the most important properties of the Pauli matrices. The Pauli or spin matrices are defined as

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{88}$$

They are hermitian, i.e. $\sigma_i^\dagger = \sigma_i$, and $\sigma_i^2 = 1$. Therefore their eigenvalues are ± 1. The eigenvectors of σ_z are $|m_z\rangle$, $m_z = \pm 1$:

$$|+1\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad |-1\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{89}$$

For these vectors we find

$$\sigma_x|\,m_z\rangle = |\,-m_z\rangle \quad \sigma_y|\,m_z\rangle = i m_z|\,-m_z\rangle \quad \sigma_z|\,m_z\rangle = m_z|\,m_z\rangle. \tag{90}$$

The products of the Pauli matrices are $\sigma_x\,\sigma_y = i\sigma_z$, where the indices can be permuted cyclically. From this follows for the commutator

$$[\sigma_x, \sigma_y] = 2i\sigma_z, \tag{91}$$

while the anticommutator vanishes:

$$\{\sigma_x, \sigma_y\} = 0 \tag{92}$$

Finally a rotation by an angle φ about the axis $\hat{n}$ changes the spin matrices

$$R_{\hat{n}}(\varphi) = e^{-i\hat{n}\cdot\vec{\sigma}\,\varphi/2} = \cos(\varphi/2) - i\sin(\varphi/2)\,\hat{n}\cdot\vec{\sigma}\,. \tag{93}$$

References

[1] R. Peierls: *Surprises in Theoretical Physics* (Princeton University Press, 1979)

[2] J.H. van Vleck: *Quantum Mechanics: The Key to Understanding Magnetism*, in S. Lundqvist (ed.): *Nobel Lectures, Physics 1971-1980* (World Scientific, Singapore, 1992) http://www.nobelprize.org/nobel_prizes/physics/laureates/1977

[3] Patrik Fazekas: *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, Singapore, 1999)

[4] A. Auerbach: *Interacting Electrons and Quantum Magnetism* (Springer, Heidelberg, 1994)

[5] R. Eder: *Multiplets in Transition-Metal Ions and Introduction to Multiband Hubbard Models*, in [18]

[6] M. Weissbluth: *Atoms and Molecules* (Academic Press, San Diego, 1978)

[7] P.W. Anderson: *Theory of Magnetic Exchange Interactions*, in F. Seitz and D. Turnbull (eds.): *Solid State Physics* **14**, 99 (1963)

[8] D.C. Mattis: *The Theory of Magnetism Made Simple* (World Scientific, Singapore, 2006)

[9] S. Blundell: *Magnetism in Condensed Matter* (Oxford University Press, 2001)

[10] E. Koch: *Many-Electron States*, in [19]

[11] A. Szabo and N.S. Ostlund: *Modern Quantum Chemistry* (Dover Publications, 1996)

[12] J.B. Goodenough: *Goodenough-Kanamori rule*, Scholarpedia **3**, 7382 (2008) http://scholarpedia.org/article/Goodenough-Kanamori_rule

[13] K.I. Kugel and D.I. Khomskii, Zh. Eksp. Teor. Fiz. **64**, 1429 (1973) [Sov. Phys. JETP **37**, 725 (1973)]

[14] A. Montorsi (Ed.): *The Hubbard Model*, (World Scientific, Singapore, 1992)

[15] F. Gebhard: *The Mott Metal-Insulator Transition* (Springer, Heidelberg, 1997)

[16] A. Messiah: *Méchanique Quantique* (Dunond, Paris, 1964)

[17] E. Pavarini, E. Koch, D. Vollhardt, and A.I. Lichtenstein: *The LDA+DMFT approach to strongly correlated materials* Reihe Modeling and Simulation, Vol. 1 (Forschungszentrum Jülich, 2011) http://www.cond-mat.de/events/correl11

[18] E. Pavarini, E. Koch, J. van den Brink, and G. Sawatzky:
Quantum Materials: Experiments and Theory
Reihe Modeling and Simulation, Vol. 6 (Forschungszentrum Jülich, 2016)
http://www.cond-mat.de/events/correl16

[19] E. Pavarini, E. Koch, and U. Schollwöck:
Emergent Phenomena in Correlated Matter
Reihe Modeling and Simulation, Vol. 3 (Forschungszentrum Jülich, 2013)
http://www.cond-mat.de/events/correl13

5 Orbital Physics

Andrzej M. Oleś
Marian Smoluchowski Institute of Physics
Jagiellonian University
Prof. S. Łojasiewicza 11, Kraków, Poland

Contents

1 Introduction: strong correlations at orbital degeneracy **2**

2 Orbital and compass models **6**

3 Superexchange models for active e_g orbitals **11**
3.1 General structure of the spin-orbital superexchange 11
3.2 Kugel-Khomskii model for $KCuF_3$ and K_2CuF_4 14
3.3 Spin-orbital superexchange model for $LaMnO_3$ 18

4 Superexchange for active t_{2g} orbitals **21**
4.1 Spin-orbital superexchange model for $LaTiO_3$ 21
4.2 Spin-orbital superexchange model for $LaVO_3$ 22

5 Spin-orbital complementarity and entanglement **24**
5.1 Goodenough-Kanamori rules 24
5.2 Spin-orbital entanglement 26
5.3 Fractionalization of orbital excitations 28

6 t-J-like model for ferromagnetic manganites **29**

7 Conclusions and outlook **30**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction: strong correlations at orbital degeneracy

Strong local Coulomb interactions lead to electron localization in Mott or charge transfer correlated insulators. The simplest model of a Mott insulator is the non-degenerate Hubbard model, where the large intraorbital Coulomb interaction U suppresses charge fluctuations due to the kinetic energy $\propto t$. As a result, the physical properties of a Mott insulator are determined by an interplay of kinetic exchange $\propto J$, with

$$J = \frac{4t^2}{U}, \tag{1}$$

derived from the Hubbard model at $U \gg t$, and the motion of holes in the restricted Hilbert space without double occupancies, as described by the t-J model [1]. Although this generic model captures the essential idea of strong correlations, realistic correlated insulators arise in transition metal oxides (or fluorides) and the degeneracy of their partly filled and nearly degenerate $3d$ (or $4d$) strongly correlated states has to be treated explicitly. Quite generally, strong local Coulomb interactions lead then to a multitude of quite complex behavior with often puzzling transport and magnetic properties [2]. The theoretical understanding of this class of compounds, including the colossal magneto-resistance (CMR) manganites as a prominent example [3], has to include not only spins and holes but in addition orbital degrees of freedom, which have to be treated on equal footing with the electron spins [4]. For a Mott insulator with transition metal ions in d^m configurations, charge excitations along the bond $\langle ij \rangle$, $d_i^m d_j^m \rightleftharpoons d_i^{m+1} d_j^{m-1}$, lead to spin-orbital superexchange which couples two neighboring ions at sites i and j.

It is important to realize that modeling of transition metal oxides can be performed on different levels of sophistication. We shall present some effective orbital and spin-orbital superexchange models for the correlated $3d$-orbitals depicted in Fig. 1 coupled by hopping t between nearest neighbor ions on a perovskite lattice, while the hopping for other lattices may be generated by the general rules formulated by Slater and Koster [5]. The orbitals have particular shapes and belong to two irreducible representations of the O_h cubic point group:

(i) a two-dimensional (2D) representation of e_g-orbitals $\{3z^2 - r^2, x^2 - y^2\}$, and

(ii) a three-dimensional (3D) representation of t_{2g}-orbitals $\{xy, yz, zx\}$.

In the absence of any tetragonal distortion or crystal-field due to surrounding oxygens, the $3d$-orbitals are degenerate within each irreducible representation of the O_h point group and have typically a large splitting $10D_q \sim 2$ eV between them. When such degenerate orbitals are party filled, electrons (or holes) have both spin and orbital degrees of freedom. The kinetic energy H_t in a perovskite follows from the hybridization between $3d$- and $2p$-orbitals. In an effective d-orbital model the oxygen $2p$-orbitals are not included explicitly and we define the hopping element t as the largest hopping element obtained for two orbitals of the same type which belong to the nearest neighboring $3d$ ions.

We begin with the conceptually simpler t_{2g}-orbitals, where finite hopping t results from the d–p hybridization along π-bonds which couples each a pair of identical orbitals active along a given bond. Each t_{2g}-orbital is active along two cubic axes and the hopping is forbidden along the one perpendicular to the plane of this orbital, e.g., the hopping between two xy-orbitals is

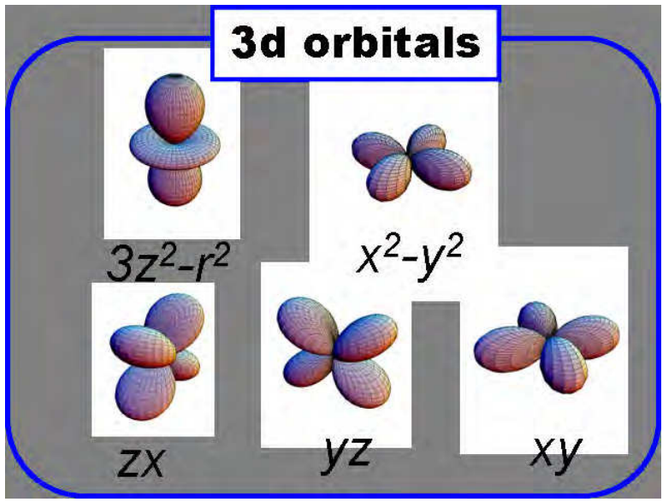


Fig. 1: *Schematic representation of* $3d$*-orbitals: Top — two* e_g*-orbitals* $\{3z^2 - r^2, x^2 - y^2\}$*; Bottom — three* t_{2g}*-orbitals* $\{zx, yz, xy\}$*. (Image courtesy of Yoshinori Tokura)*

not allowed along the c axis (due to the cancellations caused by orbital phases). It is therefore convenient to introduce the following short-hand notation for the orbital degrees of freedom [6],

$$|a\rangle \equiv |yz\rangle, \qquad |b\rangle \equiv |zx\rangle, \qquad |c\rangle \equiv |xy\rangle. \tag{2}$$

The labels $\gamma = a, b, c$ thus refer to the cubic axes where the hopping is absent for orbitals of a given type,

$$H_t(t_{2g}) = -t \sum_{\alpha} \sum_{\langle ij \rangle \| \gamma \neq \alpha} a^{\dagger}_{i\alpha\sigma} a_{j\alpha\sigma}, \tag{3}$$

Here $a^{\dagger}_{i\alpha\sigma}$ is an electron creation operator in a t_{2g}-orbital $\alpha \in \{xy, yz, zx\}$ with spin $\sigma = \uparrow, \downarrow$ at site i, and the local electron density operator for a spin-orbital state is $n_{i\alpha\sigma} = a^{\dagger}_{i\alpha\sigma} a_{i\alpha\sigma}$. Not only spin but also orbital flavor is conserved in the hopping process $\propto t$.

The hopping Hamiltonian for e_g-electrons couples two directional e_g-orbitals $\{|i\zeta_\alpha\rangle, |j\zeta_\alpha\rangle\}$ along a σ-bond $\langle ij \rangle$ [7],

$$H_t(e_g) = -t \sum_{\alpha} \sum_{\langle ij \rangle \| \alpha, \sigma} a^{\dagger}_{i\zeta_\alpha\sigma} a_{j\zeta_\alpha\sigma}. \tag{4}$$

Indeed, the hopping with amplitude $-t$ between sites i and j occurs only when an electron with spin σ transfers between two directional orbitals $|\zeta_\alpha\rangle$ oriented along the bond $\langle ij \rangle$ direction, i.e., $|\zeta_\alpha\rangle \propto 3x^2 - r^2$, $3y^2 - r^2$, and $3z^2 - r^2$ along the cubic axis $\alpha = a$, b, and c. We will similarly denote by $|\xi_\alpha\rangle$ the orbital which is orthogonal to $|\zeta_\alpha\rangle$ and is oriented perpendicular to the bond $\langle ij \rangle$ direction, i.e., $|\xi_\alpha\rangle \propto y^2 - z^2$, $z^2 - x^2$, and $x^2 - y^2$ along the axis $\alpha = a$, b, and c. For the moment we consider only electrons with one spin, $\sigma = \uparrow$, to focus on the orbital problem. While such a choice of an over-complete basis $\{\zeta_a, \zeta_b, \zeta_c\}$ is convenient, for writing down the

kinetic energy a particular orthogonal basis is needed. The usual choice is to take

$$|z\rangle \equiv \tfrac{1}{\sqrt{6}}(3z^2-r^2), \qquad |\bar{z}\rangle \equiv \tfrac{1}{\sqrt{2}}(x^2-y^2), \tag{5}$$

called *real e_g-orbitals* [7]. However, this basis is the natural one only for the bonds parallel to the c axis but not for those in the (a,b) plane, and for $\uparrow$-spin electrons the hopping reads (here for clarity we omit the spin index σ)

$$H_t^{\uparrow}(e_g) = -\frac{1}{4}t\sum_{\langle ij\rangle\parallel ab}\left[3a_{i\bar{z}}^{\dagger}a_{j\bar{z}}+a_{iz}^{\dagger}a_{jz}\mp\sqrt{3}\left(a_{i\bar{z}}^{\dagger}a_{jz}+a_{iz}^{\dagger}a_{j\bar{z}}\right)\right]-t\sum_{\langle ij\rangle\parallel c}a_{iz}^{\dagger}a_{jz}, \tag{6}$$

and although this expression is of course cubic invariant, it does not manifest this symmetry but takes a very different appearance depending on the bond direction. However, the symmetry is better visible using the basis of *complex e_g-orbitals* at each site j [7],

$$|j+\rangle = \tfrac{1}{\sqrt{2}}\big(|jz\rangle - i|j\bar{z}\rangle\big), \qquad |j-\rangle = \tfrac{1}{\sqrt{2}}\big(|jz\rangle + i|j\bar{z}\rangle\big), \tag{7}$$

corresponding to "up" and "down" pseudospin flavors, with the local pseudospin operators defined as

$$\tau_i^{+} \equiv c_{i+}^{\dagger}c_{i-}, \qquad \tau_i^{-} \equiv c_{i-}^{\dagger}c_{i+}, \qquad \tau_i^{z} \equiv \tfrac{1}{2}(c_{i+}^{\dagger}c_{i+}-c_{i-}^{\dagger}c_{i-}) = \tfrac{1}{2}(n_{i+}-n_{i-}). \tag{8}$$

The three directional $|i\zeta_\alpha\rangle$ and three planar $|i\xi_\alpha\rangle$ orbitals at site i, associated with the three cubic axes ($\alpha = a$, b, c), are the real orbitals,

$$|i\zeta_\alpha\rangle = \tfrac{1}{\sqrt{2}}\left[e^{-i\vartheta_\alpha/2}|i+\rangle + e^{+i\vartheta_\alpha/2}|i-\rangle\right] = \cos(\vartheta_\alpha/2)|iz\rangle - \sin(\vartheta_\alpha/2)|i\bar{z}\rangle, \tag{9}$$

$$|i\xi_\alpha\rangle = \tfrac{1}{\sqrt{2}}\left[e^{-i\vartheta_\alpha/2}|i+\rangle - e^{+i\vartheta_\alpha/2}|i-\rangle\right] = \sin(\vartheta_\alpha/2)|iz\rangle + \cos(\vartheta_\alpha/2)|i\bar{z}\rangle, \tag{10}$$

with the phase factors $\vartheta_{ia} = -4\pi/3$, $\vartheta_{ib} = +4\pi/3$, and $\vartheta_{ic} = 0$, and thus correspond to the pseudospin lying in the equatorial plane and pointing in one of the three equilateral "cubic" directions defined by the angles $\{\vartheta_{i\alpha}\}$.

Using the above complex-orbital representation (7) we can write the *orbital Hubbard model* for e_g-electrons with only one spin flavor $\sigma =\uparrow$ in a form similar to the spin Hubbard model,

$$\mathcal{H}^{\uparrow}(e_g) = -\frac{1}{2}t\sum_{\alpha}\sum_{\langle ij\rangle\parallel\alpha}\left[\left(a_{i+}^{\dagger}a_{j+}+a_{i-}^{\dagger}a_{j-}\right)+\gamma\left(e^{-i\chi_\alpha}a_{i+}^{\dagger}a_{j-}+e^{+i\chi_\alpha}a_{i-}^{\dagger}a_{j+}\right)\right]+\bar{U}\sum_i n_{i+}n_{i-}, \tag{11}$$

with $\chi_a = +2\pi/3$, $\chi_b = -2\pi/3$, and $\chi_c = 0$, and where the parameter γ, explained below, takes for e_g-orbitals the value $\gamma = 1$. The appearance of the phase factors $e^{\pm i\chi_\alpha}$ is characteristic of the orbital problem—they occur because the orbitals have an actual shape in real space so that each hopping process depends on the bond direction and may change the orbital flavor. The interorbital Coulomb interaction $\propto \bar{U}$ couples the electron densities in basis orbitals $n_{i\alpha} = a_{i\mu}^{\dagger}a_{i\mu}$, with $\mu \in \{+,-\}$; its form is invariant under any local basis transformation to a pair of orthogonal orbitals, i.e., it gives an energy $\bar{U}$ for a double occupancy either when

two real orbitals are simultaneously occupied, $\bar{U}\sum_i n_{iz}n_{i\bar{z}}$, or when two complex orbitals are occupied, $\bar{U}\sum_i n_{i+}n_{i-}$.

In general, on-site Coulomb interactions between two interacting electrons in $3d$-orbitals depend both on spin and orbital indices and the interaction Hamiltonian takes the form of the degenerate Hubbard model. Note that the electron interaction parameters in this model are effective ones, i.e., the $2p$-orbital parameters of O (F) ions renormalize on-site Coulomb interactions for $3d$-orbitals. The general form which includes only two-orbital interactions and the anisotropy of Coulomb and exchange elements is [8]

$$H_{int} = U\sum_{i\alpha} n_{i\alpha\uparrow}n_{i\alpha\downarrow} + \sum_{i,\alpha<\beta}\left(U_{\alpha\beta}-\frac{1}{2}J_{\alpha\beta}\right)n_{i\alpha}n_{i\beta} - 2\sum_{i,\alpha<\beta}J_{\alpha\beta}\,\vec{S}_{i\alpha}\cdot\vec{S}_{i\beta} + \sum_{i,\alpha<\beta}J_{\alpha\beta}\left(a^{\dagger}_{i\alpha\uparrow}a^{\dagger}_{i\alpha\downarrow}a_{i\beta\downarrow}a_{i\beta\uparrow} + a^{\dagger}_{i\beta\uparrow}a^{\dagger}_{i\beta\downarrow}a_{i\alpha\downarrow}a_{i\alpha\uparrow}\right). \tag{12}$$

Here $a^{\dagger}_{i\alpha\sigma}$ is an electron creation operator in any $3d$-orbital $\alpha \in \{xy, yz, zx, 3z^2-r^2, x^2-y^2\}$ and $\bar{\sigma} \equiv -\sigma$, with spin states $\sigma =\uparrow,\downarrow$ at site i. The parameters $\{U, U_{\alpha\beta}, J_{\alpha\beta}\}$ depend in the general case on the three Racah parameters A, B and C [9] which may be derived from somewhat screened atomic values. While the intraorbital Coulomb element is identical for all $3d$-orbitals,

$$U = A + 4B + 3C\,, \tag{13}$$

the interorbital Coulomb $U_{\alpha\beta}$ and exchange $J_{\alpha\beta}$ elements are anisotropic and depend on the involved pair of orbitals; the values of $J_{\alpha\beta}$ are given in Table 1. The Coulomb $U_{\alpha\beta}$ and exchange $J_{\alpha\beta}$ elements are related to the intraorbital element U by a relation which guarantees the invariance of interactions in the orbital space,

$$U = U_{\alpha\beta} + 2J_{\alpha\beta}\,. \tag{14}$$

In all situations where only the orbitals belonging to a single irreducible representation of the cubic group (e_g or t_{2g}) are partly filled, as, e.g., in the titanates, vanadates, nickelates, or copper fluorides, the filled (empty) orbitals do not contribute, and the relevant exchange elements $J_{\alpha\beta}$

Table 1: *On-site interorbital exchange elements $J_{\alpha\beta}$ for $3d$ orbitals as functions of the Racah parameters B and C (for more details see Ref. [9]).*

$3d$-orbital	xy	yz	zx	x^2-y^2	$3z^2-r^2$
xy	0	$3B+C$	$3B+C$	C	$4B+C$
yz	$3B+C$	0	$3B+C$	$3B+C$	$B+C$
zx	$3B+C$	$3B+C$	0	$3B+C$	$B+C$
x^2-y^2	C	$3B+C$	$3B+C$	0	$4B+C$
$3z^2-r^2$	$4B+C$	$B+C$	$B+C$	$4B+C$	0

are all the same (see Table 1), i.e., for t_{2g} (e_g) orbitals,

$$J_H^t = 3B + C, \tag{15}$$
$$J_H^e = 4B + C. \tag{16}$$

Then one may use a simplified *degenerate* Hubbard model with isotropic form of on-site interactions (for a given subset of $3d$-orbitals) [10],

$$\begin{aligned} H_{int}^{(0)} &= U \sum_{i\alpha} n_{i\alpha\uparrow} n_{i\alpha\downarrow} + \left(U - \frac{5}{2} J_H\right) \sum_{i,\alpha<\beta} n_{i\alpha} n_{i\beta} - 2J_H \sum_{i,\alpha<\beta} \vec{S}_{i\alpha} \cdot \vec{S}_{i\beta} \\ &+ J_H \sum_{i,\alpha<\beta} \left(a_{i\alpha\uparrow}^\dagger a_{i\alpha\downarrow}^\dagger a_{i\beta\downarrow} a_{i\beta\uparrow} + a_{i\beta\uparrow}^\dagger a_{i\beta\downarrow}^\dagger a_{i\alpha\downarrow} a_{i\alpha\uparrow} \right). \end{aligned} \tag{17}$$

It has two Kanamori parameters: the Coulomb intraorbital element U (13) and Hund's exchange J_H standing either for J_H^t (15) or for J_H^e (16). Now $\bar{U} \equiv (U - 3J_H)$ in Eq. (11). We emphasize that in a general case when both types of orbitals are partly filled (as in the CMR manganites) and both thus participate in charge excitations, the above Hamiltonian with a single Hund's exchange element J_H is insufficient and the full anisotropy given in Eq. (17) has to be used instead to generate correct charge excitation spectra of a given transition metal ion [9].

2 Orbital and compass models

If the spin state is ferromagnetic (FM) as, e.g., in the ab planes of $KCuF_3$ (or $LaMnO_3$), charge excitations $d_i^m d_j^m \rightleftharpoons d_i^{m+1} d_j^{m-1}$ with $m = 9$ (or $m = 4$) concern only high-spin (HS) ${}^3A_1$ (or ${}^6A_1$) states and the superexchange interactions reduce to an orbital superexchange model [11]. Thus we begin with an orbital model for e_g-holes in $KCuF_3$, with a local basis at site i defined by two real e_g-orbitals, see Eq. (5), being a local e_g-orbital basis at each site. The basis consists of a directional orbital $|i\zeta_c\rangle \equiv |iz\rangle$ and the planar orbital $|i\xi_c\rangle \equiv |i\bar{z}\rangle$. Other equivalent orbital bases are obtained by rotation of the above pair of orbitals by an angle ϑ to

$$\begin{aligned} |i\vartheta\rangle &= \cos(\vartheta/2)\,|iz\rangle - \sin(\vartheta/2)\,|i\bar{z}\rangle, \\ |i\bar{\vartheta}\rangle &= \sin(\vartheta/2)\,|iz\rangle + \cos(\vartheta/2)\,|i\bar{z}\rangle, \end{aligned} \tag{18}$$

i.e., to a pair $\{|i\vartheta\rangle, |i, \vartheta + \pi\rangle\}$. For angles $\vartheta = \pm 4\pi/3$ one finds equivalent pairs of directional and planar orbitals in a 2D model, $\{|i\zeta_a\rangle, |i\xi_a\rangle\}$ and $\{|i\zeta_b\rangle, |i\xi_b\rangle\}$, to the usually used e_g-orbital real basis given by Eq. (5).

Consider now a bond $\langle ij\rangle \parallel \gamma$ along one of the cubic axes $\gamma = a, b, c$, and a charge excitation generated by a hopping process $i \to j$. The hopping t couples two directional orbitals $\{|i\zeta_\gamma\rangle, |j\zeta_\gamma\rangle\}$. Local projection operators onto these active and the complementary inactive $\{|i\xi_\gamma\rangle, |j\xi_\gamma\rangle\}$ orbitals are

$$\mathcal{P}_{i\zeta}^\gamma = |i\zeta_\gamma\rangle\langle i\zeta_\gamma| = \left(\frac{1}{2} + \tau_i^{(\gamma)}\right), \qquad \mathcal{P}_{i\xi}^\gamma = |i\xi_\gamma\rangle\langle i\xi_\gamma| = \left(\frac{1}{2} - \tau_i^{(\gamma)}\right), \tag{19}$$

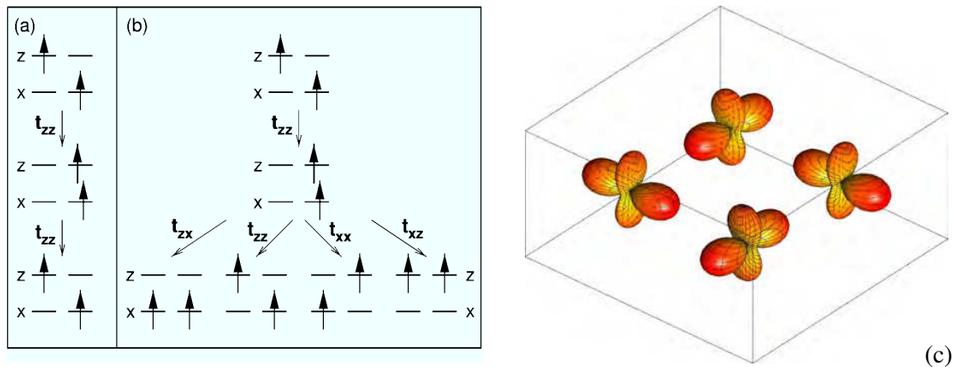


Fig. 2: *Virtual charge excitations leading to the e_g-orbital superexchange model for a strongly correlated system with $|z\rangle$ and $|x\rangle \equiv |\bar{z}\rangle$ real e_g-orbitals (5) in the subspace of $\uparrow$-spin states: (a) for a bond along the c axis $\langle ij\rangle \parallel c$; (b) for a bond in the ab plane $\langle ij\rangle \parallel ab$. In a FM plane of $KCuF_3$ ($LaMnO_3$) the superexchange (27) favors (c) AO state of $|\mathrm{AO}\pm\rangle$-orbitals (28). (Images (a-b) reproduced from Ref. [11]; image (c) courtesy of Krzysztof Bieniasz)*

where

$$\tau_i^{(\gamma)} \equiv \frac{1}{2}\left(|i\zeta_\gamma\rangle\langle i\zeta_\gamma| - |i\xi_\gamma\rangle\langle i\xi_\gamma|\right), \tag{20}$$

and these operators are represented in the fixed $\{|iz\rangle, |i\bar{z}\rangle\}$ basis as follows:

$$\tau_i^{(a)} = -\frac{1}{4}\left(\sigma_i^z - \sqrt{3}\sigma_i^x\right), \qquad \tau_i^{(b)} = -\frac{1}{4}\left(\sigma_i^z + \sqrt{3}\sigma_i^x\right), \qquad \tau_i^{(c)} = \frac{1}{2}\sigma_i^z. \tag{21}$$

A charge excitation between two transition metal ions with partly filled e_g-orbitals will arise by a hopping process between two active orbitals, $|i\zeta_\gamma\rangle$ and $|j\zeta_\gamma\rangle$. To capture such processes we introduce two projection operators on the orbital states for each bond,

$$\mathcal{P}_{\langle ij\rangle}^{(\gamma)} \equiv \left(\frac{1}{2}+\tau_i^{(\gamma)}\right)\left(\frac{1}{2}-\tau_j^{(\gamma)}\right) + \left(\frac{1}{2}-\tau_i^{(\gamma)}\right)\left(\frac{1}{2}+\tau_j^{(\gamma)}\right), \tag{22}$$

$$\mathcal{Q}_{\langle ij\rangle}^{(\gamma)} \equiv 2\left(\frac{1}{2}-\tau_i^{(\gamma)}\right)\left(\frac{1}{2}-\tau_j^{(\gamma)}\right). \tag{23}$$

Unlike for a spin system, the charge excitation $d_i^m d_j^m \rightleftharpoons d_i^{m+1} d_j^{m-1}$ is allowed only in one direction when one orbital is directional $|\zeta_\gamma\rangle$ and the other is planar $|\xi_\gamma\rangle$ on the bond $\langle ij\rangle \parallel \gamma$, i.e., $\left\langle \mathcal{P}_{\langle ij\rangle}^{(\gamma)} \right\rangle = 1$; such processes generate both HS and low-spin (LS) contributions. On the contrary, when both orbitals are directional, i.e., one has $\left\langle \mathcal{Q}_{\langle ij\rangle}^{(\gamma)} \right\rangle = 2$, only LS terms contribute. To write the superexchange model we need the charge excitation energy which for the HS channel is,

$$\varepsilon_1 \equiv E_1(d^{m+1}) + E_0(d^{m-1}) - 2E_0(d^m) = U - 3J_H = \bar{U}, \tag{24}$$

where $E_0(d^m)$ in the ground state energy for an ion with m electrons. Note that this energy is the same for $KCuF_3$ and $LaMnO_3$ [8], so the e_g-orbital model presented here is universal.

Second order perturbation theory shown in Figs. 2(a-b) gives [11],

$$\mathcal{H}^{\uparrow}(e_g) = -\frac{t^2}{\varepsilon_1} \sum_{\langle ij \rangle \| \gamma} \mathcal{P}^{(\gamma)}_{\langle ij \rangle}. \tag{25}$$

For convenience we define the dimensionless Hund's exchange parameter η

$$\eta \equiv \frac{J_H}{U}. \tag{26}$$

The value of J defines the superexchange energy scale and is the same as in the t-J model [1], while the parameter η (26) characterizes the multiplet structure when LS states are included as well, see below. The e_g-orbital model (25) (for HS states) takes the form

$$\mathcal{H}^{\uparrow}(e_g) = \frac{1}{2} J r_1 \sum_{\langle ij \rangle \| \gamma} \left(\tau^{(\gamma)}_i \tau^{(\gamma)}_j - \frac{1}{4} \right) + E_z \sum_i \tau^{(c)}_i, \tag{27}$$

where $r_1 = U/\varepsilon_{\rm HS} = U/\bar{U} = 1/(1-3\eta)$. Here we include the crystal-field term $\propto E_z$ which splits off the e_g orbitals. The same effective model is obtained from the e_g Hubbard model Eq. (11) at half-filling in the regime of $\bar{U} \gg t$. It favors, consistently with its derivation, pairs of orthogonal orbitals along the axis γ, with the energy gain for such a configuration $-\frac{1}{4}Jr_1$. When both orbitals would be instead selected as directional along the bond, $\langle \tau^{(\gamma)}_i \tau^{(\gamma)}_j \rangle = \frac{1}{4}$, the energy gain vanishes as this orbital configuration corresponds to the situation incompatible with the HS excited states considered here and the superexchange is blocked. The ground state in the 2D ab plane has alternating orbital (AO) order between the sublattices $i \in A$ and $j \in B$,

$$|i+\rangle = \tfrac{1}{\sqrt{2}}\big(|iz\rangle + |i\bar{z}\rangle\big), \qquad |j-\rangle = \tfrac{1}{\sqrt{2}}\big(|jz\rangle - |j\bar{z}\rangle\big), \tag{28}$$

of orbitals occupied by holes in $KCuF_3$ and by electrons in $LaMnO_3$, see Fig. 2(c).
Here we are interested in the low temperature range $T < 0.1J$ and the 2D (and 3D) e_g-orbital model orders at finite temperature $T < T_c$ [12], i.e., below $T_c = 0.3566J$ for a 2D model [13], so we assume perfect orbital order given by a classical *Ansatz* for the ground state,

$$|\Phi_0\rangle = \prod_{i \in A} |i\theta_A\rangle \prod_{j \in B} |j\theta_B\rangle, \tag{29}$$

with the orbital states, $|i\theta_A\rangle$ and $|j\theta_B\rangle$, characterized by opposite angles ($\theta_A = -\theta_B$) and alternating between two sublattices A and B in the ab planes. The orbital state at site i

$$|i\theta\rangle = \cos(\theta/2)\,|iz\rangle + \sin(\theta/2)\,|i\bar{z}\rangle, \tag{30}$$

is here parameterized by an angle θ which defines the amplitudes of the orbital states defined in Eq. (5). The AO state specified in Eq. (29) is thus defined by

$$\begin{aligned} |i\theta_A\rangle &= \cos(\theta/2)\,|iz\rangle + \sin(\theta/2)\,|ix\rangle, \\ |j\theta_B\rangle &= \cos(\theta/2)\,|jz\rangle - \sin(\theta/2)\,|jx\rangle, \end{aligned} \tag{31}$$

with $\theta_A = \theta$ and $\theta_B = -\theta$.

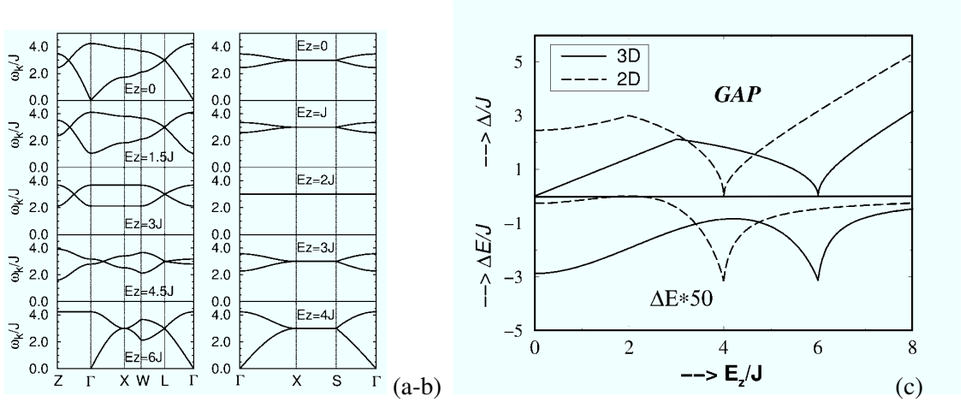


Fig. 3: *(a-b) Orbital-wave excitations obtained for different values of the crystal-field splitting E_z for a 3D (left) and a 2D (right) orbital superexchange model (27), with $Jr_1 \equiv J$. The result shown for the 3D model at $E_z = 0$ actually corresponds to the limit $E_z \to 0$. (c) Gap Δ/J in the orbital excitation spectrum and energy quantum correction $\Delta E/J$ as functions of the crystal-field splitting E_z/J, for the 3D (2D) model shown by full (dashed) lines. (Images reproduced from Ref. [11])*

The excitations from the ground state of the orbital model (27) are orbital waves (orbitons) which may be obtained in a similar way to magnons in a quantum antiferromagnet. An important difference is that the orbitons have two branches which are in general nondegenerate, see Fig. 3(a-b). In the absence of a crystal field ($E_z = 0$) the spectrum for the 2D e_g-orbital model has a gap and the orbitons have weak dispersion, so the quantum corrections to the order parameter are rather small. They are much larger in the 3D model but still smaller than in an antiferromagnet [11]. The gap closes in the 3D model at $E_z = 0$, but the quantum corrections are smaller than in the Heisenberg model. Note that the shape of the occupied orbitals changes at finite crystal field, and the orbitons have a remarkable evolution, both in the 3D and 2D model, see Figs. 3(a-b). Increasing $E_z > 0$ first increases the gap but when the field overcomes the interactions and polarizes the orbitals (at $E_z = 4J$ in 2D and $E_z = 6J$ in 3D model), the gap closes, see Fig. 3(c). This point marks a transition from the AO order to uniform ferro-orbital (FO) order. Note that in agreement with intuition the quantum corrections $\Delta E/J$ are maximal when the gap closes and low-energy orbitons contribute.

To see the relation of the 2D e_g-orbital model to the compass model [14] we introduce a 2D *generalized* compass model (GCM) with pseudospin interactions on a square lattice in the ab plane ($J_{\rm cm} > 0$) [15],

$$\mathcal{H}(\theta) = -J_{\rm cm} \sum_{\{ij\}\in ab} \left(\sigma^a_{ij}(\theta)\sigma^a_{i+1,j}(\theta) + \sigma^b_{ij}(\theta)\sigma^b_{i,j+1}(\theta) \right). \tag{32}$$

The interactions occur along nearest neighbor bonds and are balanced along both lattice directions a and b. Here $\{ij\}$ labels lattice sites in the ab plane and $\{\sigma^a_{ij}(\theta), \sigma^b_{ij}(\theta)\}$ are linear

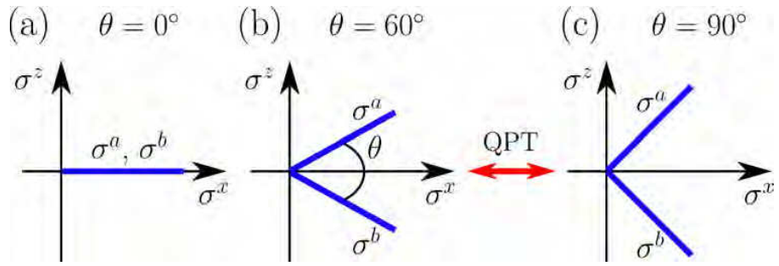


Fig. 4: *Artist's view of the evolution of orbital interactions in the generalized compass model Eq. (32) with increasing angle θ. Heavy (blue) lines indicate favored spin directions induced by interactions along two nonequivalent lattice axes a and b. Different panels show: (a) the Ising model at $\theta = 0°$, (b) the 2D e_g-orbital model at $\theta = 60°$, and (c) the OCM at $\theta = 90°$. Spin order follows the interactions in the Ising limit, while it follows one of the equivalent interactions, σ^a or σ^b, in the OCM. This results in the symmetry breaking quantum phase transition (QPT) which occurs between (b) and (c). (Image reproduced from Ref. [15])*

combinations of Pauli matrices describing interactions for $T = \frac{1}{2}$ pseudospins

$$\sigma^a_{ij}(\theta) \;=\; \cos(\theta/2)\,\sigma^x_{ij} + \sin(\theta/2)\,\sigma^z_{ij}\,, \\ \sigma^b_{ij}(\theta) \;=\; \cos(\theta/2)\,\sigma^x_{ij} - \sin(\theta/2)\,\sigma^z_{ij}\,. \tag{33}$$

The interactions in Eq. (32) include the classical Ising model for σ^x_{ij} operators at $\theta = 0°$ and become gradually more frustrated with increasing angle $\theta \in (0°, 90°]$ — they interpolate between the Ising model (at $\theta = 0°$) and the isotropic compass model (at $\theta = 90°$), see Fig. 4. The latter case is equivalent by a standard unitary transformation to the 2D compass model with standard interactions, $\sigma^x_{ij}\sigma^x_{i,j+1}$ along the a and $\sigma^z_{ij}\sigma^z_{i+1,j}$ along the b axis [15],

$$\mathcal{H}(\pi/2) = -J_{\rm cm}\sum_{\langle ij\rangle\|a}\sigma^x_{ij}\sigma^x_{i+1,j} - J_{\rm cm}\sum_{\langle ij\rangle\|b}\sigma^z_{ij}\sigma^z_{i,j+1}\,. \tag{34}$$

The model (32) includes as well the 2D e_g-orbital model as a special case, i.e., at $\theta = 60°$. Increasing the angle θ between the interacting orbital-like components (33) in Fig. 4 is equivalent to increasing frustration which becomes maximal in the 2D compass model. As a result, a second order quantum phase transition from Ising to nematic order [16] occurs at $\theta_c \simeq 84.8°$ which is surprisingly close to the compass point $\theta = 90°$, i.e., only when the interactions are sufficiently strongly frustrated. The ground state has high degeneracy $d = 2^{L+1}$ for a 2D cluster $L \times L$ of one-dimensional (1D) nematic states which are entirely different from the 2D AO order in the e_g-orbital model depicted in Fig. 4(c), yet it is stable in a range of temperatures below $T_c \simeq 0.06\,J_{\rm cm}$ [17].

3 Superexchange models for active e_g orbitals

3.1 General structure of the spin-orbital superexchange

We consider the case of partly filled degenerate $3d$-orbitals and large Hund's exchange J_H. In the regime of $t \ll U$, electrons localize and effective low-energy superexchange interactions consist of all the contributions which originate from possible virtual charge excitations, $d_i^m d_j^m \rightleftharpoons d_i^{m+1} d_j^{m-1}$ — they take the form of a spin-orbital model, see Eq. (37) below. The charge excitation n costs the energy

$$\varepsilon_n = E_n(d^{m+1}) + E_0(d^{m-1}) - 2E_0(d^m), \tag{35}$$

where the d^m ions are in the initial HS ground states with spins $S = \frac{m}{2}$ and have the Coulomb interaction energy $E_0(d^m) = \binom{m}{2}(U - 3J_H)$ each (if $m < 5$, else if $m > 5$ one has to consider here m holes instead, while the case of $m = 5$ is special and will not be considered here as in the $t_{2g}^3 e_g^2$ configuration the orbital degree of freedom is quenched). The same formula for the ground state energy applies as well to Mn^{3+} ions in d^4 configuration with spin $S = 2$ HS ground state, see Sec. 3.3. By construction also the ion with fewer electrons (holes) for $m < 5$ is in the HS state and $E_0(d^{m-1}) = \binom{m-1}{2}(U - 3J_H)$. The excitation energies (35) are thus defined by the multiplet structure of an ion with more electrons (holes) in the configuration d^{m+1}, see Fig. 5. The lowest energy excitation is given by Eq. (24) — it is obtained from the HS state of the $3d^{m+1}$ ion with total spin $\mathcal{S} = S + \frac{1}{2}$ and energy $E_1(d^{m+1}) = \binom{m+1}{2}(U - 3J_H)$. Indeed, one recovers the lowest excitation energy in the HS subspace, see Eq. (24), with J_H being Hund's exchange element for the electron (hole) involved in the charge excitation, either e_g or t_{2g}. We emphasize that this lowest excitation energy ε_1 (24) is universal and is found both in t_{2g} and e_g systems, i.e., it does not depend on the electron valence m. In contrast, the remaining energies $\{\varepsilon_n\}$ for $n > 1$ are all for LS excitations and are specific to a given valence m of the considered insulator with d^m ions. They have to be determined from the full local Coulomb interaction Hamiltonian (12), in general including also the anisotropy of the $\{U_{\alpha\beta}\}$ and $\{J_{\alpha\beta}\}$ elements.

Effective interactions in a Mott (or charge transfer) insulator with orbital degeneracy take the form of spin-orbital superexchange [4, 18]. Its general structure is given by the sum over all the nearest neighbor bonds $\langle ij \rangle \parallel \gamma$ connecting two transition metal ions and over the excitations n possible for each of them as

$$\mathcal{H} = -\sum_n \frac{t^2}{\varepsilon_n} \sum_{\langle ij \rangle \parallel \gamma} P_{\langle ij \rangle}(\mathcal{S})\, \mathcal{O}^{\gamma}_{\langle ij \rangle}, \tag{36}$$

where $P_{\langle ij \rangle}(\mathcal{S})$ is the projection on the total spin $\mathcal{S} = S \pm \frac{1}{2}$ and $\mathcal{O}^{\gamma}_{\langle ij \rangle}$ is the projection operator on the orbital state at the sites i and j of the bond. Following this general procedure, one finds a spin-orbital model with Heisenberg spin interaction for spins $S = \frac{m}{2}$ of SU(2) symmetry coupled to the orbital operators which have much lower cubic symmetry, with the general structure of spin-orbital superexchange $\propto J$ (1) [8]

$$\mathcal{H}_J = J \sum_{\gamma} \sum_{\langle ij \rangle \parallel \gamma} \left\{ \hat{\mathcal{K}}_{ij}^{(\gamma)} \left(\vec{S}_i \cdot \vec{S}_j + S^2 \right) + \hat{\mathcal{N}}_{ij}^{(\gamma)} \right\}. \tag{37}$$

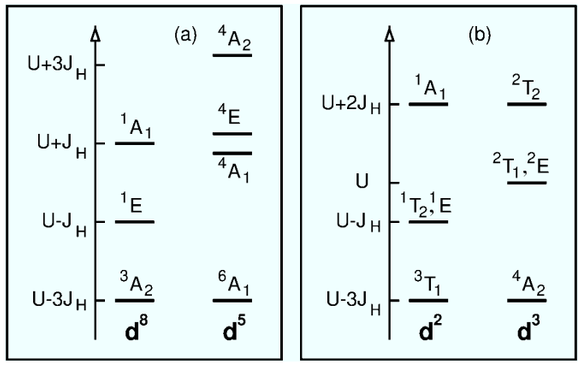


Fig. 5: *Energies of charge excitations ε_n (35) for selected cubic transition metal oxides, for: (a) e_g excitations to Cu^{3+} (d^8) and Mn^{2+} (d^5) ions; (b) t_{2g} excitations to Ti^{2+} (d^2) and V^{2+} (d^3) ions. The splittings between different states are due to Hund's exchange element J_H which refers to a pair of e_g- and t_{2g}-electrons in (a) and (b). (Image reproduced from Ref. [8])*

It connects ions at sites i and j along the bond $\langle ij\rangle \parallel \gamma$ and involves orbital operators, $\hat{\mathcal{K}}_{ij}^{(\gamma)}$ and $\hat{\mathcal{N}}_{ij}^{(\gamma)}$ which depend on the bond direction $\gamma = a, b, c$ for the three *a priori* equivalent directions in a cubic crystal. The spin scalar product, $\left(\vec{S}_i \cdot \vec{S}_j\right)$, is coupled to orbital operators $\hat{\mathcal{K}}_{ij}^{(\gamma)}$ which together with the other "decoupled" orbital operators, $\hat{\mathcal{N}}_{ij}^{(\gamma)}$, determine the orbital state in a Mott insulator. The form of these operators depends on the type of orbital degrees of freedom in a given model. They involve active orbitals on each bond $\langle ij\rangle \parallel \gamma$ along direction γ. Thus the orbital interactions are directional and have only the cubic symmetry of a (perovskite) lattice provided the symmetry in the orbital sector is not broken by other interactions, for instance by crystal-field or Jahn-Teller terms.

The magnetic superexchange constants along each cubic axis J_{ab} and J_c in the effective spin model

$$H = J_{ab} \sum_{\langle ij\rangle \parallel ab} \vec{S}_i \cdot \vec{S}_j + J_c \sum_{\langle ij\rangle \parallel c} \vec{S}_i \cdot \vec{S}_j, \tag{38}$$

are obtained from the spin-orbital model (37) by decoupling spin and orbital operators and next averaging the orbital operators over a given orbital (ordered or disordered) state. It gives effective magnetic exchange interactions: J_c along the c axis, and J_{ab} within the ab planes. The latter J_{ab} ones could in principle still be different between the a and b axes in case of finite lattice distortions due to the Jahn-Teller effect or octahedra tilting, but we limit ourselves to idealized structures with J_{ab} being the same for both planar directions. We show below that the spin-spin correlations along the c axis and within the ab planes

$$s_c = \langle \vec{S}_i \cdot \vec{S}_j \rangle_c, \qquad s_{ab} = \langle \vec{S}_i \cdot \vec{S}_j \rangle_{ab}, \tag{39}$$

next to the orbital correlations, play an important role in the intensity distribution in optical spectroscopy.

In correlated insulators with partly occupied degenerate orbitals not only the structure of the superexchange (37) is complex, but also the optical spectra exhibit strong anisotropy and temperature dependence near the magnetic transitions, as found, e.g., in $LaMnO_3$ [28] or in the cubic vanadates $LaVO_3$ and YVO_3 [29]. In such systems several excitations contribute to the excitation spectra, so one may ask how the spectral weight redistributes between individual subbands originating from these excitations. The spectral weight distribution is in general anisotropic already when orbital order sets in and breaks the cubic symmetry, but even more so when A-type or C-type AF spin order occurs below the Néel temperature $T_{\rm N}$.

At orbital degeneracy the superexchange consists of the terms $H_n^{(\gamma)}(ij)$ as a superposition of individual contributions on each bond $\langle ij\rangle$ due to charge excitation n (35) [19]

$$\mathcal{H} = J \sum_n \sum_{\langle ij\rangle \| \gamma} H_n^{(\gamma)}(ij), \tag{40}$$

with the energy unit for each individual $H_n^{(\gamma)}(ij)$ term given by the superexchange constant J (1). It follows from d–d charge excitations with an effective hopping element t between neighboring transition metal ions and is the same as that obtained in a Mott insulator with nondegenerate orbitals in the regime of $U \gg t$. The spectral weight in the optical spectroscopy is determined by the kinetic energy, and reflects the onset of magnetic order and/or orbital order [19]. In a correlated insulator the electrons are almost localized and the only kinetic energy which is left is associated with the same virtual charge excitations that contribute also to the superexchange. Therefore, the individual kinetic energy terms $K_n^{(\gamma)}$ may be directly determined from the superexchange (40) using the Hellmann-Feynman theorem,

$$K_n^{(\gamma)} = -2J \left\langle H_n^{(\gamma)}(ij)\right\rangle . \tag{41}$$

For convenience, we define here the $K_n^{(\gamma)}$ as positive quantities. Each term $K_n^{(\gamma)}$ (41) originates from a given charge excitation n along a bond $\langle ij\rangle \parallel \gamma$. These terms are directly related to the *partial optical sum rule* for individual Hubbard subbands, which reads [19]

$$\frac{a_0\hbar^2}{e^2} \int_0^\infty \sigma_n^{(\gamma)}(\omega) d\omega = \frac{\pi}{2} K_n^{(\gamma)}, \tag{42}$$

where $\sigma_n^{(\gamma)}(\omega)$ is the contribution of band n to the optical conductivity for polarization along the γ axis, a_0 is the distance between transition metal ions, and a tight-binding model with nearest neighbor hopping is implied. Using Eq. (41) one finds that the intensity of each band is indeed determined by the underlying orbital order together with the spin-spin correlation along the direction corresponding to the polarization.

One has to distinguish the above partial sum rule (42) from the full sum rule for the total spectral weight in optical spectroscopy for polarization along a cubic direction γ, involving

$$K^{(\gamma)} = -2J \textstyle\sum_n \left\langle H_n^{(\gamma)}(ij)\right\rangle, \tag{43}$$

which stands for the total intensity in the optical d–d excitations. This quantity is usually of less interest as it does not allow for a direct insight into the nature of the electronic structure being a sum over several excitations with different energies ε_n (35) and has a much weaker temperature dependence. In addition, it might be also more difficult to deduce from experiment.

3.2 Kugel-Khomskii model for $KCuF_3$ and K_2CuF_4

The simplest and seminal spin-orbital model is obtained when a fermion has two flavors, spin and orbital, and both have two components, i.e., spin and pseudospin are $S = T = \frac{1}{2}$. The physical realization is found in cuprates with degenerate e_g-orbitals, such as $KCuF_3$ or K_2CuF_4 [4], where Cu^{2+} ions are in the d^9 electronic configuration, so charge excitations $d_i^9 d_j^9 \rightleftharpoons d_i^{10} d_j^8$ are made by holes. By considering the degenerate Hubbard model for two e_g-orbitals one finds that d^8 ions have an equidistant multiplet structure, with three excitation energies which differ by $2J_H$ [here J_H stands for J_H^e in Eq. (16)], see Table 2. We emphasize that the correct spectrum has a doubly degenerate energy $(U - J_H)$ and the highest non-degenerate energy is $(U + J_H)$, see Fig. 5(a). Note that this result follows from the diagonalization of the local Coulomb interactions in the relevant subspaces—it reflects the fact that a double occupancy ($|z\uparrow z\downarrow\rangle$ or $|\bar{z}\uparrow\bar{z}\downarrow\rangle$) in either orbital state ($|z\rangle$ or $|\bar{z}\rangle$) is not an eigenstate of the degenerate Hubbard in the atomic limit (17), so the excitation energy U is absent in the spectrum, see Table 2.

The total spin state on the bond corresponds to $\mathcal{S} = 1$ or 0, so the spin projection operators $P_{\langle ij\rangle}(1)$ and $P_{\langle ij\rangle}(0)$ are easily deduced, see Table 2. The orbital configuration which corresponds to a given bond $\langle ij\rangle$ is given by one of the orbital operators in Sec. 2, either $\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$ for the doubly occupied states involving different orbitals, or $\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$ for a double occupancy in a directional orbital at site i or j. This gives a rather transparent structure of one HS and three LS excitations in Table 2. The 3D Kugel-Khomskii (KK) model then follows from Eq. (36) [20,21]

$$\begin{aligned}\mathcal{H}(d^9) &= \sum_{\gamma}\sum_{\langle ij\rangle\parallel\gamma}\Bigg[-\frac{t^2}{U-3J_H}\left(\vec{S}_i\cdot\vec{S}_j+\frac{3}{4}\right)\mathcal{P}^{(\gamma)}_{\langle ij\rangle}+\frac{t^2}{U-J_H}\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\mathcal{P}^{(\gamma)}_{\langle ij\rangle}\\ &\quad+\left(\frac{t^2}{U-J_H}+\frac{t^2}{U+J_H}\right)\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}\Bigg]+E_z\sum_i\tau_i^c. \quad (44)\end{aligned}$$

The last term $\propto E_z$ is the crystal field which splits off the degenerate e_g-orbitals when a Jahn-Teller lattice distortion occurs, and is together with Hund's exchange η a second parameter to

Table 2: *Elements needed for the construction of the Kugel-Khomskii model from charge excitations on the bond $\langle ij\rangle$: excitation n, its type (HS or LS) and energy ε_n, total spin state (triplet or singlet) and the spin projection operator $P_{\langle ij\rangle}(\mathcal{S})$, and the orbital state as well as the corresponding orbital projection operator.*

charge excitation			spin state		orbital state	
n	type	ε_n	$\mathcal{S}$	$P_{\langle ij\rangle}(\mathcal{S})$	orbitals on $\langle ij\rangle \parallel \gamma$	projection
1	HS	$U-3J_H$	1	$\left(\vec{S}_i\cdot\vec{S}_j+\frac{3}{4}\right)$	$\|i\zeta_\gamma\rangle\,\|j\xi_\gamma\rangle\,(\|i\xi_\gamma\rangle\,\|j\zeta_\gamma\rangle)$	$\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$
2	LS	$U-J_H$	0	$-\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)$	$\|i\zeta_\gamma\rangle\,\|j\xi_\gamma\rangle\,(\|i\xi_\gamma\rangle\,\|j\zeta_\gamma\rangle)$	$\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$
3	LS	$U-J_H$	0	$-\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)$	$\|i\zeta_\gamma\rangle\,\|j\zeta_\gamma\rangle$	$\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$
4	LS	$U+J_H$	0	$-\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)$	$\|i\zeta_\gamma\rangle\,\|j\zeta_\gamma\rangle$	$\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$

construct phase diagrams, see below. Here it refers to holes, i.e., large $E_z > 0$ favors hole occupation in $|\bar{z}\rangle \equiv |x^2 - y^2\rangle/\sqrt{2}$ orbitals, as in La_2CuO_4. On the other hand, while $E_z \simeq 0$, both orbitals have almost equal hole density.

Another form of the Hamiltonian (44) is obtained by introducing the coefficients

$$r_1 = \frac{1}{1-3\eta}, \qquad r_2 = r_3 = \frac{1}{1-\eta}, \qquad r_4 = \frac{1}{1+\eta}, \tag{45}$$

and defining the superexchange constant J in the same way as in the $t-J$ model Eq. (1). With the explicit representation of the orbital operators $\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$ and $\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$ in terms of $\left\{\tau^{(\gamma)}_i\right\}$ one finds

$$\begin{aligned}\mathcal{H}(d^9) \;=\; & \frac{1}{2}J\sum_{\gamma}\sum_{\langle ij\rangle\|\gamma}\left\{\left[-r_1\left(\vec{S}_i\cdot\vec{S}_j+\frac{3}{4}\right)+r_2\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\right]\left(\frac{1}{4}-\tau^{(\gamma)}_i\tau^{(\gamma)}_j\right)\right. \\ & \left.+(r_3+r_4)\left(\vec{S}_i\cdot\vec{S}_j-\frac{1}{4}\right)\left(\tau^{(\gamma)}_i+\frac{1}{2}\right)\left(\tau^{(\gamma)}_j+\frac{1}{2}\right)\right\}+E_z\sum_i\tau^c_i. \end{aligned} \tag{46}$$

In the FM state spins are integrated out and one finds from the first term just the superexchange in the e_g-orbital model analyzed before in Sec. 2.

The magnetic superexchange constants J_{ab} and J_c in the effective spin-orbital model (46) are obtained by decoupling spin and orbital operators and next averaging the orbital operators $\langle\hat{\mathcal{K}}^{(\gamma)}_{ij}\rangle$ over the classical state $|\Phi_0\rangle$ as given by Eq. (29). The relevant averages are given in Table 3, and they lead to the following expressions for the superexchange constants in Eq. (38)

$$J_c = \frac{1}{8}J\Big[-r_1\sin^2\theta+(r_2+r_3)(1+\cos\theta)+r_4(1+\cos\theta)^2\Big], \tag{47}$$

$$J_{ab} = \frac{1}{8}J\left[-r_1\left(\frac{3}{4}+\sin^2\theta\right)+(r_2+r_3)\left(1-\frac{1}{2}\cos\theta\right)+r_4\left(\frac{1}{2}-\cos\theta\right)^2\right], \tag{48}$$

which depend on two parameters: J (1) and η (26), and on the orbital order (31) specified by the orbital angle θ. It is clear that the FM term $\propto r_1$ competes with all the other AF LS terms. Nevertheless, in the ab planes, where the occupied hole e_g-orbitals alternate, the larger FM contribution dominates and makes the magnetic superexchange J_{ab} weakly FM ($J_{ab} \lesssim 0$) (when

Table 3: *Averages of the orbital projection operators standing in the spin-orbital interactions in the KK model (46) and determining the spin interactions in H_s (38) for the C-type orbital order of occupied e_g-orbitals which alternate in ab planes, as given by Eqs. (31). Nonequivalent cubic directions are labeled by $\gamma = ab, c$.*

operator	average	ab	c
$\mathcal{Q}^{(\gamma)}_{\langle ij\rangle}$	$2\left\langle\left(\frac{1}{2}-\tau^{(\gamma)}_i\right)\left(\frac{1}{2}-\tau^{(\gamma)}_j\right)\right\rangle$	$\frac{1}{2}\left(\frac{1}{2}-\cos\theta\right)^2$	$\frac{1}{2}\left(1+\cos\theta\right)^2$
$\mathcal{P}^{(\gamma)}_{\langle ij\rangle}$	$\left\langle\frac{1}{4}-\tau^{(\gamma)}_i\tau^{(\gamma)}_j\right\rangle$	$\frac{1}{4}\left(\frac{3}{4}+\sin^2\theta\right)$	$\frac{1}{4}\sin^2\theta$
$\mathcal{R}^{(\gamma)}_{\langle ij\rangle}$	$2\left\langle\left(\frac{1}{2}+\tau^{(\gamma)}_i\right)\left(\frac{1}{2}+\tau^{(\gamma)}_j\right)\right\rangle$	$\frac{1}{2}\left(\frac{1}{2}+\cos\theta\right)^2$	$\frac{1}{2}\left(1-\cos\theta\right)^2$

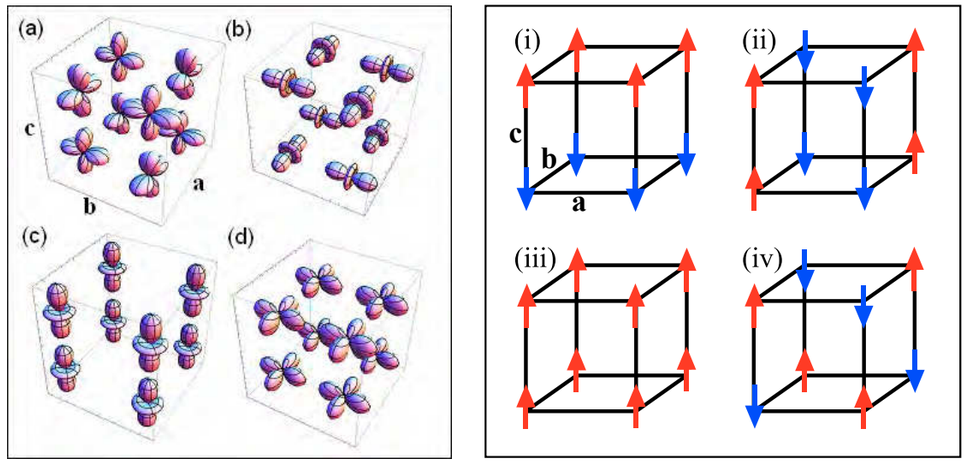


Fig. 6: *Left: schematic view of the four simplest orbital configurations on a representative cube of the 3D lattice: (a) AO order with* $\langle \tau_i^{a(b)} \rangle = \pm\frac{1}{2}$ *changing from site to site and* $\langle \tau_i^c \rangle = \frac{1}{4}$*, obtained for* $E_z < 0$*, (b) AO order with* $\langle \tau_i^{a(b)} \rangle = -\frac{1}{2}$ *changing from site to site and* $\langle \tau_i^c \rangle = -\frac{1}{4}$*, obtained for* $E_z > 0$*, (c) FO order with occupied* z*-orbitals and* $\langle \tau_i^c \rangle = \frac{1}{2}$ *(cigar-shaped orbitals), and (d) FO order with occupied* $\bar{z}$*-orbitals and* $\langle \tau_i^c \rangle = -\frac{1}{2}$ *(clover-shaped orbitals). Right: schematic view of four spin configurations (arrows stand for up or down spins) in phases with spin order: (i) A-AF, (ii) C-AF, (iii) FM, and (iv) G-AF. (Images reproduced from Ref. [24])*

$\sin^2\theta \simeq 1$), while the stronger AF superexchange along the c axis ($J_c \gg |J_{ab}|$) favors quasi one-dimensional (1D) spin fluctuations. Thus $KCuF_3$ exhibits spinon excitations for $T > T_{\rm N}$.

Consider first the 2D KK model on a square lattice, with $\gamma = a, b$ in Eq. (46), as in K_2CuF_4. In the absence of Hund's exchange, interactions between $S = \frac{1}{2}$ spins are AF. However, they are quite different depending on which of the two e_g-orbitals are occupied by holes: $J_{ab}^z = \frac{1}{16}J$ for $|z\rangle$ and $J_{ab}^{\bar{z}} = \frac{9}{16}J$ for $|\bar{z}\rangle$ hole orbitals. As a result, the AF phases with spin order in Fig. 6(iv) and the FO order shown in Figs. 6(c) and 6(d) are degenerate at finite crystal field $E_z = -\frac{1}{2}J$. This defines a quantum critical point $Q_{\rm 2D} = (-0.5, 0)$ in the $(E_z/J, \eta)$ plane. Actually, at this point also one more phase has the same energy—the FM spin phase of Fig. 6(i) with AO order of $|\pm\rangle$ orbitals (28) shown in Fig. 6(a) [21].

To capture the corrections due to quantum fluctuations, one may construct a plaquette mean field approximation or entanglement renormalization *ansatz* (ERA) [22]. One finds important corrections to a mean field phase diagram near the quantum critical point $Q_{\rm 2D}$, and a plaquette valence bond (PVB) state is stable in between the above three phases with long range order, with spin singlets on the bonds $\parallel a$ ($\parallel b$), stabilized by the directional orbitals $|\zeta_a\rangle$ ($|\zeta_b\rangle$). A novel ortho-AF phase appears as well when the magnetic interactions change from AF to FM ones due to increasing Hund's exchange η, and for $E_z/J < -1.5$, see Fig. 7(a). Since the nearest neighbor magnetic interactions are very weak, exotic four-sublattice ortho-AF spin order is stabilized by second and third nearest neighbor interactions, shown in Fig. 7(b). Such further neighbor interactions follow from spin-orbital excitations shown in Fig. 7(c). Note that both

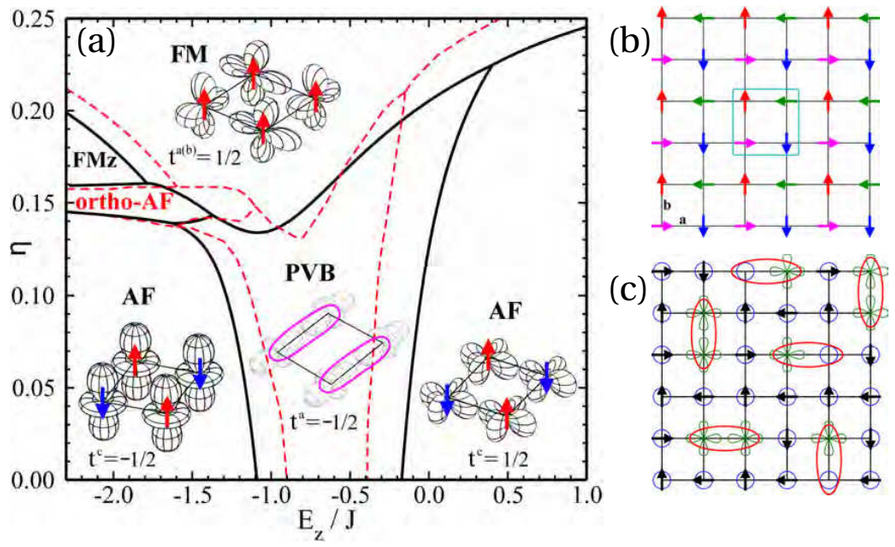


Fig. 7: *Spin-orbital phase diagram and entanglement in the 2D KK model: (a) phase diagram in the plaquette mean field (solid lines) and ERA (dashed lines) variational approximation, with insets showing representative spin and orbital configurations on a* 2×2 *plaquette — $\bar{z}$-like* $\left(t^c = -\langle \tau_i^c \rangle = \frac{1}{2}\right)$ *and z-like* $\left(t^{a,c} = -\langle \tau_i^{c(a)} \rangle = -\frac{1}{2}\right)$ *orbitals are accompanied either by AF long range order (arrows) or by spin singlets on bonds in the PVB phase; (b) view of an exotic four-sublattice ortho-AF phase near the onset of FM (or FMz) phase; (c) artist's view of the ortho-AF phase — spin singlets (ovals) are entangled with either one or two orbital excitations* $|z\rangle \to |\bar{z}\rangle$ *(clovers). (Images reproduced from Ref. [22])*

approximate methods employed in Ref. [22] (plaquette mean field approximation and ERA) give very similar range of stability of ortho-AF phase.

In the 3D KK model the exchange interaction in the ab planes (48) and along the c axis (47) are exactly balanced at the orbital degeneracy ($E_z = 0$) and the quantum critical point where several classical phases meet in mean field approximation is $Q_{\rm 3D} = (0, 0)$, see Fig. 8(a). While finite E_z favors one or the other G-AF phase, finite Hund's exchange η favors AO order stabilizing A-AF spin order, see Fig. 6(i). This phase is indeed found in $KCuF_3$ at low temperature $T < T_{\rm N}$ and is also obtained from electronic structure calculations [23]. We remark that for unrealistically large $\eta > 0.2$, spin order changes to FM.

Large qualitative changes in the phase diagram are found when spin correlations on bonds are treated in cluster mean field approximation (using plaquettes or linear clusters [24]), see Fig. 8(b). Phases with long range spin order (G-AF, A-AF, and FM) are again separated by exotic types of magnetic order which arise by a similar mechanism to that described above for an ab monolayer, i.e., nearest neighbor exchange changes sign along one cubic direction. Near the QCP $Q_{\rm 3D}$ one finds again a PVB phase, as in the 2D KK model. In addition to the phase diagram of Fig. 7(a), the transitions between G-AF and PVB phases are continuous and mixed PVB-AF phases arise.

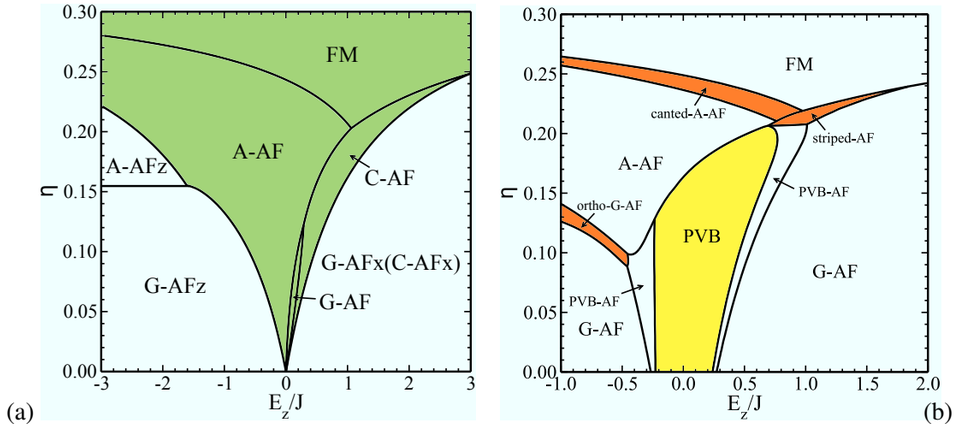


Fig. 8: *Phase diagram of the 3D KK model obtained in two mean field methods: (a) the single-site mean field, and (b) the cluster mean field. The shaded (green) area indicates phases with AO order while the remaining magnetic phases are accompanied by FO order with fully polarized orbitals, either $\bar{z}$ (x) (for $E_z > 0$) or z (for $E_z < 0$). In this approach a plaquette valence-bond (PVB) phase with alternating spin singlets in the ab planes (yellow) separates the phases with magnetic long range order, see Fig. 6. Phases with exotic magnetic order are shown in orange. Note the different ranges of E_z/J shown. (Images reproduced from Ref. [24])*

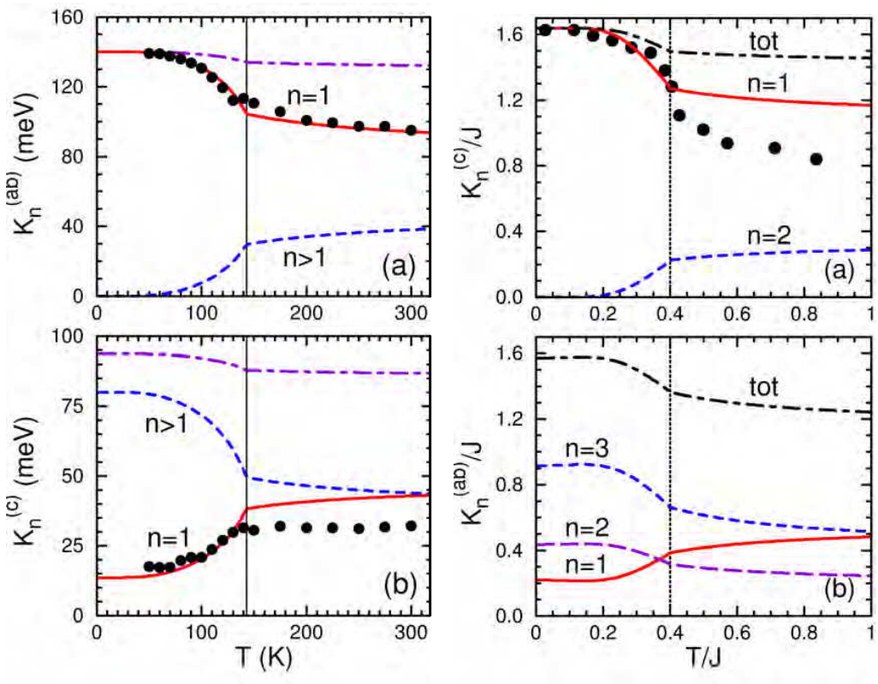
3.3 Spin-orbital superexchange model for $LaMnO_3$

Electronic structure calculations give A-AF spin order, in agreement with experiment. It follows from the spin-orbital superexchange for spins $S = 2$ in $LaMnO_3$, $\mathcal{H}_e$, due to the excitations involving e_g-electrons. The energies of the five possible excited states [9] shown in Fig. 5(a) are: (i) the HS ($\mathcal{S} = \frac{5}{2}$) $^6A_1$ state, and (ii) the LS ($\mathcal{S} = \frac{3}{2}$) states: $^4A_1$, ^4E ($^4E_\epsilon$, $^4E_\theta$), and $^4A_2$, will be parameterized again by the intraorbital Coulomb element U and by Hund's exchange J_H^e between a pair of e_g-electrons in a Mn^{2+} (d^5) ion, defined in Eq. (16). The Racah parameters $B = 0.107$ eV and $C = 0.477$ eV justify an approximate relation $C \simeq 4B$, and we find the LS excitation spectrum: $\varepsilon(^4A_1) = U + \frac{3}{4}J_H$, $\varepsilon(^4E) = U + \frac{5}{4}J_H$ (twice), and $\varepsilon(^4A_2) = U + \frac{13}{4}J_H$. Using the spin algebra (Clebsch-Gordan coefficients) and considering again two possible e_g-orbital configurations, see Eqs. (22) and (23), and charge excitations by t_{2g}-electrons, one finds a compact expression [25],

$$\mathcal{H}_e = \frac{1}{16}\sum_\gamma \sum_{\langle ij\rangle\parallel\gamma} \left\{ -\frac{8}{5}\frac{t^2}{\varepsilon(^6A_1)}\left(\vec{S}_i\cdot\vec{S}_j+6\right)\mathcal{P}_{\langle ij\rangle}^{(\gamma)} + \left[\frac{t^2}{\varepsilon(^4E)}+\frac{3}{5}\frac{t^2}{\varepsilon(^4A_1)}\right]\left(\vec{S}_i\cdot\vec{S}_j-4\right)\mathcal{P}_{\langle ij\rangle}^{(\gamma)} + \left[\frac{t^2}{\varepsilon(^4E)}+\frac{t^2}{\varepsilon(^4A_2)}\right]\left(\vec{S}_i\cdot\vec{S}_j-4\right)\mathcal{Q}_{\langle ij\rangle}^{(\gamma)}\right\} + E_z\sum_i \tau_i^c. \tag{49}$$

$$\mathcal{H}_t = \frac{1}{8}J\beta r_t\left(\vec{S}_i\cdot\vec{S}_j-4\right). \tag{50}$$

Here $\beta = (t_\pi/t)^2$ follows from the difference between the effective d–d hopping elements along the σ and π bonds, i.e., $\beta \simeq \frac{1}{9}$, while the coefficient r_t stands for a superposition of all



(A) (B)

Fig. 9: *Kinetic energies per bond $K_n^{(\gamma)}$ (41) for increasing temperature T obtained from the respective spin-orbital models for FM (top) and AF (bottom) bonds along the axis γ:*
(A) $LaMnO_3$ (with $J = 150$ meV, $\eta \simeq 0.18$ [8], end experimental points [28]);
(B) $LaVO_3$ with $\eta = 0.13$ [19] and experimental points [29].
The kinetic energies in HS states ($n = 1$, red lines) are compared with the experiment (filled circles). Vertical dotted lines indicate the value of T_N. (Images reproduced from Ref. [8])

t_{2g} excitations involved in the t_{2g} superexchange [8]. Note that spin-projection operators for high (low) total spin $\mathcal{S} = 2$ ($\mathcal{S} = 1$) cannot be used, but again the HS term stands for a FM contribution which dominates over the other LS terms when $\langle \mathcal{P}_{\langle ij \rangle}^{(\gamma)} \rangle \simeq 1$. Charge excitations by t_{2g}-electrons give double occupancies in active t_{2g}-orbitals, so $\mathcal{H}_t$ is AF but this term is small—as a result FM interactions may dominate but again only along two spatial directions. Indeed, this happens for the realistic parameters of $LaMnO_3$ for the ab planes where spin order is FM and coexists with AO order, while along the c axis spin order is AF accompanied by FO order, i.e., spin-orbital order is A-AF/C-AF. Indeed, this type of order is found both from the theory for realistic parameters and from the electronic structure calculations [26]. One concludes that Jahn-Teller orbital interactions are responsible for the enhanced value of the orbital transition temperature [27].

The optical spectral weight due to HS states in $LaMnO_3$ may be easily derived from the present model (49), following the general theory, see Eq. (41). One finds a very satisfactory agree-

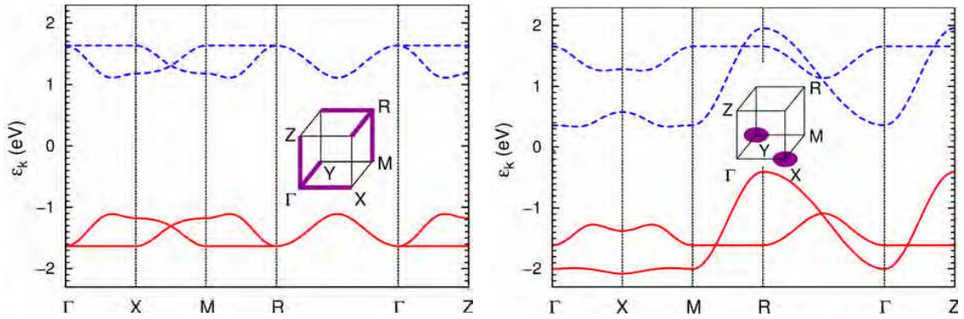


Fig. 10: *Band structure along the high symmetry directions in: (a) G-AF phase at $x = 0$ and (b) C-AF phase at $x = 0.05$. Spin majority (minority) bands are shown by solid (dashed) lines. Parameters: $t = 0.4$ eV, $J_H = 0.74$ eV, $g = 3$ eV. Insets shows the Fermi surfaces at low doping. Special k-points: $\Gamma = (0, 0, 0)$, $X = (\pi, 0, 0)$, $M = (\pi, \pi, 0)$, $R = (\pi, \pi, \pi)$, $Z = (0, 0, \pi)$. (Images reproduced from Ref. [30])*

ment between the present theory and the experimental results of [28], as shown in Fig. 9(A). We emphasize, that no fit is made here, i.e., the kinetic energies (41) are calculated using the same parameters as those used for the magnetic exchange constants [8]. Therefore, such a good agreement with experiment suggests that indeed the spin-orbital superexchange may be disentangled, as also verified later [27].

To give an example of a phase transition triggered by e_g-electron doping of $Sr_{1-x}La_xMnO_3$ we show the results obtained with a double exchange model for degenerate e_g-electrons extended by the coupling to the lattice [30],

$$\mathcal{H} = -\sum_{ij,\alpha\beta,\sigma} t^{ij}_{\alpha\beta} a^{\dagger}_{i\alpha\sigma} a_{j\beta\sigma} - 2J_H \sum_i \vec{S}_i \cdot \vec{s}_i + J \sum_{\langle ij \rangle} \vec{S}_i \cdot \vec{S}_j - gu \sum_i (n_{iz} - n_{i\bar{z}}) + \frac{1}{2} NKu^2. \quad (51)$$

It includes the hopping of e_g-electrons between orbitals $\alpha = z, \bar{z}$ as in Eq. (6). The tetragonal distortion u is finite only in the C-AF phase. Here we define it as proportional to a difference between two lattice constants a and c along the respective axis, $u \equiv 2(c - a)/(c + a)$, and N is the number of lattice sites. The microscopic model that explains the mechanism of the magnetic transition in electron doped manganites from canted G-AF to collinear C-AF phase at low doping $x \simeq 0.04$. The double exchange supported by the cooperative Jahn-Teller effect leads then to dimensional reduction from an isotropic 3D G-AF phase to a quasi-1D order of partly occupied $3z^2 - r^2$-orbitals in the C-AF phase [30]. We emphasize that this theory prediction relies on the shape of the Fermi surface which is radically different in the G-AF and the C-AF phase. Due to the Fermi surface topology, spin canting is suppressed in the C-AF phase, in agreement with the experiment.

4 Superexchange for active t_{2g} orbitals

4.1 Spin-orbital superexchange model for $LaTiO_3$

$LaTiO_3$ would be the electron-hole symmetric compound to $KCuF_3$, if not the orbital degree of freedom was t_{2g} here. This changes the nature of orbital operators from the projections for each bond to scalar products of pseudospin $T = \frac{1}{2}$ operators. The superexchange spin-orbital model (37) in the perovskite titanates couples $S = \frac{1}{2}$ spins and $T = \frac{1}{2}$ pseudospins arising from the t_{2g} orbital degrees of freedom at nearest neighbor Ti^{3+} ions, e.g., in $LaTiO_3$ or $YTiO_3$ [6]. Due to the large intraorbital Coulomb element U electrons localize and the densities satisfy the local constraint at each site i,

$$n_{ia} + n_{ib} + n_{ic} = 1. \tag{52}$$

The charge excitations lead to one of four different excited states [9], shown in Fig. 5(b):
(i) the high-spin ${}^3T_1$ state at energy $U - 3J_H$, and
(ii) three low-spin states — degenerate ${}^1T_2$ and ${}^1E$ states at energy $(U - J_H)$, and
(iii) an ${}^1A_1$ state at energy $(U + 2J_H)$.
As before, the excitation energies are parameterized by η, defined by Eq. (26), and we introduce the coefficients

$$r_1 = \frac{1}{1-3\eta}, \qquad r_2 = \frac{1}{1-\eta}, \qquad r_3 = \frac{1}{1+2\eta}. \tag{53}$$

One finds the following compact expressions for the terms contributing to superexchange $\mathcal{H}_J(d^1)$, Eq. (40) [6],

$$H_1^{(\gamma)} = \frac{1}{2}Jr_1\left(\vec{S}_i\cdot\vec{S}_j + \frac{3}{4}\right)\left(A_{ij}^{(\gamma)} - \frac{1}{2}n_{ij}^{(\gamma)}\right), \tag{54}$$

$$H_2^{(\gamma)} = \frac{1}{2}Jr_2\left(\vec{S}_i\cdot\vec{S}_j - \frac{1}{4}\right)\left(A_{ij}^{(\gamma)} - \frac{2}{3}B_{ij}^{(\gamma)} + \frac{1}{2}n_{ij}^{(\gamma)}\right), \tag{55}$$

$$H_3^{(\gamma)} = \frac{1}{3}Jr_3\left(\vec{S}_i\cdot\vec{S}_j - \frac{1}{4}\right)B_{ij}^{(\gamma)}, \tag{56}$$

where

$$A_{ij}^{(\gamma)} = 2\left(\vec{\tau}_i\cdot\vec{\tau}_j + \frac{1}{4}n_in_j\right)^{(\gamma)}, \quad B_{ij}^{(\gamma)} = 2\left(\vec{\tau}_i\otimes\vec{\tau}_j + \frac{1}{4}n_in_j\right)^{(\gamma)}, \quad n_{ij}^{(\gamma)} = n_i^{(\gamma)} + n_j^{(\gamma)}. \tag{57}$$

As in Sec. 3.2, the orbital (pseudospin) operators $\left\{A_{ij}^{(\gamma)}, B_{ij}^{(\gamma)}, n_{ij}^{(\gamma)}\right\}$ depend on the direction of the $\langle ij\rangle \parallel \gamma$ bond. Their form follows from two active t_{2g}-orbitals (flavors) along the cubic axis γ, e.g., for $\gamma = c$ the active orbitals are a and b, and they give two components of the pseudospin $T = \frac{1}{2}$ operator $\vec{\tau}_i$. The operators $\left\{A_{ij}^{(\gamma)}, B_{ij}^{(\gamma)}\right\}$ describe the interactions between these two active orbitals, which include the quantum fluctuations, and take either the form of a scalar product $\vec{\tau}_i\cdot\vec{\tau}_j$ in $A_{ij}^{(\gamma)}$, or lead to a similar expression

$$\vec{\tau}_i\otimes\vec{\tau}_j = \tau_i^x\tau_j^x - \tau_i^y\tau_j^y + \tau_i^z\tau_i^z \tag{58}$$

in $B_{ij}^{(\gamma)}$. These latter terms enhance orbital fluctuations by double excitations due to the $\tau_i^+\tau_j^+$ and $\tau_i^-\tau_j^-$ terms. The interactions along the axis γ are tuned by the number of electrons occupying active orbitals, $n_i^{(\gamma)}=1-n_{i\gamma}$, which is fixed by the number of electrons in the inactive orbital $n_{i\gamma}$ by the constraint (52). The cubic titanates are known to have particularly pronounced quantum spin-orbital fluctuations [18], and their proper treatment requires a rather sophisticated approach. Therefore, in contrast to the AF long range order found in e_g-orbital systems, spin-orbital disordered state may occur in titanium perovskites, as suggested for $LaTiO_3$ [6].

4.2 Spin-orbital superexchange model for $LaVO_3$

As the last cubic system we present the spin-orbital model for V^{3+} ions in d^2 configurations in the vanadium perovskite RVO$_3$ (R=La,...,Lu). Due to Hund's exchange one has $S=1$ spins and three ($n=1,2,3$) charge excitations ε_n arising from the transitions to [see Fig. 5(b)]:
(i) a high-spin state $^4A_2$ at energy $(U-3J_H)$,
(ii) two degenerate low-spin states $^2T_1$ and ^2E at U, and
(iii) a $^2T_2$ low-spin state at $(U+2J_H)$ [31].
Using η (26) we parameterize this multiplet structure by

$$r_1=\frac{1}{1-3\eta},\qquad r_3=\frac{1}{1+2\eta}. \tag{59}$$

The cubic symmetry is broken and the crystal field induces orbital splitting in RVO$_3$, hence $\langle n_{ic}\rangle=1$ and the orbital degrees of freedom are given by the doublet $\{a,b\}$ which defines the pseudospin operators $\vec{\tau}_i$ at site i. One derives a HS contribution $H_1^{(c)}(ij)$ for a bond $\langle ij\rangle$ along the c axis, and $H_1^{(ab)}(ij)$ for a bond in the ab plane

$$H_1^{(c)}(ij) = -\frac{1}{3}Jr_1\left(\vec{S}_i\cdot\vec{S}_j+2\right)\left(\tfrac{1}{4}-\vec{\tau}_i\cdot\vec{\tau}_j\right), \tag{60}$$

$$H_1^{(ab)}(ij) = -\frac{1}{6}Jr_1\left(\vec{S}_i\cdot\vec{S}_j+2\right)\left(\tfrac{1}{4}-\tau_i^z\tau_j^z\right). \tag{61}$$

In Eq. (60) the pseudospin operators $\vec{\tau}_i$ describe the low-energy dynamics of (initially degenerate) $\{xz,yz\}$ orbital doublet at site i; this dynamics is quenched in $H_1^{(ab)}$ (61). Here $\frac{1}{3}(\vec{S}_i\cdot\vec{S}_j+2)$ is the projection operator on the HS state for $S=1$ spins. The terms $H_n^{(c)}(ij)$ for LS excitations ($n=2,3$) contain instead the spin operator $(1-\vec{S}_i\cdot\vec{S}_j)$ (which guarantees that these terms cannot contribute for fully polarized spins $\langle\vec{S}_i\cdot\vec{S}_j\rangle=1$)

$$\begin{aligned}H_2^{(c)}(ij) &= -\frac{1}{12}J\left(1-\vec{S}_i\cdot\vec{S}_j\right)\left(\tfrac{7}{4}-\tau_i^z\tau_j^z-\tau_i^x\tau_j^x+5\tau_i^y\tau_j^y\right),\\ H_3^{(c)}(ij) &= -\frac{1}{4}Jr\left(1-\vec{S}_i\cdot\vec{S}_j\right)\left(\tfrac{1}{4}+\tau_i^z\tau_j^z+\tau_i^x\tau_j^x-\tau_i^y\tau_j^y\right),\end{aligned} \tag{62}$$

while again the terms $H_n^{(ab)}(ij)$ differ from $H_n^{(c)}(ij)$ only by orbital operators

$$\begin{aligned}H_2^{(ab)}(ij) &= -\frac{1}{8}J\left(1-\vec{S}_i\cdot\vec{S}_j\right)\left(\tfrac{19}{12}\mp\tfrac{1}{2}\tau_i^z\mp\tfrac{1}{2}\tau_j^z-\tfrac{1}{3}\tau_i^z\tau_j^z\right),\\ H_3^{(ab)}(ij) &= -\frac{1}{8}Jr\left(1-\vec{S}_i\cdot\vec{S}_j\right)\left(\tfrac{5}{4}\mp\tfrac{1}{2}\tau_i^z\mp\tfrac{1}{2}\tau_j^z+\tau_i^z\tau_j^z\right),\end{aligned} \tag{63}$$

where upper (lower) sign corresponds to bonds along the $a(b)$ axis.
First we present a mean field approximation for the spin and orbital bond correlations which are determined self-consistently after decoupling them from each other in $\mathcal{H}_J$ (37). Spin interactions in Eq. (38) are given by two exchange constants

$$\begin{aligned} J_c &= \frac{1}{2}J\left\{\eta r_1-(r_1-\eta r_1-\eta r_3)(\tfrac{1}{4}+\langle\vec{\tau}_i\cdot\vec{\tau}_j\rangle)-2\eta r_3\langle\tau_i^y\tau_j^y\rangle\right\}, \\ J_{ab} &= \frac{1}{4}J\left\{1-\eta r_1-\eta r_3+(r_1-\eta r_1-\eta r_3)(\tfrac{1}{4}+\langle\tau_i^z\tau_j^z\rangle)\right\}, \end{aligned} \tag{64}$$

determined by orbital correlations $\langle\vec{\tau}_i\cdot\vec{\tau}_j\rangle$ and $\langle\tau_i^\alpha\tau_j^\alpha\rangle$. By evaluating them one finds $J_c<0$ and $J_{ab}>0$ supporting C-AF spin order. In the orbital sector one finds

$$H_\tau=\sum_{\langle ij\rangle_c}\left[J_c^\tau\vec{\tau}_i\cdot\vec{\tau}_j-J(1-s_c)\eta r_3\tau_i^y\tau_j^y\right]+J_{ab}^\tau\sum_{\langle ij\rangle_{ab}}\tau_i^z\tau_j^z, \tag{65}$$

with

$$\begin{aligned} J_c^\tau &= \frac{1}{2}J\left[(1+s_c)r_1+(1-s_c)\eta(r_1+r_3)\right], \\ J_{ab}^\tau &= \frac{1}{4}J\left[(1-s_{ab})r_1+(1+s_{ab})\eta(r_1+r_3)\right], \end{aligned} \tag{66}$$

depending on spin correlations: $s_c=\langle\vec{S}_i\cdot\vec{S}_j\rangle_c$ and $s_{ab}=-\langle\vec{S}_i\cdot\vec{S}_j\rangle_{ab}$. In a classical C-AF state ($s_c=s_{ab}=1$) this mean field procedure becomes exact, and the orbital problem maps to Heisenberg pseudospin chains along the c axis, weakly coupled (as $\eta\ll 1$) along a and b bonds

$$H_\tau^{(0)}=Jr_1\left[\sum_{\langle ij\rangle_c}\vec{\tau}_i\cdot\vec{\tau}_j+\frac{1}{2}\eta\left(1+\frac{r_3}{r_1}\right)\sum_{\langle ij\rangle_{ab}}\tau_i^z\tau_j^z\right], \tag{67}$$

releasing large zero-point energy. Thus, spin C-AF and G-AO order with quasi-1D orbital quantum fluctuations support each other in RVO$_3$. Orbital fluctuations play here a prominent role and amplify the FM exchange J_c, making it even stronger than the AF exchange J_{ab} [31].
Having the individual terms $H_n^{(\gamma)}$ of the spin-orbital model, one may derive the spectral weights of the optical spectra (41). The HS excitations have a remarkable temperature dependence and the spectral weight decreases in the vicinity of the magnetic transition at $T_{\rm N}$, see Fig. 9(B). The observed behavior is reproduced in the theory only when spin-orbital interactions are treated in a cluster approach, i.e., they *cannot* be disentangled, see Sec. 5.2.
Unlike in LaMnO$_3$ where the spin and orbital phase-transitions are well separated, in the RVO$_3$ (R=Lu,Yb,...,La) the two transitions are close to each other [33]. It is not easy to reproduce the observed dependence of the transition temperatures $T_{\rm OO}$ and Néel T_{N1} on the ionic radius r_R (in the RVO$_3$ compounds with small r_R there is also another magnetic transition at T_{N2} [34] which is not discussed here). The spin-orbital model was extended by the coupling to the lattice to unravel a nontrivial interplay between superexchange, the orbital-lattice coupling due to the GdFeO$_3$-like rotations of the VO$_6$ octahedra, and orthorhombic lattice distortions [32]. One finds that the lattice strain affects the onset of the magnetic and orbital order by partial

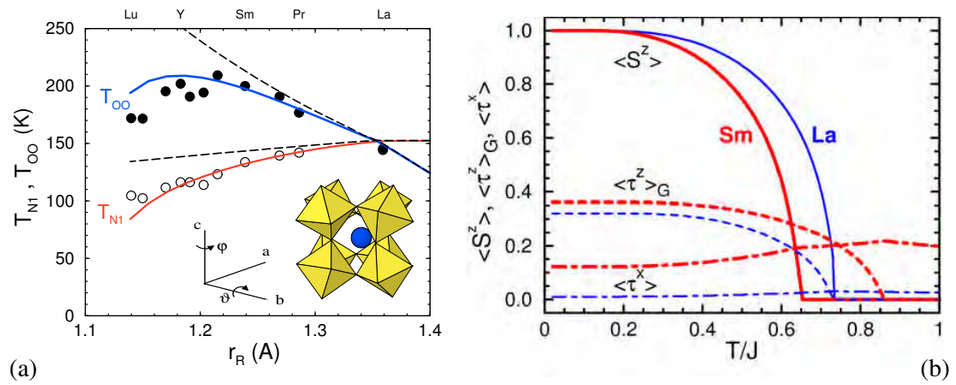


Fig. 11: *Phase transitions in the vanadium perovskites RVO_3: (a) phase diagram with the orbital $T_{\rm OO}$ and Néel T_{N1} transition temperature obtained from theory with and without orbital-lattice coupling (solid and dashed lines) [32], and from experiment (circles) [33]; (b) spin $\langle S_i^z\rangle$ (solid) and G-type orbital $\langle \tau_i^z\rangle_G$ (dashed) order parameters, vanishing at $T_{\rm OO}$ and T_{N1}, and the transverse orbital polarization $\langle \tau_i^x\rangle$ (dashed-dotted lines) for $LaVO_3$ and $SmVO_3$ (thin and heavy lines). (Images reproduced from Ref. [32])*

suppression of orbital fluctuations, and the dependence of $T_{\rm OO}$ is non-monotonous in Fig. 11(a). Thereby the orbital polarization $\propto \langle \tau^x\rangle$ increases with decreasing ionic radius r_R, and the value of T_{N1} is reduced, see Fig. 11(b). The theoretical approach demonstrates that orbital-lattice coupling is very important and reduces both $T_{\rm OO}$ and Néel T_{N1} for small ionic radii.

5 Spin-orbital complementarity and entanglement

5.1 Goodenough-Kanamori rules

While a rather advanced many-body treatment of the quantum physics characteristic for spin-orbital models is required in general, we want to present here certain simple principles which help to understand the heart of the problem and to give simple guidelines for interpreting experiments and finding relevant physical parameters of the spin-orbital models of *undoped* cubic insulators. We will argue that such an approach based upon classical orbital order is well justified in many known cases, as quantum phenomena are often quenched by the Jahn-Teller (JT) coupling between orbitals and the lattice distortions, which are present below structural phase transitions and induce orbital order both in spin-disordered and in spin-ordered or spin-liquid phases.

From the derivation of the Kugel-Khomskii model in Sec. 3.2, we have seen that pairs of directional orbitals on neighboring ions $\{|i\zeta_\gamma\rangle, |j\zeta_\gamma\rangle\}$ favor AF spin order while pairs of orthogonal orbitals such as $\{|i\zeta_\gamma\rangle, |j\xi_\gamma\rangle\}$ favor FM spin order. This is generalized by the classical Goodenough-Kanamori rules (GKR) [35] that state that AF spin order is accompanied by FO

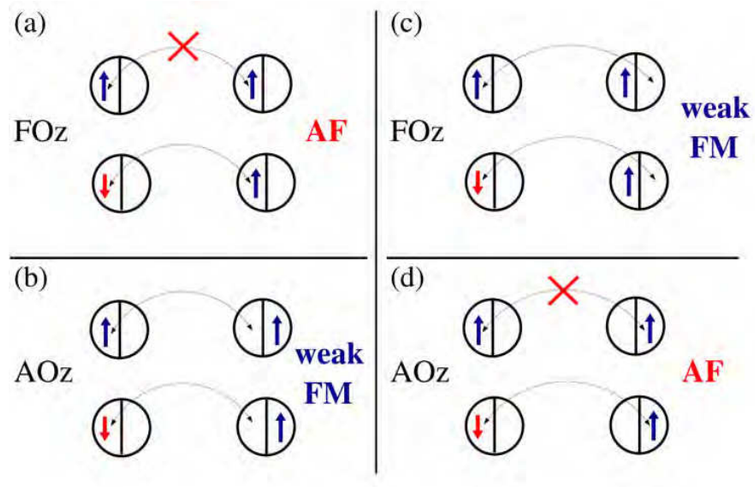


Fig. 12: *Artist's view of the GKR [35] for: (a) FOz and AF spin order and (b) AOz and FM spin order in a system with orbital flavor conserving hopping as is alkali RO_2 hyperoxides (R=K,Rb,Cs) [36]. The charge excitations generated by interorbital hopping fully violate the GKR and support the states with the same spin-orbital order: (c) FOz and FM spin order and (d) AOz and AF spin order. (Image reproduced from Ref. [36])*

order, while FM spin order is accompanied by AO order, see Figs. 12(a) and 12(b). Indeed, these rules emphasizing the complementarity of spin-orbital correlations are frequently employed to explain the observed spin-orbital order in several systems, particularly in those where spins are large, like in CMR manganites [3]. They agree with the general structure of spin-orbital superexchange in the Kugel-Khomskii model where it is sufficient to consider the flavor-conserving hopping between pairs of directional orbitals $\{|i\zeta_\gamma\rangle, |j\zeta_\gamma\rangle\}$. The excited states are then double occupancies in one of the directional orbitals while no effective interaction arises for two parallel spins (in triplet states), so the superexchange is AF. In contrast, for a pair of orthogonal orbitals, e.g., $\{|i\zeta_\gamma\rangle, |j\xi_\gamma\rangle\}$, two different orbitals are singly occupied and the FM term is stronger than the AF one as the excitation energy is lower. Therefore, configurations with AO order support FM spin order.

The above complementarity of spin-orbital order is frustrated by interorbital hopping, or may be modified by spin-orbital entanglement, see below. In such cases the order in both channels could be the same, either FM/FO, see Fig. 12(c), or AF/AO, see Fig. 12(d). Again, when different orbitals are occupied in the excited state, the spin superexchange is weak FM and when the same orbital is doubly occupied, the spin superexchange is stronger and AF. The latter AF exchange coupling dominates because antiferromagnetism, which is due to the Pauli principle, does not have to compete here with ferromagnetism. On the contrary, FM exchange is caused by the energy difference $\propto \eta$ between triplet and singlet excited states with two different orbitals occupied.

The presented modification of the GKR is of importance in alkali RO_2 hyperoxides (R=K,Rb,Cs) [36]. The JT effect is crucial for this generalization of the GKR—without it large interorbital hopping orders the T^x-orbital-mixing pseudospin component instead of the T^z component in a single plane. More generally, such generalized GKR can arise whenever the orbital order on a bond is not solely stabilized by the same spin-orbital superexchange interaction that determines the spin exchange. On a geometrically frustrated lattice, another route to this behavior can occur when the ordered orbital component preferred by superexchange depends on the direction and the relative strengths fulfill certain criteria.

5.2 Spin-orbital entanglement

A quantum state consisting of two different parts of the Hilbert space is entangled if it cannot be written as a product state. Similar to it, two operators are entangled if they give entangled states, i.e., they cannot be factorized into parts belonging to different subspaces. This happens precisely in spin-orbital models and is the source of spin-orbital entanglement [37].
To verify whether entanglement occurs it suffices to compute and analyze the spin, orbital, and spin-orbital (four-operator) correlation functions for a bond $\langle ij\rangle$ along γ axis, given, respectively, by

$$S_{ij} \equiv \frac{1}{d}\sum_n \langle n|\,\vec{S}_i\cdot\vec{S}_j\,|n\rangle\ , \tag{68}$$

$$T_{ij} \equiv \frac{1}{d}\sum_n \left\langle n\left|(\vec{T}_i\cdot\vec{T}_j)^{(\gamma)}\right|n\right\rangle\ , \tag{69}$$

$$\begin{aligned} C_{ij} &\equiv \frac{1}{d}\sum_n \left\langle n\left|(\vec{S}_i\cdot\vec{S}_j-S_{ij})(\vec{T}_i\cdot\vec{T}_j-T_{ij})^{(\gamma)}\right|n\right\rangle \\ &= \frac{1}{d}\sum_n \left\langle n\left|(\vec{S}_i\cdot\vec{S}_j)(\vec{T}_i\cdot\vec{T}_j)^{(\gamma)}\right|n\right\rangle -\frac{1}{d}\sum_n \left\langle n\left|\vec{S}_i\cdot\vec{S}_j\right|n\right\rangle \frac{1}{d}\sum_m \left\langle m\left|(\vec{T}_i\cdot\vec{T}_j)^{(\gamma)}\right|m\right\rangle\ , \end{aligned} \tag{70}$$

where d is the ground state degeneracy, and the pseudospin scalar product in Eqs. (69) and (70) is relevant for a model with active t_{2g} orbital degrees of freedom. As a representative example we evaluate here such correlations for a 2D spin-orbital model derived for a $NaTiO_2$ plane [39], with the local constraint (52) as in $LaTiO_3$; other situations with spin-orbital entanglement are discussed in Ref. [37].
To explain the physical origin of the spin-orbital model for $NaTiO_2$ [39] we consider a representative bond along the c axis shown in Fig. 13. For the realistic parameters of $NaTiO_2$ the t_{2g}-electrons are almost localized in the d^1 configurations of Ti^{3+} ions, hence their interactions with neighboring sites can be described by the effective superexchange and kinetic exchange processes. Virtual charge excitations between the neighboring sites, $d_i^1d_j^1 \rightleftharpoons d_i^2d_j^0$, generate magnetic interactions which arise from two different hopping processes for active t_{2g}-orbitals: (i) the effective hopping $t=t_{pd}^2/\Delta$ which occurs via oxygen $2p_z$-orbitals with the charge transfer excitation energy Δ, in the present case along the 90° bonds, and (ii) the direct hopping t' which couples the t_{2g}-orbitals along the bond and gives kinetic exchange interaction, as in the

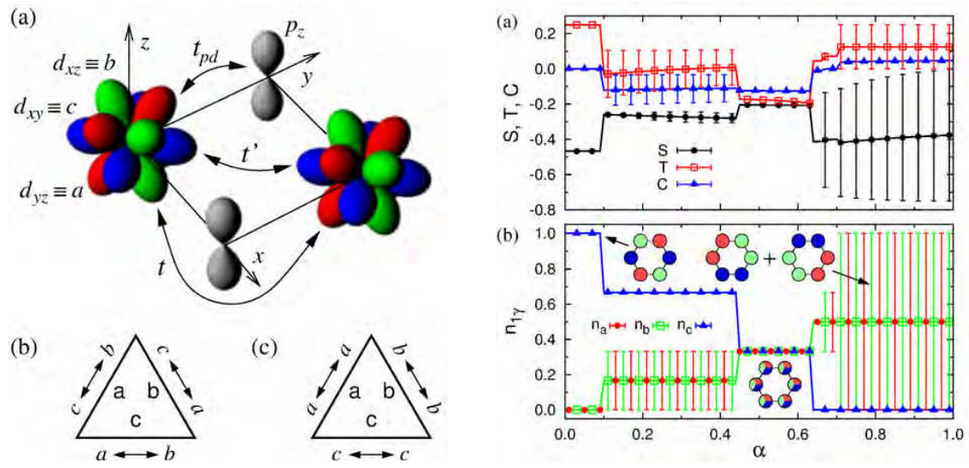


Fig. 13: *Left: (a) Hopping processes between t_{2g}-orbitals along a bond parallel to the c axis in $NaTiO_2$: (i) t_{pd} between Ti(t_{2g}) and O($2p_z$) orbitals—two t_{pd} transitions define an effective hopping t, and (ii) direct d–d hopping t'. The t_{2g}-orbitals shown by different colors are labeled as a, b, and c, see Eq. (2). The bottom part gives the hopping processes along $\gamma = a, b, c$ axes in the triangular lattice that contribute to Eq. (71): (b) superexchange and (c) direct exchange. Right: Ground state for a free hexagon as a function of α (71): (a) bond correlations—spin S_{ij} Eq. (68) (circles), orbital T_{ij} Eq. (69) (squares), and spin–orbital C_{ij} Eq. (70) (triangles); (b) orbital electron densities $n_{1\gamma}$ at a representative site $i = 1$ (left-most site): n_{1a} (circles), n_{1b} (squares), n_{1c} (triangles). The insets indicate the orbital configurations favored by superexchange ($\alpha = 0$), by mixed $0.44 < \alpha < 0.63$, and by direct exchange ($\alpha = 1$). The vertical lines indicate an exact range due to the degeneracy. (Images reproduced from Ref. [40])*

Hubbard model (1). Note that the latter processes couple orbitals with the same flavor, while the former ones couple different orbitals (for this geometry) so the occupied orbitals may be interchanged as a result of a virtual charge excitation—these processes are shown in Fig. 13. The effective spin-orbital model considered here reads [39]

$$\mathcal{H} = J \left\{ (1-\alpha)\, \mathcal{H}_s + \sqrt{(1-\alpha)\alpha}\, \mathcal{H}_m + \alpha\, \mathcal{H}_d \right\} . \tag{71}$$

The parameter α in Eq. (71) is given by the hopping elements as follows

$$\alpha = \frac{t'^2}{t^2 + t'^2} \tag{72}$$

and interpolates between the superexchange $\mathcal{H}_s$ ($\alpha = 0$) and kinetic exchange $\mathcal{H}_d$ ($\alpha = 1$), while in between mixed exchange contributes as well; these terms are explained in Ref. [39]. This model is considered here in the absence of Hund's exchange η (26), i.e., at $\eta = 0$. One finds that all the orbitals contribute equally in the entire range of α, and each orbital state is occupied at two out of six sites in the entire regime of α, see Fig. 13. The orbital state changes under increasing α and one finds four distinct regimes, with abrupt transitions between them. In

the superexchange model ($\alpha = 0$) there is precisely one orbital at each site which contributes, e.g. $n_{1c} = 1$ as the c-orbital is active along both bonds. Having a frozen orbital configuration, the orbitals decouple from spins and the ground state is disentangled, with $C_{ij} = 0$, and one finds that the spin correlations $S_{ij} = -0.4671$, as for the AF Heisenberg ring of $L = 6$ sites. Orbital fluctuations increase gradually with increasing α and this results in finite spin-orbital entanglement $C_{ij} \simeq -0.12$ for $0.10 < \alpha < 0.44$; simultaneously spin correlations weaken to $S_{ij} \simeq -0.27$.

In agreement with intuition, when $\alpha = 0.5$ and all interorbital transitions shown in Fig. 13 have equal amplitude, there is large orbital mixing which is the most prominent feature in the intermediate regime of $0.44 < \alpha < 0.63$. Although spins are coupled by AF exchange, the orbitals fluctuate here strongly and reduce further spin correlations to $S_{ij} \simeq -0.21$. The orbital correlations are negative, $T_{ij} < 0$, the spin-orbital entanglement is finite, $C_{ij} \simeq -0.13$, and the ground state is unique ($d = 1$). Here all the orbitals contribute equally and $n_{1\gamma} = 1/3$ which may be seen as a precursor of the spin-orbital liquid state which dominates the behavior of the triangular lattice. The regime of larger values of $\alpha > 0.63$ is dominated by the kinetic exchange in Eq. (71), and the ground state is degenerate with $d = 2$ [40], with strong scattering of possible electron densities $\{b_{i\gamma}\}$, see Fig. 13. Weak entanglement is found for $\alpha > 0.63$, where $C_{ij} \simeq\neq 0$. Summarizing, except for the regimes of $\alpha < 0.09$ and $\alpha > 0.63$ the ground state of a single hexagon is strongly entangled, i.e., $C_{ij} < -0.10$, see Fig. 13.

5.3 Fractionalization of orbital excitations

As a rule, even when spin and orbital operators disentangle in the ground state, some of the excited states are characterized by spin-orbital entanglement. It is therefore even more subtle to separate spin-orbital degrees of freedom to introduce orbitons as independent orbital excitations, in analogy to the purely orbital model and the result presented in Fig. 3 [41]. This problem is not yet completely understood and we show here that in a 1D spin-orbital model the orbital excitation fractionalizes into freely propagating spinon and orbiton, in analogy to charge-spinon separation in the 1D t-J model.

While a hole doped to the FM chain propagates freely, it creates a spinon and a holon in an AF background described by the t-J model. A similar situation occurs for an orbital excitation in an AF/FO spin-orbital chain [41]. An orbital excitation may propagate through the system only after creating a spinon in the first step, see Figs. 14(a) and 14(b). The spinon itself moves via spin flips $\propto J > t$, faster than the orbiton, and the two excitations get well separated, see Fig. 14(c). The orbital-wave picture of Sec. 2, on the other hand, would require the orbital excitation to move without creating the spinon in the first step. Note that this would be only possible for imperfect Néel AF spin order. Thus one concludes that the symmetry between the spin and the orbital sector is broken also for this reason and orbitals are so strongly coupled to spin excitations in realistic spin-orbital models with AF/FO order that the mean field picture separating these two sectors of the Hilbert space breaks down.

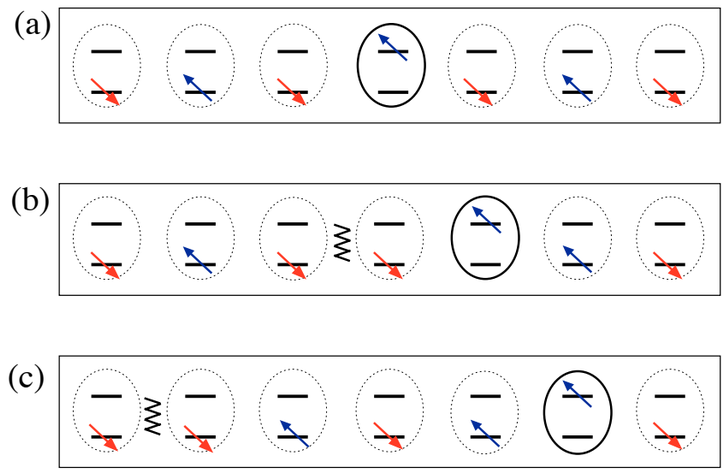


Fig. 14: *Schematic representation of the orbital motion and the spin-orbital separation in a 1D spin-orbital model. The first hop of the excited state (a)→(b) creates a spinon (wavy line) that moves via spin exchange $\propto J$. The next hop (b)→(c) gives an orbiton freely propagating as a holon with an effective hopping $t \sim J/2$. (Image reproduced from Ref. [41])*

6 t-J-like model for ferromagnetic manganites

Even more complex situations arise when charge degrees of freedom are added to spin-orbital models. The spectral properties of such models are beyond the scope of this discussion, we shall only point out that macroscopic doping changes radically the spin-orbital superexchange by adding to it ferromagnetic exchange triggered by the e_g-orbital liquid realized in hole doped manganites. As a result, the CMR effect is observed and the spin order changes to FM [3].

Similar to the spin case, the orbital Hubbard model Eq. (11) gives at large $\bar{U} \gg t$ the e_g t-J model [42], i.e., e_g-electrons may hop only in the restricted space without doubly occupied e_g^2 sites. The kinetic energy is gradually released with increasing doping x in doped manganese oxides $La_{1-x}A_xMnO_3$, with A =Sr,Ca,Pb, which is a driving mechanism for an effective FM interaction generated by the kinetic energy $\propto \tilde{H}_t^{\uparrow}(e_g)$ in the double exchange [3]. It competes with AF exchange which eventually becomes frustrated in the FM metallic phase, arising typically at sufficient hole doping $x \simeq 0.17$. The evolution of magnetic order with increasing doping results from the above frustration: at low doping $x \sim 0.1$ AF spin order becomes stable and first changes to a FM insulating phase, see Fig. 15(a). Only at larger doping x, an insulator-to-metal transition takes place which explains the CMR effect [3].

In the FM metallic phase the magnon excitation energy is derived from the manganite t-J model and consists of two terms [42]: (i) superexchange being AF for the orbital liquid and (ii) FM double exchange $J_{\rm DE}$, proportional to the kinetic energy of e_g-electrons (6)

$$J_{\rm DE} = \frac{1}{2z\mathcal{S}^2} \left| \left\langle \tilde{H}_t^{\uparrow}(e_g) \right\rangle \right| . \tag{73}$$

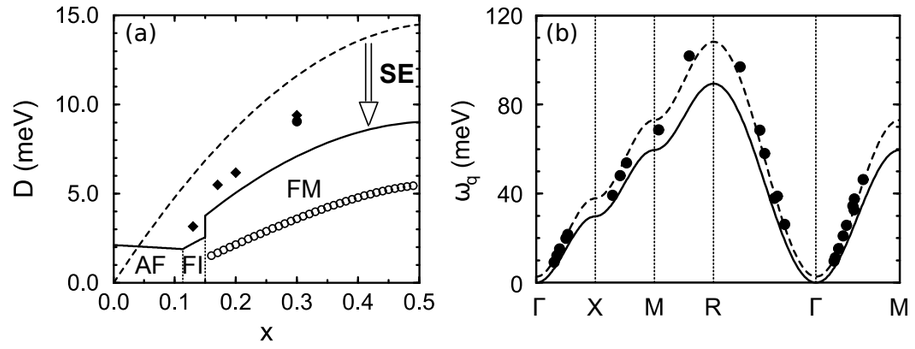


Fig. 15: *Theoretical predictions for magnon spectra in the FM metallic phase in manganites: (a) spin-wave stiffness D (solid line) as a function of hole doping x given by double exchange (dashed) reduced by superexchange (SE) for: A-AF, FM insulating (FI), and FM metallic (FM) phases, and experimental points for $La_{1-x}Sr_xMnO_3$ (diamonds) and $La_{0.7}Pb_{0.3}MnO_3$ (circle); empty circles for the hypothetical AO $|\pm\rangle$ state unstable against the e_g-orbital liquid; (b) magnon dispersion $\omega_{\vec{q}}$ obtained at $x = 0.30$ (solid line) and the experimental points for $La_{0.7}Pb_{0.3}MnO_3$ [43] (circles and dashed line). Parameters: $U = 5.9$, $J_H^e = 0.7$, $t = 0.41$, all in eV. (Images reproduced from Ref. [42])*

Here z is the number of neighbors ($z = 6$ for the cubic lattice), and $2\mathcal{S} = 4 - x$ is the average spin in a doped manganese oxide. The kinetic energy $|\langle \tilde{H}_t^{\uparrow}(e_g)\rangle|$ measures directly the band narrowing due to the strong correlations in the e_g-orbital liquid. This explains why the spin-wave stiffness D is: (i) reduced by the frustrating AF superexchange $J_{\rm SE}$ but (ii) increases proportionally to the hole doping x in the FM metallic phase, see Fig. 15(a). As a result, the magnon dispersion in the FM metallic phase is given by

$$\omega_{\vec{q}} = (J_{\rm DE} - J_{\rm SE})\, z\mathcal{S}^2(1 - \gamma_{\vec{q}}), \tag{74}$$

where $\gamma_{\vec{q}} = \frac{1}{z}\sum_{\vec{\delta}} e^{i\vec{q}\cdot\vec{\delta}}$, and $\vec{\delta}$ is a vector which connects the nearest neighbors.
An experimental proof that indeed the e_g-orbital liquid is responsible for isotropic spin excitations in the FM metallic phase of doped manganites, we show the magnon spectrum observed in $La_{0.7}Pb_{0.3}MnO_3$, with rather large stiffness constant $D = 7.25$ meV, see Fig. 15(b). Note that D would be much smaller in the phase with AO order of $|\pm\rangle$ orbitals (28). Summarizing, the isotropy of the spin excitations in the simplest manganese oxides with FM metallic phase is naturally explained by the *orbital liquid* state of disordered e_g-orbitals.

7 Conclusions and outlook

Spin-orbital physics is a very challenging field in which only certain and mainly classical aspects have been understood so far. We have explained the simplest principles of spin-orbital models determining the physical properties of strongly correlated transition metal oxides with active orbital degrees of freedom. In the correlated insulators exchange interactions are usually

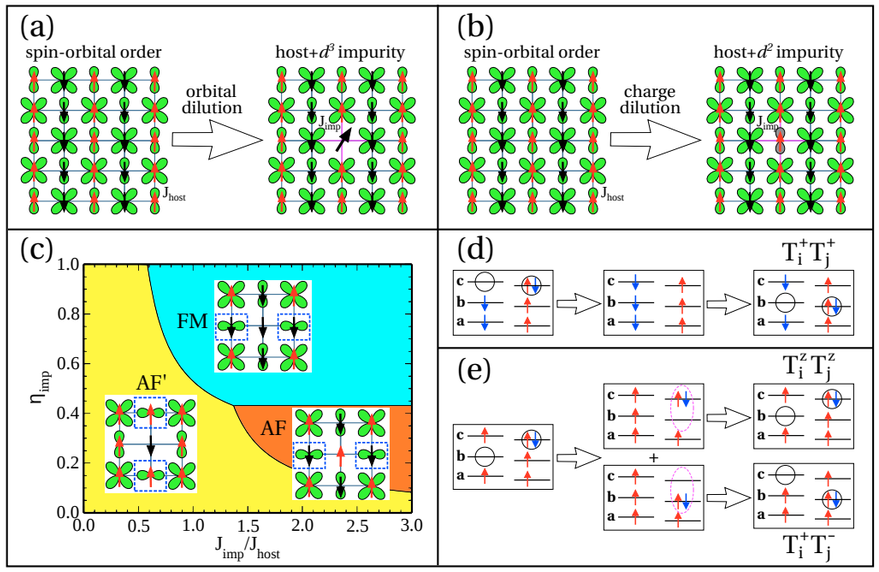


Fig. 16: *Top: Doping by transition metal ions in an ab plane with C-AF/G-AO order of $\{a, c\}$ orbitals found in d^4 Mott insulators (ruthenates) with: (a) orbital dilution by the d^3 impurity with $S = 3/2$ spin, and (b) charge dilution by the d^2 impurity with $S = 1$ spin. Host $S = 1$ spins (red/black arrows) interact by $J_{\rm host}$ and doublons in a (c) orbitals shown by green symbols. Here doping occurs at a doublon site and spins are coupled by $J_{\rm imp}$ along hybrid (red) bonds. Bottom: (c) phase diagram for a single d^3 impurity replacing a doublon in c-orbital in the C-AF host [46], with changes in the orbital order indicated by dashed boxes (note $a \to b$ orbital flips); (d-e) orbital fluctuations promoted on d^2–d^4 hybrid bonds with (d) AF and (e) FM spin correlations. In the latter case (e) the doublons at two orbitals are coupled in excited states (doublon and hole in ovals), and one obtains orbital flips $\propto T_i^-T_j^+$ accompanied by Ising terms $\propto T_i^zT_j^z$, while double excitations $\propto T_i^+T_j^+$ occur on AF bonds (d) even in the absence of Hund's exchange and are amplified by finite η. (Image reproduced from Ref. [47])*

frustrated and this frustration is released by a certain type of spin-orbital order, with the complementarity of spin and orbital correlations at AF/FO or FM/AO bonds, as explained by the Goodenough-Kanamori rules [35].

One of the challenges is spin-orbital entanglement, which becomes visible both in the ground and excited states. The coherent excitations such as magnons or orbitons are frequently not independent and also composite spin-orbital excitations are possible. Such excitations are not yet understood, except for some simplest cases as, e.g., the 1D spin-orbital model with SU(4) symmetry where all these excitations are on an equal footing and contribute to the entropy in the same way [44]. Such a perfect symmetry does not occur in nature, however, and the orbital excitations are more complex due to finite Hund's exchange interaction and, at least in some systems, orbital-lattice couplings. They may be a decisive factor explaining why spin-orbital liquids do not occur in certain models. For the same reason, in the absence of geometrical frustration, the orbital liquid seems easier to obtain than the spin liquid.

Doping of spin-orbital systems leads to very rich physics with phase transitions induced by moving charge carriers, as for instance in the well known example of the CMR manganites. Yet, the holes doped into correlated insulators with spin-orbital order may be of quite different nature. Charge defects may prevent the holes from coherent propagation [45] and as a result the spin-orbital order will persist to rather high doping level.

Recently doping by transition metal ions with different valence was explored [46]—in such t_{2g} systems local or global changes of spin-orbital order result from the complex interplay of spin-orbital degrees of freedom at *orbital dilution*, see Fig. 16(a). In general, the observed order in the doped system will then depend on the coupling between the ions with different valence compared with that within the host $J_{\rm imp}/J_{\rm host}$, and on Hund's exchange at doped ions $\eta_{\rm imp}$. Not only a crossover between AF and FM spin correlations is expected with increasing $\eta_{\rm imp}$, but also the orbital state will change from inactive orbitals to orbital polarons on the hybrid bonds with increasing $J_{\rm imp}$, see Fig. 16(c). Quite a different case is given when double occupancies are replaced by empty orbitals in *charge doping* as shown in Fig. 16(b). Here orbital fluctuations are remarkably enhanced by the novel double excitation $\propto T_i^+T_j^+$ terms, see Figs. 16(d-e). On the one hand, large spin-orbital entanglement is expected in such cases when Hund's exchange is weak, while on the other hand the superexchange will reduce to the orbital model in the FM regime. By mapping of this latter model to fermions one may expect interesting topological states in low dimension that are under investigation at present.

Acknowledgments

We kindly acknowledge support by Narodowe Centrum Nauki (NCN, National Science Centre, Poland), under Project MAESTRO No. 2012/04/A/ST3/00331.

References

[1] K.A. Chao, J. Spałek, and A.M. Oleś, J. Phys. C **10**, L271 (1977)

[2] M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)

[3] E. Dagotto, T. Hotta, and A. Moreo, Phys. Rep. **344**, 1 (2001)

[4] K.I. Kugel and D.I. Khomskii, Sov. Phys. Usp. **25**, 231 (1982)

[5] C. Slater and G.F. Koster, Phys. Rev. **94**, 1498 (1954)

[6] G. Khaliullin and S. Maekawa, Phys. Rev. Lett. **85**, 3950 (2000)

[7] L.F. Feiner and A.M. Oleś, Phys. Rev. B **71**, 144422 (2005)

[8] A.M. Oleś, G. Khaliullin, P. Horsch, and L.F. Feiner, Phys. Rev. B **72**, 214431 (2005)

[9] J.S. Griffith, *The Theory of Transition Metal Ions*
(Cambridge University Press, Cambridge, 1971)

[10] A.M. Oleś, Phys. Rev. B **28**, 327 (1983)

[11] J. van den Brink, P. Horsch, F. Mack, and A.M. Oleś, Phys. Rev. B **59**, 6795 (1999)

[12] A. van Rynbach, S. Todo, and S. Trebst, Phys. Rev. Lett. **105**, 146402 (2010)

[13] P. Czarnik, J. Dziarmaga, and A.M. Oleś, Phys. Rev. B **96**, 014420 (2017)

[14] Z. Nussinov and J van den Brink, Rev. Mod. Phys. **87**, 1 (2015)

[15] L. Cincio, J. Dziarmaga, and A.M. Oleś, Phys. Rev. B **82**, 104416 (2010)

[16] S. Wenzel and A.M. Läuchli, Phys. Rev. Lett. **106**, 197201 (2011)

[17] P. Czarnik, J. Dziarmaga, and A.M. Oleś, Phys. Rev. B **93**, 184410 (2016)

[18] G. Khaliullin, Prog. Theor. Phys. Suppl. **160**, 155 (2005)

[19] G. Khaliullin, P. Horsch, and A.M. Oleś, Phys. Rev. B **70**, 195103 (2004)

[20] L.F. Feiner, A.M. Oleś, and J. Zaanen, Phys. Rev. Lett. **78**, 2799 (1997)
J. Phys.: Condens. Matter **10**, L555 (1998)

[21] A.M. Oleś, L.F. Feiner, and J. Zaanen, Phys. Rev. B **61**, 6257 (2000)

[22] W. Brzezicki, J. Dziarmaga, and A.M. Oleś, Phys. Rev. Lett. **109**, 237201 (2012)

[23] E. Pavarini, E. Koch, and A.I. Lichtenstein, Phys. Rev. Lett. **101**, 266405 (2008)

[24] W. Brzezicki, J. Dziarmaga, and A.M. Oleś, Phys. Rev. B **87**, 064407 (2013)

[25] L.F. Feiner and A.M. Oleś, Phys. Rev. B **59**, 3295 (1999)

[26] E. Pavarini and E. Koch, Phys. Rev. Lett. **104**, 086402 (2010)

[27] M. Snamina and A.M. Oleś, Phys. Rev. B **94**, 214426 (2016)

[28] N.N. Kovaleva, A.M. Oleś, A.M. Balbashov, A. Maljuk, D.N. Argyriou, G. Khaliullin, and B. Keimer, Phys. Rev. B **81**, 235130 (2010)

[29] S. Miyasaka, Y. Okimoto, and Y. Tokura, J. Phys. Soc. Jpn. **71**, 2086 (2002)

[30] A.M. Oleś and G. Khaliullin, Phys. Rev. B **84**, 214414 (2011)

[31] G. Khaliullin, P. Horsch, and A.M. Oleś, Phys. Rev. Lett. **86**, 3879 (2001)

[32] P. Horsch, A.M. Oleś, L.F. Feiner, and G. Khaliullin, Phys. Rev. Lett. **100**, 167205 (2008)

[33] S. Miyasaka, Y. Okimoto, M. Iwama, and Y. Tokura, Phys. Rev. B **68**, 100406 (2003)

[34] J. Fujioka, T. Yasue, S. Miyasaka, Y. Yamasaki, T. Arima, H. Sagayama, T. Inami, K. Ishii, and Y. Tokura, Phys. Rev. B **82**, 144425 (2010)

[35] J.B. Goodenough, *Magnetism and the Chemical Bond* (Interscience, New York, 1963)

[36] K. Wohlfeld, M. Daghofer, and A.M. Oleś, Europhys. Lett. (EPL) **96**, 27001 (2011)

[37] A.M. Oleś, J. Phys.: Condens. Matter **24**, 313201 (2012)

[38] P. Fazekas: *Lecture Notes on Electron Correlation and Magnetism* (World Scientific, Singapore, 1999)

[39] B. Normand and A.M. Oleś, Phys. Rev. B **78**, 094427 (2008)

[40] J. Chaloupka and A.M. Oleś, Phys. Rev. B **83**, 094406 (2011)

[41] K. Wohlfeld, M. Daghofer, S. Nishimoto, G. Khaliullin, and J. van den Brink, Phys. Rev. Lett. **107**, 147201 (2011)

[42] A.M. Oleś and L.F. Feiner, Phys. Rev. B **65**, 052414 (2002)

[43] J.A. Fernandez-Baca, P. Dai, H.Y. Hwang, C. Kloc, and S.-W. Cheong, Phys. Rev. Lett. **80**, 4012 (1998)

[44] B. Frischmuth, F. Mila, and M. Troyer, Phys. Rev. Lett. **82**, 835 (1999)

[45] A. Avella, P. Horsch, and A.M. Oleś, Phys. Rev. Lett. **115**, 206403 (2015)

[46] W. Brzezicki, A.M. Oleś, and M. Cuoco, Phys. Rev. X **5**, 011037 (2015)

[47] W. Brzezicki, M. Cuoco, and A.M. Oleś, J. Supercond. Novel Magn. **30**, 129 (2017)

6 Introduction to the Hubbard Model

Robert Eder
Institut für Festkörperphysik
Karlsruhe Institute of Technology

Contents

1 The Hubbard model **2**

2 Some notation and review of Green functions **3**

3 The Hubbard dimer: solution by exact diagonalization **5**

4 Spin density wave theory **8**

5 The Hubbard-I approximation **14**

6 The Gutzwiller wave function **20**

7 Summary and discussion **27**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 The Hubbard model

The Hubbard model was proposed in the 1960s to describe electrons in $3d$ transition metals. In these elements, the radial wave function of the $3d$-electrons has a very small spatial extent. Therefore, when the $3d$ shell is occupied by several electrons, these are forced to be close to one another on the average so that the electrostatic energy is large. The energy of a given transition metal ion therefore varies strongly with the number of electrons it contains. To study the motion of conduction electrons under the influence of this strong Coulomb repulsion Hubbard [1], Kanamori [2], and Gutzwiller [3] proposed a simplified model. Thereby both, the five-fold degeneracy of the $3d$-orbital and the presence of other bands in the solid are neglected. Rather, one considers a lattice of sites – whereby the geometry of the lattice is not really specified – with one s-like orbital at each site. Orbitals on different sites are assumed to be orthogonal, but for not too distant sites i and j there are nonvanishing matrix elements $t_{i,j}$ of the Hamiltonian between the orbitals centered on these sites. The Coulomb interaction between electrons in orbitals on different sites is neglected, but if two electrons – which then necessarily have opposite spin – occupy the same orbital the energy is assumed to increase by the large amount U to simulate the strong dependence of the energy on the occupation number. If we denote the creation operator for an electron of spin σ in the orbital at lattice site i by $c^\dagger_{i,\sigma}$ the model thus can be written as

$$H = \sum_{i,j} \sum_{\sigma} t_{i,j}\, c^\dagger_{i,\sigma} c_{j,\sigma} + U \sum_i n_{i,\uparrow} n_{i,\downarrow} = H_t + H_U. \tag{1}$$

Here $n_{i,\sigma} = c^\dagger_{i,\sigma} c_{i,\sigma}$ counts the number of electrons with spin σ in the orbital at site i.
After the discovery of the cuprate superconductors in 1987 and after Zhang and Rice demonstrated [4] that the CuO_2 planes in these compounds can be described by the so-called t-J model – which is equivalent to the Hubbard model in the limit $U/t \gg 1$ – there was renewed interest in the 2-dimensional Hubbard model. However, the lightly doped Mott-insulator – which most probably is the system to be understood in order to solve the many puzzles posed by the cuprate superconductors – is still far from being solved. Accordingly, the purpose of this lecture is to present some of the basic approximations developed for this model.
We consider (1) for a two-dimensional square lattice with N sites, lattice constant $a = 1$, and periodic boundary conditions. For hopping integrals $-t$ between nearest ($(1,0)$-like) neighbors and $-t'$ between 2^{nd} nearest ($(1,1)$-like) neighbors the dispersion relation is

$$\varepsilon_{\boldsymbol{k}} = -2t \left(\cos(k_x) + \cos(k_y)\right) - 4t' \cos(k_x)\cos(k_y). \tag{2}$$

The number of electrons with spin σ in the system is denoted by N_σ – whereby we are mostly interested in the nonmagnetic case $N_\uparrow = N_\downarrow$ – so that the number of electrons is $N_e = N_\uparrow + N_\downarrow$. In the following, densities per site will be denoted n, e.g., $n_\uparrow = N_\uparrow/N$. For $n_e = 1$ we have $N_\uparrow = N_\downarrow = N/2$ so that precisely half of the $\mathbf{k}$-points for each spin direction are occupied and we have a half-filled band, i.e., a metal in conventional band theory. Instead it will be shown below that for sufficiently large U/t the Hubbard model describes an insulator, the so-called Mott-insulator. The region of primary interest for cuprate superconductors is $n_e \geq 0.8$, i.e., the lightly doped Mott-insulator, and $U/t \approx 10$.

2 Some notation and review of Green functions

We first introduce some notation that will be used frequently and give a brief review of imaginary time Green functions. There are numerous excellent textbooks on the use of field theory for condensed matter physics where more details can be found [5–7].
The thermal average of any operator $\hat{O}$ is given by

$$\langle \hat{O} \rangle_{\rm th} = \frac{1}{Z}\,{\rm Tr}\left(e^{-\beta(H-\mu N)}\hat{O}\right) \tag{3}$$

with the grand partition function

$$Z = {\rm Tr}\Big(e^{-\beta(H-\mu N)}\Big). \tag{4}$$

Introducing the imaginary-time Heisenberg operator (with $K = H - \mu N$)

$$c_\alpha(\tau) = e^{\tau K}\, c_\alpha\, e^{-\tau K} \qquad \Rightarrow \qquad -\frac{\partial c_\alpha(\tau)}{\partial \tau} = [c_\alpha(\tau), K] \tag{5}$$

the *imaginary time Green function* is defined as

$$\begin{aligned} G_{\alpha,\beta}(\tau) &= -\left\langle T\, c_\alpha(\tau)\, c_\beta^\dagger \right\rangle_{\rm th} = -\Theta(\tau)\, \left\langle c_\alpha(\tau)\, c_\beta^\dagger \right\rangle_{\rm th} + \Theta(-\tau)\, \left\langle c_\beta^\dagger\, c_\alpha(\tau) \right\rangle_{\rm th} \qquad (6)\\ &= \frac{1}{Z}\Bigg(-\Theta(\tau) \sum_{i,j} e^{-\beta(E_i-\mu N_i)}\, e^{\tau(E_i-E_j+\mu)}\, \big\langle i\big|\, c_\alpha \big|j\big\rangle\, \big\langle j\big|\, c_\beta^\dagger\, \big|i\big\rangle \\ &\quad + \Theta(-\tau) \sum_{i,j} e^{-\beta(E_i-\mu N_i)}\, e^{\tau(E_j-E_i+\mu)}\, \big\langle i\big|\, c_\beta^\dagger\, \big|j\big\rangle\, \big\langle j\big|\, c_\alpha \big|i\big\rangle \Bigg), \qquad (7) \end{aligned}$$

where α, β denote some set of quantum numbers, $|i\rangle$ are the exact eigenstates of the system with energies E_i and particle number N_i. Using $\frac{\partial \Theta(\tau)}{\partial \tau} = \delta(\tau)$ it follows from (5) and (6) that the Green function obeys the equation of motion

$$-\frac{\partial}{\partial \tau}\, G_{\alpha,\beta}(\tau) = \delta(\tau)\, \left\langle \Big\{ c_\alpha, c_\beta^\dagger \Big\} \right\rangle_{\rm th} - \left\langle T\, [c_\alpha(\tau), K]\; c_\beta^\dagger \right\rangle_{\rm th}\ . \tag{8}$$

It follows from (7) that G is well-defined only if $\tau \in [-\beta, \beta]$ (the reason is that the $E_i - \mu N_i$ are unbounded from above [8]) and that for $\tau \in [-\beta, 0]$ one has $G(\tau + \beta) = -G(\tau)$. It follows that $G(\tau)$ has the Fourier transform (see equation (25.10) in [6] with $\hbar = 1$)

$$G(\tau) = \frac{1}{\beta} \sum_{\nu=-\infty}^{\infty} e^{-i\omega_\nu \tau} G(i\omega_\nu), \qquad G(i\omega_\nu) = \int_0^\beta d\tau\, e^{i\omega_\nu \tau}\, G(\tau), \qquad \omega_\nu = \frac{(2\nu+1)\pi}{\beta}. \tag{9}$$

The $i\omega_\nu$ are the (Fermionic) Matsubara frequencies. Inserting (7) into (9) one obtains

$$G_{\alpha\beta}(i\omega_\nu) = \frac{1}{Z} \sum_{i,j} \frac{e^{-\beta(E_i-\mu N_i)} + e^{-\beta(E_j-\mu N_j)}}{i\omega_\nu + \mu - (E_j - E_i)}\, \big\langle i\big|\, c_\alpha \big|j\big\rangle \big\langle j\big|\, c_\beta^\dagger\, \big|i\big\rangle. \tag{10}$$

We specialize to a single band and assume that the z-component of the spin is a good quantum number so that the Green function is a scalar and $\alpha = \beta = (\mathbf{k}, \sigma)$. The function $G(\mathbf{k}, i\omega_\nu)$

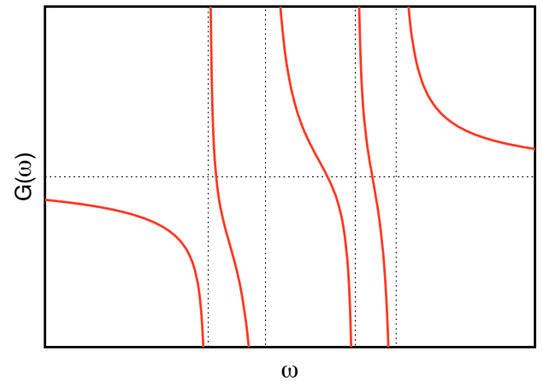


Fig. 1: *The Green function $G(\omega)$ for real ω. The dashed vertical lines give the poles, ω_i.*

can be analytically continued to the whole complex frequency plane by replacing $i\omega_\nu \to z$ where z is any complex number. As can be seen from (10) the function $G(z)$ is analytic in the complex z-plane except for the real axis where it has simple poles at $z = E_j - E_i - \mu$. It is this property on which the usefulness of the imaginary-time Green function is based: the analytic continuation of its Fourier transform $G(z)$ can be evaluated along a line infinitesimally above the real axis and then gives the Fourier transform of the retarded real-time Green function – from which the single-particle spectral function $A(\mathbf{k}, \omega)$, i.e., the combined photoemission and inverse photoemission spectrum of a system can be obtained:

$$A(\mathbf{k}, \omega) = -\frac{1}{\pi} \operatorname{Im} G(\mathbf{k}, \omega + i0^+) \,.$$

For this reason the Fourier transform (10) is often called 'the' Green function. Equation (10) is the *Lehmann representation* of the Green function.
It is shown in textbooks of field theory [5–7] that the imaginary-time Green function can be expanded in Feynman diagrams (whereas such an expansion is *not* possible for the real-time Green function at finite temperature) and the self-energy $\mathbf{\Sigma}(\mathbf{k}, z)$ can be introduced as usual:

$$\big(z + \mu - \varepsilon_{\mathbf{k}} - \mathbf{\Sigma}(\mathbf{k}, z)\big)\, G(\mathbf{k}, z) = 1 \,. \tag{11}$$

Next we discuss the analytical structure of the Green function and the self energy. It can be seen from (10) that the Fourier transform of the Green function has the general form

$$G(z) = \sum_i \frac{\alpha_i}{z - \omega_i} \,,$$

where the α_i and ω_i are real numbers. Along the real axis $G(\omega)$ therefore looks like in Fig. 1. This shows that in between any two successive poles ω_i and ω_{i+1} the Green function crosses zero with a negative slope

$$G(\omega) \approx -\beta_i(\omega - \zeta_i).$$

Near the crossing point ζ_i we thus have

$$\Sigma(\omega) = -G^{-1}(\omega) + \omega + \mu - \varepsilon_{\mathbf{k}} = \frac{\sigma_i}{\omega - \zeta_i} + \dots$$

where $\sigma_i = 1/\beta_i$. The self-energy thus has poles on the real axis as well, and these poles are 'sandwiched' in between the poles of the Green function. Luttinger has shown [9] that $\Sigma(\omega)$ is essentially determined by these poles and their residua in that it can be written in the whole complex frequency plane as

$$\Sigma(z) = V_{HF} + \sum_i \frac{\sigma_i}{z - \zeta_i} \tag{12}$$

where V_{HF} is equal to the Hartree-Fock potential (or rather its Fourier coefficient with the proper momentum).

3 The Hubbard dimer: solution by exact diagonalization

As a first example we consider the Hubbard model on a dimer

$$H = -t \sum_\sigma \left(c^\dagger_{1,\sigma} c_{2,\sigma} + c^\dagger_{2,\sigma} c_{1,\sigma} \right) + \sum_{i=1}^{2} n_{i,\uparrow} n_{i,\downarrow}. \tag{13}$$

This can be solved by *exact diagonalization* , i.e., by constructing a basis of the entire Hilbert space, setting up the Hamilton matrix in this basis and diagonalizing it. The Hamiltonian (13) is invariant under the exchange of the site indices, $1 \leftrightarrow 2$, so that we can classify eigenstates by their parity P under this operation. Alternatively, we might view (13) as describing a '2-site ring with periodic boundary conditions' and hopping integral $t/2$. Such a 2-site ring has two allowed momenta, $k = 0$ and $k = \pi$. The exchange of sites is equivalent to a translation by one lattice site, and since the definition of a Bloch state ψ_k with momentum k is $T_R\ \psi_k = e^{ikR}\ \psi_k$ it follows that $P = 1$ is equivalent to $k = 0$, whereas $P = -1$ means $k = \pi$. In the following we will always consider parity and momentum as interchangeable.
The Hilbert space can be decomposed into sectors with fixed electron number N_e, z-component of spin S_z, and parity P. We first consider the sector with $N_e = 2$ and $S_z = 0$, i.e., two electrons with opposite spin, which is equivalent to 'half filling'. The basis states with $P = \pm 1$ are:

$$\begin{aligned} |1_\pm\rangle &= \frac{1}{\sqrt{2}} \left(c^\dagger_{1,\uparrow} c^\dagger_{2,\downarrow} \pm c^\dagger_{2,\uparrow} c^\dagger_{1,\downarrow} \right) |0\rangle, \\ |2_\pm\rangle &= \frac{1}{\sqrt{2}} \left(c^\dagger_{1,\uparrow} c^\dagger_{1,\downarrow} \pm c^\dagger_{2,\uparrow} c^\dagger_{2,\downarrow} \right) |0\rangle. \end{aligned} \tag{14}$$

For the ground state with $P = +1$ we make the ansatz $|\psi_0\rangle = u|1_+\rangle + v|2_+\rangle$. The relevant matrix elements are $\langle 1_+|H|2_+\rangle = -2t$, $\langle 1_+|H|1_+\rangle = 0$ and $\langle 2_+|H|2_+\rangle = U$, so that the ground state energy E_0 and the coefficients u and v can be obtained by solving the eigenvalue problem

$$\begin{pmatrix} 0 & -2t \\ -2t & U \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = E_0 \begin{pmatrix} u \\ v \end{pmatrix}.$$

We find $E_0 = \frac{U}{2} - W$ with $W = \sqrt{\left(\frac{U}{2}\right)^2 + 4t^2}$ and

$$u = -\sqrt{\frac{W + \frac{U}{2}}{2W}}, \qquad v = \frac{-2t}{\sqrt{2W(W + \frac{U}{2})}}. \tag{15}$$

Both $|1_+\rangle$ and $|2_+\rangle$ are spin singlets so that $|\psi_0\rangle$ is a singlet as well. For the states with $P = -1$ we have $\langle 1_-|H|2_-\rangle = 0$, whence the eigenstates are $|1_-\rangle$ with energy 0 and $|2_-\rangle$ with energy U. $|1_-\rangle$ is a spin triplet and the remaining two members of the triplet are $c^\dagger_{1,\uparrow}c^\dagger_{2,\uparrow}|0\rangle$ and $c^\dagger_{1,\downarrow}c^\dagger_{2,\downarrow}|0\rangle$. Both states are in fact the only basis states in the sectors $N_e = 2$, $S_z = \pm 1$ and $P = -1$ and hence are eigenstates by construction with energy 0. In the limit $U/t \to \infty$ we have $W \to \frac{U}{2} + \frac{4t^2}{U}$. The low energy spectrum of the dimer with two electrons then consists of four states: the spin singlet with energy $-\frac{4t^2}{U}$ and the three members of the triplet with energy 0. We can thus describe the dimer by an effective low energy Hamiltonian of the form

$$H \;=\; J\left(\boldsymbol{S}_1 \cdot \boldsymbol{S}_2 - \frac{n_1 n_2}{4}\right)$$

where $n_i = \sum_\sigma c^\dagger_{i,\sigma}c_{i,\sigma}$ is the operator of electron number and $\boldsymbol{S}_i = \frac{1}{2}\sum_{\sigma,\sigma'} c^\dagger_{i,\sigma}\boldsymbol{\tau}_{\sigma,\sigma'}c_{i,\sigma}$ the operator of electron spin at site i ($\boldsymbol{\tau}$ denotes the vector of Pauli matrices) and $J = \frac{4t^2}{U}$ – this is nothing but the *Heisenberg antiferromagnet*.
We can also construct states with $N_e = 1, 3$ and given parity and z-spin $\sigma = \pm\frac{1}{2}$:

$$\begin{aligned}
|3_\pm, \sigma\rangle &= \frac{1}{\sqrt{2}}\left(c^\dagger_{1,\sigma} \pm c^\dagger_{2,\sigma}\right)|0\rangle, \\
|4_\pm, \sigma\rangle &= \frac{1}{\sqrt{2}}\left(c^\dagger_{1,\sigma}c^\dagger_{2,\uparrow}c^\dagger_{2,\downarrow} \pm c^\dagger_{2,\sigma}c^\dagger_{1,\uparrow}c^\dagger_{1,\downarrow}\right)|0\rangle.
\end{aligned}$$

Since these states are the only ones in their respective (N_e, S_z, P) sector they are again eigenstates by construction and indeed $H|3_\pm, \sigma\rangle = \mp t|3_\pm, \sigma\rangle$ and $H|4_\pm, \sigma\rangle = (U \pm t)|4_\pm, \sigma\rangle$. The energies of the single-electron states can be written in the familiar form $\varepsilon_k = 2\,\frac{t}{2}\cos(k)$ expected for a two-site ring with periodic boundary conditions and hopping integral $\frac{t}{2}$. Having found all eigenstates $|j\rangle$ and their energies E_j we can write down the Green function $G(k, \omega)$ using the Lehmann representation (10). Thereby we simplify matters by taking the limit of low temperature and assuming that the chemical potential μ has been chosen such that the thermal occupation factor $e^{-\beta E_j}/Z$ is unity for $|j\rangle = |\psi_0\rangle$ and zero for all other states (this can be achieved by setting, e.g., $\mu = \frac{U}{2}$). Defining electron operators with definite parity by $c_{\pm,\sigma} = \frac{1}{\sqrt{2}}(c_{1,\sigma} \pm c_{2,\sigma})$ we find

$$\begin{aligned}
c_{\pm,\sigma}|\psi_0\rangle &= \frac{1}{2}\left(c^\dagger_{1,\sigma} \pm c^\dagger_{2,\sigma}\right)\left(u\left(c^\dagger_{1,\uparrow}c^\dagger_{2,\downarrow} + c^\dagger_{2,\uparrow}c^\dagger_{1,\downarrow}\right) + v\left(c^\dagger_{1,\uparrow}c^\dagger_{1,\downarrow} + c^\dagger_{2,\uparrow}c^\dagger_{2,\downarrow}\right)\right)|0\rangle \\
&= \frac{\pm u + v}{\sqrt{2}}\,|3_\pm, \bar{\sigma}\rangle, \\
c^\dagger_{\pm,\sigma}|\psi_0\rangle &= \frac{\mp u + v}{\sqrt{2}}\,|4_\pm, \sigma\rangle,
\end{aligned}$$

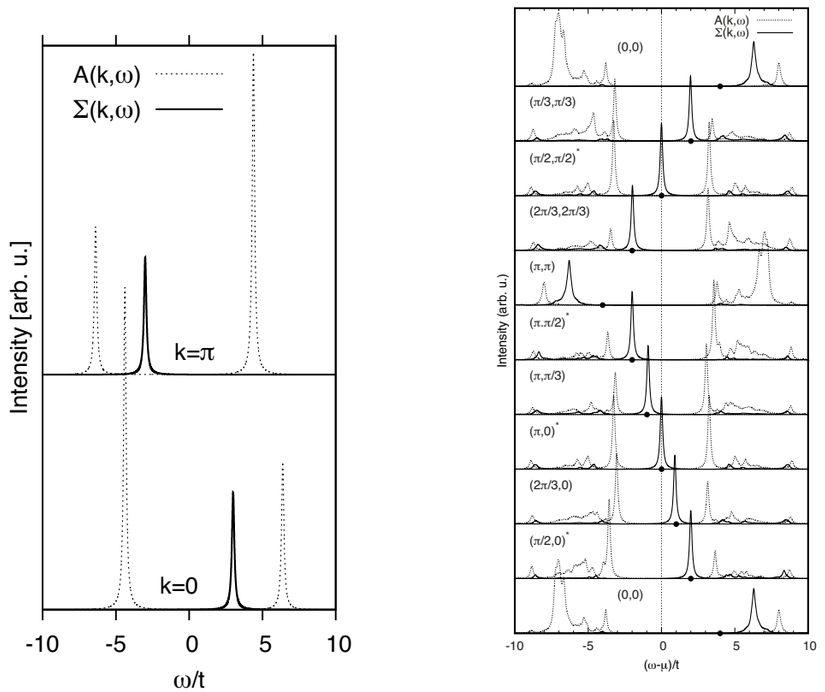


Fig. 2: *Single particle spectral function and imaginary part of the self-energy for the Hubbard model with $U/t = 10$, δ-functions have been replaced by Lorentzians. Left: for the dimer, right for the 16 and 18-site cluster with periodic boundary conditions (reprinted with permission from [10], Copyright 2011 by the American Physical society). The dots correspond to $\zeta_{\boldsymbol{k}} = \frac{U}{2} - \mu - \varepsilon_{\boldsymbol{k}}$.*

and using (15) for u and v we find the squared matrix elements

$$|\langle 3_\pm, \bar{\sigma}|c_{\pm,\sigma}|\psi_0\rangle|^2 = \frac{1}{2} \pm \frac{t}{W}\,, \qquad |\langle 4_\pm, \sigma|c^\dagger_{\pm,\sigma}|\psi_0\rangle|^2 = \frac{1}{2} \mp \frac{t}{W}.$$

Using (10) the Fourier transform of the Green function $G_\pm(\omega) = -\big\langle\, T\, c_\pm(\tau) c^\dagger_\pm \big\rangle$ is

$$G_\pm(\omega) = \frac{\frac{1}{2} \pm \frac{t}{W}}{\omega + \mu - (E_0 \pm t)} + \frac{\frac{1}{2} \mp \frac{t}{W}}{\omega + \mu - (U \pm t - E_0)}.$$

Here the first term corresponds to $|j\rangle = |\psi_0\rangle$, whereas the second term corresponds to $|i\rangle = |\psi_0\rangle$ in (10). Equating $G^{-1}_\pm(\omega) = \omega + \mu \pm t - \Sigma_\pm(\omega)$ we can now solve for the self energy:

$$\Sigma_\pm(\omega) \;\; = \;\; \frac{U}{2} + \frac{\left(\frac{U}{2}\right)^2}{\omega + \mu \mp 3t - \frac{U}{2}}.$$

$\Sigma_\pm(\omega)$ is indeed consistent with the general form (12) and the additive constant $U/2$ is indeed the Hartree-Fock potential. For the dimer $\Sigma_\pm(\omega)$ has only a single pole of strength $(U/2)^2$ which does have a substantial dispersion, in that its positions for $k = 0$ and $k = \pi$ differ by a full $6t$. Plots of $A(k, \omega)$ and the imaginary part of $\Sigma(k, \omega)$ are given in Figure 2. The part

of $A(k,\omega)$ for $\omega < 0$ gives the photoemission spectrum (PES) whereas the part for $\omega > 0$ is the inverse photoemission spectrum (IPES). PES and IPES form two 'bands' separated by an energy gap of $\approx U$ – these are the lower and upper Hubbard band. The pole of the self energy is within the Hubbard gap and its dispersion is downward as one goes from $k = 0$ to $k = \pi$, that means 'inverted' as compared to that of the energy ε_k. Interestingly, some of the features observed in the dimer generalize to larger systems. Figure 2 also shows $A(k,\omega)$ and the imaginary part of $\Sigma(k,\omega)$ for the allowed momenta of the 16- and 18-site cluster which were obtained by Lanczos diagonalization [10]. In $A(k,\omega)$ one can again recognize the lower and upper Hubbard band – for the larger clusters these 'bands' consist not only of sharp peaks but contain extended incoherent continua. As was the case for the dimer, $\Sigma(k,\omega)$ has an isolated peak of strength $\propto (U/2)^2$ within the Hubbard gap. Over a large part of the Brillouin zone this peak has a smooth dispersion which closely follows the 'inverted' band dispersion, $\zeta_{\boldsymbol{k}} = \frac{U}{2} - \varepsilon_{\boldsymbol{k}}$.

4 Spin density wave theory

Next, we discuss spin density wave theory which is a mean-field theory and which is reasonably successful in describing the antiferromagnetic phase of the Hubbard model. By comparing the result of applying both sides to the four possible states of a single atom – $|0\rangle$, $|\uparrow\rangle$, $|\downarrow\rangle$ and $|\uparrow\downarrow\rangle$ – it is easy to verify that

$$U\, n_{i,\uparrow} n_{i,\downarrow} = \frac{n_i U}{2} - \frac{2U}{3}\, \boldsymbol{S}_i^2 \,.$$

This equation shows that the system can lower its Coulomb energy by forming *magnetic moments*, $\langle \boldsymbol{S}_i^2 \rangle \neq 0$, where $\langle \dots \rangle$ denotes the expectation value. These magnetic moments may be either static – so that $\langle \boldsymbol{S}_i \rangle \neq 0$ – or fluctuating, i.e., $\langle \boldsymbol{S}_i \rangle = 0$. An example for fluctuating moments is provided by the Hubbard dimer discussed above. Namely for the ground state $|\psi_0\rangle$ of the dimer

$$\begin{aligned} S_1^z\,|\psi_0\rangle &= S_1^z \left(\frac{u}{\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{2,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right) + \frac{v}{\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{1,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{2,\downarrow}^\dagger \right) \right) \\ &= \frac{u}{2\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{2,\downarrow}^\dagger - c_{2,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right) \\ \boldsymbol{S}_1^2\,|\psi_0\rangle &= \frac{3u}{4\sqrt{2}} \left(c_{1,\uparrow}^\dagger c_{2,\downarrow}^\dagger + c_{2,\uparrow}^\dagger c_{1,\downarrow}^\dagger \right), \end{aligned}$$

so that indeed $\langle\psi_0|S_1^z|\psi_0\rangle = 0$ but $\langle\psi_0|\boldsymbol{S}_1^2|\psi_0\rangle = \frac{3u^2}{4}$ (it is easy to shows that $\langle\psi_0|S_1^x|\psi_0\rangle = \langle\psi_0|S_1^y|\psi_0\rangle = 0$ as well and everything also holds true for site 2).

Spin density wave theory, on the other hand, assumes static moments, $\langle \boldsymbol{S}_i \rangle \neq 0$. Based on the results for the Hubbard dimer, which showed that the spins have antiferromagnetic correlations, we expect that these static magnetic moments prefer to be antiparallel on neighboring sites. For a 2D square lattice this requirement defines the Néel state: we choose

$$\begin{aligned} \langle n_{i,\uparrow} \rangle &= \frac{n_e}{2} + \frac{m}{2}\, e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i} \\ \langle n_{i,\downarrow} \rangle &= \frac{n_e}{2} - \frac{m}{2}\, e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i} \end{aligned} \qquad \Rightarrow \qquad \langle n_i \rangle = n_e \quad \text{and} \quad \langle S_i^z \rangle = \frac{m}{2}\, e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i}.$$

Here $\boldsymbol{Q} = (\pi, \pi)$ and the exponential $e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i}$ is $+1$ (-1) if the sum of x- and y-component of the site i is even (odd) which defines the sublattices A (with $e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i} = 1$) and B (with $e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i} = -1$). We consider the trivial decomposition $n_{i,\sigma} = \langle n_{i,\sigma}\rangle + \delta n_{i,\sigma}$, where the *operator* $\delta n_{i,\sigma} = n_{i,\sigma} - \langle n_{i,\sigma}\rangle$ describes fluctuations of n_i around its mean value The basic assumption of all mean-field theories is that these fluctuations are 'small'. Accordingly, when forming the product $U n_{i,\uparrow} n_{i,\downarrow}$ the '2^{nd} order term' $\delta n_{i,\uparrow} \cdot \delta n_{i,\downarrow}$ is discarded, so that the interaction term becomes

$$\begin{aligned} U \sum_i n_{i,\uparrow} n_{i,\downarrow} &\rightarrow U \sum_i \Big(n_{i,\uparrow} \langle n_{i,\downarrow}\rangle + n_{i,\downarrow} \langle n_{i,\uparrow}\rangle - \langle n_{i,\uparrow}\rangle \langle n_{i,\downarrow}\rangle \Big) \\ &= \frac{n_e U}{2} \sum_{i,\sigma} n_{i,\sigma} - \frac{mU}{2} \sum_i e^{i\boldsymbol{Q}\cdot\boldsymbol{R}_i} (n_{i,\uparrow} - n_{i,\downarrow}) - NU \, \frac{n_e^2 - m^2}{4} \, . \end{aligned}$$

Switching to Fourier transformed operators the mean-field Hamiltonian $K_{MF} = H_{MF} - \mu N$ becomes

$$K_{MF} = \sum_{\mathbf{k},\sigma} \tilde{\varepsilon}_{\mathbf{k}} \, c^\dagger_{\mathbf{k},\sigma} c_{\mathbf{k},\sigma} - \Delta \sum_{\boldsymbol{k}} \Big(c^\dagger_{\mathbf{k}+\mathbf{Q},\uparrow} c_{\mathbf{k},\uparrow} - c^\dagger_{\mathbf{k}+\mathbf{Q},\downarrow} c_{\mathbf{k},\downarrow} \Big) - NU \, \frac{n^2 - m^2}{4} \, , \qquad (16)$$

where $\tilde{\varepsilon}_{\mathbf{k}} = \varepsilon_{\mathbf{k}} + \frac{n_e U}{2} - \mu$ and $\Delta = \frac{mU}{2}$. The term $\propto \Delta$ appears to be non-Hermitian at first glance. However, by shifting the summation variable $\mathbf{k} \rightarrow \mathbf{k} + \mathbf{Q}$ and noting that $\mathbf{k} + 2\mathbf{Q} = \mathbf{k}$ one can see that the term is in fact its own Hermitian conjugate.

The Hamiltonian (16) is quadratic in the Fermion operators and thus can be diagonalized by a unitary transformation. In a second step, the value of m is determined by recomputing $\langle n_{i,\sigma}\rangle$ from this solution and demanding self-consistency. We will now carry out this program thereby following Gorkov's re-derivation of BCS theory [11] in terms of the imaginary time Green functions introduced in the first section.

We define the Green function $G_\sigma(\mathbf{k}, \tau) = -\langle T \, c_{\mathbf{k},\sigma}(\tau) c^\dagger_{\mathbf{k},\sigma}\rangle$ and with the Hamiltonian (16) its equation of motion is

$$\begin{aligned} -\frac{\partial}{\partial\tau} G_\sigma(\mathbf{k},\tau) &= \delta(\tau) \langle \{ c^\dagger_{\mathbf{k},\sigma}, c_{\mathbf{k},\sigma} \} \rangle - \langle T \, [c_{\mathbf{k},\sigma}(\tau), K_{MF}] c^\dagger_{\mathbf{k},\sigma} \rangle \\ &= \delta(\tau) + \tilde{\varepsilon}_{\mathbf{k}} \, G_\sigma(\mathbf{k},\tau) \mp \Delta \, \tilde{G}_\sigma(\mathbf{k},\tau) \, , \end{aligned}$$

where the upper sign holds for $\sigma = \uparrow$ and we have introduced the *anomalous Green function* $\tilde{G}_\sigma(\mathbf{k},\tau) = -\langle T \, c_{\mathbf{k}+\mathbf{Q},\sigma}(\tau) c^\dagger_{\mathbf{k},\sigma}\rangle$. In a nonmagnetic system $\tilde{G}_\sigma(\mathbf{k},\tau)$ would be zero due to momentum conservation but in a magnetic system it can be different from zero. Its equation of motion is

$$\begin{aligned} -\frac{\partial}{\partial\tau} \tilde{G}_\sigma(\mathbf{k},\tau) &= \delta(\tau) \langle \{ c^\dagger_{\mathbf{k},\sigma}, c_{\mathbf{k}+\mathbf{Q},\sigma} \} \rangle - \langle T \, [c_{\mathbf{k}+\mathbf{Q},\sigma}(\tau), K_{MF}] c^\dagger_{\mathbf{k},\sigma} \rangle \\ &= \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}} \, \tilde{G}_\sigma(\mathbf{k},\tau) \mp \Delta \, G_\sigma(\mathbf{k},\tau) \, . \end{aligned}$$

The system of equations of motion therefore closes and upon Fourier transformation with respect to τ – whereby $-\partial_\tau \rightarrow i\omega_\nu$ – and becomes

$$\begin{pmatrix} i\omega_\nu - \tilde{\varepsilon}_{\mathbf{k}} & \pm\Delta \\ \pm\Delta & i\omega_\nu - \tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}} \end{pmatrix} \begin{pmatrix} G_\sigma(\mathbf{k}, i\omega_\nu) \\ \tilde{G}_\sigma(\mathbf{k}, i\omega_\nu) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} . \qquad (17)$$

Next we recall [12] that for any complex number a and any complex vector $\boldsymbol{b}$:

$$(a\mathbf{1}+\boldsymbol{b}\cdot\boldsymbol{\tau})(a\mathbf{1}-\boldsymbol{b}\cdot\boldsymbol{\tau})=a^2-\boldsymbol{b}^2 \qquad \Rightarrow \qquad (a\mathbf{1}+\boldsymbol{b}\cdot\boldsymbol{\tau})^{-1}=\frac{a\mathbf{1}-\boldsymbol{b}\cdot\boldsymbol{\tau}}{a^2-\boldsymbol{b}^2}.$$

Defining

$$\zeta_{\boldsymbol{k}}=\frac{1}{2}\left(\tilde{\varepsilon}_{\mathbf{k}}+\tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}}\right), \qquad \eta_{\boldsymbol{k}}=\frac{1}{2}\left(\tilde{\varepsilon}_{\mathbf{k}}-\tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}}\right), \tag{18}$$

the 2×2 coefficient matrix H on the left hand side of (17) can be brought to the form $a\mathbf{1}+\boldsymbol{b}\cdot\boldsymbol{\tau}$ with $a=i\omega_\nu-\zeta_{\boldsymbol{k}}$ and $\boldsymbol{b}=(\pm\Delta,0,-\eta_{\boldsymbol{k}})$. Multiplying both sides of (17) by H^{-1} we obtain

$$\begin{aligned}
G_\sigma(\mathbf{k},i\omega_\nu)&=\frac{i\omega_\nu-\tilde{\varepsilon}_{\mathbf{k}+\mathbf{Q}}}{(i\omega_\nu-\zeta_{\boldsymbol{k}})^2-W_{\mathbf{k}}^2}=\frac{Z_{\boldsymbol{k}}^{(-)}}{i\omega_\nu-E_{\boldsymbol{k}}^{(-)}}+\frac{Z_{\boldsymbol{k}}^{(+)}}{i\omega_\nu-E_{\boldsymbol{k}}^{(+)}},\\
\tilde{G}_\sigma(\mathbf{k},i\omega_\nu)&=\frac{\mp\Delta}{(i\omega_\nu-\zeta_{\boldsymbol{k}})^2-W_{\mathbf{k}}^2}=\frac{\mp\Delta}{(i\omega_\nu-E_{\boldsymbol{k}}^{(-)})(i\omega_\nu-E_{\boldsymbol{k}}^{(+)})},
\end{aligned} \tag{19}$$

where $W_{\mathbf{k}}^2=\eta_{\boldsymbol{k}}^2+\Delta^2$ and we have introduced the quasiparticle energies and weights

$$E_{\boldsymbol{k}}^{(\pm)}=\zeta_{\boldsymbol{k}}\pm W_{\mathbf{k}} \qquad Z_{\boldsymbol{k}}^{(\pm)}=\frac{1}{2}\left(1\pm\frac{\eta_{\boldsymbol{k}}}{W_{\mathbf{k}}}\right). \tag{20}$$

From the expression for G we obtain the single particle spectral function

$$A(\mathbf{k},\omega) \;=\; Z_{\boldsymbol{k}}^{(-)}\,\delta(\omega-E_{\boldsymbol{k}}^{(-)})+Z_{\boldsymbol{k}}^{(+)}\,\delta(\omega-E_{\boldsymbol{k}}^{(+)}).$$

Rather than a single band with dispersion $\varepsilon_{\boldsymbol{k}}$ and weight 1, SDW theory thus predicts two bands with a reduced and k-dependent spectral weight. Figure (3) shows $A(\boldsymbol{k},\omega)$ obtained by evaluating (19) for different values of Δ. Increasing Δ opens a gap in the original band. For small Δ the spectral weight of the bands has a substantial $\boldsymbol{k}$-dependence and for the photoemission spectrum ($\omega<0$) drops sharply upon crossing the noninteracting Fermi surface at $(\frac{\pi}{2},\frac{\pi}{2})$. So far we have carried out the first step, the solution of the mean-field Hamiltonian. In the second step we have to recompute $\langle\boldsymbol{S}_i\rangle$ from this solution. We note that for *any* site i

$$n_e=\langle n_{i,\uparrow}+n_{i,\downarrow}\rangle \quad \text{and} \qquad m=\langle n_{i,\uparrow}-n_{i,\downarrow}\rangle\; e^{i\mathbf{Q}\cdot\mathbf{R}_i}\,.$$

We sum this over i, divide by N, and switch to Fourier transformed c-operators:

$$\begin{aligned}
n_e&=\frac{1}{N}\sum_{\mathbf{k}}\left\langle c_{\mathbf{k},\uparrow}^\dagger c_{\mathbf{k},\uparrow}+c_{\mathbf{k},\downarrow}^\dagger c_{\mathbf{k},\downarrow}\right\rangle \qquad =\frac{1}{N}\sum_{\mathbf{k}}\Big(G_\uparrow(\mathbf{k},\tau=0^-)+G_\downarrow(\mathbf{k},\tau=0^-)\Big)\\
&=\frac{2}{N}\sum_{\mathbf{k}}G_\uparrow(\mathbf{k},\tau=0^-)\\
m&=\frac{1}{N}\sum_{\mathbf{k}}\left\langle c_{\mathbf{k}+\mathbf{Q},\uparrow}^\dagger c_{\mathbf{k},\uparrow}-c_{\mathbf{k}+\mathbf{Q},\downarrow}^\dagger c_{\mathbf{k},\downarrow}\right\rangle=\frac{1}{N}\sum_{\mathbf{k}}\Big(\tilde{G}_\uparrow(\mathbf{k},\tau=0^-)-\tilde{G}_\downarrow(\mathbf{k},\tau=0^-)\Big)\\
&=\frac{2}{N}\sum_{\mathbf{k}}\tilde{G}_\uparrow(\mathbf{k},\tau=0^-)\,,
\end{aligned} \tag{21}$$

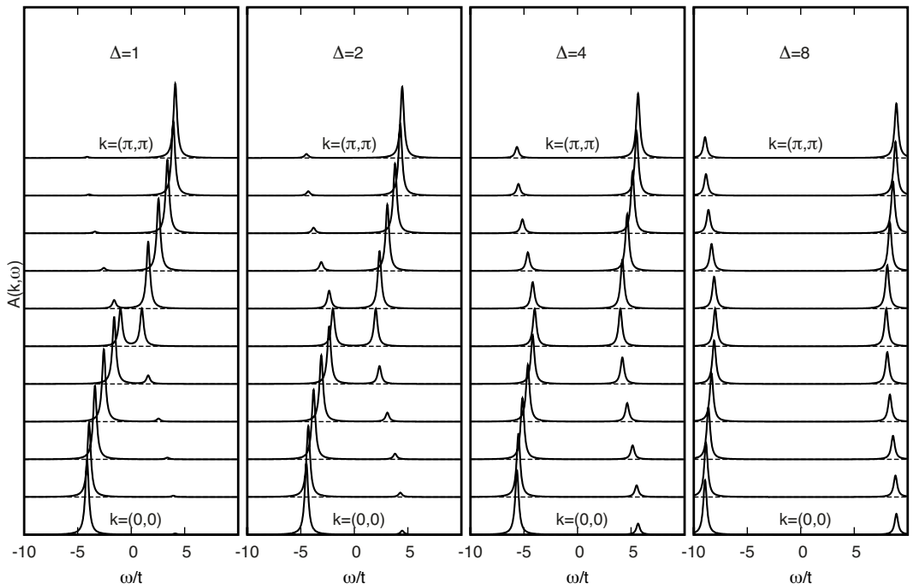


Fig. 3: *Single particle spectral function $A(\mathbf{k},\omega)$ along $(0,0) \rightarrow (\pi,\pi)$ as obtained by spin density wave mean-field theory for different values of $\Delta = \frac{mU}{2}$ and half-filling. The original band dispersion is that of eqn. (2) with $t = 1$ and $t' = 0$. The part for $\omega < 0$ ($\omega > 0$) gives the photoemission (inverse photoemission) spectrum.*

where we shifted the summation variable $\mathbf{k} \rightarrow \mathbf{k} + \mathbf{Q}$ and used (19) in the equation for m. Inserting the Fourier expansion (9) and using (19) for the Green functions we obtain

$$n_e = \frac{2}{N} \sum_{\mathbf{k}} \left(-\frac{1}{\beta} \right) \sum_{\nu} e^{i\omega_\nu 0^+} \left(\frac{-Z_{\boldsymbol{k}}^{(-)}}{i\omega_\nu - E_{\boldsymbol{k}}^{(-)}} + \frac{-Z_{\boldsymbol{k}}^{(+)}}{i\omega_\nu - E_{\boldsymbol{k}}^{(+)}} \right),$$

$$m = \frac{2}{N} \sum_{\mathbf{k}} \left(-\frac{1}{\beta} \right) \sum_{\nu} e^{i\omega_\nu 0^+} \frac{\Delta}{(i\omega_\nu - E_{\boldsymbol{k}}^{(-)})(i\omega_\nu - E_{\boldsymbol{k}}^{(+)})}.$$

We now replace the sum over Matsubara frequencies by a contour integration, which is a standard technique in field theory (see, e.g., section 25 of [6]). Namely for any function $F(\omega)$ which is analytic along the imaginary axis we have

$$-\frac{1}{\beta} \sum_{\nu} F(i\omega_\nu) = \frac{1}{2\pi i} \oint_{\mathcal{C}} d\omega \; f(\omega) \; F(\omega)$$

where

$$f(\omega) = \frac{1}{e^{\beta\omega} + 1},$$

is the Fermi function and the contour $\mathcal{C}$ encircles the imaginary axis in counterclockwise fashion, see Fig. 4(a). This replacement makes use of the theorem of residues and the easily verified fact that the Fermi function $f(\omega)$ has simple poles with residuum $-1/\beta$ at all Matsubara frequencies $i\omega_\nu$. Next we note that the integrals along the two clover-shaped contours in Fig. 4(b)

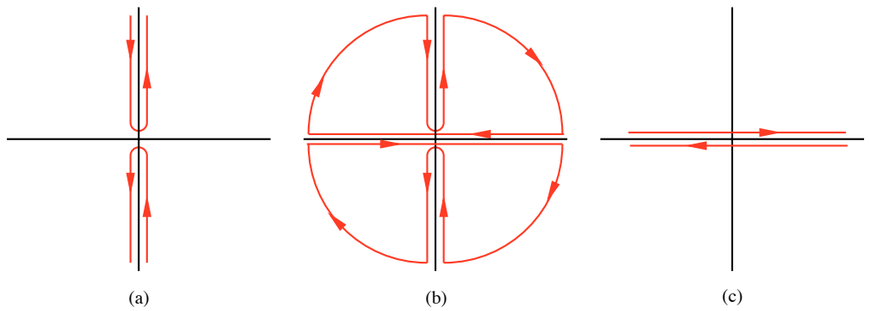


Fig. 4: *Since the integrals along the two contours in (b) are zero and the contributions from the circular arcs vanish, the integral along the contour in (a) is equal to that over contour (c).*

are zero, *provided* the integrand is analytic in the interior of the two curves. Since the Fermi function $f(\omega)$ has all of its poles along the imaginary axis, whereas the Green function has its poles on the real axis, both of which are outside of the curves in Fig. 4(b), this is certainly true. Next, Jordan's lemma can be invoked to establish that the integrals along the large semicircles vanish. Here the Fermi function $f(\omega)$ guarantees that the contribution from the semicircle with $\mathrm{Re}(\omega) > 0$ vanishes, whereas the factor $e^{\omega 0^+}$ does the same for the semicircle with $\mathrm{Re}(\omega) < 0$ (which also shows that it was necessary to keep this factor). It follows that the integral along the contour $\mathcal{C}$ in Fig. 4(a) is equal to that along the contour $\mathcal{C}'$ in Fig. 4(c) (note the inverted direction of the curves in (c) as compared to (b)!). The contour in Fig. 4(c) encircles the real axis clockwise – it follows from the theorem of residues that the contour integral is $(-2\pi i)$ times the sum of the residua of the respective $f(\omega)F(\omega)$ at its two poles at $E_{\mathbf{k}}^{(\pm)}$. Using the definition $\Delta = \frac{mU}{2}$ we thus obtain

$$n_e = \frac{2}{N} \sum_{\mathbf{k}} \left(Z_{\mathbf{k}}^{(-)} f(E_{\mathbf{k}}^{(-)}) + Z_{\mathbf{k}}^{(+)} f(E_{\mathbf{k}}^{(+)}) \right)$$
$$1 = \frac{U}{N} \sum_{\mathbf{k}} \frac{1}{2W_{\boldsymbol{k}}} \left(f(E_{\mathbf{k}}^{-}) - f(E_{\mathbf{k}}^{+}) \right).$$

It follows from (2) that $\zeta_{\boldsymbol{k}} = 4t' \cos(k_x)\cos(k_y) + \frac{U}{2} - \mu$ and $\eta_{\boldsymbol{k}} = -2t\left(\cos(k_x) + \cos(k_y)\right)$. Moreover, $\zeta_{\boldsymbol{k}+\boldsymbol{Q}} = \zeta_{\boldsymbol{k}}$ and $\eta_{\boldsymbol{k}+\boldsymbol{Q}} = -\eta_{\boldsymbol{k}}$, so that $W_{\boldsymbol{k}+\boldsymbol{Q}} = W_{\boldsymbol{k}}$. Next, (20) shows that $Z_{\boldsymbol{k}+\boldsymbol{Q}}^{(\pm)} = Z_{\boldsymbol{k}}^{(\mp)}$, $Z_{\boldsymbol{k}+\boldsymbol{Q}}^{(\pm)} + Z_{\boldsymbol{k}}^{(\pm)} = 1$ and $E_{\boldsymbol{k}+\boldsymbol{Q}}^{(\pm)} = E_{\boldsymbol{k}}^{(\pm)}$. These relations makes it possible to restrict the sum over momenta in (22) and (23) to the antiferromagnetic Brillouin zone (AFBZ) and we obtain

$$n_e = \frac{2}{N} \sum_{\mathbf{k}\in AFBZ} \left(f(E_{\mathbf{k}}^{(-)}) + f(E_{\mathbf{k}}^{(+)}) \right), \tag{22}$$
$$1 = \frac{U}{N} \sum_{\mathbf{k}\in AFBZ} \frac{1}{W_{\boldsymbol{k}}} \left(f(E_{\mathbf{k}}^{-}) - f(E_{\mathbf{k}}^{+}) \right). \tag{23}$$

For given $\varepsilon_{\boldsymbol{k}}$, U and n_e we now have a complete description of the system. For a qualitative discussion, let us assume that the lower band is completely occupied (i.e. $f(E_{\mathbf{k}}^{(-)}) = 1$) and the

upper one completely empty (i.e. $f(E^{(+)}_{\mathbf{k}}) = 0$). Since the number of momenta in the AFBZ is $\frac{N}{2}$ this implies $n_e = 1$ – or 'half-filling'. In the paramagnetic phase $n_e = 1$ would correspond to a half-filled band whereas SDW theory gives an antiferromagnetic insulator. Since $\Delta = \frac{mU}{2}$ we expect $\Delta/t \to \infty$ as $U/t \to \infty$. It follows that $W_{\boldsymbol{k}} \to \Delta$ so from (23) we obtain $m \to 1$ in this limit, which means the fully polarized Néel state. For an electron density $n_e = 1 - \delta$ and in the limit $T \to 0$ the fraction of momenta in the AFBZ which are occupied is $1 - \delta$ as well – put another way, we have hole pockets which cover a fraction of δ of the AFBZ, and consequently, in the original zone scheme, also a fraction of δ of the full Brillouin zone. Interestingly, in a doped semiconductor – with electron density $n_e = 2 - \delta$ – the hole pockets would cover only a fraction of $\delta/2$ of the Brillouin zone.

The equation for the temperature where m starts to deviate from zero – the so-called Néel temperature T_N – can be obtained by taking the limit $m \to 0$ whence $W_{\boldsymbol{k}} \to |\eta_{\boldsymbol{k}}|$. Inserting this one finds that T_N is determined by the equation

$$1 = -\frac{U}{N} \sum_{\mathbf{k}} \frac{f(\varepsilon_{\boldsymbol{k}+\boldsymbol{Q}}) - f(\varepsilon_{\boldsymbol{k}})}{\varepsilon_{\boldsymbol{k}+\boldsymbol{Q}} - \varepsilon_{\boldsymbol{k}}} \,.$$

Note that the left hand side is guaranteed to be positive because $f(\omega)$ is a monotonously decreasing function of its argument and that the temperature appears implicitly in the Fermi functions. Let us assume that we keep the temperature constant and increase U starting from zero. We ask under what conditions ordering sets in for an as low U as possible. In order for the $\mathbf{k}$-sum to be large, there must be many pairs $(\boldsymbol{k}, \boldsymbol{k} + \boldsymbol{Q})$ such that their energies are close to one another and such that both momenta are close to the Fermi surface – because only then the difference of Fermi functions can be appreciable. This gives us the condition of *Fermi surface nesting*: the ordering vector $\boldsymbol{Q}$ must connect as long sections of the Fermi surface as possible.

For the general case the system of equations (22) and (23) can be solved by starting with some value m_{in}, determining μ such that (22) gives the correct electron density, and then evaluating m_{out} from (23). By scanning m_{in} one can then determine the value where $m_{out} = m_{in}$. Fig. 5 shows some results obtained in this way. A detailed discussion of the resulting phase diagram of the Hubbard model for various band fillings and values of the hopping integrals t and t' can be found, e.g., in Refs. [14] – here we do not discuss this in detail.

As already mentioned the above derivation was originally invented by Gorkov to re-derive the BCS theory of superconductivity [11] and can be easily generalized to any mean-field theory. The formulation in terms of Green functions makes it easy to include the effects of disorder or spatial variations and the various techniques applied above – such as the use of equations of motion, expressing the order parameter in terms of the anomalous Green function and the evaluation of sums by contour integration – are applied again and again in many papers on advanced problems in superconductivity.

SDW theory describes some features of undoped cuprate superconductors and related compounds correctly, but fails even qualitatively in many aspects. For example, the 'parent compound' compounds La_2CuO_4 indeed is an antiferromagnetic insulator with a Néel temperature of around 300 K. SDW theory also qualitatively reproduces the phenomenon of the 'remnant

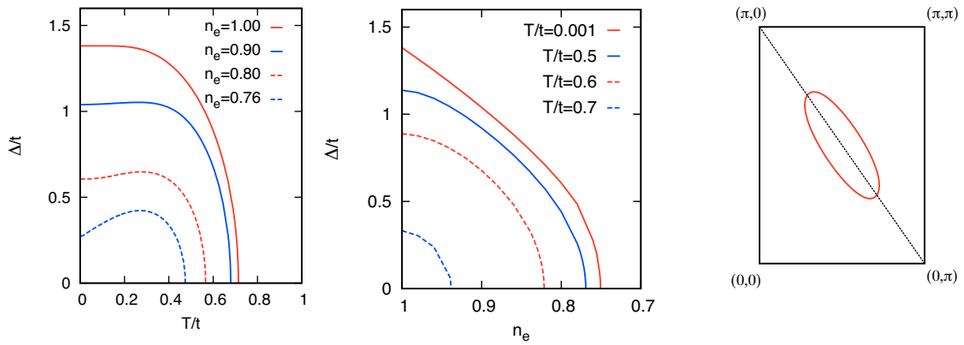


Fig. 5: *Dependence of the SDW gap on temperature T for different electron densities n_e (left) and on electron density n_e for different temperatures T (center). Right: Fermi surface for an electron density $n_e = 0.9$. Parameter values are $U/t = 4$ and $t'/t = 0.2$.*

Fermi surface' [13]. The phrase describes the experimental observation that the intensity of the quasiparticle band observed in ARPES experiments on undoped CuO-compounds such as $Sr_2CuO_2Cl_2$ drops sharply when crossing the noninteracting Fermi surface, precisely as seen in the spectra in Fig. 3. However, to reproduce the sharpness of the drop seen in experiment would require values of Δ which would be much too small to reproduce the magnitude of the insulating gap.

Moreover, contrary to the predictions of SDW theory the insulating gap does not go to zero at the Néel temperature but is essentially temperature independent - which is the hallmark of a true Mott-insulator. Moreover, for doped compounds such as $La_{2-x}Ba_xCuO_4$ static antiferromagnetic order disappears already for hole concentrations of a few percent whereby again the Hubbard gap observed in the insulator persists with practically unchanged magnitude. What we therefore really need to describe is a *paramagnetic system* with a Hubbard gap and this will be the objective of the next section.

5 The Hubbard-I approximation

This is the 'defining approximation' of the Mott-insulator by which Hubbard for the first time introduced central concepts of strongly correlated electron systems such as the two Hubbard bands [1]. In the following we first give a sloppy re-derivation which is meant to clarify the physical content of the Hubbard-I approximation and then present Hubbard's rigorous derivation in terms of Green functions.

We consider the limit of finite U and $t_{i,j} = 0$, $N_\uparrow = N_\downarrow = N/2$ so that $N_e = N$. The ground state has one electron per lattice site and the energy is $E = 0$. Since the spin of the electron at any given site is arbitrary this ground state is highly degenerate. We ignore this degeneracy and assume that there is a unique state $|\Psi_0\rangle$ which may be thought of as a suitable superposition of all these degenerate states and which we assume to be 'disordered' – it will become clear in a moment what this means.

Next we assume that a small but finite $t_{i,j}$ is switched on. Then, an electron of spin σ can be transferred from a site j to another site i resulting in an empty site at j and a double occupancy at site i. The energy thereby increases by U. The hopping process is possible only if the electron which was originally at the site i has the spin $-\sigma$ and since our initial state $|\Psi_0\rangle$ is 'disordered' the probability for this to be the case is $1/2$ – which is the definition of 'disordered'. We now interpret the original state $|\Psi_0\rangle$ as the vacuum – denoted by $|0\rangle$ – of our theory and the state created by the hopping process as containing a Fermionic hole-like particle at j and a Fermionic double-occupancy-like particle at site i: $d^\dagger_{i,\sigma} h^\dagger_{j,-\sigma}|0\rangle$. The order of the Fermionic operators in this state is due to the fact that in the original hopping term the annihilation operator $c_{j,\sigma}$ which creates the hole stands to the right of the creation operator $c^\dagger_{i,\sigma}$ which creates the double occupancy. Moreover we assign the negative spin to the operator which creates the hole because replacement of, e.g., an $\uparrow$-electron by a hole decreases the z-spin by $1/2$. We obtain the following Hamiltonian to describe the holes and double occupancies:

$$H_{\text{eff},1} = \frac{1}{2} \sum_{i,j} \sum_{\sigma} \left(t_{i,j}\, d^\dagger_{i,\sigma} h^\dagger_{j,-\sigma} + H.c. \right) + U \sum_{i,\sigma} d^\dagger_{i,\sigma} d_{i,\sigma}\,. \tag{24}$$

Once a hole and a double occupancy have been created, each of these particles may be transported further by the hopping term. If we assume that the surplus or missing electron retains its spin – which means that the double occupancies and holes propagate without 'leaving a trace' of inverted spins – for example a surplus $\uparrow$-electron can hop from site i to site j only if the spin at site j is $\downarrow$ – we again have the probability $1/2$ for this process. We therefore can write down the second term terms for the effective Hamiltonian

$$H_{\text{eff},2} = \frac{1}{2} \sum_{i,j} \sum_{\sigma} t_{i,j} \left(d^\dagger_{i,\sigma} d_{j,\sigma} - h^\dagger_{i,-\sigma} h_{j,-\sigma} \right). \tag{25}$$

The negative sign of the hopping term for holes is due to the fact that the original hopping term has to be rewritten as $-t_{i,j} c_{j,\sigma} c^\dagger_{i,\sigma}$ to describe the propagation of a hole. Addition of (24) and (25) and Fourier transformation gives

$$H_{\text{eff}} = \sum_{\mathbf{k},\sigma} \left(\left(\frac{\varepsilon_{\mathbf{k}}}{2} + U \right) d^\dagger_{\mathbf{k},\sigma} d_{\mathbf{k},\sigma} - \frac{\varepsilon_{\mathbf{k}}}{2} h^\dagger_{\mathbf{k},\sigma} h_{\mathbf{k},\sigma} \right) + \sum_{\mathbf{k},\sigma} \frac{\varepsilon_{\mathbf{k}}}{2} \left(d^\dagger_{\mathbf{k},\sigma} h^\dagger_{-\mathbf{k},-\sigma} + H.c. \right), \tag{26}$$

where $\varepsilon_{\mathbf{k}}$ is the Fourier transform of $t_{i,j}$. Note that this now is a quadratic form where the Coulomb interaction is described by the extra energy of U for the double-occupancy-like 'particle'. Via the unitary transformation

$$\begin{aligned} \gamma_{-,\mathbf{k},\sigma} &= \quad u_{\mathbf{k}}\, d_{\mathbf{k},\sigma} + v_{\mathbf{k}}\, h^\dagger_{-\mathbf{k},-\sigma} \\ \gamma_{+,\mathbf{k},\sigma} &= -v_{\mathbf{k}}\, d_{\mathbf{k},\sigma} + u_{\mathbf{k}}\, h^\dagger_{-\mathbf{k},-\sigma} \end{aligned} \tag{27}$$

this can be solved, resulting in the dispersion relations for the lower and upper Hubbard band

$$E^{(\pm)}_{\boldsymbol{k}} = \frac{1}{2} \left(\varepsilon_{\mathbf{k}} + U \pm \sqrt{\varepsilon^2_{\mathbf{k}} + U^2} \right). \tag{28}$$

In the limit $U/t \gg 1$ this simplifies to $E_{\mathbf{k},-} = \frac{\varepsilon_\mathbf{k}}{2}$, $E_{\mathbf{k},+} = \frac{\varepsilon_\mathbf{k}}{2} + U$ so that the original band with dispersion $\varepsilon_\mathbf{k}$ is split into two bands, separated by a gap of U and each having half of the original width. Qualitatively this is similar to the exact result for the Hubbard dimer. For the case of particle-hole symmetry the chemical potential is $U/2$ so that the lower band is completely filled, the upper one completely empty. Rather than being a metal – as expected for the situation of a half-filled band – the presence of the Coulomb interaction turns the system into an insulator. From the above we can see that this is the consequence of 'expanding around' the hypothetical 'vacuum state' $|\Psi_0\rangle$ with one electron per site so that we obtain a dilute gas of hole-like and double-occupancy-like particles which are created in pairs and propagate, whereby the double-occupancies have a large 'energy of formation' of U.
Next, we derive these results in a more rigorous fashion thereby following Hubbard's original paper [1]. We split the electron operator into two components

$$c_{i,\sigma} \quad = \quad c_{i,\sigma} n_{i,-\sigma} + c_{i,\sigma}(1 - n_{i,-\sigma}) = \hat{d}_{i,\sigma} + \hat{c}_{i,\sigma}, \tag{29}$$

which obey $[\hat{d}_{i,\sigma}, H_U] = U\hat{d}_{i,\sigma}$ and $[\hat{c}_{i,\sigma}, H_U] = 0$. Next we define the four Green functions

$$G_{\alpha,\beta}(\mathbf{k}, \tau) = -\left\langle T\, \alpha_{\mathbf{k},\sigma}(\tau) \beta^\dagger_{\mathbf{k},\sigma} \right\rangle, \tag{30}$$

where $\alpha, \beta \in \{\hat{c}, \hat{d}\}$. These obey the equations of motion

$$-\frac{\partial}{\partial_\tau} G_{\alpha,\beta}(\vec{k}, \tau) = \delta(\tau) \left\langle \left\{ \beta^\dagger_{\mathbf{k},\sigma}, \alpha_{\mathbf{k},\sigma} \right\} \right\rangle - \left\langle T\, [\alpha_{\mathbf{k},\sigma}(\tau), H]\, \beta^\dagger_{\mathbf{k},\sigma} \right\rangle.$$

The commutators $[\alpha_{\mathbf{k},\sigma}, H_U]$ are trivial but the commutators with the kinetic term H_t are involved. After some algebra – thereby using the identity $n_{i,\sigma} = \frac{n_i}{2} + \sigma S^z_i$ - we find:

$$\begin{aligned}
[\hat{c}_{i,\uparrow},\, H_t] &= \sum_j t_{ij} \left[\left(1 - \frac{n_e}{2}\right) c_{j,\uparrow} + (c_{j,\uparrow} S^z_i + c_{j,\downarrow} S^-_i) - \frac{1}{2} c_{j,\uparrow}(n_i - n_e) + c^\dagger_{j,\downarrow} c_{i,\downarrow} c_{i,\uparrow} \right], \\
[\hat{d}_{i,\uparrow}, H_t] &= \sum_j t_{ij} \left[\qquad \frac{n_e}{2}\; c_{j,\uparrow} - (c_{j,\uparrow} S^z_i + c_{j,\downarrow} S^-_i) + \frac{1}{2} c_{j,\uparrow}(n_i - n_e) - c^\dagger_{j,\downarrow} c_{i,\downarrow} c_{i,\uparrow} \right].
\end{aligned} \tag{31}$$

The first term on the r.h.s describes the 'simple' propagation of the hole. The second term is the contraction of the spin-1 operator $\mathbf{S}_i$ and the spinor $c_{j,\sigma}$ into a spin-1/2 object. It describes how a hole moves to site j but leaves behind a spin-excitation at site the i. Similarly, the third term describes hopping combined with the creation of a density excitation at site j whereas the last term describes the coupling of a pair-excitation (this would be important for negative U). The Hubbard-I approximation is obtained by keeping only the first term in each of the square brackets on the respective right hand side – obviously a rather crude approximation. After Fourier transformation we obtain

$$\begin{aligned}
[\hat{c}_{\mathbf{k},\uparrow},\, H] &\approx \left(1 - \frac{n_e}{2}\right) \varepsilon_\mathbf{k}\, (\hat{c}_{\mathbf{k},\uparrow} + \hat{d}_{\mathbf{k},\uparrow}) \qquad\qquad - \mu\, \hat{c}_{\mathbf{k},\uparrow} \\
[\hat{d}_{\mathbf{k},\uparrow}, H] &\approx \qquad \frac{n_e}{2}\; \varepsilon_\mathbf{k}\, (\hat{c}_{\mathbf{k},\uparrow} + \hat{d}_{\mathbf{k},\uparrow}) + U\hat{d}_{\mathbf{k},\uparrow} - \mu\, \hat{d}_{\mathbf{k},\uparrow}.
\end{aligned}$$

Using the anticommutation relations $\{\hat{d}^\dagger_{i,\sigma}, \hat{d}_{i,\sigma}\} = n_{i-\sigma}$, $\{\hat{c}^\dagger_{i,\sigma}, \hat{c}_{i,\sigma}\} = (1-n_{i-\sigma})$, $\{\hat{d}^\dagger_{i,\sigma}, \hat{c}_{i,\sigma}\} = \{\hat{c}^\dagger_{i,\sigma}, \hat{d}_{i,\sigma}\} = 0$ and putting $\langle n_{i,\sigma}\rangle = \frac{n_e}{2}$ we obtain the Fourier transformed equations of motion:

$$\begin{pmatrix} i\omega_\nu + \mu - (1-\frac{n_e}{2})\varepsilon_{\mathbf{k}} & -(1-\frac{n_e}{2})\varepsilon_{\mathbf{k}} \\ -\frac{n_e}{2}\,\varepsilon_{\mathbf{k}} & i\omega_\nu + \mu - \frac{n_e}{2}\,\varepsilon_{\mathbf{k}} - U \end{pmatrix} \begin{pmatrix} G_{\hat{c},\hat{c}} & G_{\hat{c},\hat{d}} \\ G_{\hat{d},\hat{c}} & G_{\hat{d},\hat{d}} \end{pmatrix} = \begin{pmatrix} 1-\frac{n_e}{2} & 0 \\ 0 & \frac{n_e}{2} \end{pmatrix}.$$

We now use the identity (which holds for any 2×2 matrix)

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

to solve for the Green function matrix $\mathbf{G}(\mathbf{k},\omega)$. Since $c_{\mathbf{k},\sigma} = \hat{c}_{\mathbf{k},\sigma} + \hat{d}_{\mathbf{k},\sigma}$ the usual electron Green function $G(\mathbf{k},\omega)$ is $G = G_{\hat{c},\hat{c}} + G_{\hat{c},\hat{d}} + G_{\hat{d},\hat{c}} + G_{\hat{d},\hat{d}}$, which can be brought to the form

$$G(\mathbf{k},\omega) = \frac{Z^{(-)}_{\boldsymbol{k}}}{\omega - E^{(-)}_{\boldsymbol{k}}} + \frac{Z^{(+)}_{\boldsymbol{k}}}{\omega - E^{(+)}_{\boldsymbol{k}}} = \frac{1}{\omega + \mu - \varepsilon_{\mathbf{k}} - \Sigma(\omega)} \tag{32}$$

whereby (with $W_{\boldsymbol{k}} = \sqrt{U^2 + \varepsilon^2_{\boldsymbol{k}} - 2(1-n_e)\varepsilon_{\boldsymbol{k}}U}$):

$$E^{(\pm)}_{\boldsymbol{k}} = \frac{1}{2}\left(U + \varepsilon_{\mathbf{k}} \pm W_{\boldsymbol{k}}\right) - \mu\,, \qquad Z^{(\pm)}_{\boldsymbol{k}} = \frac{1}{2}\left(1 \pm \frac{\varepsilon_{\mathbf{k}} - (1-n_e)U}{W}\right),$$

and the self-energy in Hubbard-I approximation is given by

$$\Sigma(\omega) = \frac{n_e}{2}\,U + \frac{n_e}{2}\left(1-\frac{n_e}{2}\right)\frac{U^2}{\omega + \mu - (1-\frac{n_e}{2})U}\,.$$

The additive constant $\frac{n_e}{2}U$ indeed is the Hartree-Fock potential and there is a single dispersionless pole whose strength is $\left(\frac{U}{2}\right)^2$ at half-filling, $n_e = \frac{1}{2}$. The pole strength thus agrees with the exact solution for the dimer but the lack of dispersion shows that the Hubbard-I approximation would fail already for the Hubbard dimer.
If we specialize to half-filling and put $n_e = 1$ the quasiparticle energies become

$$E^{(\pm)}_{\boldsymbol{k}} = \frac{1}{2}\left(U + \varepsilon_{\mathbf{k}} \pm \sqrt{\varepsilon^2_{\mathbf{k}} + U^2}\right) - \mu.$$

This is (up to the term μ) consistent with the result (28) obtained from the heuristic Hamiltonian (26), thus demonstrating the equivalence of this simple picture with the Hubbard-I approximation. In order to fix the chemical potential for electron densities $n_e < 1$ we write the operator of electron number as

$$\hat{N}_e = 2\sum_i n_{i,\uparrow}n_{i,\downarrow} + \sum_i \left(n_{i,\uparrow}(1-n_{i,\downarrow}) + n_{i,\downarrow}(1-n_{i,\uparrow})\right) = \sum_{i,\sigma}\left(\hat{d}^\dagger_{i,\sigma}\hat{d}_{i,\sigma} + \hat{c}^\dagger_{i,\sigma}\hat{c}_{i,\sigma}\right).$$

The electron density n_e thus can be expressed in terms of the Green functions (30) as

$$n_e = \frac{2}{N}\sum_{\mathbf{k}}\left(G_{\hat{d},\hat{d}}(\mathbf{k},\tau=0^-) + G_{\hat{c},\hat{c}}(\mathbf{k},\tau=0^-)\right), \tag{33}$$

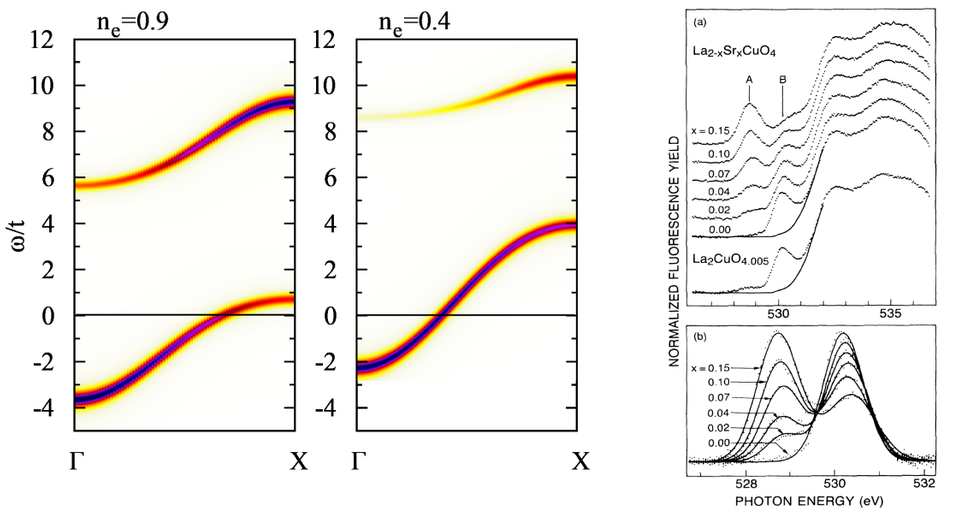


Fig. 6: *Left: Single particle spectral function obtained from the Greens function (32) for two different electron densities. The Fermi energy is zero. Right: X-ray luminescence spectra for $La_{2-x}Sr_xCuO_4$ show the unoccupied part of the lower Hubbard-band (A) and the upper Hubbard-band (B). With decreasing $n_e \approx 1-x$ the upper Hubbard-band rapidly looses intensity. Reprinted with permission from [15], Copyright 1991 by the American Physical society.*

which is analogous to the expression (21) and can be evaluated by the same procedure. To simplify matters we specialize to the limit $t/U \ll 1$ and close to half-filling, $n_e = 1-\delta$ with $\delta \ll 1$. The dispersion relation of the lower band then is

$$E^{(-)}_{\boldsymbol{k}} \approx \frac{1+\delta}{2}\varepsilon_{\boldsymbol{k}} - \frac{\varepsilon^2_{\boldsymbol{k}}}{4U} + \dots$$

where $\dots$ denotes terms of higher order in the small quantities. For not too large t' this has its maximum at (π,π). Assuming that the upper band is high above the chemical potential, the electron density becomes

$$n_e = \frac{2}{N}\sum_{\mathbf{k}} f(E^{(-)}_{\boldsymbol{k}}) \left(\frac{1+\delta}{2} - \frac{\delta\varepsilon^2_{\boldsymbol{k}}}{4U^2} + \dots \right).$$

To discuss the Fermi surface we consider the limit $T \to 0$. For half-filling $n_e = 1$ so that $\delta = 0$ and the lower band must be completely filled – as one would expect for an insulator. Neglecting the term $\propto U^{-2}$ we find that for $\delta > 0$ the fraction of the Brillouin zone where the lower Hubbard band is occupied is $1-2\delta$, or, put another way, there are hole pockets around (π,π) whose volume is a fraction 2δ of the Brillouin zone. As a quantitative example, Figure 6 shows the spectral density obtained from the Green function (32) for $U/t = 8$ and different band fillings, whereas Figure 7 shows the resulting Fermi surfaces and the dependence of the Fermi surface volume on electron density. In Fig. 6 one can recognize the two Hubbard bands separated by an appreciable energy gap. For $n_e = 0.9$ – i.e. close to half-filling – the Fermi energy intersects the lower Hubbard band close to (π,π) resulting in the hole pocket around

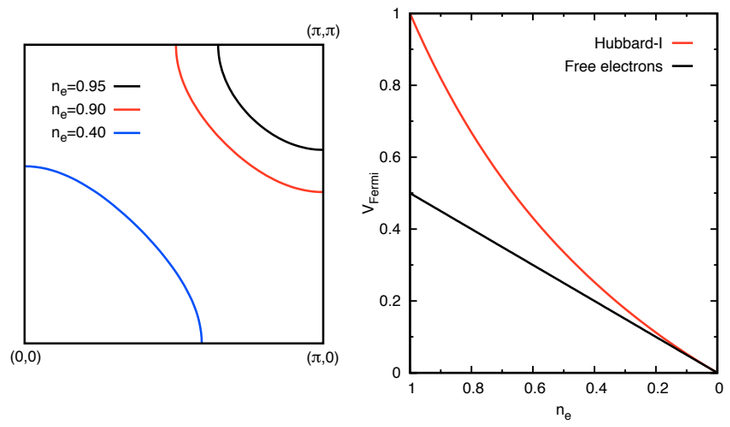


Fig. 7: *Left: Fermi surface for different electron densities. Right: Fermi surface volume as a function of electron density* n_e.

$X = (\pi, \pi)$ discussed above. An interesting feature seen in Fig. 6 is the transfer of spectral weight from the upper to the lower Hubbard band upon hole doping: as the electron density n_e decreases the upper Hubbard band persists but looses weight, whereas the lower Hubbard band becomes more intense. To understand this we note first that for $n_e \leq 1$ the upper Hubbard band always belongs to the inverse photoemission or electron addition spectrum. Also, we have seen in the simplified derivation that the upper band mainly has double-occupancy character.

As electrons are removed from the system, however, the probability that an added electron is placed at an occupied site to create a double occupancy becomes smaller and consequently the weight of the upper band diminishes. This doping dependent intensity of what would be the conduction band in an ordinary semiconductor or insulator is one of the fingerprints of strong correlations and can be observed experimentally in cuprate superconductors – an example is shown in Fig. 6 [15]. It should be noted, however, that the Hubbard-I approximation considerably underestimates the decrease of the intensity of the upper Hubbard band with doping.

Figure 7 also shows the dependence of the Fermi surface volume $V_{\rm Fermi}$ on electron density n_e. More precisely, this is the fraction of the Brillouin zone where the lower Hubbard band is below E_F, i.e. 'occupied'. Also shown is $V_{\rm Fermi}$ for free electrons – where $V_{\rm Fermi} = n_e/2$. The Hubbard-I approximation gives $V_{\rm Fermi} \to 1$ – that means a completely filled band – as $n_e \to 1$, predicts $V_{\rm Fermi} = 1 - 2\delta = 2n_e - 1$ for n_e slightly less than 1 and approaches the free electron behavior for small n_e. This leads to a peculiar nonlinear dependence $V_{\rm Fermi}(n_e)$ which most probably is unphysical.

Let us now compare the Hubbard-I approximation to numerical simulations. As we saw in our simplified derivation, an important assumption of the Hubbard-I approximation is the 'disordered' ground state. This is best realized at high temperatures, more precisely at a temperature much higher than the characteristic energy of spin excitations, $J = 4t^2/U$. Figure 8 shows the result of a Quantum Monte-Carlo calculation of the spectral density for an 8×8 cluster at the

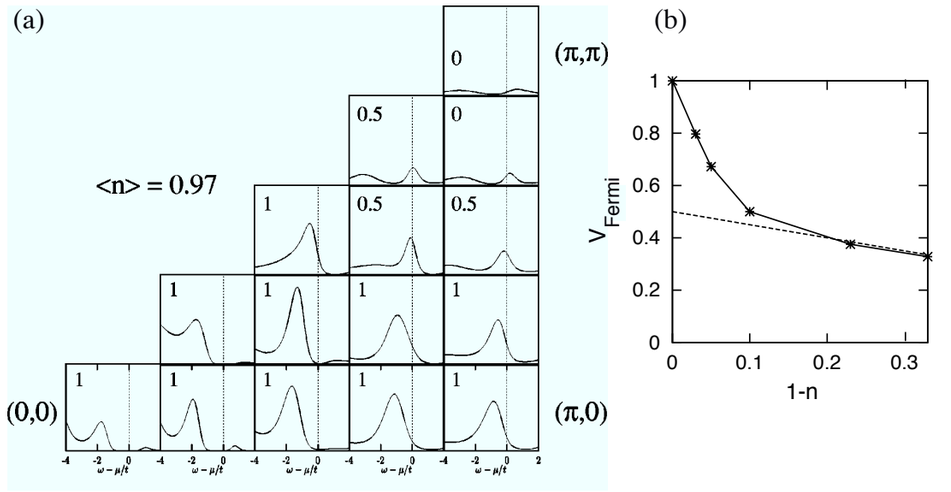


Fig. 8: *Right: Single particle spectral function $A(\mathbf{k},\omega)$ obtained by Quantum Monte-Carlo simulations on an 8×8 cluster at $k_BT = t$. Left: Fermi surface volume (34) deduced from $A(\mathbf{k},\omega)$ versus electron density. The dashed line corresponds to free electrons. Reprinted with permission from [16], Copyright 2000 by the American Physical society.*

rather high temperature $k_BT = t$. The 8×8 cluster has the allowed momenta $(\frac{n\pi}{4}, \frac{m\pi}{4})$ with integer m and n and Fig. 8 shows the part of the spectral density near the chemical potential μ for all allowed momenta in the irreducible wedge of the Brillouin zone for electron densities close to $n_e = 1$. Close to (π,π) a relatively well-defined peak passes through μ as (π,π) is approached and forms a relatively small hole pocket around (π,π) – similar to the prediction of the Hubbard-I approximation in Fig. 6 for $n_e = 0.9$. To study $V_{\rm Fermi}$ an 'occupation number' $n_{\mathbf{k}}$ of 1, 0.5 or 0 was assigned to each momentum $\mathbf{k}$, depending of whether the dispersive peak is below, more or less on, or above the chemical potential at $\mathbf{k}$. The fractional Fermi surface volume then is

$$V_{\rm Fermi} = \frac{1}{64}\sum_{\mathbf{k}} n_{\mathbf{k}}, \tag{34}$$

where 64 is the number of momenta in the 8×8 cluster. The obtained estimate for $V_{\rm Fermi}$ is also shown in Figure 8 as a function of electron density and indeed has a rough similarity to the result for the Hubbard-I approximation.

6 The Gutzwiller wave function

This is the second 'classic' approximation for the Hubbard model. It starts from the Fermi sea $|FS\rangle$ – i.e. the ground state for $U = 0$ – and reduces the number of double occupancies by acting with a suitable projecting operator. More precisely the Gutzwiller wave function reads [3]

$$|\Phi_G\rangle = \prod_{i=1}^{N}(1-\lambda\, n_{i,\uparrow}n_{i,\downarrow})\,|FS\rangle,$$

where λ is a variational parameter to be determined by minimizing the energy E_G. To see what this means, let us go back to the Hubbard dimer. There, the ground state for $U = 0$ is

$$`|FS\rangle' = c^\dagger_{+,\uparrow} c^\dagger_{+,\downarrow} |0\rangle = \frac{1}{2}\left[\left(c^\dagger_{1,\uparrow} c^\dagger_{2,\downarrow} + c^\dagger_{2,\uparrow} c^\dagger_{1,\downarrow}\right) + \left(c^\dagger_{1,\uparrow} c^\dagger_{1,\downarrow} + c^\dagger_{2,\uparrow} c^\dagger_{2,\downarrow}\right)\right] |0\rangle.$$

The first term in the square bracket does not contain any double occupancy, so that this term is annihilated by acting with $n_{i,\uparrow} n_{i,\downarrow}$ – this term therefore is simply reproduced by the Gutzwiller projector. On the other hand, each contribution in the second term has one double occupancy and thus gets a factor of $1 \cdot (1 - \lambda)$. The Gutzwiller wave function therefore is

$$|\Phi_G\rangle \;\; = \;\; \frac{1}{\sqrt{2}} |1_+\rangle + \frac{1-\lambda}{\sqrt{2}} |2_+\rangle,$$

with $|1_+\rangle$ and $|2_+\rangle$ as defined in (14). Therefore, if we choose (with u, v in (15))

$$1 - \lambda = \frac{v}{u} \qquad \Rightarrow \qquad \lambda = 1 - \frac{2t}{W + \frac{U}{2}},$$

the Gutzwiller wave function is – up to a normalization factor – identical to the exact ground state wave function of the dimer! From the above it also becomes apparent that the Gutzwiller projector reduces the weight of states containing double occupancies.

We proceed to the infinite lattice and first rewrite the Fermi sea as a superposition of real space configurations. Suppressing the spin index (and denoting permutations by σ) we have

$$\begin{aligned}
\prod_{j=1}^{M} c^\dagger_{\mathbf{k}_j} |0\rangle &= \frac{1}{\sqrt{N}^M} \sum_{i_1, i_2, i_3, \ldots i_M} \exp\left(i \sum_{j=1}^{M} \mathbf{k}_j \cdot \mathbf{R}_{i_j}\right) \prod_{j=1}^{M} c^\dagger_{i_j} |0\rangle \\
&= \frac{1}{\sqrt{N}^M} \sum_{i_1 > i_2 > i_3 \cdots > i_M} \sum_{\sigma} \exp\left(i \sum_{j=1}^{M} \mathbf{k}_j \cdot \mathbf{R}_{i_{\sigma(j)}}\right) \prod_{j=1}^{M} c^\dagger_{i_{\sigma(j)}} |0\rangle
\end{aligned}$$

In the second line we used the fact that instead of summing over all M-tuples of indices we may as well sum only over ordered M-tuples of indices and then sum over all $M!$ permutations σ of the M indices.

Next, in each of the products $\prod_{j=1}^{M} c^\dagger_{i_{\sigma(j)}}$ we permute the $c^\dagger_i$ operators back to the ordered sequence $c^\dagger_{i_1} c^\dagger_{i_2} \ldots c^\dagger_{i_M}$. The permutation which brings $\sigma(i) \to i$ obviously is σ^{-1} and since the Fermi sign of σ^{-1} is equal to that of σ we obtain

$$\begin{aligned}
&\frac{1}{\sqrt{N}^M} \sum_{i_1 > i_2 > i_3 \cdots > i_M} \sum_{\sigma} (-1)^\sigma \exp\left(i \sum_{j=1}^{M} \mathbf{k}_j \cdot \mathbf{R}_{i_{\sigma(j)}}\right) c^\dagger_{i_1} c^\dagger_{i_2} \cdots c^\dagger_{i_M} |0\rangle \\
= \;\; &\frac{1}{\sqrt{N}^M} \sum_{i_1 > i_2 > i_3 \cdots > i_M} D(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_M | i_1, i_2, \ldots i_M) \quad c^\dagger_{i_1} c^\dagger_{i_2} \cdots c^\dagger_{i_M} |0\rangle \,,
\end{aligned}$$

where the second line is the definition of the symbol $D(\mathbf{k}_j | i_j)$. From the above we see that the Fermi sea may be thought of as a superposition of real space configurations

$$c^\dagger_{i_1,\uparrow} c^\dagger_{i_2,\uparrow} c^\dagger_{i_3,\uparrow} \cdots c^\dagger_{i_{N_\uparrow},\uparrow} \, c^\dagger_{j_1,\downarrow} c^\dagger_{j_2,\downarrow} c^\dagger_{j_3,\downarrow} \cdots c^\dagger_{j_{N_\downarrow},\downarrow} |0\rangle$$

which are multiplied by two determinants D, one for each spin direction. Each of these real space configurations has a certain number N_d of doubly occupied sites and therefore gets an additional factor of $(1-\lambda)^{N_d} < 1$ in the Gutzwiller wave function so that states with a larger number of double occupancies have a smaller weight as compared to the original Fermi sea. Next, the Gutzwiller function can be decomposed into components with fixed N_d:

$$|\Phi_G\rangle = \sum_{N_d} |\Phi(N_d)\rangle$$

where $|\Phi(N_d)\rangle$ is the sum over all real-space configurations with N_d double occupancies, each multiplied by its proper prefactor. The total norm $\langle\Phi_G|\Phi_G\rangle$ can be rewritten as the sum over N_d terms of the form $W(N_d) = \langle\Phi(N_d)|\Phi(N_d)\rangle$ and we now consider which N_d gives the largest contribution in this sum. To compute norms, we need to evaluate expressions such as

$$\begin{aligned} D^*(\mathbf{k_j}|i_j)\ D(\mathbf{k_j}|i_j) = & \sum_{\sigma,\sigma'} (-1)^{\sigma}(-1)^{\sigma'} \exp\left(i \sum_{j=1}^{M} \mathbf{k}_j \cdot (\mathbf{R}_{i_{\sigma(j)}} - \mathbf{R}_{i_{\sigma'(j)}}) \right) \\ = & M! + \sum_{\sigma\neq\sigma'} (-1)^{\sigma}(-1)^{\sigma'} \exp\left(i \sum_{j=1}^{M} \mathbf{k}_j \cdot (\mathbf{R}_{i_{\sigma(j)}} - \mathbf{R}_{i_{\sigma'(j)}}) \right). \end{aligned} \tag{35}$$

where in the first term we have collected the $M!$ terms with $\sigma = \sigma'$. At this point, we make an important approximation: (35) still has to be summed over $i_1, i_2, i_3 \ldots i_M$. The terms for $\sigma \neq \sigma'$ thereby have a rapidly oscillating phase and a large degree of cancellation will occur in the summation. Accordingly we retain only the first term, i.e., we replace

$$D^*(\mathbf{k_j}|i_j)\ D(\mathbf{k_j}|i_j) \approx M!.$$

With this approximation the contribution of the states with N_d double occupancies becomes

$$W(N_d) = \frac{N_\uparrow!\ N_\downarrow!}{N^{N_\uparrow+N_\downarrow}}\ (1-\lambda)^{2N_d}\ C(N_\uparrow, N_\downarrow, N_d)\,,$$

where $C(N_\uparrow, N_\downarrow, N_d)$ is the number of ways in which $N_\uparrow$ electrons with spin $\uparrow$ and $N_\downarrow$ electrons with spin $\downarrow$ can be distributed over the N lattice sites such as to generate N_d double occupancies. This is a straightforward combinatorical problem with the result

$$C(N_\uparrow, N_\downarrow, N_d) = \frac{N!}{N_d!(N_\uparrow - N_d)!\ (N_\downarrow - N_d)!\ (N - N_\uparrow - N_\downarrow + N_d)!}.$$

Next, we take the logarithm of $W(N_d)$, use the Stirling formula $\log(N!) \approx N\log(N) - N$ and differentiate with respect to N_d. Introducing the densities $n_d = N_d/N$ etc. we obtain

$$\begin{aligned} \frac{d}{dN_d}\ \log\left(W(N_d)\right) &= \log\left((1-\lambda)^2\ \frac{(n_\uparrow - n_d)\ (n_\downarrow - n_d)}{n_d\ (1 - n_\uparrow - n_\downarrow + n_d)} \right), \\ \frac{d^2}{dN_d^2}\ \log\left(W(N_d)\right) &= -\frac{1}{N}\left(\frac{1}{n_d} + \frac{1}{n_\uparrow - n_d} + \frac{1}{n_\downarrow - n_d} + \frac{1}{1 - n_\uparrow - n_\downarrow + n_d} \right) = -\frac{c}{N}\,, \end{aligned}$$

where $c > 0$ in the last line is of order unity. The first of these equations gives us the n_d where the contribution to the norm, $W(N_d)$ is a maximum. For the general case the formula is somewhat involved, so we specialize to the case $n_\uparrow = n_\downarrow = \frac{1}{2}$ where

$$n_{d,0} \;=\; \frac{1-\lambda}{2(2-\lambda)}. \tag{36}$$

For the noninteracting case $\lambda \to 0$ this gives $n_{d,0} = 1/4$ as it has to be. From the second equation we find

$$\log\left(W(N_d)\right) = \log\left(W(N_{d,0})\right) - \frac{c}{2N}\,(N - N_{d,0})^2 + \ldots$$
$$W(N_d) = W(N_{d,0}) \exp\left(-\frac{c}{2N}(N_d - N_{d,0})^2\right) = W(N_{d,0}) \exp\left(-\frac{Nc}{2}(n_d - n_{d,0})^2\right),$$

which shows that as a function of n_d the weight $W(N_d)$ is a Gaussian of width $\propto N^{-1/2}$. This means, however, that in the thermodynamical limit only states with $n_d = n_{d,0}$ have an appreciable weight in the Gutzwiller wave function and variation of λ simply shifts this sharp peak of $W(N_d)$ to a different $n_{d,0}$ An immediate consequence is that the computation of the expectation value of the interaction Hamiltonian becomes trivial, namely $\langle H_U \rangle = N\,U\,n_{d,0}$.
The expectation value of the kinetic energy and is more involved. The above discussion showed that the Gutzwiller wave function is composed of real-space configurations for which the number of double occupancies is close to a certain value $N_{d,0}$ which is smaller than for the noninteracting Fermi sea. This means, however, that the expectation value of the kinetic energy is smaller as well. Namely using again the operators $\hat{d}$ and $\hat{c}$ we have

$$c^\dagger_{i,\sigma} c_{j,\sigma} = \hat{d}^\dagger_{i,\sigma} \hat{d}_{j,\sigma} + \hat{c}^\dagger_{i,\sigma} \hat{d}_{j,\sigma} + \hat{d}^\dagger_{i,\sigma} \hat{c}_{j,\sigma} + \hat{c}^\dagger_{i,\sigma} \hat{c}_{j,\sigma}\,.$$

If the number of double occupancies is decreased, the expectation value of the first term on the r.h.s. clearly must decrease. Second, since the number of electrons is constant, reducing the number of double occupancies necessarily results in a reduction of the number of empty sites by the same number so that the expectation value of the last term on the r.h.s. also must decrease.
The Gutzwiller approximation assumes, that these effects can be taken into account by reducing the expectation value of the kinetic energy of the uncorrelated Fermi sea by suitable renormalization factors η:

$$\frac{\langle \Phi_G | H_t | \Phi_G \rangle}{\langle \Phi_G | \Phi_G \rangle} = \sum_\sigma \eta_\sigma \langle FS, \sigma | H_t | FS, \sigma \rangle$$

where $|FS, \sigma\rangle$ is the Fermi sea for σ-electrons (if $N_\uparrow = N_\downarrow$ the two terms are of course identical). These renormalization factors η_σ thereby are evaluated for an 'auxiliary wave function' in which the determinants $D(\mathbf{k}_1, \mathbf{k}_2, \ldots, \mathbf{k}_M | i_1, i_2, \ldots i_M)$ are replaced by a constant (which would have to be $\sqrt{M!}$ if the auxiliary wave function is supposed to have the same norm as the Gutzwiller wave function) and where the Fermi sign is ignored in the calculation of all matrix elements of the hopping term (this is because the Fermi sign is supposed to be taken care of already by the filling of the uncorrelated Fermi sea according to the Pauli principle!). The

evaluation of the η by combinatorical considerations is discussed in a very transparent way by Ogawa, Kanda, and Matsubara [17]. Here we use an even simpler way of calculating η by introducing four 'book-keeping kets' for every site i: $|i,0\rangle$, $|i,\uparrow\rangle$, $|i,\downarrow\rangle$ and $|i,\uparrow\downarrow\rangle$. They represent in an obvious way the four possible configurations of the site i. Then we define

$$B_i = \frac{|i,0\rangle + \alpha_\uparrow |i,\uparrow\rangle + \alpha_\downarrow |i,\downarrow\rangle + \beta |i,\uparrow\downarrow\rangle}{\sqrt{1+\alpha_\uparrow^2+\alpha_\downarrow^2+\beta^2}} \qquad \text{and} \quad |\Psi\rangle = \prod_i B_i \, |0\rangle$$

with real α_σ and β. The state $|\Psi\rangle$ has norm 1 and if it were translated into a true state of electrons, the numbers of electrons and double occupancies would be

$$\begin{aligned} \langle N_\sigma \rangle &= N \, \frac{\alpha_\sigma^2+\beta^2}{1+\alpha_\uparrow^2+\alpha_\downarrow^2+\beta^2}, \\ \langle N_d \rangle &= N \, \frac{\beta^2}{1+\alpha_\uparrow^2+\alpha_\downarrow^2+\beta^2}. \end{aligned} \tag{37}$$

These equations can be reverted to give

$$\alpha_\sigma = \sqrt{\frac{n_\sigma - n_d}{1-n_\uparrow-n_\downarrow+n_d}} \qquad \text{and} \quad \beta = \sqrt{\frac{n_d}{1-n_\uparrow-n_\downarrow+n_d}}\,. \tag{38}$$

On the other hand $|\Psi\rangle$ does not correspond to a state with a fixed number of electrons so we introduce

$$|\Psi'\rangle = \mathcal{P}(N_\uparrow, N_\downarrow, N_d)\ |\Psi\rangle,$$

where the projection operator $\mathcal{P}$ projects onto the component of $|\Psi\rangle$ which has precisely $\langle N_\uparrow\rangle$ $\uparrow$-electrons etc. Next, the representation of the electron annihilation operator $c_{i,\sigma}$ is

$$\tilde{c}_{i,\sigma} \quad = \quad |i,0\rangle \, \langle i,\sigma| + |i,-\sigma\rangle \, \langle i,\uparrow\downarrow|.$$

Here a subtle detail should be noted: in the expression on the r.h.s. it is assumed that a double occupancy always is converted into the state $|i,-\sigma\rangle$ *with a positive sign*. This would not be the case for the true Fermion operator, where the sign would depend on the sequence of the two electron creation operators on the doubly occupied site. This is precisely the neglect of the Fermi sign that was mentioned above. Then, to estimate the reduction of the kinetic energy due to the reduction of the number of doubly occupied and empty sites we evaluate

$$r(\sigma, n_\uparrow, n_\downarrow, n_d) = \frac{\langle\Psi'|\tilde{c}_{i,\sigma}^\dagger \tilde{c}_{j,\sigma}|\Psi'\rangle}{\langle\Psi'|\Psi'\rangle}\,. \tag{39}$$

So far our auxiliary wave function has not brought about much simplification because the presence of the projection operator $\mathcal{P}$ makes the computation of r very tedious. It is straightforward to see, however, that if $|\langle\Psi\rangle$ is decomposed into components of fixed $N_\uparrow$, $N_\downarrow$ and N_d only those components with values of $N_\uparrow$, $N_\downarrow$ and N_d which deviate by at most $N^{-1/2}$ from the average values (37) have an appreciable weight. This means, however, that $\mathcal{P}$ simply can be dropped so

that we replace $|\Psi'\rangle \to |\Psi\rangle$ in (39). Then, since $|\Psi\rangle$ is normalized the denominator in (39) can be dropped. Moreover, since $|\Psi\rangle$ is a product state the expectation value of the two operators factorizes and since all sites are equivalent and the coefficients α_σ and β are real the expectation values of $\tilde{c}^\dagger_{i,\sigma}$ and $\tilde{c}_{j,\sigma}$ are identical so that

$$\begin{aligned} r(\sigma, n_\uparrow, n_\downarrow, n_d) = \langle\Psi|\tilde{c}^\dagger_{i,\sigma}|\Psi\rangle^2 &= \left(\frac{\alpha_\sigma + \alpha_{-\sigma}\beta}{1+\alpha_\uparrow^2+\alpha_\downarrow^2+\beta^2}\right)^2 \\ &= \left(\sqrt{n_\sigma - n_d}\sqrt{1-n_\uparrow-n_\downarrow+n_d} + \sqrt{n_d}\,\sqrt{n_{-\sigma}-n_d}\right)^2, \end{aligned}$$

where the second line has been obtained by inserting (38). In this way we have expressed $r(n_\sigma, n_d)$ in terms of n_d which in turn is given as a function of λ by (36). Lastly, we divide r by its value for $U \to 0$, where $n_d = n_\uparrow \cdot n_\downarrow$, so as to obtain the proper limiting value of $\eta = 1$ for $U = 0$, and finally get

$$\eta(\sigma, n_\uparrow, n_\downarrow, n_d) = \left(\frac{\sqrt{n_\sigma - n_d}\sqrt{1-n_\uparrow-n_\downarrow+n_d} + \sqrt{n_d}\,\sqrt{n_{-\sigma}-n_d}}{\sqrt{n_\sigma(1-n_\sigma)}}\right)^2. \tag{40}$$

In varying the energy it is actually easier to switch from λ to n_d as variational parameter. Specializing to the paramagnetic case $n_\uparrow = n_\downarrow$ the energy per site thus becomes

$$e_G = \eta(n_\sigma, n_d)\, t_0 + n_d\, U, \tag{41}$$

where $e_G = E_G/N$ and t_0 is the (kinetic) energy of the Fermi sea per site. Using (40) this is now readily minimized with respect to n_d.

The Gutzwiller wave function gives us, strictly speaking, only the ground state energies and some ground state expectation values, but not a band structure. However, we may consider states like

$$|\Phi_G(\mathbf{k})\rangle = \prod_i (1 - \lambda\, n_{i,\uparrow} n_{i,\downarrow}) c_{\mathbf{k},\uparrow}\, |FS\rangle,$$

i.e., a state with one hole in the Fermi sea (it is understood that $\mathbf{k}$ is an occupied momentum). The Fermi sea with a hole has energy $E_{FS} - \varepsilon_{\mathbf{k}}$. It thus might seem plausible that the energy of $|\Phi_G(\mathbf{k})\rangle$ is $E_G - \tilde{\varepsilon}_{\mathbf{k}}$ i.e., the energy of the Gutzwiller wave function minus the 'quasiparticle energy'. Performing the variational procedure for $|\Phi_G(\mathbf{k})\rangle$ amounts to replacing $e_G \to e_G - \frac{1}{N}\tilde{\varepsilon}_{\mathbf{k}}$ $t_0 \to t_0 - \frac{1}{N}\varepsilon_{\mathbf{k}}$, $n_\uparrow \to n_\uparrow - \frac{1}{N}$, and $n_d \to n_d + \frac{1}{N}\delta n_d$ where δn_d is the as yet unknown shift of n_d. Inserting into (41) and expanding we find

$$e - \frac{1}{N}\,\tilde{\varepsilon}_{\mathbf{k}} = \left(\eta - \frac{1}{N}\frac{\partial\eta}{\partial n_\uparrow} + \frac{1}{N}\frac{\partial\eta}{\partial n_d}\delta n_d\right)\left(t_0 - \frac{1}{N}\,\varepsilon_{\mathbf{k}}\right) + n_d\, U + \frac{1}{N}\delta n_d\, U.$$

The terms of zeroth order in $1/N$ cancel due to (41) and collecting the first order terms gives

$$\tilde{\varepsilon}_{\mathbf{k}} \;=\; \eta\varepsilon_{\mathbf{k}} + t_0\,\frac{\partial\eta}{\partial n_\uparrow} - \left(\frac{\partial\eta}{\partial n_d}\,t_0 + U\right)\delta n_d.$$

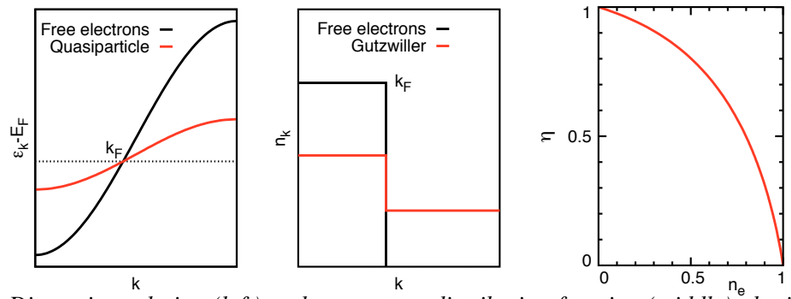


Fig. 9: *Dispersion relation (left) and momentum distribution function (middle) obtained from the Gutzwiller wave function compared to the free electron case. The right part shows the dependence $\eta(n_e)$ for the half-filled two-dimensional Hubbard model with $U/t = 16$.*

The last term on the r.h.s. vanishes because the expression in the bracket is $\frac{de_G}{dn_d}$. The second term on the r.h.s. gives the change of the kinetic energy of all other electrons due to removal of the single electron. This is a $\mathbf{k}$-independent shift which can be absorbed into a shift of E_F. The quasiparticle dispersion $\tilde{\varepsilon}_{\mathbf{k}}$ therefore follows the original dispersion, but renormalized by the factor $\eta < 1$. This is an effect known as 'correlation narrowing'.

Next we consider the ground state momentum distribution function, i.e., the ground state expectation value $n_{\mathbf{k}} = 2\langle c^\dagger_{\mathbf{k},\uparrow} c_{\mathbf{k},\uparrow}\rangle$. This can be obtained from the ground state energy by means of the Hellmann-Feynman theorem: $n_{\mathbf{k}} = \frac{\partial E_G}{\partial \varepsilon_{\mathbf{k}}}$. More generally, it is the functional derivative of the ground state energy with respect to $\varepsilon_{\mathbf{k}}$, that means under a change $t_{ij} \to t_{ij} + \delta t_{ij}$ so that $\varepsilon_{\mathbf{k}} \to \varepsilon_{\mathbf{k}} + \delta\varepsilon_{\mathbf{k}}$ the change of the ground state energy is

$$e_G \to e_G + 2\sum_{\mathbf{k}} n_{\mathbf{k}}\, \delta\varepsilon_{\mathbf{k}}\,.$$

From (41) we obtain the variation of e_G as

$$\delta e_G = 2\eta \sum_{\mathbf{k}} n^{(0)}_{\mathbf{k}}\, \delta\varepsilon_{\mathbf{k}} + \delta n_d \left(\frac{\partial \eta}{\partial n_d}\, t_0 + U\right).$$

where $n^{(0)}_{\mathbf{k}} = \Theta(E_F - \varepsilon_{\mathbf{k}})$ is the momentum distribution of the Fermi sea. Again, the second term on the r.h.s. vanishes due to the extremum condition for n_d so that $n_{\mathbf{k}} = \eta\ n^{(0)}_{\mathbf{k}}$. This cannot be entirely correct, however, because we have the sum-rule $2\sum_{\mathbf{k}} n_{\mathbf{k}} = N_e$ and since this is fulfilled by $n^{(0)}_{\mathbf{k}}$ and $\eta < 1$ it cannot be fulfilled for $n_{\mathbf{k}}$. The solution is that the 'missing $n_{\mathbf{k}}$' takes the form of a $\mathbf{k}$-independent additive constant which then has to be $(1-\eta)\frac{n_e}{2}$. In fact, for any $\varepsilon_{\mathbf{k}}$ which can be represented by hopping integrals $t_{i,j}$ one has $\sum_{\mathbf{k}} \varepsilon_{\mathbf{k}} = 0$ so that such a $\mathbf{k}$-independent additive constant would not contribute to the variation of e_G. The momentum distribution obtained by the Gutzwiller approximation thus has a step of magnitude η at the position of the original Fermi surface. Let us now consider in some more detail the case $n_\sigma = \frac{1}{2}$ where the Mott-insulator should be realized for large U/t. We find from (40)

$$\eta(n_d) = 16 n_d \left(\frac{1}{2} - n_d\right).$$

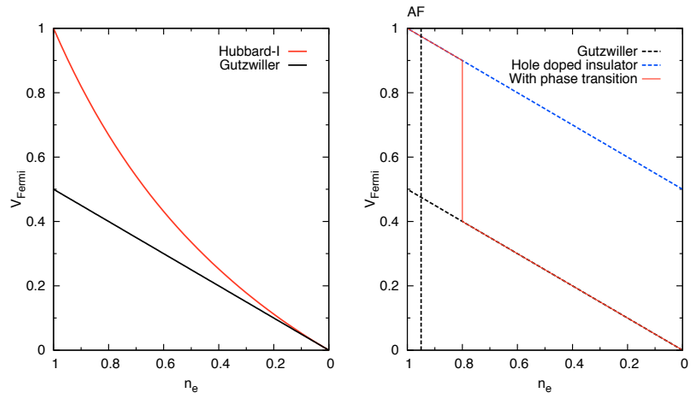


Fig. 10: *Left: Comparison of the Fermi surface volume as obtained from the Hubbard-I approximation and the Gutzwiller wave function. Right: a possible compromise with a phase transition between two phases with different* $V_{\rm Fermi}(n_e)$.

Minimizing (41) this gives

$$n_d = \frac{1}{4} - \frac{U}{32|t_0|}\,,$$

whereby we have taken into account that $t_0 < 0$ for a half-filled band. Starting from the noninteracting value $\frac{1}{4}$, n_d decreases linearly with U and reaches zero at the critical value $U_c = 8|t_0|$. For $n_d = 0$ we have $\eta = 0$, so that the bandwidth of the quasiparticles becomes zero, i.e., the band mass diverges, and the step in the momentum distribution vanishes as well. This is commonly interpreted as a metal-to-insulator transition as a function of increasing U, the so-called Brinkman-Rice transition [18]. Brinkman and Rice also could show that the magnetic susceptibility diverges at the transition as one would expect it for a diverging effective mass.
Let us now consider the two-dimensional model with nearest neighbor hopping $-t$. Then, $t_0 = -1.621\ t$ so that the critical $U_c = 12.969\ t$. Figure 9 then shows the dependence of η on n_e for $U/t = 16$ i.e., for $U > U_c$. As $n_e \to 1$ the renormalization factor $\eta \to 0$ so that both, the bandwidth and the step in $n_{\mathbf{k}}$ vanish for the half-filled band. The Hubbard-I approximation and the Gutzwiller wave function thus give completely different predictions about what happens when the half-filled band case is approached by increasing the electron density for constant U/t: whereas the Hubbard-I approximation predicts the lower Hubbard band with (almost) constant bandwidth and a hole-pocket-like Fermi surface with a volume $\propto (1 - n_e)$ so that the Fermi surface vanishes at $n_e \to 1$, the Gutzwiller wave function predicts a Fermi surface with a volume equal to that obtained for free electrons, but with a vanishing bandwidth and spectral weight as $n_e \to 1$.

7 Summary and discussion

Since a generally accepted theory of the lightly doped 2-dimensional Hubbard model does not seem to exist so far so that also the cuprate superconductors are not really understood as yet, maybe the best one can do at present is to outline the problems that would have to be resolved.

The first one of these is the Fermi surface close to half-filling. As we have seen the Hubbard-I approximation and the Gutzwiller wave function predict completely different behavior close to half-filling: a hole-like Fermi surface with a volume $\propto n_h = 1 - n_e$ in the lower Hubbard-band whose volume tends to zero as $n_e \to 1$ versus a free-electron-like Fermi surface with volume $\frac{n_e}{2}$ formed by a band whose mass diverges as $n_e \to 1$. A possible compromise between the two approximations could be as shown in Figure 10: near $n_e = 1$ there are hole pockets with a volume that is strictly proportional to the hole number $n_h = 1 - n_e$ – i.e. the doped Hubbard-band – and then at some critical density a phase transition occurs to a phase where the Fermi surface volume is $n_e/2$. This might be the true behavior which Hubbard-I approximation 'tries to reproduce'. Viewed this way, the 'pseudogap phase' of cuprate superconductors actually should be identified with the hole-pocket phase and the quantum critical point which is surrounded by the superconducting dome corresponds to the transition to the free-electron-like Fermi surface. A theory which is supposed to describe this, first of all, must reproduce the two Hubbard bands – otherwise the hole-doped lower Hubbard band cannot be captured. Next, the two different phases would have to be reproduced which is a considerable problem because there is no obvious order parameter for the transition between a paramagnetic small Fermi surface and a paramagnetic large Fermi surface.

References

[1] J. Hubbard, Proc. Roy. Soc. London A **276** 238 (1963)

[2] A. Kanamori, Progr. Theor. Phys. **30**, 275 (1963)

[3] M.C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963)

[4] F.C. Zhang and T.M. Rice, Phys. Rev. B **37**, 3759 (1988)

[5] A.A. Abrikosov, L.P. Gorkov and I.E. Dzyaloshinksi:
Methods of Quantum Field Theory in Statistical Physics (Prentice-Hall, New Jersey 1964)

[6] A.L. Fetter and J.D. Walecka: *Quantum Theory of Many-Particle Systems*
(McGraw-Hill, San Francisco, 1971)

[7] J.W. Negele and H. Orland: *Quantum Many-Particle Systems*
(Addison-Wesley, Redwood, 1988)

[8] G. Baym and N.D. Mermin, J. Math. Phys. **2**,232 (1961)

[9] J.M. Luttinger, Phys. Rev. **121**, 942 (1961)

[10] R. Eder, K. Seki, and Y. Ohta Phys. Rev. B **83**, 205137 (2011)

[11] L.P. Gorkov, Soviet Phys. JETP **7** 505 (1958)

[12] L.D. Landau and E.M. Lifshitz: *Lehrbuch der Theoretischen Physik*
(Akademie Verlag Berlin, 1988)

[13] A. Damascelli, Z. Hussain, and Z.-X. Shen, Rev. Mod. Phys. **75** 473 (2003)

[14] J.E. Hirsch Phys. Rev. B **31**, 4403 (1985); H.Q. Lin and J.E. Hirsch,
Phys. Rev. B **35**, 3359 (1987)

[15] C.T. Chen, F. Sette, Y. Ma, M.S. Hybertsen, E.B. Stechel, W.M.C. Foulkes, M. Schulter, S-W. Cheong, A.S. Cooper, L.W. Rupp, B. Batlogg, Y.L. Soo, Z.H. Ming, A. Krol, and Y.H. Kao, Phys. Rev. Lett. 66, **104** (1991)

[16] C. Gröber, R. Eder, and W. Hanke, Phys. Rev. B **62** 4336 (2000)

[17] T. Ogawa, K. Kanda, and T. Matsubara, Progr. Theor. Phys. **53**, 614 (1975)

[18] W.F. Brinkman and T.M. Rice, Phys. Rev. B **2**, 4302 (1970)

7 The Finite Temperature Lanczos Method and its Applications

Peter Prelovšek
J. Stefan Institute
Jamova 39, SI-1000 Ljubljana, Slovenia

Contents

1 Introduction **2**

2 Exact diagonalization and Lanczos method **3**
2.1 Models, geometries, and system sizes 3
2.2 Lanczos diagonalization technique 5

3 Ground state properties and dynamics **6**

4 Static properties and dynamics at $T > 0$ **8**
4.1 Finite-temperature Lanczos method: Static quantities 8
4.2 Finite-temperature Lanczos method: Dynamical response 11
4.3 Finite temperature Lanczos method: Implementation 12
4.4 Low-temperature Lanczos method 14
4.5 Microcanonical Lanczos method 16
4.6 Statical and dynamical quantities at $T > 0$: Applications 17

5 Real-time dynamics using the Lanczos method **20**

6 Discussion **21**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

Models of strongly correlated systems have been one of the most intensively studied theoretical subjects in the last two decades, stimulated at first by the discovery of compounds superconducting at high-temperatures and ever since by the emergence of various novel materials and phenomena which could be traced back to strongly correlated electrons in these systems. Recently, cold atoms in optical lattices offer a different realization of strongly correlated quantum entities, whereby these systems can be even tuned closer to theoretical models.

One of the most straightforward methods to numerically deal with the lattice (discrete) models of correlated particles, which are inherently many-body (MB) quantum systems, is exact diagonalization (ED) of small-size systems. In view of the absence of well-controlled analytical methods, the ED method has been employed intensively to obtain results for static and dynamical properties of various models with different aims: a) to search and confirm novel phenomena specific to strongly correlated systems, b) to test theoretical ideas and analytical results, c) to get reference results for more advanced numerical techniques.

MB quantum lattice models of interacting particles are characterized by the dimension of the Hilbert space given by the number of basis states $N_{st} \propto K^N$ that is in turn exponentially increasing with the lattice size N, where K is the number of local quantum states. It is therefore clear that ED methods can treat fully only systems with limited N_{st}, i.e., both K and N must be quite modest.

Among the ED approaches the full ED within the Hilbert space of the model Hamiltonian, yielding all eigenenergies and eigenfunctions, is the simplest to understand, most transparent, and easy to implement. In principle it allows the evaluation of any ground state (g.s.) property as well as finite temperature $T > 0$ static or dynamic quantities, at the expense of a very restricted N_{st}. In spite of that, it represents a very instructive approach and also remains essentially the only practical method when all exact levels are needed, e.g., for studies of level statistics.

Lanczos-based ED methods have already a long history of applications since Cornelius Lanczos [1] proposed the diagonalization of sparse matrices using the iterative procedure, allowing for much bigger Hilbert spaces N_{st} relative to full ED. The Lanczos diagonalization technique is at present a part of standard numerical linear algebra procedures [2, 3] and as such in solid state physics mainly used to obtain the g.s. energy and wavefunction and the corresponding expectation values. The approach has been quite early-on extended to calculation of the dynamical response functions within the g.s. [4]. The method has been in the last 20 years extensively used in connection with models related to high-T_c materials, for which we can refer to an earlier overview [5].

Here we focus on recent developments of ED-based and Lanczos-based methods. The basics of the Lanczos method are presented in Sec. 2 and its application for g.s. properties in Sec. 3. One of the already established generalizations is the finite-temperature Lanczos method (FTLM) [6–8], reviewed in Sec. 4, which allows for the evaluation of $T > 0$ static and dynamic properties within simplest models. Several extensions and modifications of the latter have been introduced more recently, in particular the low-temperature Lanczos method (LTLM) [9] and

the microcanonical Lanczos method (MCLM) [10], particularly applicable within the high-T regime. Recently, there is also quite an intensive activity on studies of real-time evolution of correlated systems, both under equilibrium and non-equilibrium conditions that can be simulated using the ED and Lanczos-based methods, as discussed in Sec. 5.

2 Exact diagonalization and Lanczos method

2.1 Models, geometries, and system sizes

ED-based methods are mostly restricted to simple models with only few local quantum states K per lattice site in order to reach reasonable system sizes N. Consequently, there are only few classes of MB models that so far exhaust the majority of ED and Lanczos-method studies, clearly also motivated and influenced by the challenging physics and relevance to novel materials and related experiments.

To get some feeling for the available sizes reachable within ED-based approaches, it should be kept in mind that in full ED routines the CPU time scales with the number of operations $Op \propto N_{st}^3$, while the memory requirement is related to the storage of the whole Hamiltonian matrix and all eigenvectors, i.e., $Mem \propto N_{st}^2$. This limits, at the present stage of computer facilities, the full ED method to $N_{st} < 2.\ 10^4$ MB states. On the other hand, using Lanczos-based iterative methods for the diagonalization of sparse matrices (Hamiltonians), CPU and memory requirements scale as $Op, Mem \propto N_{st}$, at least in their basic application, to calculate the g.s. and its wavefunction. In present-day applications this allows the consideration of much larger basis sets, i.e., $N_{st} < 10^9$. Still, lattice sizes N reached using the Lanczos technique remain rather modest, compared to some other numerical approaches such as DMRG and quantum-Monte-Carlo QMC methods, if the full Hilbert basis space relevant for the model is used.

The simplest nontrivial class of MB lattice models are spin models, the prototype being the anisotropic Heisenberg model for coupled $S = 1/2$ spins,

$$H = \sum_{\langle ij\rangle\alpha} J_{ij}^{\alpha\alpha}\, S_i^\alpha S_j^\alpha, \tag{1}$$

where the sum $\langle ij\rangle$ runs over pairs of lattice sites with an arbitrary interaction $J_{ij}^{\alpha\alpha}$ (being in principle anisotropic) and S_i^α are the components of the local $S = 1/2$ operator. The model has just $K = 2$ quantum states per lattice site and therefore allows for biggest possible N in the ED-based approaches, where $N_{st} \propto 2^N$ basis states. To reduce N_{st} as many symmetries and good quantum numbers as practically possible are used to decompose the Hamiltonian into separate blocks. Evident choices are sectors with the (z-component of) total spin S_{tot}^z and the wavevector $\mathbf{q}$ for systems with periodic boundary conditions, but also rotational symmetries of particular lattices have been used. In this way system sizes up to $N \sim 36$ (for the largest and most interesting sector $S_{tot}^z = 0$) have been reached so far using the Lanczos technique without any basis reduction.

On the basis of this simple model one can already discuss the feasibility of Lanczos-based methods with respect to other numerical quantum MB methods. For the g.s. of 1D spin systems

more powerful methods allowing for much bigger systems are DMRG and related approaches. For unfrustrated models in D > 1 QMC methods are superior for the evaluation of static quantities at any T. Still, Lanczos-based methods become competitive or at least are not superseded for frustrated spin models (where QMC can run into the minus-sign problem) or for dynamical properties at $T > 0$.

Next in complexity and very intensively studied is the t-J model, representing strongly correlated itinerant electrons with an antiferromagnetic (AFM) interaction between their spins

$$H = -\sum_{\langle ij\rangle s} \left(t_{ij}\, \tilde{c}^{\dagger}_{js} \tilde{c}_{is} + \text{H.c.} \right) + J \sum_{\langle ij\rangle} \mathbf{S}_i \cdot \mathbf{S}_j \,, \tag{2}$$

where due to the strong on-site repulsion doubly occupied sites are forbidden and one is dealing with projected fermion operators $\tilde{c}_{is} = c_{is}(1 - n_{i,-s})$. The model can be considered as a good microscopic model for superconducting cuprates, which are doped Mott insulators, and has therefore been one of the most studied models using the Lanczos method [5]. For a theoretical and experimental overview of Mott insulators and metal-insulator transitions see Ref. [11]. It has $K = 3$ quantum states per lattice site and, besides S^z_{tot} and $\mathbf{q}$, also the number of electrons N_e (or more appropriately the number of holes $N_h = N - N_e$) are the simplest quantum numbers to implement. Since the model reveals an interesting physics in D > 1, the effort was in connection with high-T_c cuprates mostly concentrated on the 2D square lattice. Here the alternative numerical methods have more drawbacks (e.g., the minus sign problem in QMC methods due to the itinerant character of the fermions) so that Lanczos-based methods are still competitive, in particular for getting information on $T > 0$ dynamics and transport. The largest systems considered with the Lanczos method so far are 2D square lattices with $N = 32$ sites and $N_h = 4$ holes [12].

Clearly, one of the most investigated problems within the MB community is the standard single-band Hubbard model, which has $K = 4$ states per lattice site. Due to the complexity $N_{st} \propto 4^N$ the application of ED and Lanczos-based method is already quite restricted reaching so far $N = 20$ sites [13] requiring already $N_{st} \sim 10^9$ basis states. The model is also the subject of numerous studies using more powerful QMC method and various cluster dynamical-mean-field-theory (DMFT) methods for much larger lattices so Lanczos-based approaches have here more specific goals.

Since reachable lattices sizes for the above mentioned models are rather small it is important to properly choose their geometries. This is not a problem for 1D models, but becomes already essential for 2D lattices, analyzed in connection with novel materials, in particular high-T_c cuprates and related materials. In order to keep periodic boundary conditions for 2D square lattices the choice of Pythagorean lattices with $N = \lambda_x^2 + \lambda_y^2$ with λ_x, λ_y [14] has significantly extended available sizes. Some frequently used ones are presented in Fig. 1. Taking into account only even N, such lattices include $N =$ 8, 10, 16, 18, 20, 26, 32, and 36 sites. While the unit cells of such lattices are squares, it has been observed that they are not always optimal with respect to the number of next-nearest and further nearest neighbors. It has been claimed and partly tested that better result are obtained with slightly deformed lattices (still with periodic boundary conditions) which at the same time offer an even larger choice of sizes [15].

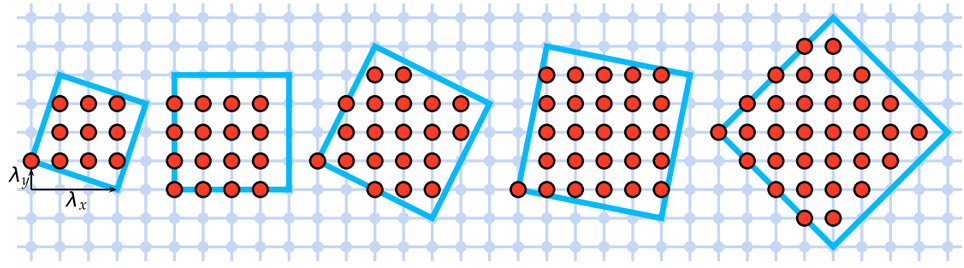


Fig. 1: *Some tilted clusters used in 2D square-lattice studies*

2.2 Lanczos diagonalization technique

The Lanczos technique is a general procedure to transform and reduce a symmetric $N_{st} \times N_{st}$ matrix A to a symmetric $M \times M$ tridiagonal matrix T_M. From the chosen initial N_{st}-dimensional vector $\mathbf{v_1}$ one generates an orthogonal basis of $\{\mathbf{v_1}, \cdots, \mathbf{v_M}\}$ vectors which span the Krylov space $\{\mathbf{v_1}, \mathbf{A v_1}, \cdots, \mathbf{A^{M-1} v_1}\}$ [1–3, 16].
In usual applications for the quantum MB system defined with Hamiltonian H the Lanczos algorithm starts with a normalized vector $|\phi_0\rangle$, chosen as a random vector in the relevant Hilbert space with N_{st} basis states. The procedure generates orthogonal Lanczos vectors $L_M = \{|\phi_m\rangle \mid m = 0 \ldots M\}$ spanning the Krylov space $\{|\phi_0\rangle, H|\phi_0\rangle, \cdots, H^M|\phi_0\rangle\}$. Steps are as follows: H is applied to $|\phi_0\rangle$ and the resulting vector is split in components parallel to $|\phi_0\rangle$, and normalized $|\phi_1\rangle$ orthogonal to it, respectively,

$$H|\phi_0\rangle = a_0|\phi_0\rangle + b_1|\phi_1\rangle. \tag{3}$$

Since H is Hermitian, $a_0 = \langle\phi_0|H|\phi_0\rangle$ is real, while the phase of $|\phi_1\rangle$ can be chosen so that b_1 is also real. In the next step H is applied to $|\phi_1\rangle$,

$$H|\phi_1\rangle = b_1'|\phi_0\rangle + a_1|\phi_1\rangle + b_2|\phi_2\rangle, \tag{4}$$

where $|\phi_2\rangle$ is orthogonal to $|\phi_0\rangle$ and $|\phi_1\rangle$. It follows also $b_1' = \langle\phi_0|H|\phi_1\rangle = b_1$. Proceeding with the iteration one gets in i steps

$$H|\phi_i\rangle = b_i|\phi_{i-1}\rangle + a_i|\phi_i\rangle + b_{i+1}|\phi_{i+1}\rangle, \qquad 1 \le i \le M, \tag{5}$$

where in Eq. (5) by construction there are no terms involving $|\phi_{i-2}\rangle$ etc. By stopping the iteration at $i = M$ and setting $b_{M+1} = 0$, the Hamiltonian can be represented in the basis of orthogonal Lanczos functions $|\phi_i\rangle$ as the tridiagonal matrix H_M with diagonal elements a_i, $i = 0 \ldots M$, and off-diagonal ones b_i, $i = 1 \ldots M$. Such a matrix is easily diagonalized using standard numerical routines to obtain approximate eigenvalues ε_j and corresponding orthonormal eigenvectors $|\psi_j\rangle$,

$$|\psi_j\rangle = \sum_{i=0}^{M} v_{ji}|\phi_i\rangle, \qquad j = 0 \ldots M. \tag{6}$$

It is important to realize that $|\psi_j\rangle$ are (in general) not exact eigenfunctions of H, but show a remainder. On the other hand, it is evident from the diagonalization of H_M that matrix elements

$$\langle\psi_i|H|\psi_j\rangle = \varepsilon_j\delta_{ij}, \qquad i,j = 0\ldots M, \tag{7}$$

are diagonal independently of L_M (but provided $i,j \leq M$), although the values ε_j can be only approximate.

If in the equation (5) $b_{M+1} = 0$, we have found an $(M+1)$-dimensional eigenspace where H_M is already an exact representation of H. This inevitably happens when $M = N_{st} - 1$, but for $M < N_{st} - 1$ it can only occur if the starting vector is orthogonal to some invariant subspace of H which we avoid by choosing the input vector $|\phi_0\rangle$ as a random one.

It should be recognized that the Lanczos approach is effective only for sparse Hamiltonians, characterized by the connectivity of each basis state with $K_n \ll N_{st}$ basis states. All prototype discrete tight-binding models discussed in Sec. 2.1 are indeed of such a type in the local MB basis. Estimating the computation requirements, the number of operations Op needed to perform M Lanczos iterations scales as $Op \propto K_n M N_{st}$. The main restriction is still in memory requirements due to the large N_{st}. A straightforward application of Eq. (5) would require the fast storage of all $|\phi_i\rangle$, $i = 0\ldots M$, i.e., also the memory capacity $Mem \propto MN_{st}$. However, for the evaluation of the eigenvalues alone during the iteration, Eq. (5), only three $|\phi_i\rangle$ are successively required, so this leads to $Mem \propto 3N_{st}$. If the Hamiltonian matrix is not evaluated on the fly, then also $Mem \propto K_n N_{st}$ for the nonzero Hamilton matrix elements is needed.

The Lanczos diagonalization is in essence an iterative power method which is known to converge fast for the extreme lower and upper eigenvalues [2, 3]. In physical application most relevant is the search for the g.s. energy E_0 and the corresponding wavefunction $|\Psi_0\rangle$. Typically, $M > 50$ are enough to reach very high accuracy for both. It is evident that for such modest $M \ll N_{st}$ one cannot expect any reliable results for eigenstates beyond the few at the bottom and the top of the spectrum. On the other hand, the Lanczos procedure is subject to roundoff errors, introduced by the finite-precision arithmetics which usually only becomes severe at larger $M > 100$ after the convergence of extreme eigenvalues, and is seen as the loss of orthogonality of the vectors $|\phi_i\rangle$. It can be remedied by successive reorthogonalization [2,3,16] of new states $|\phi_i'\rangle$, plagued with errors, with respect to previous ones. However this procedure requires $Op \sim M^2 N_{st}$ operations, and can become computationally more demanding than the Lanczos iterations themselves. This effect also prevents one from using the Lanczos method, e.g., to efficiently tridiagonalize large dense matrices [3].

3 Ground state properties and dynamics

After $|\Psi_0\rangle$ is obtained, the static properties of the g.s. can be evaluated in principle for any operator A as

$$\bar{A}_0 = \langle\Psi_0|A|\Psi_0\rangle. \tag{8}$$

Clearly, the procedure (8) is effective for large a basis only if the operator A is sparse in the same basis, as is the case for most operators of interest.

It is an important advantage of the Lanczos procedure that dynamical g.s. functions can easily be calculated [4]. Let us consider the dynamical (autocorrelation) response function

$$C(\omega) = \langle \Psi_0 | A^\dagger \frac{1}{\omega^+ + E_0 - H} A | \Psi_0 \rangle \tag{9}$$

for the observable given by the operator A, where $\omega^+ = \omega + i\delta$ with $\delta > 0$. To calculate $C(\omega)$ one has to run a second Lanczos procedure with a new initial function $|\tilde{\phi}_0\rangle$,

$$|\tilde{\phi}_0\rangle = \frac{1}{\alpha} A|\Psi_0\rangle, \qquad \alpha = \sqrt{\langle \Psi_0 | A^\dagger A | \Psi_0 \rangle}. \tag{10}$$

Starting with $|\tilde{\phi}_0\rangle$ one generates another Lanczos subspace $\tilde{L}_{\tilde{M}} = \{|\tilde{\phi}_j\rangle,\ j = 0, \tilde{M}\}$ with (approximate) eigenvectors $|\tilde{\psi}_j\rangle$ and eigenenergies $\tilde{\varepsilon}_j$. The matrix for H in the new basis is again a tridiagonal one with $\tilde{a}_j$ and $\tilde{b}_j$ elements, respectively. Terminating the Lanczos procedure at a given $\tilde{M}$, one can evaluate Eq. (9) as a resolvent of the $H_{\tilde{M}}$ matrix expressed in the continued-fraction form [17, 4, 5],

$$C(\omega) = \cfrac{\alpha^2}{\omega^+ + E_0 - \tilde{a}_0 - \cfrac{\tilde{b}_1^2}{\omega^+ + E_0 - \tilde{a}_1 - \cfrac{\tilde{b}_2^2}{\omega^+ + E_0 - \tilde{a}_2 - \ldots}}}, \tag{11}$$

terminating with $\tilde{b}_{\tilde{M}+1} = 0$, although other termination functions can also be employed and can be well justified.

We note that frequency moments of the spectral function

$$\mu_l = -\frac{1}{\pi} \int_{-\infty}^{\infty} \omega^l \, \mathrm{Im}\, C(\omega)\, d\omega = \langle \Psi_0 | A^\dagger (H - E_0)^l A | \Psi_0 \rangle = \alpha^2 \langle \tilde{\phi}_0 | (H - E_0)^l | \tilde{\phi}_0 \rangle \tag{12}$$

are exact for given $|\Psi_0\rangle$ provided $l \leq \tilde{M}$, since the operator H^l, $l < \tilde{M}$, is exactly reproduced within the Lanczos (or corresponding Krylov) space $\tilde{L}_{\tilde{M}}$.

Finally, $C(\omega)$ (11) can be presented as a sum of $j = 0$, $\tilde{M}$ poles at $\omega = \tilde{\varepsilon}_j - E_0$ with corresponding weights w_j. As a practical matter we note that in analogy to Eq. (6)

$$w_j = |\langle \tilde{\psi}_j | A | \Psi_0 \rangle|^2 = \alpha^2 |\langle \tilde{\psi}_j | \tilde{\phi}_0 \rangle|^2 = \alpha^2 \tilde{v}_{j0}^2 \,, \tag{13}$$

hence no matrix elements need to be evaluated within this approach. In contrast to the autocorrelation function (11), the procedure allows also the treatment of general correlation functions $C_{AB}(\omega)$, with $B \neq A^\dagger$. In this case matrix elements $\langle \Psi_0 | B | \tilde{\psi}_j \rangle$ have to be evaluated explicitly. It should be also mentioned that at least the lowest poles of $C(\omega)$, Eq. (11), should coincide with eigenenergies $\omega = E_i - E_0$ if $|\tilde{\phi}_0\rangle$ is not orthogonal to $|\Psi_0\rangle$. However, using $\tilde{M} > 50$, spurious poles can emerge (if no reorthogonalization is used) which, however, carry no weight as is evident from exact moments (12).

In this chapter we do not intend to present an overview of applications of the full ED and Lanczos-type studies of g.s. static and dynamical properties of correlated systems. There have been numerous such investigations even before the high-T_c era, intensified strongly with studies of prototype models relevant for high-T_c cuprates [5] and other novel materials with correlated electrons. Although a variety of models has been investigated they are still quite restricted in the number of local degrees and sizes.

4 Static properties and dynamics at $T > 0$

Before describing the finite temperature Lanczos method (FTLM) we should note that the Lanczos basis is a very useful and natural basis for evaluating matrix elements of the type

$$W_{kl} = \langle n|H^k B H^l A|n\rangle, \tag{14}$$

where $|n\rangle$ is an arbitrary normalized vector, and A, B are general operators. One can calculate this expression exactly by performing two Lanczos procedures with $M = \max(k, l)$ steps. The first one, starting with the vector $|\phi_0\rangle = |n\rangle$, produces the Lanczos basis L_M along with approximate eigenstates $|\psi_j\rangle$ and ε_j. The second Lanczos procedure is started with the normalized vector $|\tilde{\phi}_0\rangle \propto A|\phi_0\rangle = A|n\rangle$, Eq. (10), and generates $\tilde{L}_M$ with corresponding $|\tilde{\psi}_j\rangle$ and $\tilde{\varepsilon}_j$. We can now define projectors onto limited subspaces

$$P_M = \sum_{i=0}^{M} |\psi_i\rangle\langle\psi_i|, \quad \tilde{P}_M = \sum_{i=0}^{M} |\tilde{\psi}_i\rangle\langle\tilde{\psi}_i|. \tag{15}$$

Provided that $(l, k) < M$ projectors P_M and $\tilde{P}_M$ span the whole relevant basis for the operators H^k and H^l, respectively, so that one can rewrite W_{kl} in Eq. (14) as

$$W_{kl} = \langle\phi_0|P_M H P_M H \ldots H P_M B \tilde{P}_M H \ldots \tilde{P}_M H \tilde{P}_M A|\phi_0\rangle. \tag{16}$$

Since H is diagonal in the basis $|\psi_j\rangle$ and $|\tilde{\psi}_j\rangle$, respectively, one can write finally

$$W_{kl} = \sum_{i=0}^{M}\sum_{j=0}^{M} \langle\phi_0|\psi_i\rangle\langle\psi_i|B|\tilde{\psi}_j\rangle\langle\tilde{\psi}_j|A|\phi_0\rangle \ (\varepsilon_i)^k (\tilde{\varepsilon}_j)^l. \tag{17}$$

It is important to note that expression (17) for the matrix element is exact, independently of how (in)accurate the representation $|\psi_i\rangle$, ε_i and $|\tilde{\psi}_j\rangle$, ε_j, respectively, are for true system eigenvalues. The only condition is that number of Lanczos steps is sufficient, i.e., $M > (l, k)$.

4.1 Finite-temperature Lanczos method: Static quantities

A straightforward calculation of the canonical thermodynamic average of an operator A at finite temperature $T > 0$ (in a finite system) requires the knowledge of all eigenstates $|\Psi_n\rangle$ and corresponding energies E_n, obtained, e.g., by the full ED of H

$$\langle A\rangle = \sum_{n=1}^{N_{st}} e^{-\beta E_n}\langle\Psi_n|A|\Psi_n\rangle \Big/ \sum_{n=1}^{N_{st}} e^{-\beta E_n}, \tag{18}$$

where $\beta = 1/k_B T$. Such a direct evaluation is both CPU time and storage demanding for larger systems and is at present accessible only for $N_{st} \sim 20000$.
In a general orthonormal basis $|n\rangle$ for finite system with N_{st} basis states one can express the canonical expectation value $\langle A\rangle$ as

$$\langle A\rangle = \sum_{n=1}^{N_{st}} \langle n|e^{-\beta H} A|n\rangle \Big/ \sum_{n=1}^{N_{st}} \langle n|e^{-\beta H}|n\rangle, \tag{19}$$

The FTLM for $T > 0$ is based on the evaluation of the expectation value in Eq. (19) for each starting $|n\rangle$ using the Lanczos basis. We note that such a procedure guarantees the correct high-T expansion series (for given finite system) to high order. Let us perform the high-T expansion of Eq. (19),

$$\begin{aligned} \langle A\rangle &= Z^{-1}\sum_{n=1}^{N_{st}}\sum_{k=0}^{\infty}\frac{(-\beta)^k}{k!}\langle n|H^kA|n\rangle, \\ Z &= \sum_{n=1}^{N_{st}}\sum_{k=0}^{\infty}\frac{(-\beta)^k}{k!}\langle n|H^k|n\rangle. \end{aligned} \tag{20}$$

Terms in the expansion $\langle n|H^kA|n\rangle$ can be calculated exactly using the Lanczos procedure with $M \geq k$ steps (using $|\phi_0^n\rangle = |n\rangle$ as the starting function) since this is a special case of the expression (14). Using relation (17) with $l = 0$ and $B = 1$, we get

$$\langle n|H^kA|n\rangle = \sum_{i=0}^{M}\langle n|\psi_i^n\rangle\langle\psi_i^n|A|n\rangle(\varepsilon_i^n)^k. \tag{21}$$

Working in a restricted basis $k \leq M$, we can insert the expression (21) into sums (20), extending them to $k > M$. The final result can be expressed as

$$\begin{aligned} \langle A\rangle &= Z^{-1}\sum_{n=1}^{N_{st}}\sum_{i=0}^{M}e^{-\beta\varepsilon_i^n}\langle n|\psi_i^n\rangle\langle\psi_i^n|A|n\rangle, \\ Z &= \sum_{n=1}^{N_{st}}\sum_{i=0}^{M}e^{-\beta\varepsilon_i^n}\langle n|\psi_i^n\rangle\langle\psi_i^n|n\rangle, \end{aligned} \tag{22}$$

and the error of the approximation is $O(\beta^{M+1})$.

Evidently, within a finite system Eq. (22), expanded as a series in β, reproduces exactly the high-T series to the order M. In addition, in contrast to the usual high-T expansion, Eq. (22) remains accurate also for $T \to 0$. Let us assume for simplicity that the g.s. $|\Psi_0\rangle$ is nondegenerate. For initial states $|n\rangle$ not orthogonal to $|\Psi_0\rangle$, already at modest $M \sim 50$ the lowest eigenstate $|\psi_0^n\rangle$ converges to $|\Psi_0\rangle$. We thus have for $\beta \to \infty$,

$$\langle A\rangle = \sum_{n=1}^{N_{st}}\langle n|\Psi_0\rangle\langle\Psi_0|A|n\rangle \Big/ \sum_{n=1}^{N_{st}}\langle n|\Psi_0\rangle\langle\Psi_0|n\rangle = \langle\Psi_0|A|\Psi_0\rangle/\langle\Psi_0|\Psi_0\rangle, \tag{23}$$

where we have taken into account the completeness of the set $|n\rangle$. Thus we obtain just the usual g.s. expectation value of operator A.

The computation of static quantities (22) still involves the summation over the complete set of N_{st} states $|n\rangle$, which is clearly not feasible in practice. To obtain a useful method, a further essential approximation replaces the full summation over $|n\rangle$ by a partial one over a much smaller set of random states [18, 19]. Such an approximation is analogous to Monte Carlo methods and leads to a statistical error which can be well estimated and is generally quite small.

Let us first consider only the expectation value (19) with respect to a single random state $|r\rangle$, which is a linear combination of basis states

$$|r\rangle = \sum_{n=1}^{N_{st}} \eta_{rn} |n\rangle, \tag{24}$$

where the η_{rn} are assumed to be distributed randomly. Then the random quantity can be expressed as

$$\tilde{A}_r = \frac{\langle r|e^{-\beta H}A|r\rangle}{\langle r|e^{-\beta H}|r\rangle} = \sum_{n,m=1}^{N_{st}} \eta^*_{rn}\eta_{rm}\langle n|e^{-\beta H}A|m\rangle \Big/ \sum_{n,m=1}^{N_{st}} \eta^*_{rn}\eta_{rm}\langle n|e^{-\beta H}|m\rangle. \tag{25}$$

Assuming that due to the random sign (phase), offdiagonal terms with $\eta^*_{rn}\eta_{rm}$, $m \neq n$ cancel on average for large N_{st}, we remain with

$$\bar{A}_r = \sum_{n=1}^{N_{st}} |\eta_{rn}|^2 \langle n|e^{-\beta H}A|n\rangle \Big/ \sum_{n=1}^{N_{st}} |\eta_{rn}|^2 \langle n|e^{-\beta H}|n\rangle. \tag{26}$$

We can express $|\eta_{rn}|^2 = 1/N_{st} + \delta_{rn}$. The random deviations δ_{rn} should not be correlated with matrix elements $\langle n|e^{-\beta H}|n\rangle = Z_n$ and $\langle n|e^{-\beta H}A|n\rangle = Z_n A_n$, therefore $\bar{A}_r$ is close to $\langle A\rangle$ with an statistical error related to the effective number of terms $\bar{Z}$ in the thermodynamic sum, i.e.

$$\bar{A}_r = \langle A\rangle(1 + \mathcal{O}(1/\sqrt{\bar{Z}})), \tag{27}$$

$$\bar{Z} = e^{\beta E_0}\sum_n Z_n = \sum_{n=1}^{N_{st}} \langle n|e^{-\beta(H-E_0)}|n\rangle. \tag{28}$$

Note that for $T \to \infty$ we have $\bar{Z} \to N_{st}$ and therefore at large N_{st} a very accurate average (28) can be obtained even from a single random state [18, 19]. On the other hand, at finite $T < \infty$ the statistical error of $\tilde{A}_r$ increases with decreasing $\bar{Z}$.
To reduce statistical errors, in particular at modest $T > 0$, within the FTLM we sum in addition over R different randomly chosen $|r\rangle$, so that in the final application Eq. (22) leads to

$$\langle A\rangle = \frac{N_{st}}{ZR}\sum_{r=1}^{R}\sum_{j=0}^{M} e^{-\beta\varepsilon^r_j}\langle r|\psi^r_j\rangle\langle\psi^r_j|A|r\rangle,$$

$$Z = \frac{N_{st}}{R}\sum_{r=1}^{R}\sum_{j=0}^{M} e^{-\beta\varepsilon^r_j}|\langle r|\psi^r_j\rangle|^2. \tag{29}$$

Random states $|r\rangle = |\phi^r_0\rangle$ serve as initial functions for the Lanczos iteration, resulting in M eigenvalues ε^r_j with corresponding $|\psi^r_j\rangle$. The relative statistical error is reduced by sampling (both for $\langle A\rangle$ and Z) and behaves as

$$\delta\langle A\rangle/\langle A\rangle = \mathcal{O}(1/\sqrt{R\bar{Z}}). \tag{30}$$

For a general operator A the calculation of $|\psi_j^r\rangle$ and the corresponding matrix elements $\langle\psi_j^r|A|r\rangle$ is needed. On the other hand, the computational effort is significantly reduced if A is a conserved quantity, i.e., $[H, A] = 0$, and can be diagonalized simultaneously with H. Then

$$\langle A\rangle = \frac{N_{st}}{ZR}\sum_{r=1}^{R}\sum_{j=0}^{M} e^{-\beta\varepsilon_j^r}|\langle r|\psi_j^r\rangle|^2 A_j^r. \tag{31}$$

In this case the evaluation of eigenfunctions is not necessary since the element $\langle r|\psi_j^r\rangle = v_{j0}^r$, Eq. (6), is obtained directly from the eigenvectors of the tridiagonal matrix H_M^r. There are several quantities of interest which can be evaluated in this way, in particular thermodynamic properties such as internal energy, specific heat, entropy, as well as uniform susceptibilities etc. [7, 8].

Taking into account all mentioned assumptions, the approximation $\langle A\rangle$ (29) yields a good estimate of the thermodynamic average at all T. For low T the error is expected to be of the order of $\mathcal{O}(1/\sqrt{R})$, while for high T the error is expected to scale even as $\mathcal{O}(1/\sqrt{N_{st}R})$. Since arguments leading to these estimates are not always easy to verify, it is essential to test the method for particular cases.

4.2 Finite-temperature Lanczos method: Dynamical response

The essential advantage of the FTLM with respect to other methods is in the calculation of dynamical quantities. Let us consider the dynamical susceptibility as given by the autocorrelation function $C(\omega)$ (the procedure for a general correlation function $C_{AB}(\omega)$ is given in Ref. [7]),

$$\chi''(\omega) = \pi(1 - e^{-\beta\omega})C(\omega), \qquad C(\omega) = \frac{1}{\pi}\mathrm{Re}\int_0^{+\infty} dt\, e^{i\omega t}C(t), \tag{32}$$

with

$$C(t) = \langle A^\dagger(t)A(0)\rangle = \frac{1}{Z}\sum_n \langle n|e^{(-\beta+it)H}A^\dagger e^{-iHt}A|n\rangle. \tag{33}$$

Expanding the exponentials in analogy to static quantities, Eq. (20), we get

$$C(t) = Z^{-1}\sum_{n=1}^{N_{st}}\sum_{k,l=0}^{\infty}\frac{(-\beta+it)^k}{k!}\frac{(-it)^l}{l!}\langle n|H^k A^\dagger H^l A|n\rangle. \tag{34}$$

The expansion coefficients in Eq. (34) can be again obtained via the Lanczos method, as discussed in Sec. 4.1. Performing two Lanczos iterations with M steps, starting from the normalized vectors $|\phi_0^n\rangle = |n\rangle$ and $|\tilde{\phi}_0^n\rangle \propto A|n\rangle$, respectively, we calculate the coefficients W_{kl} following equation (17). We again note that (within the full basis $|n\rangle$) the series are, via the W_{kl}, exactly evaluated within the Lanczos basis up to order $l, k \leq M$. The latter yields through Eq. (34) a combination of (β, t) expansion, i.e., a combination of a high-T and short-t (in frequency high-ω) expansion to very high order. Extending and resumming the series in k and l into exponentials, we get in analogy with Eq. (22)

$$C(t) = Z^{-1}\sum_{n=1}^{N_{st}}\sum_{i,j=0}^{M} e^{-\beta\varepsilon_i^n}e^{it(\varepsilon_i^n - \tilde{\varepsilon}_j^n)}\langle n|\psi_i^n\rangle\langle\psi_i^n|A^\dagger|\tilde{\psi}_j^n\rangle\langle\tilde{\psi}_j^n|A|n\rangle. \tag{35}$$

Finally replacing the full summation with the random sampling from the FTLM recipe for the correlation function, we obtain

$$C(\omega) = \frac{N_{st}}{ZR} \sum_{r=1}^{R} \sum_{i,j=1}^{M} e^{-\beta\varepsilon_i} \langle r|\psi_i^r\rangle\langle\psi_i^r|A^\dagger|\tilde{\psi}_j^r\rangle\langle\tilde{\psi}_j^r|r\rangle \delta(\omega - \tilde{\varepsilon}_j^r + \varepsilon_i^r). \tag{36}$$

We check the nontrivial $T = 0$ limit of the above expression. If the $|n\rangle$ are not orthogonal to the g.s., $|\Psi_0\rangle$, then for large enough M the lowest-lying state converges to $\varepsilon_0^n \sim E_0$ and $|\psi_0^n\rangle \sim |\Psi_0\rangle$, respectively. In this case we have

$$C(\omega, T = 0) \approx \frac{N_{st}}{R} \sum_{r=1}^{R} \sum_{j=0}^{M} \langle\Psi_0|A^\dagger|\tilde{\psi}_j^n\rangle\langle\tilde{\psi}_j^n|A|r\rangle\langle r|\Psi_0\rangle \delta(\omega + E_0 - \tilde{\varepsilon}_j^n) \tag{37}$$

At $T \sim 0$ one needs in general $M \gg 100$ so that at least the low-lying states relevant to Eq. (37) approach $|\tilde{\psi}_j^n\rangle \to |\Psi_j\rangle$ and $\tilde{\varepsilon}_j^n \to E_j$. Also a considerable number of samples $R > 1$ is required to get correct also amplitudes of separate peaks in the spectrum of Eq. (37), which are a subject to statistical errors due to the incomplete projection on the different random $|r\rangle$ in $\langle\tilde{\psi}_j^n|A|r\rangle\langle r|\Psi_0\rangle$. Similar statistical errors can in fact appear also for static quantities in Eq. (29).

4.3 Finite temperature Lanczos method: Implementation

Most straightforward is the implementation of the FTLM for static quantities, Eq. (29). In particular for conserved quantities, Eq. (31), the computation load is essentially that of the g.s. Lanczos iteration, repeated R times, and only minor changes are needed within the usual g.s. Lanczos code.

On the other hand, for the dynamical correlation function (36) the memory requirement as well as the CPU time is dominated mostly by the evaluation of the matrix element $\langle\psi_i^r|A^\dagger|\tilde{\psi}_j^r\rangle$ where the operations scale as $Op \propto RM^2N_{st}$ and memory as $Mem \propto MN_{st}$. This effectively limits the application of the FTLM to $50 < M < 500$ where the lower bound is determined by the convergence of the g.s. $|\Psi_0\rangle$. Still, it should be noted that the calculation can be done simultaneously (without any additional cost) for all desired T, since matrix elements are evaluated only once. Evidently, one should use as much as possible symmetries of the Hamiltonian, e.g., N_e, S_{tot}^z, q to reduce the effective N_{st} by splitting the sampling over different symmetry sectors.

The effect of finite M is less evident. Since $M \sim 100$ is enough to converge well a few lowest levels, it is also generally satisfactory for reliable dynamical correlation functions at low T. At high T, however, one can observe very regular oscillations which are an artifact of the Lanczos iterations with $M \ll N_{st}$. Namely, the procedure generates between the extreme eigenvalues a spectrum of quasi-states with quite equidistant level spacing $\Delta\varepsilon \sim \Delta E/M$, where ΔE is the full energy span of MB eigenstates. The effect is well visible in Fig. 2 where the high-T result for the spin structure factor $S(q = \pi, \omega)$ os the 1D Heisenberg model, Eq. (1), is presented for various M. It is evident that for the presented case ($N = 24$ and $\Delta E \sim 16J$) $M > 200$ is sufficient to obtain smooth spectra even for high $T \gg J$. However, larger M are advisable if sharper structures persist at high T.

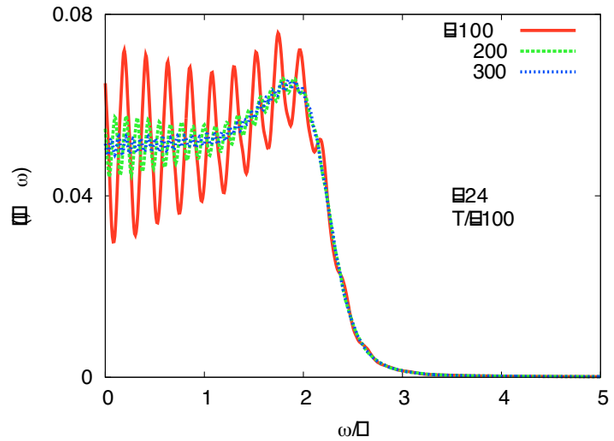


Fig. 2: *High-T spin structure factor $S(q = \pi, \omega)$ for the 1D Heisenberg model, as calculated with different numbers of Lanczos steps M.*

The role of random sampling R is less important for intermediate and high T since the relative error is largely determined via $\bar{Z}$ as evident from Eq. (28). Larger $R \gg 1$ is necessary only for the correct limit $T \to 0$ (for given system size) and for off-diagonal operators A.

One can claim that the FTLM in general obtains for all reachable systems results which are at any T very close to exact (full ED) results for the same finite (given N) system and the accuracy can be improved by increasing M and R. Still, it remains nontrivial but crucial to understand and have in control finite size effects.

At $T = 0$ both static and dynamical quantities are calculated from the g.s. $|\Psi_0\rangle$, which can be quite dependent on the size and the shape of the system. At least in 1D for static quantities the finite-size scaling $N \to \infty$ can be performed in a controlled way, although in this case more powerful methods as, e.g., DMRG are mostly available. In higher dimensional lattices, e.g., in 2D systems, finite-size scaling is practically impossible due to the very restricted choice of small sizes and different shapes. Also g.s. ($T = 0$) dynamical quantities are often dominated by few (typically $N_p < M$) peaks which are finite-size dependent [5]. On the other hand, $T > 0$ generally introduces the thermodynamic averaging over a large number of eigenstates. This directly reduces finite-size effects for static quantities, whereas for dynamical quantities spectra become denser. From Eq. (36) it follows that we get in spectra at elevated $T > 0$ typically $N_p \propto RM^2$ different peaks resulting in nearly continuous spectra. This is also evident from the example of a high-T result in Fig. 2.

It is plausible that finite-size effects at $T > 0$ become weaker. However, it should be recognized that there could exist several characteristic length scales in the physical (and model) system, e.g. the antiferromagnetic (AFM) correlation length ξ, the transport mean free path l_s etc. These lengths generally decrease with increasing T and results for related quantities get a macroscopic relevance provided that $\xi(T), l_s(T) < L$ where $L \propto N^{1/D}$ is the linear size of the system. However, there exist also anomalous cases, e.g., in an integrable system l_s can remain infinite even at $T \to \infty$ [20, 21].

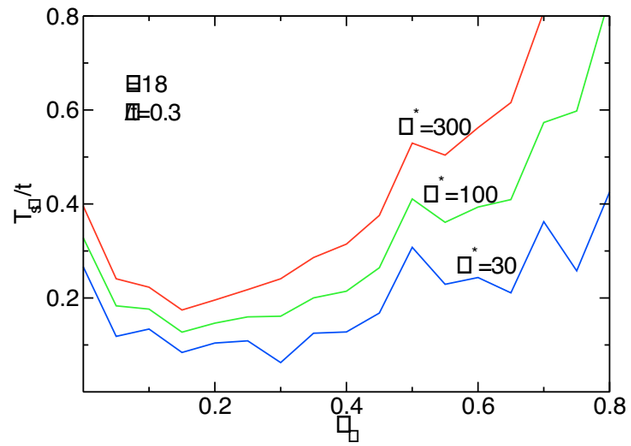


Fig. 3: *Finite-size temperature T_{fs} vs. hole doping c_h in the 2D t-J model with $J/t = 0.3$, as calculated with the FTLM in a system of $N = 18$ sites [7].*

As a simple criterion for finite size effects one can use the normalized thermodynamic sum $\bar{Z}(T)$, Eq. (28), which provides the effective number of MB states contributing at chosen T (note that for a system with a nondegenerate g.s. $\bar{Z}(T=0)=1$). A finite-size temperature T_{fs} can be thus defined with the relation $\bar{Z}(T_{fs}) = Z^*$ where in practice the range $10 < Z^* < 50$ is reasonable. Clearly, the FTLM is best suited just for systems with a large density of low lying MB states, i.e., for large $\bar{Z}$ at low T.

Since $\bar{Z}(T)$ is directly related to the entropy density s and the specific heat C_v of the system, large $\bar{Z}$ at low T is the signature of frustrated quantum MB systems, which are generally difficult to cope with using other methods (e.g., the QMC method). Typically examples of such strongly correlated electrons with an inherent frustration are the doped AFM and the t-J model, Eq. (2), in the strong correlation regime $J < t$. As an example, we present in Fig. 3 the variation of T_{fs} in the 2D t-J model with the hole doping $c_h = N_h/N$, as calculated for different $Z^* = 30$–300 for the fixed system of $N = 18$ sites and $J/t = 0.3$ as relevant for high-T_c cuprates. It is indicative that T_{fs} reaches a minimum for intermediate (optimum) doping $c_h = c_h^* \sim 0.15$, where we are able to reach $T_{fs}/t \sim 0.1$. Away from the optimum doping T_{fs} is larger, i.e., low-energy spectra are quite sparse both for the undoped AFM and even more so for the effectively noninteracting electrons far away from half-filling (for nearly empty or full band).

4.4 Low-temperature Lanczos method

The standard FTLM suffers at $T \to 0$ from statistical errors due to finite sampling R, both for static quantities, Eqs. (29), (30), as well as for dynamical correlations, Eqs. (36), (37). The discrepancy can be easily monitored by the direct comparison with the g.s. Lanczos method, Eqs. (8), (11). To avoid this problem, a variation of the FTLM method, called Low-temperature Lanczos method (LTLM) has been proposed [9] which obtains correct g.s. result (for finite systems) independent of the sampling R.

The idea of LTLM is to rewrite Eq. (19) in a symmetric form

$$\langle A\rangle = \frac{1}{Z}\sum_{n=1}^{N_{st}}\langle n|e^{-\beta H/2}Ae^{-\beta H/2}|n\rangle, \tag{38}$$

and insert the Lanczos basis in analogy with the FTLM, Eq. (19), now represented with a double sum

$$\langle A\rangle = \frac{N_{st}}{ZR}\sum_{r=1}^{R}\sum_{j,l=0}^{M} e^{-\beta(\varepsilon_j^r+\varepsilon_l^r)/2}\langle r|\psi_j^r\rangle\langle\psi_j^r|A|\psi_l^r\rangle\langle\psi_l^r|r\rangle, \tag{39}$$

The advantage of the latter form is that it satisfies the correct $T=0$ limit provided that the g.s. is well converged, i.e., $|\psi_0^r\rangle \sim |\Psi_0\rangle$. It then follows from Eq. (39),

$$\langle A\rangle = \sum_{r=1}^{R}\langle r|\Psi_0\rangle\langle\Psi_0|A|\Psi_0\rangle\langle\Psi_0|r\rangle \Big/ \sum_{r=1}^{R}\langle\Psi_0|r\rangle\langle r|\Psi_0\rangle = \langle\Psi_0|A|\Psi_0\rangle, \tag{40}$$

for any chosen set of $|r\rangle$. For the dynamical correlations $C(t)$ one can in a straightforward way derive the corresponding expression in the Lanczos basis

$$C(\omega) = \frac{N_{st}}{ZR}\sum_{r=1}^{R}\sum_{i,j,l=0}^{M} e^{-\beta(\varepsilon_i^r+\varepsilon_l^r)/2}\langle r|\psi_i^r\rangle\langle\psi_i^r|A^\dagger|\tilde{\psi}_j^{rl}\rangle\langle\tilde{\psi}_j^{rl}|A|\psi_l^r\rangle\langle\psi_l^r|r\rangle\,\delta(\omega-\tilde{\varepsilon}_j^{rl}+\frac{1}{2}(\varepsilon_i^r+\varepsilon_l^r)). \tag{41}$$

It is again evident that for $T\to 0$ the sampling does not influence results, being correct even for $R=1$ if the g.s. $|\Psi_0\rangle$ is well converged for all starting $|r\rangle$. The payoff is in an additional summation over the new Lanczos basis sets $|\tilde{\psi}_j^{rl}\rangle$, which needs to be started from each $A|\psi_l^r\rangle$ in Eq. (41) separately. Since the LTLM is designed for lower T, one can effectively restrict summations in (i,l) in Eq. (41) to much smaller $M'\ll M$, where only lowest states with $\varepsilon_i^r,\varepsilon_l^r\sim E_0$ contribute [9], and in addition use smaller $M_1\ll M$ for the basis $|\tilde{\psi}_j^{rl}\rangle$.
An alternative version for a Lanczos-type approach [22] to dynamical quantities is not to start the second Lanczos run from $A|r\rangle$ [7] or from $A|\psi_l^r\rangle$ [9], but from

$$|\widetilde{Ar}\rangle = \sum_{l=0}^{M} A|\psi_l^r\rangle e^{-\beta\varepsilon_l^r/2}\langle\psi_l^r|r\rangle. \tag{42}$$

In this way one obtains with the second Lanczos run the Lanczos eigenstates $|\widetilde{\psi}_k^r\rangle$, which cover the relevant Hilbert space for starting random vector $|r\rangle$ and the inverse temperature β. The resulting dynamical autocorrelation function is

$$C(\omega) = \frac{N_{\rm st}}{RZ}\sum_{r=1}^{R}\sum_{i,k=0}^{M} e^{-\beta\varepsilon_i^r/2}\langle r|\psi_i^r\rangle\langle\psi_i^r|A^\dagger|\widetilde{\psi}_k^r\rangle\langle\widetilde{\psi}_k^r|\widetilde{Ar}\rangle\,\delta(\omega-\tilde{\varepsilon}_k^r+\varepsilon_i^r). \tag{43}$$

In this way the sufficiency of only one random vector in the $T=0$ limit is reproduced, while at $T>0$ the algorithm has the same time efficiency as the FTLM, but with much smaller random sampling needed to reach the same accuracy (at least for low T). However, the price paid is that results for each T need to be calculated separately, while within the FTLM all T (or T up to a certain value within the LTLM) are evaluated simultaneously.

4.5 Microcanonical Lanczos method

While most investigations in strongly correlated systems focus on the low-T regime, there are systems where dynamical properties are nontrivial even at high T. A well known such case is the spin diffusion constant $D_s(T)$ in the isotropic Heisenberg model, Eq. (1), whose value is not known, and even its existence at any $T > 0$ is uncertain. Similar although somewhat less controversial is the case of transport quantities, both for integrable or generic nonintegrable models. Whereas the FTLM seems well adapted for studies of transport response functions, oscillations due to a limited M can compromise the crucial low-ω resolution, cf. Fig. 2.

At elevated T it is therefore an advantage to use the microcanonical Lanczos method (MCLM) [10], employing the fact from statistical physics that in the thermodynamic limit (for large system) the microcanonical ensemble should yield the same results as the canonical one. Shortcomings of the MCLM are due to the fact that in finite systems statistical fluctuations are much larger within the microcanonical ensemble. Still, reachable finite-size systems have a very high density of states in the core of the MB spectrum as probed by high T. Hence, statistical fluctuations are at high T effectively smoothed out in contrast to low-T properties dominated by a small number of low lying MB states.

The implementation of the MCLM is quite simple and straightforward. One first determines the target energy $\lambda = \langle H \rangle(T)$ which represents the microcanonical energy equivalent to the canonical average energy for chosen T and system size N. Since λ is a parameter within the MCLM, one can relate it to T by performing either FTLM (simplified due to H being a conserved quantity) on the same system, or extrapolating full ED results (with linear dependence on N) on small lattices. Next we find a representative microcanonical state $|\Psi_\lambda\rangle$ for the energy λ. One convenient way within the Lanczos-type approach is to use the new operator

$$V = (H - \lambda)^2. \tag{44}$$

Performing Lanczos iterations with the operator V yields again the extremum eigenvalues, in particular the lowest one close to $V \sim 0$. In contrast to the g.s. procedure, the convergence to a true eigenstate cannot be reached in system sizes of interest even with $M_1 \gg 100$. The reason is the extremely small eigenvalue spacing of the operator V, scaling as $\Delta V_n \propto (\Delta E/N_{st})^2$, with ΔE being the whole energy span within the given system. Fortunately such a convergence is not necessary (nor even desired) since the essential parameter is the small energy uncertainty σ_E, given by

$$\sigma_E^2 = \langle \Psi_\lambda | V | \Psi_\lambda \rangle. \tag{45}$$

For small energy spread $\sigma_E/\Delta E < 10^{-3}$ typically $M_1 \sim 1000$ is needed. Again, to avoid storing M_1 Lanczos wavefunctions $|\phi_i\rangle$ the Lanczos procedure is performed twice as described in Sec. 2.2, i.e., the second time with known tridiagonal matrix elements to calculate finally $|\Psi_\lambda\rangle$ in analogy with Eq. (6). The latter is then used to evaluate any static expectation average $\langle A \rangle$ or the dynamical correlation function as in Eq. (9),

$$C(\omega, \lambda) = \langle \Psi_\lambda | A^\dagger \frac{1}{\omega^+ + \lambda - H} A | \Psi_\lambda \rangle. \tag{46}$$

The latter is evaluated again using Lanczos iterations with M_2 steps starting with the initial wavefunction $|\tilde{\phi}_0\rangle \propto A|\Psi_\lambda\rangle$ and $C(\omega,\lambda)$ is represented in terms of a continued fraction. Since the MB levels are very dense and correlation functions smooth at $T \gg 0$, large $M_2 \gg 100$ are needed but as well easily reachable to achieve high-ω resolution in $C(\omega,\lambda)$.
It is evident that the computer requirement for the MCLM both regarding the CPU and memory are essentially the same as for the g.s. dynamical calculations except that typically $M_1, M_2 \gg 100$. In particular, requirements are less demanding than using the FTLM with $M > 100$. A general experience is that for systems with large $N_{st} \gg 10000$ the MCLM dynamical results agree very well with FTLM results for the same system. It should also be noted that the actual frequency resolution $\delta\omega$ in $C(\omega,\lambda)$, Eq. (46), is limited by $\delta\omega \sim \sigma_E$ which is, however, straightforward to improve by increasing M_1, M_2 with typical values $M_1, M_2 > 1000$. One can also improve the MCLM results for any T by performing an additional sampling over initial random starting $|\phi_0\rangle$ as well as over λ with a probability distribution $p(\lambda)$ simulating the canonical ensemble in a finite-size system, i.e., by replacing Eq. (46) with

$$C(\omega) = \sum_\lambda p(\lambda) C(\omega,\lambda). \tag{47}$$

4.6 Statical and dynamical quantities at $T > 0$: Applications

The FTLM has been designed to deal with the simplest tight-binding models of strongly correlated electrons, at the time mostly with challenging microscopic electronic models of high-T_c superconductors [6, 7], where besides superconductivity there is a variety of anomalous non-Fermi-liquid-like properties even in the normal state. Clearly of interest in this connection are prototype MB models as the Heisenberg model, Eq. (1), the t-J model, Eq. (2), and the Hubbard model on the 2D square lattice. The unfrustrated Heisenberg model can be numerically studied on much bigger lattices with QMC and related methods. The 2D Hubbard model was and still is mostly subject of DMFT and QMC studies, since at half-filling or close to it the Lanczos methods are quite restricted due to the large N_{st} even for modest sizes $N \sim 16$. Therefore one focus of Lanczos-based approaches was on the t-J model being, with some generalizations, a microscopic representation of electronic properties of high-T_c cuprates.
Thermodynamic quantities such as chemical potential μ, entropy density s, specific heat C_v are the easiest to implement within the FTLM. Their T- and (hole) doping c_h-dependence within the t-J model on a 2D square lattice (calculated for up to $N = 26$ sites) reveal the very anomalous behavior of doped Mott insulators [23] (as evident already from Fig. 3), confirmed also by results for the more complete Hubbard model [24].
The advantages of the FTLM and also its feasibility for the 2D t-J model are even more evident in numerous studies of spin- and charge-dynamics at $T > 0$ [7], which show good agreement with neutron scattering and NMR [25, 26], optical conductivity $\sigma(\omega)$ and resistivity $\rho(T)$ [27], as well as some other anomalous properties of the cuprates [8]. As an example of a transport quantity hardly accessible by other methods we present in Fig. 4 the universal planar resistivity $\rho(T)$, as extracted from the dynamical conductivity $\sigma(\omega \to 0) = 1/\rho$, within the t-J model

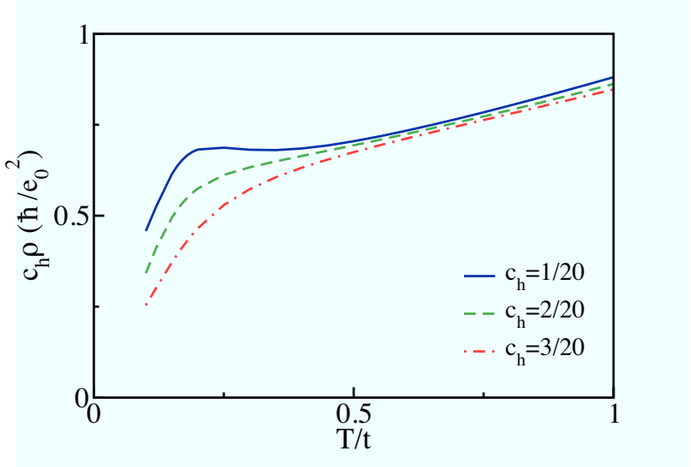


Fig. 4: *Normalized 2D resistivity $c_h\rho$ vs. T/t within the t-J model with $J/t = 0.3$ for different hole concentrations c_h [27].*

for different doping levels c_h [27]. The result in Fig. 4 clearly show a linear dependence below the pseudogap temperature T^* dependent on doping c_h. Another characteristic signature is a saturation (plateau) of $\rho(T)$ at low doping and the universal trend at high T.

Spectral properties as manifested in the single-particle spectral functions $A(\mathbf{k},\omega)$ are at the core of the understanding of cuprates, as well as of strongly correlated electrons in general. Here, even g.s. and low-T properties are a challenge for numerical studies whereby the FTLM can be viewed as a controlled way to get reliable (macroscopic-like) $T \to 0$ result, in contrast to quite finite-size plagued results obtained via the g.s. Lanczos procedure [5]. Using the FTLM at $T \sim T_{fs}$ with twisted boundary conditions we can simulate a continuous wavevector $\mathbf{k}$. Using in addition coarse graining averaging one can reach results for $A(\mathbf{k},\omega)$ [28,29] giving insights into electron vs. hole doped angle-resolved photoemission experiments, quasiparticle relaxation, and waterfall-like effects. A characteristic result of such studies is in Fig. 5 for the single-particle density of states $\mathcal{N}(\omega) = \sum_{\mathbf{k}} A(\mathbf{k},\omega)$ [28]. Here, the strength of the FTLM is visible in the high ω resolution within the most interesting low-ω window. Interesting and reproducible are also nontrivial spectral shapes as the sharp peak close to $\omega < 0$ and a broad shoulder for $\omega \ll 0$. Most important is, however, the evident pseudogap (observed also experimentally in cuprates) visible at $\omega \sim 0$ in the low-doping regime.

Besides the challenging models for cuprates there have been also studies of static and dynamical properties of multiband and multiorbital models which either reduce to the generalized t-J model [30] or to Kondo lattice models [31, 32]. While the increasing number of local basis states K clearly limits the applicability of ED-based methods, they are competitive in treating nontrivial frustrated spin models less suitable for QMC and other methods, however closely related to the physics of novel materials. Moreover, frustrated models are characterized by a large entropy density s and related low T_{fs}, essential conditions for the feasibility of FTLM results. Examples of such systems are the Shastry-Sutherland model [33, 34], the 2D J_1-J_2 model [35], and properties of frustrated magnetic molecules [36–38].

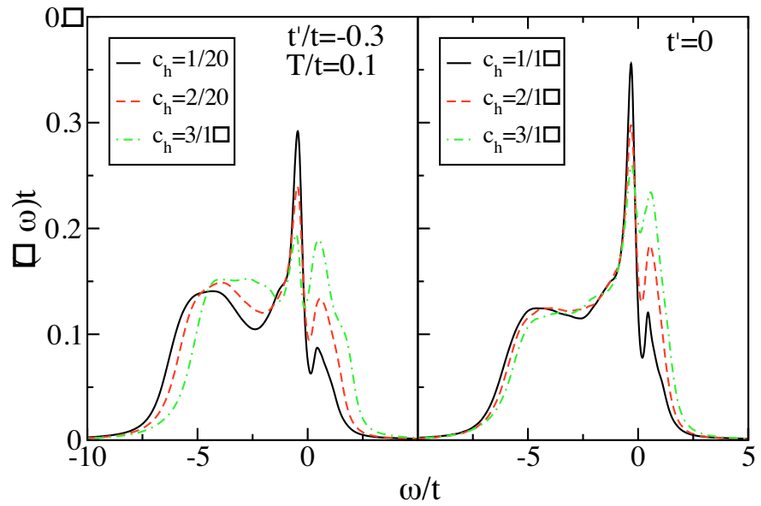


Fig. 5: *Density of states $\mathcal{N}(\omega)$ for different dopings c_h within the extended t-J model with n.n.n. hopping $t' = -0.3t$ and $t' = 0$, respectively [28].*

Another class of problems which can be quite effectively dealt with using the FTLM and MCLM approaches is the fundamental as well as experimentally relevant problem of transport in 1D systems of interacting fermions as realized, e.g., in quasi-1D spin-chain materials [39]. It has been recognized that the transport response at any $T > 0$ crucially differs between integrable and nonintegrable systems. Since the 1D isotropic as well as anisotropic Heisenberg model, Eq. (1), is integrable it opens a variety of fundamental questions of anomalous transport in such systems, the effects of perturbative terms and impurities. Such fundamental questions on transport and low-ω dynamic response remain nontrivial even at high T [20, 21], hence the MCLM is the most feasible and straightforward method. It has been in fact first probed on the anomalous transport in 1D insulators [40] but furtheron used to study interaction-induced transport at $T > 0$ in disordered 1D systems [41, 42], in particular in relation to challenging problems of many-body localization [43, 44] being inherently the question of low-frequency dynamics at $T \to \infty$.

In Fig. 6 we present as an example MCLM result for the dynamical spin conductivity in the anisotropic Heisenberg model, Eq. (1), where $J^{zz} \neq J^{xx} = J^{yy} = J$ in the Ising-like (with the spin gap in the g.s.) regime $\Delta = J^{zz}/J > 1$. Results for the high-T dynamical spin conductivity $T\sigma(\omega)$ are shown for various next-neighbor (anisotropic) couplings $\alpha = J_2^{zz}/J$. The first message is that the MCLM is well adapted for the high ω resolution (here using $M_1 = M_2 = 2000$) and reaching large $N = 30$ ($N_{st} \sim 5.10^6$ in a single $S^z = 0, q$ sector). Another conclusion is that the dynamics of such systems is very anomalous. For the integrable case $\alpha = 0$ we find $\sigma_0 = \sigma(0) \sim 0$ but also an anomalous finite-size peak at $\omega_p \propto 1/N$ [40]. At the same time breaking integrability with $\alpha > 0$ appears to lead to $\sigma_0 > 0$ still approaching an 'ideal' insulator (insulating at all T) for a weak perturbation $\sigma_0(\alpha \to 0) \to 0$ [45].

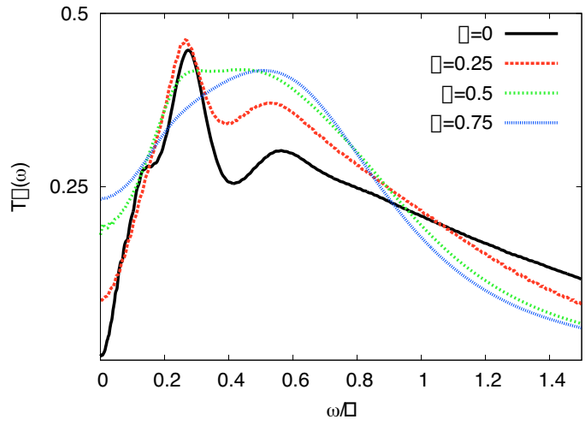


Fig. 6: *High-T dynamical spin conductivity $T\sigma(\omega)$ within the anisotropic Heisenberg model in the Ising-like regime, $\Delta = 1.5$, and various next-neighbor interactions $\alpha = J_2^{zz}/J$ as calculated with the MCLM on a chain of $N = 30$ sites.*

5 Real-time dynamics using the Lanczos method

Research in the field of non-equilibrium dynamics of complex quantum systems constitutes a formidable theoretical challenge. When dealing with ED approaches or calculations in a reduced basis, the time evolution of the time-dependent Shrödinger equation,

$$i\frac{\partial \Psi(t)}{\partial t} = H(t)\Psi(t), \tag{48}$$

can be efficiently obtained using the time-dependent Lanczos technique, as originally described in Ref. [46] and later applied and analyzed in more detail [47]. One of the straightforward reasons is that most commonly the Lanczos method is used to compute the g.s. of MB Hamiltonian. Generalizing the method to time-dependent calculation represents only a minor change to already existing codes. Even though the method is most suitable for the time evolution of the time-independent Hamiltonian, it can nevertheless be applied even to the time-dependent case. The time evolution of $|\Psi(t)\rangle$ is then calculated by a step-wise propagation in time t by small time increments δt, generating at each step a Lanczos basis of dimension M (typically $M < 10$), to follow the evolution

$$|\Psi(t+\delta t)\rangle \simeq e^{-iH(t)\delta t}|\Psi(t)\rangle \simeq \sum_{l=1}^{M} e^{-i\varepsilon_l \delta t}|\psi_l\rangle\langle\psi_l|\Psi(t)\rangle, \tag{49}$$

where $|\psi_l\rangle, \varepsilon_l, l = 0 \ldots M$ are Lanczos eigenfunctions and eigenvalues, respectively, obtained via the Lanczos iteration started with $|\phi_0\rangle = |\Psi(t)\rangle$. The advantage of the time-evolution method following Eq. (49) is that it preserves the normalization of $|\Psi(t+\delta t)$ for arbitraryly large δt. The approximation of finite M in Eq. (49) is also correct at least to the M-th Taylor-expansion order in δt. It is, however, important to stress that δt should be chosen small enough to properly take into account the time-dependence of $H(t)$. E.g., when driving the system with a constant external electric field, $\delta t/t_B \sim 10^{-3}$ where t_B is the Bloch oscillation period [48,45].

So far, investigations of correlated systems under the influence of a driving electric field in 1D using Lanczos time-evolution focused on generic systems, like the metallic and Mott-insulating regime of interacting spinless fermions [48, 45]. Even though rather small systems can be studied it has been established that a steady state can be reached without any additional coupling to a heat bath, provided that the Joule heating of the system is properly taken into account.

6 Discussion

Exact diagonalization based methods, both the full ED and the Lanczos-type ED approach, are very extensively employed in the investigations of strongly correlated MB quantum systems in solids and elsewhere. The reason for their widespread use are several: a) unbiased approach to the MB problem without any simplifications or approximations, independent of the complexity of the MB system, b) relative simplicity of generating the codes for various models and observables, c) easy and straightforward testing of codes, d) direct interpretation of the obtained quantum MB states and their possible anomalous structure and properties, e) high pedagogical impact as a quick and at the same time very nontrivial introduction into the nature of MB quantum physics. Also the Lanczos-based methods described in this review, i.e., the g.s. Lanczos method for static and dynamic quantities, and the somewhat more elaborate FTLM, MCLM, LTLM and EDLFS, require rather modest programming efforts in comparison with more complex numerical methods, e.g., QMC- and DMRG-based methods, as described in other chapters. Clearly, the main drawback of ED methods is the smallness of lattice sizes N limited by the number of basis states (at present $N_{st} < 10^9$) that can be treated with a Lanczos iteration procedure. The achievable N with ED methods appears quite modest in comparison with some established and recently developed numerical methods, such as QMC, DMRG, matrix-product-states methods, etc. Still, in spite of the intensive developments and advances of novel numerical methods in last two decades, there are several aspects of strong-correlation physics, where ED-based methods are so far either the only feasible or at least superior ones. In this chapter we have focused mostly on Lanczos-based methods and applications where they are competitive and get nontrivial results with a macroscopic validity:

a) MB g.s. and its properties: for frustrated and complex models mostly so far do not offer alternative powerful methods at least beyond $D = 1$ systems, where DMRG is efficient.

b) $T > 0$ static properties evaluated with as the FTLM and the LTLM are most powerful and reliable for frustrated and complex system, in particular in systems with high degeneracies of MB states and large entropy at low T,

c) $T > 0$ Lanczos methods for dynamical quantities, such as FTLM and MCLM, yield for many models and geometries results superior to other methods or, in several cases, even the only accessible results. Particular advantages are the high ω resolution at all T beyond the finite size limit $T > T_{fs}$, macroscopic-like results at low T with proper $T \to 0$ scaling, and the possibility of detailed studies of systems with nontrivial dynamics at any, in particular high T.

d) The Lanczos technique is the natural application for methods with a restricted MB basis sets and DMRG-type targeting, as well as for the real-time evolution of strongly correlated systems.

References

[1] C. Lanczos, J. Res. Nat. Bur. Stand., **45**, 255 (1950)

[2] B.N. Parlett: *The Symmetric Eigenvalue Problem* (Prentice Hall, Eaglewood Cliffs, 1980)

[3] J.W. Demmel: *Applied Numerical Linear Algebra* (SIAM, Philadelphia, 1997)

[4] R. Haydock, V. Heine, and M.J. Kelly, J. Phys. C: Solid State Phys. **8**, 2591 (1975)

[5] for a review, see E. Dagotto, Rev. Mod. Phys. **66**, 763 (1994)

[6] J. Jaklič and P. Prelovšek, Phys. Rev. B **49**, 5065 (1994)

[7] for a review, see J. Jaklič and P. Prelovšek, Adv. Phys. **49**, 1 (2000)

[8] for a review, see P. Prelovšek and J. Bonča, in A. Avella and F. Mancini (eds.): *Strongly Correlated Systems: Numerical Methods* (Springer, 2013)

[9] M. Aichhorn, M. Daghofer, H.G. Evertz, and W. von der Linden, Phys. Rev. B **67**, 161103(R) (2003)

[10] M.W. Long, P. Prelovšek, S. El Shawish, J. Karadamoglou, and X. Zotos, Phys. Rev. B **68**, 235106 (2003)

[11] for a review, see M. Imada, A. Fujimori, and Y. Tokura, Rev. Mod. Phys. **70**, 1039 (1998)

[12] P.W. Leung, Phys. Rev. B **73**, 14502 (2006)

[13] T. Tohyama, Y. Inoue, K. Tsutsui, and S. Maekawa, Phys. Rev. B **72**, 045113 (2005)

[14] J. Oitmaa and D.D. Betts, Can. J. Phys. **56**, 897 (1978)

[15] D.D. Betts, H.Q. lin, and J.S. Flynn, Can. J. Phys. **77**, 3535 (1999); P.R.C. Kent, M. Jarrell, T.A. Maier, and Th. Pruschke Phys. Rev. B **72**, 060411 (2005)

[16] J. Cullum and R.A. Willoughby, J. Comp. Phys. **44**, 329 (1981)

[17] H. Mori, Prog. Theor. Phys. **34**, 423 (1965)

[18] M. Imada and M. Takahashi, J. Phys. Soc. Jpn. **55**, 3354 (1986)

[19] R.N. Silver and H. Röder, Int. J. Mod. Phys. C **5**, 735 (1995)

[20] X. Zotos and P. Prelovšek, Phys. Rev. B **53**, 983 (1996)

[21] for a review, see F. Heidrich-Meisner, A. Honecker, and W. Brenig, Eur. Phys. J. Special Topics **151**, 135 (2007)

[22] J. Kokalj, Ph.D. Thesis, University of Ljubljana, 2010, unpublished.

[23] J. Jaklič and P. Prelovšek, Phys. Rev. Lett. **77**, 892 (1996)

[24] J. Bonča and P. Prelovšek, Phys. Rev. B **67**, 180502(R) (2003)

[25] J. Jaklič and P. Prelovšek, Phys. Rev. Lett. **75**, 1340 (1995)

[26] P. Prelovšek, I. Sega, and J. Bonča, Phys. Rev. Lett. **92**, 027002 (2004)

[27] M.M. Zemljič and P. Prelovšek, Phys. Rev. B **72**, 075108 (2005)

[28] M.M. Zemljič and P. Prelovšek, Phys. Rev. B **75**, 104514 (2007)

[29] M.M. Zemljič, P. Prelovšek, and T. Tohyama, Phys. Rev. Lett. **100**, 036402 (2008)

[30] P. Horsch, J. Jaklič and F. Mack, Phys. Rev. B **59**, 6217 (1999)

[31] P. Horsch, J. Jaklič and F. Mack, Phys. Rev. B **59**, 14149(R) (1999)

[32] K. Haule, J. Bonča, and P. Prelovšek, Phys. Rev. B **61**, 2482 (2000)

[33] S. El Shawish, J. Bonča, and I. Sega, Phys. Rev. B **72**, 184409 (2005)

[34] S. El Shawish, A. Ramšak, and J. Bonča, Phys. Rev. B **75**, 205442 (2007)

[35] B. Schmidt, P. Thalmeier, and N. Shannon, Phys. Rev. B **76**, 125113 (2007)

[36] J. Schnack and O. Wendland, Eur. Phys. J. B **78**, 535 (2010)

[37] for a review, see J. Schnack and J. Ummethum, Polyhedron **66**, 28 (2103)

[38] M. Haertel, J. Richter, D. Ihle, J. Schnack, and S.-L. Drechsler, Phys. Rev. B **84**, 104411 (2011)

[39] for a review, see C. Hess, Eur. Phys. J. Special Topics **151**, 73 (2007)

[40] P. Prelovšek, S. El Shawish, X. Zotos, and M. Long, Phys. Rev. B **70**, 205129 (2004)

[41] A. Karahalios, A. Metavitsiadis, X. Zotos, A. Gorczyca, and P. Prelovšek, Phys. Rev. B **79**, 024425 (2009)

[42] O.S. Barišić and P. Prelovšek, Phys. Rev. B **82**, 161106 (2010)

[43] O.S. Barišić, J. Kokalj, I. Balog, and P. Prelovšek, Phys. Rev. B **94**, 045126 (2016)

[44] P. Prelovšek, O.S. Barišić, and M. Žnidarič, Phys. Rev. B **94**, 241104 (2016)

[45] M. Mierzejewski, J. Bonča, and P. Prelovšek, Phys. Rev. Lett. **107**, 126601 (2011)

[46] T.J. Park and J.C. Light, J. Chem. Phys. **85**, 5870 (1986)

[47] N. Mohankumar and S.M. Auerbach, Comp. Phys. Comm. **175**, 473 (2006)

[48] M. Mierzejewski and P. Prelovšek, Phys. Rev. Lett. **105**, 186405 (2010)

8 Gutzwiller Density Functional Theory

Florian Gebhard
Department of Physics
Philipps Universität Marburg, Germany

Contents

1 Introduction **2**

2 Density Functional Theory **3**
2.1 Many-particle Hamiltonian and Ritz variational principle 3
2.2 Levy's constrained search 3
2.3 Single-particle reference system 5
2.4 Kohn-Sham Hamiltonian 6

3 DFT for many-particle reference systems **8**
3.1 Hubbard Hamiltonian and Hubbard density functional 8
3.2 Gutzwiller density functional 10
3.3 Gutzwiller density functional for infinite lattice coordination number 14
3.4 Local Hamiltonian and double counting for transition metals 15

4 Results for transition metals **18**
4.1 Nickel 18
4.2 Iron 22

5 Summary and outlook **24**

A Appendix **25**
A.1 Single-particle systems 25
A.2 Atomic Hamiltonian in cubic symmetry 28

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

Density Functional Theory (DFT) is the workhorse of electronic structure theory. Based on the Hohenberg-Kohn and Kohn-Sham theorems [1], the ground-state properties of an interacting many-electron system are calculated from those of an effective single-particle problem that can be solved numerically. An essential ingredient in DFT is the so-called exchange-correlation potential which, however, is unknown and sensible approximations must be devised, e.g., the local (spin) density approximation, L(S)DA. In this way, the electronic properties of metals were calculated systematically [2]. Unfortunately, the L(S)DA leads to unsatisfactory results for transition metals and their compounds. The electrons in the narrow $3d$, or $4f/5f$, bands experience correlations that are not covered by current exchange-correlation potentials.

For a more accurate description of electronic correlations in narrow bands, Hubbard-type models [3, 4] have been put forward. However, simplistic model Hamiltonians can describe limited aspects of real materials at best, while, at the same time, they reintroduce the full complexity of the many-body problem. Recently, new methods were developed that permit the (numerical) analysis of multi-band Hubbard models and, moreover, can be combined with DFT, specifically, the LDA+U method [5], the LDA+DMFT (Dynamical Mean-Field Theory) [6,7], and the Gutzwiller variational approach [8–11]. The LDA+U approach treats atomic interactions on a mean-field level so that it is computationally cheap but it ignores true many-body correlations. The DMFT becomes formally exact for infinite lattice coordination number, $Z \to \infty$, but it requires the self-consistent solution of a dynamical impurity problem that is numerically very demanding. The Gutzwiller DFT is based on a variational treatment of local many-body correlations. Expectation values can be calculated for $Z \to \infty$ without further approximations, and the remaining computational problem remains tractable.

In this chapter, we present a formal derivation of the Gutzwiller DFT as a generic extension of the DFT. Our formulae apply for general Gutzwiller-correlated wave functions and reproduce expressions used previously [9, 10] as special cases. We provide results for nickel in the face-centered cubic (fcc) structure and for iron in its body-centered cubic (bcc) ground state. The Gutzwiller DFT results for the lattice constant, the magnetic spin-only moment, and the bulk modulus agree very well with experiments. Moreover, the quasi-particle bandstructure from Gutzwiller DFT is in satisfactory agreement with data from Angle-Resolved Photo-Emission Spectroscopy (ARPES). As found earlier [8–10], the Gutzwiller DFT overcomes the limitations of DFT for the description of transition metals.

In Sect. 2 we recall the derivation of Density Functional Theory (DFT) as a variational approach to the many-body problem and its mapping to an effective single-particle reference system (Kohn-Sham scheme). In Sect. 3 we extend our derivation to many-particle reference systems. We formulate the Gutzwiller density functional whose minimization leads to the Gutzwiller–Kohn-Sham Hamiltonian. The theory is worked out in the limit of large coordination number, $Z \to \infty$, where explicit expressions for the Gutzwiller density functional are available. In Sect. 4 we present results for fcc nickel ($Z = 12$) and for bcc iron ($Z = 8$). Summary and conclusions, Sect. 5, close our presentation. Some technical details are deferred to the appendix. This work is based on Refs. [12, 13]. Further on, excerpts are taken without explicit citations.

2 Density Functional Theory

We start our presentation with a concise derivation of Density Functional Theory that can readily be extended to the Gutzwiller Density Functional Theory.

2.1 Many-particle Hamiltonian and Ritz variational principle

The non-relativistic many-particle Hamiltonian for electrons with spin $\sigma =\uparrow,\downarrow$ reads ($\hbar \equiv 1$)

$$\begin{aligned} \hat{H} &= \hat{H}_{\rm band} + \hat{H}_{\rm int} \ , \\ \hat{H}_{\rm band} &= \sum_{\sigma} \int {\rm d}\mathbf{r}\ \hat{\Psi}^{\dagger}_{\sigma}(\mathbf{r}) \left(-\frac{\Delta_{\mathbf{r}}}{2m} + U(\mathbf{r}) \right) \hat{\Psi}_{\sigma}(\mathbf{r}) \ , \\ \hat{H}_{\rm int} &= \sum_{\sigma,\sigma'} \int {\rm d}\mathbf{r} \int {\rm d}\mathbf{r}'\, \hat{\Psi}^{\dagger}_{\sigma}(\mathbf{r}) \hat{\Psi}^{\dagger}_{\sigma'}(\mathbf{r}')\, V(\mathbf{r}-\mathbf{r}')\, \hat{\Psi}_{\sigma'}(\mathbf{r}') \hat{\Psi}_{\sigma}(\mathbf{r}) \end{aligned} \tag{1}$$

with

$$V(\mathbf{r}-\mathbf{r}') = \frac{1}{2} \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \ . \tag{2}$$

The electrons experience the periodic potential of the ions, $U(\mathbf{r})$, and their mutual Coulomb interaction, $V(\mathbf{r}-\mathbf{r}')$. The total number of electrons is $N = N_{\uparrow} + N_{\downarrow}$. According to the Ritz variational principle, the ground state of a Hamiltonian $\hat{H}$ can be obtained from the minimization of the energy functional

$$E\left[\{|\Psi\rangle\}\right] = \langle \Psi | \hat{H} | \Psi \rangle \tag{3}$$

in the subset of normalized states $|\Psi\rangle$ in the Hilbert space with N electrons, $\langle \Psi | \Psi \rangle = 1$.

2.2 Levy's constrained search

The minimization of the energy functional (3) is done in two steps, the constrained search [14], Sect. 2.2.1, and the minimization of the density functional, Sect. 2.2.2. To this end, we consider the subset of normalized states $|\Psi^{(n)}\rangle$ with fixed electron densities $n_{\sigma}(\mathbf{r})$,

$$n_{\sigma}(\mathbf{r}) = \langle \Psi^{(n)} | \hat{\Psi}^{\dagger}_{\sigma}(\mathbf{r}) \hat{\Psi}_{\sigma}(\mathbf{r}) | \Psi^{(n)} \rangle \ . \tag{4}$$

In the following we accept 'physical' densities only, i.e., those $n_{\sigma}(\mathbf{r})$ for which states $|\Psi^{(n)}\rangle$ can actually be found. For the subset of states $|\Psi^{(n)}\rangle$ we define the electronic Hamiltonian

$$\hat{H}_{\rm e} = \hat{H}_{\rm kin} + \hat{V}_{\rm xc} \ , \tag{5}$$

$$\hat{H}_{\rm kin} = \sum_{\sigma} \int {\rm d}\mathbf{r}\ \hat{\Psi}^{\dagger}_{\sigma}(\mathbf{r}) \left(-\frac{\Delta_{\mathbf{r}}}{2m} \right) \hat{\Psi}_{\sigma}(\mathbf{r}) \ , \tag{6}$$

$$\begin{aligned} \hat{V}_{\rm xc} = \sum_{\sigma,\sigma'} \int {\rm d}\mathbf{r} \int {\rm d}\mathbf{r}'\, V(\mathbf{r}-\mathbf{r}') \Big[& \hat{\Psi}^{\dagger}_{\sigma}(\mathbf{r}) \hat{\Psi}^{\dagger}_{\sigma'}(\mathbf{r}') \hat{\Psi}_{\sigma'}(\mathbf{r}') \hat{\Psi}_{\sigma}(\mathbf{r}) \\ & - \hat{\Psi}^{\dagger}_{\sigma}(\mathbf{r}) \hat{\Psi}_{\sigma}(\mathbf{r}) n_{\sigma'}(\mathbf{r}') - \hat{\Psi}^{\dagger}_{\sigma'}(\mathbf{r}') \hat{\Psi}_{\sigma'}(\mathbf{r}') n_{\sigma}(\mathbf{r}) + n_{\sigma}(\mathbf{r}) n_{\sigma'}(\mathbf{r}') \Big] \ . \end{aligned} \tag{7}$$

Here, we extracted the Hartree terms from the Coulomb interaction $H_{\rm int}$ in eq. (1) so that $\hat{V}_{\rm xc}$ contains only the so-called exchange and correlation contributions. In the subset of normalized states $|\Psi^{(n)}\rangle$ we define the functional

$$F\left[\{n_\sigma(\mathbf{r})\}, \{|\Psi^{(n)}\rangle\}\right] = \langle\Psi^{(n)}|\hat{H}_{\rm e}|\Psi^{(n)}\rangle \ . \tag{8}$$

For fixed densities $n_\sigma(\mathbf{r})$, the Hamiltonian $\hat{H}_{\rm e}$ defines an electronic problem where the periodic potential of the ions and the Hartree interaction are formally absent.

2.2.1 Constrained search

The formal task is to find the minimum of the energy functional F in (8) with respect to $|\Psi^{(n)}\rangle$,

$$\bar{F}\left[\{n_\sigma(\mathbf{r})\}\right] = {\rm Min}_{\{|\Psi^{(n)}\rangle\}} F\left[\{n_\sigma(\mathbf{r})\}, \{|\Psi^{(n)}\rangle\}\right] \ . \tag{9}$$

Recall that the electron densities $n_\sigma(\mathbf{r})$ are fixed in this step. We denote the resulting optimal many-particle state by $|\Psi_0^{(n)}\rangle$. Thus, we may write

$$\bar{F}\left[\{n_\sigma(\mathbf{r})\}\right] = F\left[\{n_\sigma(\mathbf{r})\}, \left\{|\Psi_0^{(n)}\rangle\right\}\right] = \langle\Psi_0^{(n)}|\hat{H}_{\rm e}|\Psi_0^{(n)}\rangle \ . \tag{10}$$

For later use, we define the corresponding functionals for the kinetic energy

$$K\left[\{n_\sigma(\mathbf{r})\}\right] = \langle\Psi_0^{(n)}|\hat{H}_{\rm kin}|\Psi_0^{(n)}\rangle \tag{11}$$

and the exchange-correlation energy

$$E_{\rm xc}\left[\{n_\sigma(\mathbf{r})\}\right] = \langle\Psi_0^{(n)}|\hat{V}_{\rm xc}|\Psi_0^{(n)}\rangle \tag{12}$$

so that

$$\bar{F}\left[\{n_\sigma(\mathbf{r})\}\right] = K\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm xc}\left[\{n_\sigma(\mathbf{r})\}\right] \ . \tag{13}$$

2.2.2 Density functional, ground-state density, and ground-state energy

After the constrained search as a first step, we are led to the density functional that determines the ground-state energy and densities (Hohenberg-Kohn theorem [1])

$$\begin{aligned} D\left[\{n_\sigma(\mathbf{r})\}\right] &= \bar{F}\left[\{n_\sigma(\mathbf{r})\}\right] + U\left[\{n_\sigma(\mathbf{r})\}\right] + V_{\rm Har}\left[\{n_\sigma(\mathbf{r})\}\right] \\ &= K\left[\{n_\sigma(\mathbf{r})\}\right] + U\left[\{n_\sigma(\mathbf{r})\}\right] + V_{\rm Har}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm xc}\left[\{n_\sigma(\mathbf{r})\}\right] \end{aligned} \tag{14}$$

with the ionic and Hartree energy contributions

$$\begin{aligned} U\left[\{n_\sigma(\mathbf{r})\}\right] &= \sum_\sigma \int {\rm d}\mathbf{r}\, U(\mathbf{r})\, n_\sigma(\mathbf{r}) \ , \\ V_{\rm Har}\left[\{n_\sigma(\mathbf{r})\}\right] &= \sum_{\sigma,\sigma'} \int {\rm d}\mathbf{r} \int {\rm d}\mathbf{r}'\, V(\mathbf{r}-\mathbf{r}')\, n_\sigma(\mathbf{r})\, n_{\sigma'}(\mathbf{r}') \ . \end{aligned} \tag{15}$$

According to the Ritz variational principle, the ground-state energy E_0 is found from the minimization of this functional over the densities $n_\sigma(\mathbf{r})$,

$$E_0 = {\rm Min}_{\{n_\sigma(\mathbf{r})\}} D\left[\{n_\sigma(\mathbf{r})\}\right] \ . \tag{16}$$

The ground-state densities $n_\sigma^0(\mathbf{r})$ are those where the minimum of $D\left[\{n_\sigma(\mathbf{r})\}\right]$ is obtained.

2.3 Single-particle reference system

We consider the subset of single-particle product states $|\Phi^{(n)}\rangle$ that are normalized to unity, $\langle\Phi^{(n)}|\Phi^{(n)}\rangle = 1$. As before, the upper index indicates that they all lead to the same (physical) single-particle densities $n_\sigma^{\rm sp}(\mathbf{r})$,

$$n_\sigma^{\rm sp}(\mathbf{r}) = \langle\Phi^{(n)}|\hat{\Psi}_\sigma^\dagger(\mathbf{r})\hat{\Psi}_\sigma(\mathbf{r})|\Phi^{(n)}\rangle \ . \tag{17}$$

As our single-particle Hamiltonian we consider the kinetic-energy operator $\hat{H}_{\rm kin}$, see eq. (6). For fixed single-particle densities $n_\sigma^{\rm sp}(\mathbf{r})$ we define the single-particle kinetic-energy functional

$$F_{\rm sp}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}, \{|\Phi^{(n)}\rangle\}\right] = \langle\Phi^{(n)}|\hat{H}_{\rm kin}|\Phi^{(n)}\rangle \ . \tag{18}$$

2.3.1 Constrained search

As in Sect. 2.2, we carry out a constrained search in the subset of states $|\Phi^{(n)}\rangle$. The task is the minimization of the kinetic-energy functional $F_{\rm sp}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}, \{|\Phi^{(n)}\rangle\}\right]$. We denote the optimized single-particle product state by $|\Phi_0^{(n)}\rangle$ so that we find the density functional for the kinetic energy as

$$\bar{F}_{\rm sp}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right] = \langle\Phi_0^{(n)}|\hat{H}_{\rm kin}|\Phi_0^{(n)}\rangle \equiv K_{\rm sp}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right] \ . \tag{19}$$

2.3.2 Single-particle density functional

As the density functional $D_{\rm sp}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right]$ that corresponds to the single-particle problem we define

$$D_{\rm sp}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right] = K_{\rm sp}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right] + U\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right] + V_{\rm Har}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right] + E_{\rm sp,xc}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right] \tag{20}$$

with the kinetic energy term from (19), the contributions from the external potential and the Hartree terms $U\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right]$ and $V_{\rm Har}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right]$ from eq. (15), and a single-particle exchange-correlation potential $E_{\rm sp,xc}\left[\{n_\sigma^{\rm sp}(\mathbf{r})\}\right]$ that we will specify later. The functional (20) defines our single-particle reference system.

2.3.3 Noninteracting V-representability

In order to link the many-particle and single-particle approaches we make the assumption of non-interacting V-representability [1]: For any given (physical) density $n_\sigma(\mathbf{r})$ we can find a subset of normalized single-particle product states $|\Phi^{(n)}\rangle$ with N electrons such that

$$n_\sigma^{\rm sp}(\mathbf{r}) = n_\sigma(\mathbf{r}) \ . \tag{21}$$

Moreover, we demand that the density functionals $D\left[\{n_\sigma(\mathbf{r})\}\right]$ (14) for the interacting electrons and $D_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right]$ (20) for the single-particle problem agree with each other [15],

$$D_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right] = D\left[\{n_\sigma(\mathbf{r})\}\right] \ . \tag{22}$$

Then, the single-particle problem leads to the same ground-state density $n_\sigma^0(\mathbf{r})$ and ground-state energy E_0 as the interacting-particle Hamiltonian because the density variation is done with the same density functional (Kohn-Sham theorem) [1].
The condition (22) is equivalent to

$$K_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm sp,xc}\left[\{n_\sigma(\mathbf{r})\}\right] = K\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm xc}\left[\{n_\sigma(\mathbf{r})\}\right] \tag{23}$$

because the interaction with the external potential and the Hartree term only depend on the densities that are presumed equal for the interacting problem and the non-interacting reference system, see eq. (21). Eq. (23) then leads to an exact expression for the single-particle exchange-correlation energy

$$E_{\rm sp,xc}\left[\{n_\sigma(\mathbf{r})\}\right] \;=\; K\left[\{n_\sigma(\mathbf{r})\}\right] - K_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm xc}\left[\{n_\sigma(\mathbf{r})\}\right] \;. \tag{24}$$

This is our defining equation for $E_{\rm sp,xc}\left[\{n_\sigma(\mathbf{r})\}\right]$ in eq. (20).

2.4 Kohn-Sham Hamiltonian

In the following we address the single-particle energy functional directly, i.e., the Ritz variational problem without a prior constrained search,

$$E\left[\{n_\sigma(\mathbf{r})\}, \{|\Phi\rangle\}\right] = \langle\Phi|\hat{H}_{\rm kin}|\Phi\rangle + U\left[\{n_\sigma(\mathbf{r})\}\right] + V_{\rm Har}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm sp,xc}\left[\{n_\sigma(\mathbf{r})\}\right] \;. \tag{25}$$

For the extension to the Gutzwiller Density Functional Theory in Sect. 3, we expand the field operators in a basis,

$$\hat{\Psi}_\sigma(\mathbf{r}) = \sum_i \langle\mathbf{r}|i,\sigma\rangle \hat{c}_{i,\sigma} \quad , \quad \hat{\Psi}_\sigma^\dagger(\mathbf{r}) = \sum_i \hat{c}_{i,\sigma}^\dagger \langle i,\sigma|\mathbf{r}\rangle \;, \tag{26}$$

where the index i represents a combination of site (or crystal momentum) index and an orbital index. For a canonical basis we must have completeness and orthogonality,

$$\sum_{i,\sigma} |i,\sigma\rangle\langle i,\sigma| = \hat{1} \quad , \quad \langle i,\sigma|j,\sigma'\rangle = \delta_{i,j}\delta_{\sigma,\sigma'} \;. \tag{27}$$

When we insert (26) into (6), we obtain the operator for the kinetic energy in a general single-particle basis,

$$\hat{H}_{\rm kin} = \sum_{i,j,\sigma} T_{i,j;\sigma} \hat{c}_{i,\sigma}^\dagger \hat{c}_{j,\sigma} \;, \tag{28}$$

where the elements of the kinetic-energy matrix $\widetilde{T}_\sigma$ are given by

$$T_{i,j;\sigma} = \int {\rm d}\mathbf{r}\; \xi_{i,\sigma}^*(\mathbf{r}) \left(-\frac{\Delta_{\mathbf{r}}}{2m}\right) \xi_{j,\sigma}(\mathbf{r}) \;, \tag{29}$$

with $\xi_{i,\sigma}(\mathbf{r}) = \langle\mathbf{r}|i,\sigma\rangle$.

2.4.1 Energy functional

We introduce the single-particle density matrix $\tilde{\rho}$. Its elements in the general single-particle basis read

$$\rho_{(i,\sigma),(j,\sigma)} = \langle \Phi | \hat{c}^\dagger_{j,\sigma} \hat{c}_{i,\sigma} | \Phi \rangle \equiv \rho_{i,j;\sigma} \; . \tag{30}$$

Then, the densities are given by

$$n_\sigma(\mathbf{r}) = \sum_{i,j} \xi^*_{i,\sigma}(\mathbf{r}) \xi_{j,\sigma}(\mathbf{r}) \rho_{j,i;\sigma} \; . \tag{31}$$

Using these definitions, we can write the energy functional in the form

$$E\left[\{n_\sigma(\mathbf{r})\}, \tilde{\rho}\right] = \sum_{i,j} \sum_\sigma T_{i,j;\sigma} \rho_{j,i;\sigma} + U\left[\{n_\sigma(\mathbf{r})\}\right] + V_{\mathrm{Har}}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\mathrm{sp,xc}}\left[\{n_\sigma(\mathbf{r})\}\right] \; . \tag{32}$$

The fact that $|\Phi\rangle$ are normalized single-particle product states is encoded in the matrix relation

$$\tilde{\rho} \cdot \tilde{\rho} = \tilde{\rho} \; . \tag{33}$$

This is readily proven by using a unitary transformation between the operators $\hat{c}_{i,\sigma}$ and the single-particle operators $\hat{b}_{k,\sigma}$ that generate $|\Phi\rangle$, see appendix A.1.1.
When we minimize $E\left[\{n_\sigma(\mathbf{r})\}, \tilde{\rho}\right]$ with respect to $\tilde{\rho}$ we must take the condition (33) into account using a matrix $\widetilde{\Omega}$ of Lagrange multipliers $\Omega_{l,m;\sigma}$. Moreover, we use the Lagrange multipliers $\kappa_\sigma(\mathbf{r})$ to ensure eq. (31). Altogether we address $G_{\mathrm{DFT}} \equiv G_{\mathrm{DFT}}\left[\tilde{\rho}, \widetilde{\Omega}, \{n_\sigma(\mathbf{r})\}, \{\kappa_\sigma(\mathbf{r})\}\right]$

$$\begin{aligned} G_{\mathrm{DFT}} &= E\left[\{n_\sigma(\mathbf{r})\}, \tilde{\rho}\right] - \sum_{l,m,\sigma} \Omega_{l,m;\sigma} \left(\sum_p \rho_{l,p;\sigma} \rho_{p,m;\sigma} - \rho_{l,m;\sigma} \right) \\ &\quad - \sum_\sigma \int \mathrm{d}\mathbf{r} \; \kappa_\sigma(\mathbf{r}) \left(n_\sigma(\mathbf{r}) - \sum_{i,j} \xi^*_{i,\sigma}(\mathbf{r}) \xi_{j,\sigma}(\mathbf{r}) \rho_{j,i;\sigma} \right) . \end{aligned} \tag{34}$$

2.4.2 Minimization

When we minimize G_{DFT} in eq. (34) with respect to $n_\sigma(\mathbf{r})$ we find

$$\kappa_\sigma(\mathbf{r}) = U(\mathbf{r}) + V_{\mathrm{Har}}(\mathbf{r}) + v_{\mathrm{sp,xc},\sigma}(\mathbf{r}) \; , \tag{35}$$

$$V_{\mathrm{Har}}(\mathbf{r}) \equiv \sum_{\sigma'} \int \mathrm{d}\mathbf{r}' \, 2V(\mathbf{r} - \mathbf{r}') n^0_{\sigma'}(\mathbf{r}') \; , \tag{36}$$

$$\begin{aligned} v_{\mathrm{sp,xc},\sigma}(\mathbf{r}) &\equiv \left. \frac{\partial E_{\mathrm{sp,xc}}\left[\{n_{\sigma'}(\mathbf{r}')\}\right]}{\partial n_\sigma(\mathbf{r})} \right|_{n_\sigma(\mathbf{r}) = n^0_\sigma(\mathbf{r})} \\ &= \left. \frac{\partial \Big[K\left[\{n_{\sigma'}(\mathbf{r}')\}\right] - K_{\mathrm{sp}}\left[\{n_{\sigma'}(\mathbf{r}')\}\right] + E_{\mathrm{xc}}\left[\{n_{\sigma'}(\mathbf{r}')\}\right] \Big]}{\partial n_\sigma(\mathbf{r})} \right|_{n_\sigma(\mathbf{r}) = n^0_\sigma(\mathbf{r})} , \end{aligned} \tag{37}$$

where $V_{\mathrm{Har}}(\mathbf{r})$ is the Hartree interaction and $v_{\mathrm{sp,xc},\sigma}(\mathbf{r})$ is the single-particle exchange-correlation potential.

The minimization with respect to $\tilde{\rho}$ is outlined in appendix A.1.2, see also Ref. [16]. It leads to the Kohn-Sham single-particle Hamiltonian

$$\hat{H}^{\rm KS} = \sum_{i,j,\sigma} T^{\rm KS}_{i,j;\sigma} \hat{c}^\dagger_{i,\sigma} \hat{c}_{j,\sigma} \ , \tag{38}$$

where the elements of the Kohn-Sham Hamilton matrix $\widetilde{T}^{\rm KS}_\sigma$ are given by

$$T^{\rm KS}_{i,j;\sigma} = \frac{\partial E\left[\{n_\sigma(\mathbf{r})\}, \tilde{\rho}\right]}{\partial \rho_{j,i;\sigma}} + \int {\rm d}\mathbf{r}\ \kappa_\sigma(\mathbf{r}) \xi^*_{i,\sigma}(\mathbf{r}) \xi_{j,\sigma}(\mathbf{r}) \ . \tag{39}$$

Explicitly,

$$\begin{aligned} T^{\rm KS}_{i,j;\sigma} &= \int {\rm d}\mathbf{r}\ \xi^*_{i,\sigma}(\mathbf{r}) h^{\rm KS}_\sigma(\mathbf{r}) \xi_{j,\sigma}(\mathbf{r}) \ , & (40)\\ h^{\rm KS}_\sigma(\mathbf{r}) &\equiv -\frac{\Delta_{\mathbf{r}}}{2m} + V^{\rm KS}_\sigma(\mathbf{r}) \ , & (41)\\ V^{\rm KS}_\sigma(\mathbf{r}) &\equiv \kappa_\sigma(\mathbf{r}) = U(\mathbf{r}) + V_{\rm Har}(\mathbf{r}) + v_{{\rm sp,xc},\sigma}(\mathbf{r}) \ . & (42) \end{aligned}$$

Here, we defined the 'Kohn-Sham potential' $V^{\rm KS}_\sigma(\mathbf{r})$ that, in our derivation, is identical to the Lagrange parameter $\kappa_\sigma(\mathbf{r})$.

The remaining task is to find the basis in which the Kohn-Sham matrix $\widetilde{T}^{\rm KS}_\sigma$ is diagonal, for a translationally invariant system, see appendix A.1.3. This gives the dispersion $\varepsilon_{\rm n}(\mathbf{k})$ of the Kohn-Sham quasi-particles.

3 DFT for many-particle reference systems

The Kohn-Sham potential (37) cannot be calculated exactly because the functionals in eq. (24) are not known. Therefore, assumptions must be made about the form of the single-particle exchange-correlation potential, e.g., the Local Density Approximation [1]. Unfortunately, such approximations are often not satisfactory, e.g., for transition metals. Consequently, more sophisticated many-electron approaches must be employed to improve the Kohn-Sham approach.

3.1 Hubbard Hamiltonian and Hubbard density functional

3.1.1 Multi-band Hubbard model

A better description of transition metals and their compounds can be achieved by supplementing the single-particle reference system resulting from $\hat{H}_{\rm kin}$ in Sect. 2.3 by a multi-band Hubbard interaction. Then, our multi-band reference system follows from

$$\hat{H}_{\rm H} = \hat{H}_{\rm kin} + \hat{V}_{\rm loc} - \hat{V}_{\rm dc} \ , \tag{43}$$

where $\hat{V}_{\rm loc}$ describes local interactions between electrons in Wannier orbitals on the same site $\mathbf{R}$. The local single-particle operator $\hat{V}_{\rm dc}$ accounts for the double counting of their interactions in

the Hubbard term $\hat{V}_{\rm loc}$ and in the single-particle exchange-correlation energy $E_{\rm sp,xc}$. We assume that $\hat{V}_{\rm loc}$ and $\hat{V}_{\rm dc}$ do not depend on the densities $n_\sigma(\mathbf{r})$ explicitly.
For the local interaction we set

$$\begin{aligned} \hat{V}_{\rm loc} &= \sum_{\mathbf{R}} \hat{V}_{\rm loc}(\mathbf{R}) \ , \\ \hat{V}_{\rm loc}(\mathbf{R}) &= \frac{1}{2} \sum_{(c_1,\sigma_1),\ldots,(c_4,\sigma_4)} U^{(c_1,\sigma_1),(c_2,\sigma_2)}_{(c_3,\sigma_3),(c_4,\sigma_4)} \, \hat{c}^\dagger_{\mathbf{R},c_1,\sigma_1} \hat{c}^\dagger_{\mathbf{R},c_2,\sigma_2} \hat{c}_{\mathbf{R},c_3,\sigma_3} \hat{c}_{\mathbf{R},c_4,\sigma_4} \ . \end{aligned} \tag{44}$$

Note that only electrons in the small subset of correlated orbitals (index c) experience the two-particle interaction $\hat{V}_{\rm loc}$: When there are two electrons in the Wannier orbitals $\phi_{\mathbf{R},c_3,\sigma_3}(\mathbf{r})$ and $\phi_{\mathbf{R},c_4,\sigma_4}(\mathbf{r})$ centered around the lattice site $\mathbf{R}$, they are scattered into the orbitals $\phi_{\mathbf{R},c_1,\sigma_1}(\mathbf{r})$ and $\phi_{\mathbf{R},c_2,\sigma_2}(\mathbf{r})$, centered around the same lattice site $\mathbf{R}$; for the definition of basis states, see appendix A.1.3. Typically, we consider $c = 3d$ for the transition metals and their compounds.
The interaction strengths are parameters of the theory; for a comprehensive symmetry analysis, see Ref. [17]. Later, we shall employ the spherical approximation so that $U^{\cdots}_{\cdots}$ for d-electrons can be expressed in terms of three Racah parameters A, B, and C. Fixing C/B makes it possible to introduce an effective Hubbard parameter U and an effective Hund's-rule coupling J, see Sect. 3.4 and appendix A.2. Due to screening, the effective Hubbard interaction U is smaller than its bare, atomic value. In general, U and J are chosen to obtain good agreement with experiment, see Sect. 4.

3.1.2 Hubbard density functional

According to Levy's constrained search, we must find the minimum of the functional

$$F_{\rm H}\left[\{n_\sigma(\mathbf{r})\}, \left\{|\Psi^{(n)}\rangle\right\}\right] = \langle \Psi^{(n)}|\hat{H}_{\rm H}|\Psi^{(n)}\rangle \tag{45}$$

in the subset of normalized states with given (physical) density $n_\sigma(\mathbf{r})$, see eq. (4). The minimum of $F_{\rm H}\left[\{n_\sigma(\mathbf{r})\}, \left\{|\Psi^{(n)}\rangle\right\}\right]$ over the states $|\Psi^{(n)}\rangle$ is the ground state $|\Psi^{(n)}_{\rm H,0}\rangle$ of the Hamiltonian $\hat{H}_{\rm H}$ for fixed densities $n_\sigma(\mathbf{r})$. In analogy to Sect. 2.3, we define the Hubbard density functional

$$\begin{aligned} D_{\rm H}\left[\{n_\sigma(\mathbf{r})\}\right] &= K_{\rm H}\left[\{n_\sigma(\mathbf{r})\}\right] + U\left[\{n_\sigma(\mathbf{r})\}\right] + V_{\rm Har}\left[\{n_\sigma(\mathbf{r})\}\right] \\ &\quad + V_{\rm loc}\left[\{n_\sigma(\mathbf{r})\}\right] - V_{\rm dc}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm H,xc}\left[\{n_\sigma(\mathbf{r})\}\right] \ , \end{aligned} \tag{46}$$

where

$$K_{\rm H}\left[\{n_\sigma(\mathbf{r})\}\right] = \langle \Psi^{(n)}_{\rm H,0}|\hat{H}_{\rm kin}|\Psi^{(n)}_{\rm H,0}\rangle \ , \quad V_{\rm loc/dc}\left[\{n_\sigma(\mathbf{r})\}\right] = \langle \Psi^{(n)}_{\rm H,0}|\hat{V}_{\rm loc/dc}|\Psi^{(n)}_{\rm H,0}\rangle \ , \tag{47}$$

and $E_{\rm H,xc}\left[\{n_\sigma(\mathbf{r})\}\right]$ is the exchange-correlation energy for $\hat{H}_{\rm H}$. As in Sect. 2.3, the Hubbard density functional agrees with the exact density functional if we choose

$$\begin{aligned} E_{\rm H,xc}\left[\{n_\sigma(\mathbf{r})\}\right] &= K\left[\{n_\sigma(\mathbf{r})\}\right] - K_{\rm H}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\rm xc}\left[\{n_\sigma(\mathbf{r})\}\right] \\ &\quad - \left(V_{\rm loc}\left[\{n_\sigma(\mathbf{r})\}\right] - V_{\rm dc}\left[\{n_\sigma(\mathbf{r})\}\right]\right) \ . \end{aligned} \tag{48}$$

Then, the Hubbard approach provides the exact ground-state densities and ground-state energy of our full many-particle Hamiltonian (Hohenberg-Kohn–Hubbard theorem). Of course, our derivation relies on the assumption of Hubbard V-representability of the densities $n_\sigma(\mathbf{r})$.

3.1.3 Hubbard single-particle potential

When we directly apply Ritz' principle, we have to minimize the energy functional

$$E\left[\{n_\sigma(\mathbf{r})\},\{|\Psi\rangle\}\right]=\langle\Psi|\hat{H}_{\rm H}|\Psi\rangle+U\left[\{n_\sigma(\mathbf{r})\}\right]+V_{\rm Har}\left[\{n_\sigma(\mathbf{r})\}\right]+E_{\rm H,xc}\left[\{n_\sigma(\mathbf{r})\}\right]. \quad (49)$$

We include the constraints eq. (4) and the normalization condition using the Lagrange parameters $\kappa_\sigma(\mathbf{r})$ and E_0 in the functional $G_{\rm H}\equiv G_{\rm H}\left[\{|\Psi\rangle\},\{n_\sigma(\mathbf{r})\},\{\kappa_\sigma(\mathbf{r})\},E_0\right]$

$$G_{\rm H}=E\left[\{n_\sigma(\mathbf{r})\},\{|\Psi\rangle\}\right]-E_0\left(\langle\Psi|\Psi\rangle-1\right)-\sum_\sigma\int{\rm d}\mathbf{r}\;\kappa_\sigma(\mathbf{r})\left(n_\sigma(\mathbf{r})-\langle\Psi|\hat{\Psi}^\dagger_\sigma(\mathbf{r})\hat{\Psi}_\sigma(\mathbf{r})|\Psi\rangle\right)\;. \quad (50)$$

As in Sect. 2.4, see eqs. (35) and (42), the variation of $G_{\rm H}$ with respect to $n_\sigma(\mathbf{r})$ gives the single-particle potential

$$\begin{aligned} V^{\rm H}_\sigma(\mathbf{r}) &\equiv U(\mathbf{r})+V_{\rm Har}(\mathbf{r})+v_{{\rm H,xc},\sigma}(\mathbf{r})\;,\\ v_{{\rm H,xc},\sigma}(\mathbf{r}) &\equiv \left.\frac{\partial E_{\rm H,xc}\left[\{n_{\sigma'}(\mathbf{r}')\}\right]}{\partial n_\sigma(\mathbf{r})}\right|_{n_\sigma(\mathbf{r})=n^0_\sigma(\mathbf{r})}\;. \end{aligned} \quad (51)$$

The Hubbard-model approach is based on the idea that typical approximations for the exchange-correlation energy, e.g., the local-density approximation, are better suited for the Hubbard model than for the Kohn-Sham approach,

$$E_{\rm H,xc}\left[\{n_\sigma(\mathbf{r})\}\right]\approx E_{\rm LDA,xc}\left[\{n_\sigma(\mathbf{r})\}\right]\;. \quad (52)$$

Indeed, as seen from eq. (48), in the Hubbard exchange-correlation energy $E_{\rm H,xc}$ the exchange-correlation contributions in the exact $E_{\rm xc}$ are reduced by the Hubbard term $V_{\rm loc}\left[\{n_\sigma(\mathbf{r})\}\right]-V_{\rm dc}\left[\{n_\sigma(\mathbf{r})\}\right]$, reflecting a more elaborate treatment of local correlations.

Unfortunately, the minimization of (49) with respect to $|\Psi\rangle$ constitutes an unsolvable many-particle problem. Indeed, the ground state $|\Psi_0\rangle$ is the solution of the many-particle Schrödinger equation with energy E_0,

$$\left(\hat{H}_0+\hat{V}_{\rm loc}-\hat{V}_{\rm dc}\right)|\Psi_0\rangle=E_0|\Psi_0\rangle \quad (53)$$

with the single-particle Hamiltonian

$$\hat{H}_0=\sum_\sigma\int{\rm d}\mathbf{r}\;\hat{\Psi}^\dagger_\sigma(\mathbf{r})\left(-\frac{\Delta_{\mathbf{r}}}{2m}+U(\mathbf{r})+V_{\rm Har}(\mathbf{r})+v_{{\rm H,xc},\sigma}(\mathbf{r})\right)\hat{\Psi}_\sigma(\mathbf{r})\;. \quad (54)$$

The 'Kohn-Sham–Hubbard equations' (53) can be used as starting point for further approximations, for example, the Dynamical Mean-Field Theory (DMFT). In the following we will address the functional in eq. (49) directly.

3.2 Gutzwiller density functional

In the widely used LDA+U approach [5], the functional in eq. (49) is evaluated and (approximately) minimized by means of single-particle product wave functions. However, this approach treats correlations only on a mean-field level. In the more sophisticated Gutzwiller approach, we consider the functional in eq. (49) in the subset of Gutzwiller-correlated variational many-particle states.

3.2.1 Gutzwiller variational ground state

In order to formulate the Gutzwiller variational ground state [4,8], we consider the atomic states $|\Gamma\rangle_{\mathbf{R}}$ that are built from the correlated orbitals. The local Hamiltonians take the form

$$\hat{V}_{\mathrm{loc/dc}}(\mathbf{R}) = \sum_{\Gamma,\Gamma'} E^{\mathrm{loc/dc}}_{\Gamma,\Gamma'}(\mathbf{R}) |\Gamma\rangle_{\mathbf{RR}}\langle\Gamma'| = \sum_{\Gamma,\Gamma'} E^{\mathrm{loc/dc}}_{\Gamma,\Gamma'}(\mathbf{R}) \hat{m}_{\mathbf{R};\Gamma,\Gamma'} \ , \tag{55}$$

where $|\Gamma\rangle_{\mathbf{R}}$ contains $|\Gamma_{\mathbf{R}}|$ electrons. Here, we introduced

$$E^{\mathrm{loc/dc}}_{\Gamma,\Gamma'}(\mathbf{R}) = {}_{\mathbf{R}}\langle\Gamma|\hat{V}_{\mathrm{loc/dc}}(\mathbf{R})|\Gamma'\rangle_{\mathbf{R}} \tag{56}$$

and the local many-particle operators $\hat{m}_{\mathbf{R};\Gamma,\Gamma'} = |\Gamma\rangle_{\mathbf{RR}}\langle\Gamma'|$.
The Gutzwiller correlator and the Gutzwiller variational states are defined as

$$\hat{P}_{\mathrm{G}} = \prod_{\mathbf{R}} \sum_{\Gamma,\Gamma'} \lambda_{\Gamma,\Gamma'}(\mathbf{R}) \hat{m}_{\mathbf{R};\Gamma,\Gamma'} \quad , \quad |\Psi_{\mathrm{G}}\rangle = \hat{P}_{\mathrm{G}}|\Phi\rangle \ . \tag{57}$$

Here, $|\Phi\rangle$ is a single-particle product state, and $\lambda_{\Gamma,\Gamma'}(\mathbf{R})$ defines the matrix $\tilde{\lambda}(\mathbf{R})$ of, in general, complex variational parameters.

3.2.2 Gutzwiller functionals

We evaluate the energy functional (49) in the restricted subset of Gutzwiller variational states,

$$\begin{aligned} E\left[\{n_\sigma(\mathbf{r})\}, \{|\Psi_{\mathrm{G}}\rangle\}\right] &= \sum_{\mathbf{R},b,\mathbf{R}',b',\sigma} T_{(\mathbf{R},b),(\mathbf{R}',b');\sigma} \rho^{\mathrm{G}}_{(\mathbf{R}',b'),(\mathbf{R},b);\sigma} + V^{\mathrm{G}}_{\mathrm{loc}} - V^{\mathrm{G}}_{\mathrm{dc}} \\ &\quad + U\left[\{n_\sigma(\mathbf{r})\}\right] + V_{\mathrm{Har}}\left[\{n_\sigma(\mathbf{r})\}\right] + E_{\mathrm{H,xc}}\left[\{n_\sigma(\mathbf{r})\}\right], \\ V^{\mathrm{G}}_{\mathrm{loc/dc}} &= \sum_{\mathbf{R}} \sum_{\Gamma,\Gamma'} E^{\mathrm{loc/dc}}_{\Gamma,\Gamma'}(\mathbf{R}) m^{\mathrm{G}}_{\mathbf{R};\Gamma,\Gamma'} \ . \end{aligned} \tag{58}$$

Note that we work with the orbital Wannier basis, see appendix A.1.3,

$$T_{(\mathbf{R},b),(\mathbf{R}',b');\sigma} = \int \mathrm{d}\mathbf{r} \ \phi^*_{\mathbf{R},b,\sigma}(\mathbf{r}) \left(-\frac{\Delta_{\mathbf{r}}}{2m}\right) \phi_{\mathbf{R}',b',\sigma}(\mathbf{r}) \ . \tag{59}$$

The elements of the Gutzwiller-correlated single-particle density matrix are

$$\rho^{\mathrm{G}}_{(\mathbf{R}',b'),(\mathbf{R},b);\sigma} = \frac{\langle\Psi_{\mathrm{G}}|\hat{c}^\dagger_{\mathbf{R},b,\sigma}\hat{c}_{\mathbf{R}',b',\sigma}|\Psi_{\mathrm{G}}\rangle}{\langle\Psi_{\mathrm{G}}|\Psi_{\mathrm{G}}\rangle} = \frac{\langle\Phi|\hat{P}^\dagger_{\mathrm{G}}\hat{c}^\dagger_{\mathbf{R},b,\sigma}\hat{c}_{\mathbf{R}',b',\sigma}\hat{P}_{\mathrm{G}}|\Phi\rangle}{\langle\Phi|\hat{P}^\dagger_{\mathrm{G}}\hat{P}_{\mathrm{G}}|\Phi\rangle} \ , \tag{60}$$

and the densities become

$$n_\sigma(\mathbf{r}) = \sum_{\mathbf{R},b,\mathbf{R}',b'} \phi^*_{\mathbf{R},b,\sigma}(\mathbf{r}) \phi_{\mathbf{R}',b',\sigma}(\mathbf{r}) \rho^{\mathrm{G}}_{(\mathbf{R}',b'),(\mathbf{R},b);\sigma} \ . \tag{61}$$

The expectation values for the atomic operators are given by

$$m^{\mathrm{G}}_{\mathbf{R};\Gamma,\Gamma'} = \frac{\langle\Psi_{\mathrm{G}}|\hat{m}_{\mathbf{R};\Gamma,\Gamma'}|\Psi_{\mathrm{G}}\rangle}{\langle\Psi_{\mathrm{G}}|\Psi_{\mathrm{G}}\rangle} = \frac{\langle\Phi|\hat{P}^\dagger_{\mathrm{G}}\hat{m}_{\mathbf{R};\Gamma,\Gamma'}\hat{P}_{\mathrm{G}}|\Phi\rangle}{\langle\Phi|\hat{P}^\dagger_{\mathrm{G}}\hat{P}_{\mathrm{G}}|\Phi\rangle} \ . \tag{62}$$

The diagrammatic evaluation of $\rho^{\rm G}_{(\mathbf{R}',b'),(\mathbf{R},b);\sigma}$ and of $m^{\rm G}_{\mathbf{R};\Gamma,\Gamma'}$ shows that these quantities are functionals of the non-interacting single-particle density matrices $\tilde{\rho}$, see eq. (30), and of the variational parameters $\lambda_{\Gamma,\Gamma'}(\mathbf{R})$. Moreover, it turns out that the local, non-interacting single-particle density matrix $\tilde{C}(\mathbf{R})$ with the elements

$$C_{b,b';\sigma}(\mathbf{R}) \equiv \rho_{(\mathbf{R},b),(\mathbf{R},b');\sigma} \tag{63}$$

plays a prominent role in the Gutzwiller energy functional, in particular, for infinite lattice coordination number. Therefore, we may write

$$E\left[\{n_\sigma(\mathbf{r})\}, \{|\Psi_{\rm G}\rangle\}\right] \equiv E^{\rm G}\left[\tilde{\rho}, \left\{\tilde{\lambda}(\mathbf{R})\right\}, \{n_\sigma(\mathbf{r})\}, \left\{\tilde{C}(\mathbf{R})\right\}\right] . \tag{64}$$

In the Lagrange functional we shall impose the relation (63) with the help of the Hermitian Lagrange parameter matrix $\tilde{\eta}$ with entries $\eta_{(\mathbf{R},b),(\mathbf{R},b');\sigma}$. Lastly, for the analytical evaluation of eq. (64) it is helpful to impose a set of (real-valued) local constraints ($l = 1, 2, \ldots, N_{\rm con}$)

$$g_{l,\mathbf{R}}\left[\tilde{\lambda}(\mathbf{R}), \tilde{C}(\mathbf{R})\right] = 0 , \tag{65}$$

which we implement with real Lagrange parameters $\Lambda_l(\mathbf{R})$; explicit expressions are given in eqs. (74) and (75).

In the following, we abbreviate $i = (\mathbf{R}, b)$ and $j = (\mathbf{R}', b')$. Consequently, in analogy with Sect. 2.4, we address

$$G^{\rm G}_{\rm DFT} \equiv G^{\rm G}_{\rm DFT}\begin{bmatrix}\tilde{\rho}, & \{n_\sigma(\mathbf{r})\}, \left\{\tilde{C}(\mathbf{R})\right\}, \left\{\tilde{\lambda}(\mathbf{R})\right\} \\ \widetilde{\Omega}, \{\kappa_\sigma(\mathbf{r})\}, \{\tilde{\eta}(\mathbf{R})\}, & \{\Lambda_l(\mathbf{R})\}\end{bmatrix} \tag{66}$$

as our Lagrange functional,

$$\begin{aligned} G^{\rm G}_{\rm DFT} \;=\; & E^{\rm G}\left[\tilde{\rho}, \left\{\tilde{\lambda}(\mathbf{R})\right\}, \{n_\sigma(\mathbf{r})\}, \left\{\tilde{C}(\mathbf{R})\right\}\right] - \sum_{l,m,\sigma} \Omega_{l,m;\sigma}\left(\tilde{\rho}\cdot\tilde{\rho}-\tilde{\rho}\right)_{m,l;\sigma} \\ & - \sum_\sigma \int {\rm d}\mathbf{r}\; \kappa_\sigma(\mathbf{r})\left(n_\sigma(\mathbf{r}) - \sum_{i,j}\phi^*_{i,\sigma}(\mathbf{r})\phi_{j,\sigma}(\mathbf{r})\rho^{\rm G}_{j,i;\sigma}\right) \\ & + \sum_{l,\mathbf{R}} \Lambda_l(\mathbf{R}) g_{l,\mathbf{R}} - \sum_{\mathbf{R},b,b',\sigma} \eta_{b,b';\sigma}(\mathbf{R})\left(C_{b',b;\sigma}(\mathbf{R}) - \rho_{(\mathbf{R},b'),(\mathbf{R},b);\sigma}\right) , \end{aligned} \tag{67}$$

cf. eq. (34). Here, we took the condition (61) into account using Lagrange parameters $\kappa_\sigma(\mathbf{r})$ because the external potential, the Hartree term and the exchange-correlation potential in eq. (58) depend on the densities.

3.2.3 Minimization of the Gutzwiller energy functional

The functional $G^{\rm G}_{\rm DFT}$ in eq. (67) has to be minimized with respect to $n_\sigma(\mathbf{r})$, $\tilde{C}(\mathbf{R})$, $\tilde{\lambda}(\mathbf{R})$, and $\tilde{\rho}$. The variation with respect to the Lagrange parameters $\kappa_\sigma(\mathbf{r})$, $\tilde{\eta}(\mathbf{R})$, $\Lambda_l(\mathbf{R})$, and $\widetilde{\Omega}$ gives the constraints (61), (63), (65), and (33), respectively.

1. As in the derivation of the exact Schrödinger equation (53), the variation of $G^{\mathrm{G}}_{\mathrm{DFT}}$ with respect to $n_\sigma(\mathbf{r})$ generates the single-particle potential,

$$\kappa_\sigma(\mathbf{r}) = V^{\mathrm{H}}_\sigma(\mathbf{r}) \;, \tag{68}$$

see eqs. (42) and (51).

2. The minimization with respect to $\tilde{C}(\mathbf{R})$ gives

$$\begin{aligned}\eta_{b,b';\sigma}(\mathbf{R}) &= \frac{\partial E^{\mathrm{G}}}{\partial C_{b',b;\sigma}(\mathbf{R})} + \sum_l \Lambda_l(\mathbf{R}) \frac{\partial g_{l,\mathbf{R}}}{\partial C_{b',b;\sigma}(\mathbf{R})} \\ &\quad + \sum_{i,j,\sigma'} \int \mathrm{d}\mathbf{r}\; V^{\mathrm{H}}_{\sigma'}(\mathbf{r}) \phi^*_{i,\sigma'}(\mathbf{r}) \phi_{j,\sigma'}(\mathbf{r}) \frac{\partial \rho^{\mathrm{G}}_{j,i;\sigma'}}{\partial C_{b',b;\sigma}(\mathbf{R})} \;.\end{aligned} \tag{69}$$

3. The minimization with respect to the Gutzwiller correlation parameters $\tilde{\lambda}(\mathbf{R})$ results in

$$\begin{aligned}0 &= \frac{\partial E^{\mathrm{G}}}{\partial \lambda_{\Gamma,\Gamma'}(\mathbf{R})} + \sum_{l,m,\sigma} \int \mathrm{d}\mathbf{r}\; V^{\mathrm{H}}_\sigma(\mathbf{r}) \phi^*_{l,\sigma}(\mathbf{r}) \phi_{m,\sigma}(\mathbf{r}) \frac{\partial \rho^{\mathrm{G}}_{m,l,\sigma}}{\partial \lambda_{\Gamma,\Gamma'}(\mathbf{R})} \\ &\quad + \sum_l \Lambda_l(\mathbf{R}) \frac{\partial g_{l,\mathbf{R}}}{\partial \lambda_{\Gamma,\Gamma'}(\mathbf{R})} \\ &= \sum_{l,m,\sigma} h^0_{l,m;\sigma} \frac{\partial \rho^{\mathrm{G}}_{m,l,\sigma}}{\partial \lambda_{\Gamma,\Gamma'}(\mathbf{R})} + \frac{\partial \left(V^{\mathrm{G}}_{\mathrm{loc}} - V^{\mathrm{G}}_{\mathrm{dc}}\right)}{\partial \lambda_{\Gamma,\Gamma'}(\mathbf{R})} + \sum_l \Lambda_l(\mathbf{R}) \frac{\partial g_{l,\mathbf{R}}}{\partial \lambda_{\Gamma,\Gamma'}(\mathbf{R})} \;,\end{aligned} \tag{70}$$

$$h^0_{l,m;\sigma} \equiv \int \mathrm{d}\mathbf{r}\; \phi^*_{l,\sigma}(\mathbf{r}) \Big(-\frac{\Delta_{\mathbf{r}}}{2m} + U(\mathbf{r}) + V_{\mathrm{Har}}(\mathbf{r}) + v_{\mathrm{H,xc},\sigma}(\mathbf{r}) \Big) \phi_{m,\sigma}(\mathbf{r}) \tag{71}$$

for all $\lambda_{\Gamma,\Gamma'}(\mathbf{R})$. Note that, in the case of complex Gutzwiller parameters, we also have to minimize with respect to $(\lambda_{\Gamma,\Gamma'}(\mathbf{R}))^*$. Using these equations we calculate the Lagrange parameters $\Lambda_l(\mathbf{R})$ that are needed in eq. (69).

4. The minimization of $G^{\mathrm{G}}_{\mathrm{DFT}}$ with respect to $\tilde{\rho}$ generates the Landau–Gutzwiller quasi-particle Hamiltonian, see appendix A.1.2,

$$\hat{H}^{\mathrm{G}}_{\mathrm{qp}} = \sum_{i,j,\sigma} h^{\mathrm{G}}_{i,j;\sigma} \hat{c}^\dagger_{i,\sigma} \hat{c}_{j,\sigma} \tag{72}$$

with the entries

$$\begin{aligned}h^{\mathrm{G}}_{i,j;\sigma} &= \frac{\partial E^{\mathrm{G}}}{\partial \rho_{j,i;\sigma}} + \sum_{l,m,\sigma'} \int \mathrm{d}\mathbf{r}\; V^{\mathrm{H}}_{\sigma'}(\mathbf{r}) \phi^*_{l,\sigma'}(\mathbf{r}) \phi_{m,\sigma'}(\mathbf{r}) \frac{\partial \rho^{\mathrm{G}}_{m,l,\sigma'}}{\partial \rho_{j,i;\sigma}} + \sum_{\mathbf{R},b,b',\sigma'} \eta_{b,b';\sigma'}(\mathbf{R}) \frac{\partial \rho_{(\mathbf{R},b'),(\mathbf{R},b);\sigma'}}{\partial \rho_{j,i;\sigma}} \\ &= \sum_{l,m,\sigma'} h^0_{l,m;\sigma'} \frac{\partial \rho^{\mathrm{G}}_{m,l,\sigma'}}{\partial \rho_{j,i;\sigma}} + \frac{\partial \left(V^{\mathrm{G}}_{\mathrm{loc}} - V^{\mathrm{G}}_{\mathrm{dc}}\right)}{\partial \rho_{j,i;\sigma}} + \sum_{\mathbf{R},b,b'} \delta_{j,(\mathbf{R},b')} \delta_{i,(\mathbf{R},b)} \eta_{b,b';\sigma}(\mathbf{R}),\end{aligned} \tag{73}$$

where we used eqs. (58) and (71).

The single-particle state $|\Phi\rangle$ is the ground state of the Hamiltonian (72) from which the single-particle density matrix $\tilde{\rho}$ follows.

The minimization problem outlined in steps (i)–(iv) requires the evaluation of the energy $E^{\rm G}$ in eq. (58). In particular, the correlated single-particle density matrix $\tilde{\rho}^{\rm G}$, eq. (60), must be determined.

All equations derived in this section are completely general. They can, at least in principle, be evaluated by means of a diagrammatic expansion method [18–21]. The leading order of the expansion corresponds to an approximation-free evaluation of expectation values for Gutzwiller wave functions in the limit of large lattice coordination number. This limit, also known as "Gutzwiller Approximation", will be studied in the rest of this chapter.

3.3 Gutzwiller density functional for infinite lattice coordination number

For $Z \to \infty$, the Gutzwiller-correlated single-particle density matrix and the Gutzwiller probabilities for the local occupancies can be calculated explicitly without further approximations; for a formal proof, see Ref. [21]. In this section we make no symmetry assumptions (translational invariance, crystal symmetries). Note, however, that the equations do not cover the case of spin-orbit coupling; for the latter, see Refs. [22, 23].

3.3.1 Local constraints

As shown in Refs. [8, 24] it is convenient for the evaluation of Gutzwiller wave functions to impose the following (local) constraints

$$\sum_{\Gamma,\Gamma_1,\Gamma_2} \lambda^*_{\Gamma,\Gamma_1}(\mathbf{R})\lambda_{\Gamma,\Gamma_2}(\mathbf{R})\langle \hat{m}_{\mathbf{R};\Gamma_1,\Gamma_2}\rangle_\Phi = 1 \tag{74}$$

and

$$\sum_{\Gamma,\Gamma_1,\Gamma_2} \lambda^*_{\Gamma,\Gamma_1}(\mathbf{R})\lambda_{\Gamma,\Gamma_2}(\mathbf{R})\langle \hat{m}_{\mathbf{R};\Gamma_1,\Gamma_2}\hat{c}^\dagger_{\mathbf{R},b,\sigma}\hat{c}_{\mathbf{R},b',\sigma}\rangle_\Phi = \langle \hat{c}^\dagger_{\mathbf{R},b,\sigma}\hat{c}_{\mathbf{R},b',\sigma}\rangle_\Phi \ , \tag{75}$$

where we abbreviated $\langle \hat{A}\rangle_\Phi \equiv \langle \Phi|\hat{A}|\Phi\rangle$. Note that, for complex constraints, the index l in (65) labels real and imaginary parts separately.

3.3.2 Atomic occupancies

In the limit of infinite lattice coordination number, the interaction and double-counting energy can be expressed solely in terms of the local variational parameters $\tilde{\lambda}(\mathbf{R})$ and the local density matrix $\tilde{C}(\mathbf{R})$ of the correlated bands in $|\Phi\rangle$,

$$V^{\rm G}_{\rm loc/dc} = \sum_{\mathbf{R}} \sum_{\Gamma_1,\ldots,\Gamma_4} \lambda^*_{\Gamma_2,\Gamma_1}(\mathbf{R}) E^{\rm loc/dc}_{\Gamma_2,\Gamma_3}(\mathbf{R})\lambda_{\Gamma_3,\Gamma_4}(\mathbf{R})\langle \hat{m}_{\mathbf{R};\Gamma_1,\Gamma_4}\rangle_\Phi \ . \tag{76}$$

The remaining expectation values $\langle \hat{m}_{\mathbf{R};\Gamma_1,\Gamma_4}\rangle_\Phi$ are evaluated using Wick's theorem. Explicit expressions are given in Refs. [8, 25].

3.3.3 Correlated single-particle density matrix

The local part of the correlated single-particle density matrix is given by

$$\rho^{\mathrm{G}}_{(\mathbf{R},b'),(\mathbf{R},b);\sigma} = \sum_{\Gamma_1,\ldots,\Gamma_4} \lambda^{*}_{\Gamma_2,\Gamma_1}(\mathbf{R})\lambda_{\Gamma_3,\Gamma_4}(\mathbf{R})\langle \hat{m}_{\mathbf{R};\Gamma_1,\Gamma_2}\hat{c}^{\dagger}_{\mathbf{R},b,\sigma}\hat{c}_{\mathbf{R},b',\sigma}\hat{m}_{\mathbf{R};\Gamma_3,\Gamma_4}\rangle_{\Phi} \equiv C^{\mathrm{G}}_{b',b;\sigma}(\mathbf{R}). \quad (77)$$

It can be evaluated using Wick's theorem. As can be seen from eq. (77), it is a function of the variational parameters $\lambda_{\Gamma,\Gamma'}(\mathbf{R})$ and of the local non-interacting single-particle density matrix $\tilde{C}(\mathbf{R})$.
For $\mathbf{R} \neq \mathbf{R}'$, we have for the correlated single-particle density matrix

$$\rho^{\mathrm{G}}_{(\mathbf{R}',b'),(\mathbf{R},b);\sigma} = \sum_{a,a'} q^{a,\sigma}_{b,\sigma}(\mathbf{R}) \left(q^{a',\sigma}_{b',\sigma}(\mathbf{R}')\right)^{*} \rho_{(\mathbf{R}',a'),(\mathbf{R},a);\sigma} \quad (78)$$

with the orbital-dependent renormalization factors $q^{a,\sigma}_{b,\sigma}(\mathbf{R})$ for the electron transfer between different sites. Explicit expressions in terms of the variational parameters $\tilde{\lambda}(\mathbf{R})$ and of the local non-interacting single-particle density matrix $\tilde{C}(\mathbf{R})$ are given in Refs. [8, 25].

3.3.4 Implementation for translational invariant systems

For a system that is invariant under translation by a lattice vector and contains only one atom per unit cell, all local quantities become independent of the site index, e.g., $\lambda_{\Gamma,\Gamma'}(\mathbf{R}) \equiv \lambda_{\Gamma,\Gamma'}$ for the Gutzwiller variational parameters. Since $\mathbf{k}$ from the first Brillouin zone is a good quantum number, we work with the (orbital) Bloch basis, see appendix A.1.3.
The minimization of the energy functional with respect to the single-particle density matrix leads to the Gutzwiller–Kohn-Sham Hamiltonian. In the orbital Bloch basis $\phi_{\mathbf{k},b,\sigma}(\mathbf{r})$, see appendix A.1.3, the corresponding quasi-particle Hamiltonian reads,

$$\hat{H}^{\mathrm{G}}_{\mathrm{qp}} = \sum_{\mathbf{k},b,b',\sigma} h^{\mathrm{G}}_{b,b';\sigma}(\mathbf{k})\hat{c}^{\dagger}_{\mathbf{k},b,\sigma}\hat{c}_{\mathbf{k},b',\sigma} \ , \quad (79)$$

see eq. (72). The Landau-Gutzwiller quasi-particle dispersion $\varepsilon^{\mathrm{G}}_{n,\sigma}(\mathbf{k})$ follows from the diagonalization of $h^{\mathrm{G}}_{b,b';\sigma}(\mathbf{k})$. For explicit expressions for $h^{\mathrm{G}}_{b,b';\sigma}(\mathbf{k})$ and the actual numerical implementation within QUANTUMESPRESSO, see Refs. [12, 13].

3.4 Local Hamiltonian and double counting for transition metals

For a Gutzwiller DFT calculation we need to specify the Coulomb parameters in the local Hamiltonian and the form of the double-counting operator in eqs. (43) and (44).

3.4.1 Cubic symmetry and spherical approximation

In many theoretical studies one uses a local Hamiltonian with only density-density interactions,

$$\hat{V}^{\mathrm{dens}}_{\mathrm{loc}} = \frac{1}{2}\sum_{c,\sigma} U(c,c)\hat{n}_{c,\sigma}\hat{n}_{c,\bar{\sigma}} + \frac{1}{2}\sum_{c(\neq)c'}\sum_{\sigma,\sigma'} \widetilde{U}_{\sigma,\sigma'}(c,c')\hat{n}_{c,\sigma}\hat{n}_{c',\sigma'} \ . \quad (80)$$

Here, we introduced $\bar{\uparrow} = \downarrow$ ($\bar{\downarrow} = \uparrow$) and $\widetilde{U}_{\sigma,\sigma'}(c,c') = U(c,c') - \delta_{\sigma,\sigma'} J(c,c')$, where $U(c,c')$ and $J(c,c')$ are the local Hubbard and Hund's-rule exchange interactions. An additional and quite common approximation is the use of orbital-independent Coulomb parameters,

$$U(c,c) \equiv U \ , \quad \text{and} \quad U(c,c') \equiv U' , \ J(c,c') \equiv J \quad \text{for } c \neq c' . \tag{81}$$

For a system of five correlated $3d$ orbitals in a cubic environment as in nickel and iron, however, the Hamiltonian (80) is incomplete [26]. The full Hamiltonian reads

$$\hat{V}_{\rm loc}^{\rm full} = \hat{V}_{\rm loc}^{\rm dens} + \hat{V}_{\rm loc}^{\rm n.dens.} \ , \tag{82}$$

where

$$\begin{aligned} \hat{V}_{\rm loc}^{\rm n.dens.} \ = \ & \frac{1}{2} \sum_{c(\neq)c'} J(c,c') \Big(\hat{c}^\dagger_{c,\uparrow} \hat{c}^\dagger_{c,\downarrow} \hat{c}_{c',\downarrow} \hat{c}_{c',\uparrow} + \text{h.c.} \Big) + \frac{1}{2} \sum_{c(\neq)c';\sigma} J(c,c') \hat{c}^\dagger_{c,\sigma} \hat{c}^\dagger_{c',\bar{\sigma}} \hat{c}_{c,\bar{\sigma}} \hat{c}_{c',\sigma} \\ & + \Bigg[\sum_{t;\sigma,\sigma'} (T(t) - \delta_{\sigma,\sigma'} A(t)) \hat{n}_{t,\sigma} \hat{c}^\dagger_{u,\sigma'} \hat{c}_{v,\sigma'} \\ & \quad + \sum_{t,\sigma} A(t) \left(\hat{c}^\dagger_{t,\sigma} \hat{c}^\dagger_{t,\bar{\sigma}} \hat{c}_{u,\bar{\sigma}} \hat{c}_{v,\sigma} + \hat{c}^\dagger_{t,\sigma} \hat{c}^\dagger_{u,\bar{\sigma}} \hat{c}_{t,\bar{\sigma}} \hat{c}_{v,\sigma} \right) \\ & \quad + \sum_{t(\neq)t'(\neq)t''} \sum_{e,\sigma,\sigma'} S(t,t';t'',e) \hat{c}^\dagger_{t,\sigma} \hat{c}^\dagger_{t',\sigma'} \hat{c}_{t'',\sigma'} \hat{c}_{e,\sigma} + \text{h.c.} \Bigg] . \end{aligned} \tag{83}$$

Here, $t = \zeta, \eta, \xi$ and $e = u, v$ are indices for the three t_{2g} orbitals with symmetries $\zeta = xy$, $\eta = xz$, and $\xi = yz$, and the two e_g orbitals with symmetries $u = 3z^2 - r^2$ and $v = x^2 - y^2$, respectively. The parameters $A(t)$, $T(t)$, $S(t,t';t'',e)$ in eq. (83) are of the same order of magnitude as the exchange interactions $J(c,c')$ and, hence, there is no a-priori reason to neglect $V_{\rm loc}^{\rm n.dens.}$. Of all the parameters $U(c,c')$, $J(c,c')$, $A(t)$, $T(t)$, $S(t,t';t'',e)$ only ten are independent in cubic symmetry, see appendix A.2 and Ref. [17].

When we assume that all $3d$-orbitals have the same radial wave-function ('spherical approximation'), all parameters are determined by, e.g., the three Racah parameters A, B, C, see appendix A.2. For comparison with other work, we introduce the average Coulomb interaction between electrons in the same $3d$-orbitals, $U = \sum_c U(c,c)/5 = A + 4B + 3C$, the average Coulomb interaction between electrons in different orbitals, $U' = \sum_{c<c'} U(c,c')/10 = A - B + C$, and the average Hund's-rule exchange interaction, $J = \sum_{c<c'} J(c,c')/10 = 5B/2 + C$ that are related by the symmetry relation $U' = U - 2J$, see appendix A.2. Due to this symmetry relation, the three values of U, U', and J do not determine the Racah parameters A, B, C uniquely. Therefore, we make use of the relation $C/B = 4$ which is a reasonable assumption for metallic nickel [8, 26]. In this way, the three Racah parameters and, consequently, all parameters in $\hat{V}_{\rm loc}^{\rm full}$ are functions of U and J, $A = U - 32J/13$, $B = 2J/13$, $C = 8J/13$. This permits a meaningful comparison of our results for all local Hamiltonians.

3.4.2 Double counting corrections

There exists no systematic (let alone rigorous) derivation of the double-counting corrections in eq. (43). A widely used form for this operator has first been introduced in the context of the

LDA+U method. Its expectation value is given by

$$V_{\rm dc}^{\rm G} = \frac{\overline{U}}{2}\bar{n}(\bar{n}-1) - \frac{\overline{J}}{2}\sum_{\sigma}\bar{n}_{\sigma}(1-\bar{n}_{\sigma}) \;, \tag{84}$$

where $\bar{n}_{\sigma} \equiv \sum_{c=1}^{N_c} C^{\rm G}_{c,c;\sigma}$, $\bar{n} \equiv \bar{n}_{\uparrow} + \bar{n}_{\downarrow}$, and N_c is the number of correlated orbitals ($N_c = 5$ for nickel). Moreover, $\overline{U} = (U + 4U')/5$ and $\overline{J} = \overline{U} - U' + J$. Here,

$$C^{\rm G}_{c,c;\sigma} = \frac{\langle \Psi_{\rm G}|\hat{c}^{+}_{\mathbf{R},c,\sigma}\hat{c}_{\mathbf{R},c,\sigma}|\Psi_{\rm G}\rangle}{\langle \Psi_{\rm G}|\Psi_{\rm G}\rangle} = \langle \Phi|\hat{c}^{+}_{\mathbf{R},c,\sigma}\hat{c}_{\mathbf{R},c,\sigma}|\Phi\rangle \tag{85}$$

is the σ-electron density for the correlated $3d$ orbital c in the Gutzwiller wavefunction. Note that the second equality only holds in the limit of infinite dimensions for our e_g-t_{2g} orbital structure [12].

3.4.3 Size of optimal atomic parameters

Before we proceed, we briefly comment on our optimal Coulomb parameters for nickel and iron because they are substantially larger than the parameters used in other studies [27–31]. For example, for iron the values $U = 2\,\mathrm{eV} \ldots 3\,\mathrm{eV}$ and $J = 0.8\,\mathrm{eV} \ldots 1.0\,\mathrm{eV}$ are used, e.g., to describe the high-temperature regime with the transition from fcc iron to bcc iron and the Curie transition from non-magnetic to magnetic bcc iron, while more recent LDA+DMFT studies employ larger values, $\bar{U} = 4.3\,\mathrm{eV}$, $\widetilde{J} = 1.0\,\mathrm{eV}$ [32].

First of all, we note that the large spread of values of (U, J) in the literature is due to the strong sensitivity of these parameters to the energy window used for projecting, or *downfolding*, the full electronic structure to an effective many-body model [33]. It is well known that the bare Hubbard parameters U are of the order of 20 eV, or larger [3]. They apply for instantaneous charge excitations of an isolated atom, which are strongly screened in a solid. In Fe, for example, the screening reduces U to ~ 3 eV for d-only models [34, 35]. Our self-consistent DFT method is based on a projective technique to construct Wannier functions. In the present calculations, we chose a large energy window, which ensures a very good localization of the $3d$ orbitals, and a minimal dependence of the basis set on atomic positions. This large energy window translates into larger values of U, J [36]. Other calculations can typically afford to retain fewer bands.

Second, we note that the Hubbard-U in our Gutzwiller treatment parameterizes the interaction of two electrons in the same orbital, see appendix A.2. In other approaches, this quantity describes some orbital average. For example, Pourovskii et al. [32] use the Slater-Condon parameter $F^{(0)} = \bar{U}$, where $\bar{U} = (U + 4U')/5$, and $U' = A - B + C = U - 2J$ is the inter-orbital Coulomb repulsion. Naturally, the intra-orbital U is larger than an average over intra-orbital and inter-orbital Coulomb repulsions. Likewise, we work with the average Hund's-rule coupling $J = 5B/2 + C$, see appendix A.2, whereas $\widetilde{J} \equiv (F^{(2)} + F^{(4)})/14 = 7B/2 + 7C/5 = 7J/5$ [32]. Therefore, $F^{(0)} = 4.3\,\mathrm{eV}$ and $\widetilde{J} = 1.0\,\mathrm{eV}$ correspond to $J = 0.71\,\mathrm{eV}$ and $U = \bar{U} + 8\widetilde{J}/7 = 5.4\,\mathrm{eV}$ with $J/U = 0.13$. We note in passing that we work with $C/B = 4$ whereas others use $F^{(2)}/F^{(4)} = 8/5$ which corresponds to $C/B = 175/47 \approx 3.7$ [37].

Lastly, in our Gutzwiller calculations, we use parameters such as U and J to 'match' selected experimental quantities. In this way, we compensate approximations in the model setup, e.g., the neglect of non-local correlations in Hubbard-type models, and in the model analysis, e.g., the limit of infinite dimensions or an approximate variational ground state. For example, in Gutzwiller calculations, the optimal Coulomb parameters must be chosen somewhat smaller when the full atomic interaction is replaced by density-density interactions only [35]. Similarly, larger U-values are found to be optimal when the impurity solver in Quantum-Monte-Carlo is rotationally invariant [27]. Our substantial Hubbard interaction leads to noticeable bandwidth renormalizations and an increase of the quasi-particle masses at the Fermi energy, as seen in experiment [38, 39].

4 Results for transition metals

In this section we compile recent results for nickel and iron as obtained from the Gutzwiller DFT [12, 13].

4.1 Nickel

As a variational approach, the Gutzwiller DFT is expected to be most suitable for the calculation of ground-state properties such as the lattice constant, the magnetic moment, or the Fermi surface of a Fermi liquid. Although more speculative than the ground-state calculations, it is also common to interpret the eigenvalues of the Gutzwiller–Kohn-Sham Hamiltonian $\varepsilon^{\rm G}_{{\rm n},\sigma}(\mathbf{k})$ as the dispersion of the single-particle excitations [40].

4.1.1 Lattice constant, magnetic moment, and bulk modulus of nickel

In Fig. 1, we show the lattice constant and the magnetic moment as a function of U ($1\,{\rm eV} \le U \le 14\,{\rm eV}$) for four different values of J/U ($J/U = 0$, 0.05, 0.075, 0.10). As is well known, the DFT-LDA underestimates the lattice constant, $a_0^{\rm LDA} = 6.47\ a_{\rm B}$ is considerably smaller than the experimental value of $a_0 = 6.66\ a_{\rm B}$ where $a_{\rm B} = 0.529177\,$Å is the Bohr radius. Fig. 1 shows that the Hubbard interaction U increases the lattice constant whereby the Hund's-rule exchange J diminishes the slope. Apparently, a good agreement with the experimental lattice constant requires substantial Hubbard interactions, $U > 10\,{\rm eV}$. The increase of the lattice parameter as a function of the Hubbard interaction is readily understood because the $3d$ bandwidth is reduced by electronic correlations so that the $3d$ electrons contribute less to the metallic binding.

Fig. 1 shows the well-known fact that DFT-LDA reproduces the experimental value for the spin-only magnetic moment $m_{\rm so}$ very well, $m_{\rm so}^{\rm LDA} = 0.58\mu_{\rm B}$ vs. $m_{\rm so}^{\rm exp} = 0.55\mu_{\rm B}$. However, when the DFT-LDA calculation is performed for the experimental value of the lattice constant, the magnetic moment is grossly overestimated. As seen in Fig. 1, the Gutzwiller DFT allows us to reconcile the experimental findings both for the lattice constant and the magnetic moment if we work in the parameter range $11\,{\rm eV} < U < 14\,{\rm eV}$ and $0.05 < J/U < 0.07$. Note that a

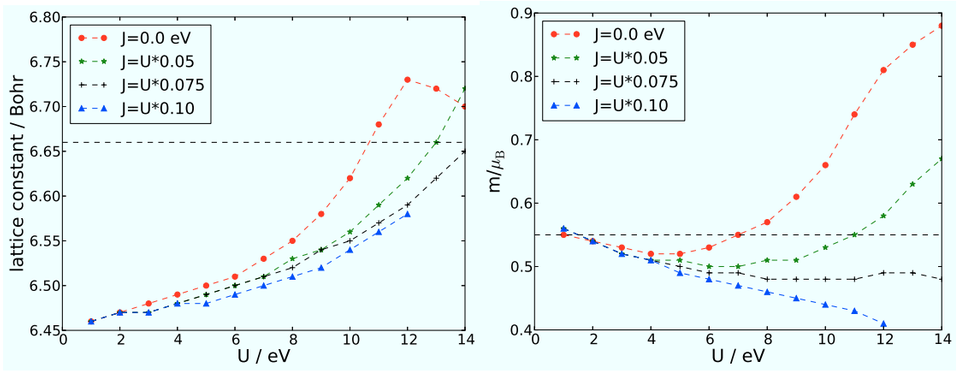


Fig. 1: *Lattice constant (left) and magnetic moment (right) of nickel as a function of U, for four different values of J/U; dashed horizontal lines: experimental values.*

'fine-tuning' of parameters is not required to obtain a reasonable agreement between theory and experiment for the lattice constant and spin-only magnetic moment.

For nickel, detailed information about the quasi-particle bands is available. The quasi-particle dispersion at various high-symmetry points in the Brillouin zone is more sensitive to the precise values of U and J. As we shall show below in more detail, we obtain a satisfactory agreement with ARPES data for the choice ($U^{\rm opt} = 13\,{\rm eV}$, $J^{\rm opt} = 0.9\,{\rm eV}$) with an uncertainty of ± 1 in the last digit. For our optimal values we show in Fig. 2 the ground-state energy per particle $E(a)/N$ as a function of the fcc lattice constant a together with a second-order polynomial fit. The minimum is obtained at $a_0 = 6.63 a_{\rm B}$, in good agreement with the experimental value $a_0^{\rm exp} = 6.66 a_{\rm B}$. For the magnetic spin-only moment we obtain $m_{\rm so} = 0.52 \mu_{\rm B}$, in good agreement with the experimental value $m_{\rm so}^{\rm exp} = 0.55 \mu_{\rm B}$.

From the curvature of $E(a)/N$ at $a = a_0$ we can extract the bulk modulus. The bulk modulus at zero temperature is defined as the second-derivative of the ground-state energy with respect to the volume,

$$K = V_0 \left. \frac{{\rm d}^2 E(V)}{{\rm d}V^2} \right|_{V=V_0} . \tag{86}$$

This implies the Taylor expansion $E(V) = E(V_0) + (KV_0/2)(V/V_0 - 1)^2 + \ldots$ for the ground-state energy as a function of the volume $V = a^3$ (Birch-Murnaghan fit). For the ground-state energy per particle we can thus write $E(a)/N = E(a_0)/N + e_2(a/a_{\rm B} - a_0/a_{\rm B})^2 + \ldots$ with

$$e_2 = \frac{9}{8} K a_{\rm B}^3 (a_0/a_{\rm B}) \; , \tag{87}$$

where we took into account that the fcc unit cell hosts four atoms, $V_0 = N a_0^3/4$. The fit leads to $K = 169\,{\rm GPa}$, in good agreement with the experimental value, $K = 182\,{\rm GPa}$ [41]. It is a well-known fact that the DFT-LDA overestimates the bulk modulus of nickel. Indeed, DFT-LDA gives $K^{\rm LDA} = 245\,{\rm GPa}$.

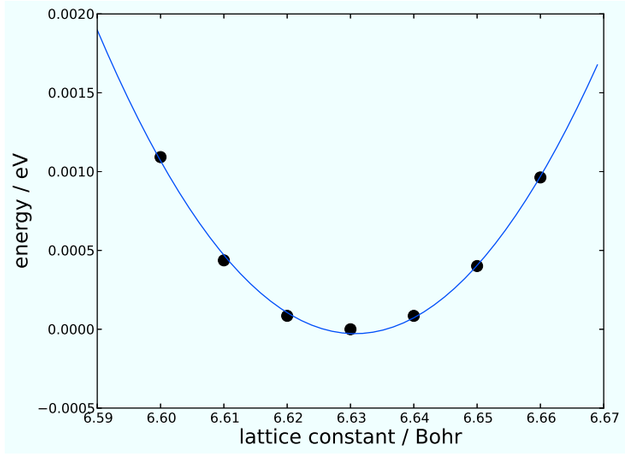


Fig. 2: *Ground-state energy per particle $E_0(a)/N$ relative to its value at $a = 6.63a_{\rm B}$ as a function of the fcc lattice parameter $a/a_{\rm B}$ in units of the Bohr radius $a_{\rm B}$ for ($U^{\rm opt} = 13\,{\rm eV}$, $J^{\rm opt} = 0.9\,{\rm eV}$). Full line: second-order polynomial fit.*

4.1.2 Quasi-particle bands of nickel

In Fig. 3 we show the quasi-particle band structure of fcc nickel for ($U^{\rm opt} = 13\,{\rm eV}$, $J^{\rm opt} = 0.9\,{\rm eV}$). The most prominent effect of the Gutzwiller correlator is the reduction of the $3d$ bandwidth. From a paramagnetic DFT-LDA calculation one can deduce $W^{\rm LDA} = 4.5\,{\rm eV}$ [42–44], whereas we find $W = 3.3\,{\rm eV}$, in agreement with experiment.

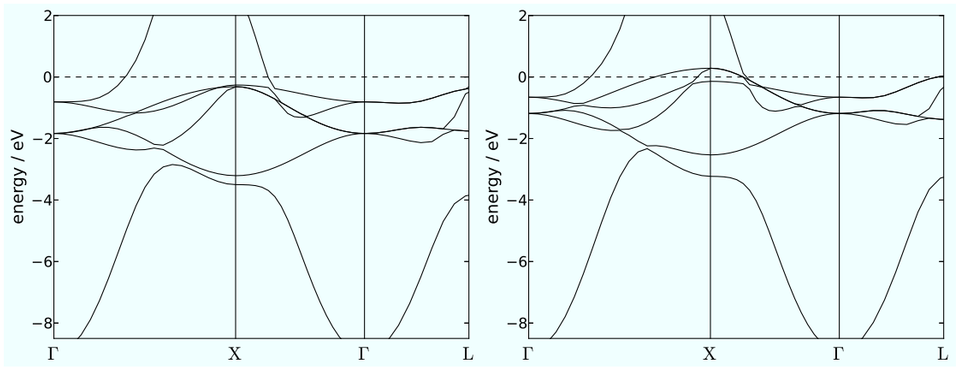


Fig. 3: *Quasi-particle band structure of fcc nickel along high-symmetry lines in the first Brillouin zone for ($U^{\rm opt} = 13\,{\rm eV}$, $J^{\rm opt} = 0.9\,{\rm eV}$). Left: majority spin; Right: minority spin. The Fermi energy is at $E_{\rm F}^{\rm G} = 0$.*

Symmetry	Experiment	Gutzwiller-DFT
$\langle\Gamma_1\rangle$	8.90 ± 0.30	8.95[0.08]
$\langle\Gamma_{25'}\rangle$	1.30 ± 0.06	1.51[0.65]
$\langle\Gamma_{12}\rangle$	0.48 ± 0.08	0.73[0.15]
$\langle X_1\rangle$	3.30 ± 0.20	3.37[0.27]
$\langle X_3\rangle$	2.63 ± 0.10	2.87[0.68]
$X_{2\uparrow}$	0.21 ± 0.03	0.26
$X_{2\downarrow}$	0.04 ± 0.03	0.14
$X_{5\uparrow}$	0.15 ± 0.03	0.32
$\Delta_{e_g}(X_2)$	0.17 ± 0.05	0.12
$\Delta_{t_{2g}}(X_5)$	0.33 ± 0.04	0.60
$\langle L_1\rangle$	3.66 ± 0.10	3.49[0.61]
$\langle L_3\rangle$	1.43 ± 0.07	1.58[0.38]
$L_{3\uparrow}$	0.18 ± 0.03	0.37
$\langle L_{2'}\rangle$	1.00 ± 0.20	0.14[0.06]
$\langle\Lambda_{3;1/2}\rangle$	$0.50[0.21 \pm 0.02]$	0.64[0.30]

Table 1: *Quasi-particle band energies for fcc nickel with respect to the Fermi energy in eV at various high-symmetry points (counted positive for occupied states).* $\langle\ldots\rangle$ *indicates the spin average, error bars in the experiments without spin resolution are given as* $\pm$*. Theoretical data show the spin average and the exchange splittings in square brackets.* $\Lambda_{3;1/2}$ *denotes the point half-way on the* Λ*-line that links the points* Γ *and* L*. The first column gives experimental data compiled in [8], the second column gives the result from Gutzwiller DFT at (*$U^{\mathrm{opt}} = 13\,\mathrm{eV}, J^{\mathrm{opt}} = 0.9\,\mathrm{eV}$*).*

A more detailed comparison of the quasi-particle band structure with experiment is given in table 1. The overall agreement between experiment and theory is quite satisfactory. In particular, only one hole ellipsoid is found at the X-point, in agreement with experiment and in contrast to the DFT-LDA result [8].

We comment on two noticeable discrepancies between theory and experiment. First, the energy of the band $L_{2'}$ at the L-point deviates by a factor of five. This is an artifact that occurs already at the DFT-LDA level and is not cured by the Gutzwiller approach. Since the level has pure $3p$ character around the L point, the origin of the discrepancy is related to the uncertainties in the partial charge densities n_{3d}, $n_{3p,3s}$ in the $3d$ and $3p/3s$ bands. Second, the Gutzwiller DFT prediction for the exchange splitting $\Delta_{t_{2g}}(X_5)$ of the t_{2g} bands at the X-point is a factor of two larger than in experiment. This deviation is related to the fact that, quite generally, all bands are slightly too low in energy. This can be cured by decreasing U and increasing J but this deteriorates the values for the lattice constant and the magnetic moment. We suspect that the deviations are partly due to the use of a heuristic double-counting correction and the neglect of the spin-orbit coupling. Moreover, we expect the results for the band structure to improve when we replace the "poor-man's Wannier" orbitals for the correlated $3d$ electrons by more sophisticated localized wave functions.

4.2 Iron

The ground state of iron poses a difficult problem because the local-density approximation to density functional theory predicts a face-centered cubic (fcc) or a hexagonal closed-packed (hcp) ground state [45, 46]. Gutzwiller-DFT finds the correct (magnetic) body-centered cubic crystal structure without resorting to generalized gradient corrections (GGA) [47, 48].

4.2.1 Lattice constant, magnetic moment, and bulk modulus of iron

When we use $(U = 9\,\mathrm{eV}, J = 0.54\,\mathrm{eV})$ for iron, we find the lattice parameter $a = 5.39 a_\mathrm{B} = 2.85\,\text{Å}$ and the magnetic moment $m = 2.24\mu_\mathrm{B}$ that agree very well with the experimental values, $a_\mathrm{exp} = 5.42 a_\mathrm{B} = 2.87\,\text{Å}$ [49] and $\mu_\mathrm{exp} = 2.22\mu_\mathrm{B}$ [50]. In Fig. 4 we show the ground-state energy per atom, $e(v) = E(V)/N$, as a function of the unit-cell volume $v = V/N = a^3/2$ in the vicinity of the optimal value $v_0 = a^3/2 = 78.3 a_\mathrm{B}^3 = 11.6\,\text{Å}^3$.

As is seen, the magnetic bcc iron phase is energetically favorable over the non-magnetic hcp phase, as found experimentally. Within the Gutzwiller DFT this effect is seen to be the consequence of electronic correlations, not of generalized gradient corrections (GGA) to the LDA functional. Moreover, Fig. 4 suggests that a transition from magnetic bcc iron to non-magnetic hcp iron is possible upon reducing the lattice volume. Such a pressure-induced first-order phase transition is indeed observed experimentally at $p_c^\mathrm{exp} = 10 \ldots 15\,\mathrm{GPa}$ [51] at room temperature. In Gutzwiller-DFT the critical pressure is higher, $p_c \approx 40\,\mathrm{GPa}$, where entropy effects from magnons and phonons are not included.

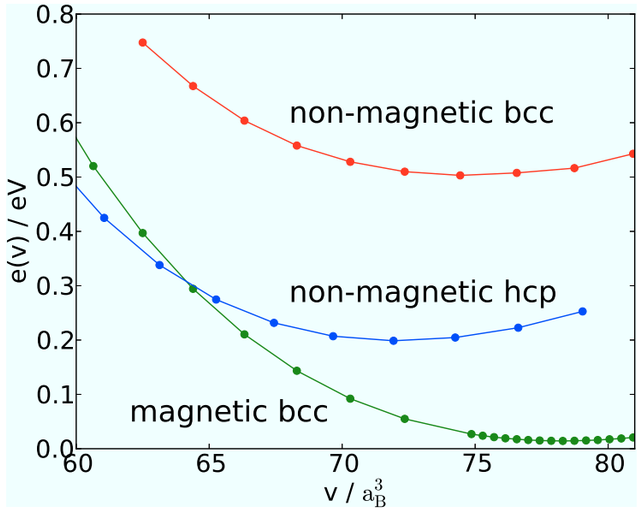


Fig. 4: *Energy per atom $e(v)$ in units of eV as a function of the unit-cell volume v in units of a_B^3 for non-magnetic and ferromagnetic bcc iron and non-magnetic hcp iron for $U = 9\,\mathrm{eV}$ and $J = 0.54\,\mathrm{eV}$, and ambient pressure. The energies are shifted by the same constant amount.*

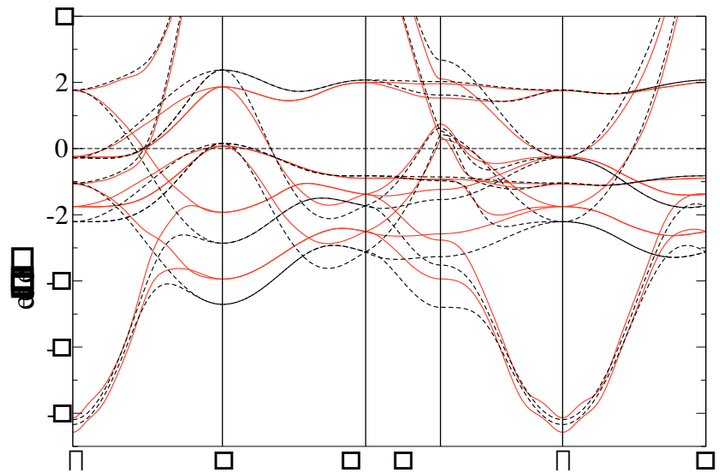


Fig. 5: *Comparison between DFT(GGA) bands (black, dashed lines) and bands from Gutzwiller-DFT (red, full lines) for ($U = 9.0\,\mathrm{eV}$, $J = 0.54\,\mathrm{eV}$) for ferromagnetic bcc iron. For clarity, we do not discriminate between majority and minority spin bands. The Fermi energy is at $E_\mathrm{F} = 0$ (dashed horizontal line).*

In Gutzwiller-DFT we find a bulk modulus of $B = 165\,\mathrm{GPa}$, in very good agreement with the experimental value, $B^\mathrm{exp} = (170 \pm 4)\,\mathrm{GPa}$ [52, 49]. The LDA+Gutzwiller value substantially improves the DFT(LDA) value of $B^\mathrm{LDA} = 227\,\mathrm{GPa}$, it is slightly better than the values from DFT(GGA) studies, $B^\mathrm{GGA} = (190 \pm 10)\,\mathrm{GPa}$ [49], and agrees with the value obtained in DMFT calculations, $B^\mathrm{DMFT} = 168\,\mathrm{GPa}$ [32].

4.2.2 Quasi-particle bands of ferromagnetic bcc iron

We show the bandstructure for ferromagnetic bcc iron from Gutzwiller-DFT for ($U = 9.0\,\mathrm{eV}$, $J = 0.54\,\mathrm{eV}$) in comparison with those from (scalar relativistic) DFT(GGA) calculations in Fig. 5. The DFT(GGA) provides the same equilibrium lattice parameter as used in Gutzwiller-DFT, $a = 5.39 a_\mathrm{B}$. The correlations in the Gutzwiller approach lead to an additional bandwidth reduction of the d bands across the Brillouin zone. The uncorrelated, $4sp$-type parts of the quasi-particle bands deep below the Fermi energy do not differ much, e.g., the lowest $4sp$-type majority bands at the Γ point, about 9 eV below the Fermi energy.

The bandwidth reduction in iron is not as strong as in nickel. Nevertheless, for selected symmetry points, the discrepancies between the quasi-particle bands from DFT and Gutzwiller-DFT are quite large, see Fig. 5. For example, at the H-point in the Brillouin zone we find a bandwidth reduction for the majority band by 36%, from $H^\mathrm{LDA}_{\mathrm{low},\uparrow} = 5.38\,\mathrm{eV}$ down to $H^\mathrm{G\text{-}DFT}_{\mathrm{low},\uparrow} = 3.94\,\mathrm{eV}$, in good agreement with experiment, $H_{\mathrm{low},\uparrow} = 3.8\,\mathrm{eV}$ [53]. Likewise, at the N-point in the Brillouin zone there is a majority spin band at $N_{\mathrm{low},\uparrow} = 4.5\,\mathrm{eV}$ below the Fermi energy in experiment [53], in comparison with $N^\mathrm{LDA}_{\mathrm{low},\uparrow} = 5.47\,\mathrm{eV}$ in DFT(LDA) and $N^\mathrm{G\text{-}DFT}_{\mathrm{low},\uparrow} = 3.90\,\mathrm{eV}$ in Gutzwiller-DFT. A comparison of the Landau-Gutzwiller quasi-particle bands and experimental ARPES data close to the Fermi energy can be found in Ref. [13].

5 Summary and outlook

In this work, we presented a detailed derivation of the Gutzwiller Density Functional Theory. Unlike previous studies, our formalism covers all conceivable cases of symmetries and Gutzwiller wave functions. Moreover, our theory is not based on the 'Gutzwiller approximation' which corresponds to an evaluation of expectation values in the limit of infinite lattice coordination number. It is only in the last step that we resort to this limit.
In particular, our derivation consists of three main steps.

1. The density functional of the full many-particle system is related to that of a reference system with local Coulomb interactions in the correlated orbitals. This generalizes the widely used Kohn-Sham scheme, where a single-particle reference system is used, to the Kohn-Sham–Hubbard approach that may be analyzed by sophisticated many-particle methods.
2. In this work, the energy functional of the Hubbard-type reference system is (approximately) evaluated by means of Gutzwiller variational wave functions.
3. Analytical results for the Gutzwiller energy functional are derived in the limit of large coordination number.

We studied the electronic properties of ferromagnetic nickel and iron. The Gutzwiller DFT resolves the main deficiencies of DFT(LDA) in describing ground-state properties such as the lattice constant, the magnetic moment, or the bulk modulus of nickel and iron. In particular, Gutzwiller-DFT gives the proper bcc ground-state structure for iron, without resorting to generalized gradient corrections. Note that our approach requires relatively large values for the local Coulomb interaction, $U = \mathcal{O}(10\,\mathrm{eV})$, to obtain good agreement with experiments.
Our results for the quasi-particle band structure are satisfactory. A perfect agreement with ARPES data would be surprising because we calculate these quantities based on Fermi-liquid assumptions that are strictly valid only in the vicinity of the Fermi surface. The quasi-particle energies strongly depend on the orbital occupations that are influenced by the double-counting corrections. We consider the arbitrariness of the double-counting corrections as the main shortcoming of the Gutzwiller DFT in its present form that should be addressed in future studies.
The method presented in this work can be developed further in various directions. On the level of the Gutzwiller approximation, a number of implementations is still on the agenda, e.g., the inclusion of spin-orbit coupling or the study of phonons. To tackle unconventional superconductivity as seen, e.g., in the cuprates, one has to go beyond the Gutzwiller approximation, see Refs. [18–20] for the single-band model and Ref. [21] for multi-band models. These approaches can be combined with the DFT as explained in Sect. 3. Finally, a time-dependent Gutzwiller method can be used for the calculation of two-particle Green functions, see Refs. [54–56], and references therein. As a long-term perspective, a combination of these methods and the Hubbard-Kohn-Sham Hamiltonian would permit a first-principles calculation of important experimental quantities such as the magnetic susceptibility.

Acknowledgements

I thank Jörg Bünemann for his critical reading of the manuscript.

A Appendix

A.1 Single-particle systems

A.1.1 Single-particle density matrix

With the help of a single-particle basis $|k\rangle$ in which a given single-particle operator $\hat{H}_{\rm sp}$ is diagonal, an eigenstate can be written as

$$|\Phi\rangle = \prod_k{}' \hat{b}_k^\dagger |{\rm vac}\rangle \ , \tag{88}$$

where the prime indicates that N single-particle states are occupied in $|\Phi\rangle$. The single-particle density matrix is diagonal in $|\Phi\rangle$,

$$\rho_{k,k'} \equiv \langle\Phi|\hat{b}_k^\dagger \hat{b}_{k'}|\Phi\rangle = \delta_{k,k'} f_k \ , \tag{89}$$

and the entries on the diagonal obey $f_k^2 = f_k$ because we have $f_k = 0, 1$. Therefore, we have shown that

$$\tilde{\rho} \cdot \tilde{\rho} = \tilde{\rho} \ . \tag{90}$$

Since the operators $\hat{c}_i^\dagger$ for any other single-particle basis and the operators $\hat{b}_k^\dagger$ are related via a unitary transformation, eq. (33) holds generally for single-particle density matrices for single-particle product states.

A.1.2 Minimization with respect to the single-particle density matrix

We consider a general real function $E(\tilde{\rho})$ of a non-interacting density matrix $\tilde{\rho}$ with the elements

$$\rho_{i,j} = \langle\Phi|\hat{c}_j^\dagger \hat{c}_i|\Phi\rangle \ . \tag{91}$$

The fact that $\tilde{\rho}$ is derived from a single-particle product wave function $|\Phi\rangle$ is equivalent to the matrix equation (90). Hence, the minimum of $E(\tilde{\rho})$ in the 'space' of all non-interacting density matrices is determined by the condition

$$\frac{\partial}{\partial \rho_{j,i}} L(\tilde{\rho}) = 0 \ , \tag{92}$$

where we introduced the 'Lagrange functional'

$$L(\tilde{\rho}) \equiv E(\tilde{\rho}) - \sum_{l,m} \Omega_{l,m} \Big(\sum_p \rho_{m,p} \rho_{p,l} - \rho_{m,l} \Big) \tag{93}$$

and the matrix $\widetilde{\Omega}$ of Lagrange parameters $\Omega_{l,m}$. Eq. (92) leads to the matrix equation

$$\tilde{H} = \tilde{\rho} \cdot \widetilde{\Omega} + \widetilde{\Omega} \cdot \tilde{\rho} - \widetilde{\Omega} \tag{94}$$

for the 'Hamilton matrix' $\tilde{H}$ with the elements

$$H_{i,j} = \frac{\partial}{\partial \rho_{j,i}} E(\tilde{\rho}) \ . \tag{95}$$

Equation (94) is satisfied if eq. (90) holds and if

$$[\tilde{H}, \tilde{\rho}] = 0 \ . \tag{96}$$

Hence, $\tilde{H}$ and $\tilde{\rho}$ must have the same basis of (single-particle) eigenvectors and, consequently, we find an extremum of $E(\tilde{\rho})$ if we choose $|\Phi\rangle$ as an eigenstate of

$$\hat{H}_{\rm sp} = \sum_{i,j} H_{i,j} \hat{c}_i^\dagger \hat{c}_j \ . \tag{97}$$

Usually, $|\Phi\rangle$ can be chosen as the ground state of $\hat{H}_{\rm sp}$.

A.1.3 Basis sets

In the following we assume that the potential is lattice periodic,

$$V_\sigma^{\rm KS}(\mathbf{r}) = U(\mathbf{r}) + V_{\rm Har}(\mathbf{r}) + v_{{\rm sp,xc},\sigma}(\mathbf{r}) = V_\sigma^{\rm KS}(\mathbf{r}+\mathbf{R}) \ , \tag{98}$$

where $\mathbf{R}$ is a lattice vector. The Fourier components are finite for reciprocal lattice vectors $\mathbf{G}$ only,

$$V_{\mathbf{G},\sigma}^{\rm KS} = \frac{1}{V} \int {\rm d}\mathbf{r} \ V_\sigma^{\rm KS}(\mathbf{r}) e^{-{\rm i}\mathbf{G}\cdot\mathbf{r}} \ , \tag{99}$$

where V is the crystal volume. As a consequence of the lattice periodicity, the crystal momentum $\mathbf{k}$ from the first Brillouin zone is a good quantum number.
As seen from eq. (41), the Kohn-Sham Hamiltonian is diagonalized for the single-particle states $\psi_{\mathbf{k},{\rm n},\sigma}(\mathbf{r}) = \langle \mathbf{r}|\mathbf{k},{\rm n},\sigma\rangle$ that obey the Kohn-Sham equations [1] (n: band index)

$$h_\sigma^{\rm KS}(\mathbf{r}) \psi_{\mathbf{k},{\rm n},\sigma}(\mathbf{r}) = \varepsilon_{{\rm n},\sigma}(\mathbf{k}) \psi_{\mathbf{k},{\rm n},\sigma}(\mathbf{r}) \ . \tag{100}$$

In its eigenbasis, the Kohn-Sham Hamiltonian takes the form

$$\hat{H}^{\rm KS} = \sum_{\mathbf{k},{\rm n},\sigma} \varepsilon_{{\rm n},\sigma}(\mathbf{k}) \hat{b}_{\mathbf{k},{\rm n},\sigma}^\dagger \hat{b}_{\mathbf{k},{\rm n},\sigma} \ . \tag{101}$$

Its ground state is given by

$$|\Phi_0\rangle = \prod_\sigma \prod_{\mathbf{k},{\rm n}}{}' \hat{b}_{\mathbf{k},{\rm n},\sigma}^\dagger |{\rm vac}\rangle \ , \tag{102}$$

where the N levels lowest in energy are occupied as indicated by the prime at the product, $\varepsilon_{{\rm n},\sigma}(\mathbf{k}) \le E_{{\rm F},\sigma}$. Then,

$$f_{\mathbf{k},{\rm n},\sigma} = \langle \Phi_0 | \hat{b}_{\mathbf{k},{\rm n},\sigma}^\dagger \hat{b}_{\mathbf{k},{\rm n},\sigma} | \Phi_0 \rangle = \Theta\left(E_{{\rm F},\sigma} - \varepsilon_{{\rm n},\sigma}(\mathbf{k})\right) \tag{103}$$

is unity for occupied levels up to the Fermi energy $E_{{\rm F},\sigma}$, and zero otherwise.
From eq. (26), the field operators read

$$\hat{\Psi}_\sigma(\mathbf{r}) = \sum_{\mathbf{k},{\rm n}} \psi_{\mathbf{k},{\rm n},\sigma}(\mathbf{r}) \hat{b}_{\mathbf{k},{\rm n},\sigma} \quad , \quad \hat{\Psi}_\sigma^\dagger(\mathbf{r}) = \sum_{\mathbf{k},{\rm n}} \psi_{\mathbf{k},{\rm n},\sigma}^*(\mathbf{r}) \hat{b}_{\mathbf{k},{\rm n},\sigma}^\dagger \ . \tag{104}$$

Therefore, the ground-state density is readily obtained as

$$\begin{aligned} n^0_\sigma(\mathbf{r}) &= \langle \Phi_0 | \hat{\Psi}^\dagger_\sigma(\mathbf{r}) \hat{\Psi}_\sigma(\mathbf{r}) | \Phi_0 \rangle = \sum_{\mathbf{k},\mathrm{n}} f_{\mathbf{k},\mathrm{n},\sigma} |\psi_{\mathbf{k},\mathrm{n},\sigma}(\mathbf{r})|^2 = \langle \mathbf{r} | \sum_{\mathbf{k}} \hat{\rho}^{(0)}_\sigma(\mathbf{k}) | \mathbf{r} \rangle \ , \\ \hat{\rho}^{(0)}_\sigma(\mathbf{k}) &= \sum_n f_{\mathbf{k},\mathrm{n},\sigma} |\mathbf{k},\mathrm{n},\sigma\rangle\langle \mathbf{k},\mathrm{n},\sigma| \ , \end{aligned} \tag{105}$$

see also eq. (31). Since this quantity enters the Kohn-Sham Hamiltonian, its solution must be achieved self-consistently.

In order to make contact with many-particle approaches based on Hubbard-type models, we need to identify orbitals that enter the local two-particle interaction. Implemented plane-wave codes provide the transformation coefficients $F_{(\mathbf{k},\mathrm{n}),(\mathbf{R},b);\sigma}$ from Bloch eigenstates $|\mathbf{k},\mathrm{n},\sigma\rangle$ to orbital Wannier states $|\mathbf{R},b,\sigma\rangle$,

$$|\mathbf{R},b,\sigma\rangle = \sum_{\mathbf{k},n} F_{(\mathbf{k},\mathrm{n}),(\mathbf{R},b);\sigma} |\mathbf{k},\mathrm{n},\sigma\rangle \ , \ F_{(\mathbf{k},\mathrm{n}),(\mathbf{R},b);\sigma} = \langle \mathbf{k},\mathrm{n},\sigma | \mathbf{R},b,\sigma \rangle \ . \tag{106}$$

The Wannier orbitals

$$\phi_{\mathbf{R},b,\sigma}(\mathbf{r}) = \langle \mathbf{r} | \mathbf{R},b,\sigma \rangle \tag{107}$$

are maximal around a lattice site $\mathbf{R}$ and the orbital index b resembles atomic quantum numbers, e.g., $b = 3s, 3p, 3d$. In the orbital Wannier basis the field operators are given by

$$\hat{\Psi}^\dagger_\sigma(\mathbf{r}) = \sum_{\mathbf{R},b} \phi^*_{\mathbf{R},b,\sigma}(\mathbf{r}) \hat{c}^\dagger_{\mathbf{R},b,\sigma} \quad , \quad \hat{\Psi}_\sigma(\mathbf{r}) = \sum_{\mathbf{R},b} \phi_{\mathbf{R},b,\sigma}(\mathbf{r}) \hat{c}_{\mathbf{R},b,\sigma} \ , \tag{108}$$

and the Kohn-Sham Hamiltonian in the orbital Wannier basis becomes

$$\hat{H}^{\mathrm{KS}} = \sum_{\mathbf{R},b,\mathbf{R}',b',\sigma} T^{\mathrm{KS}}_{(\mathbf{R},b),(\mathbf{R}',b');\sigma} \hat{c}^\dagger_{\mathbf{R},b,\sigma} \hat{c}_{\mathbf{R}',b',\sigma} \tag{109}$$

with the overlap matrix elements

$$T^{\mathrm{KS}}_{(\mathbf{R},b),(\mathbf{R}',b');\sigma} = \int \mathrm{d}\mathbf{r} \ \phi^*_{\mathbf{R},b,\sigma}(\mathbf{r}) h^{\mathrm{KS}}_\sigma(\mathbf{r}) \phi_{\mathbf{R}',b',\sigma}(\mathbf{r}) \ , \tag{110}$$

see eq. (41). These matrix elements appear in a tight-binding representation of the kinetic energy in Hubbard-type models.

We also define the orbital Bloch basis,

$$\begin{aligned} \phi_{\mathbf{k},b,\sigma}(\mathbf{r}) &= \sqrt{\frac{1}{L}} \sum_{\mathbf{R}} e^{\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \phi_{\mathbf{R},b,\sigma}(\mathbf{r}) \ , \\ \phi_{\mathbf{R},b,\sigma}(\mathbf{r}) &= \sqrt{\frac{1}{L}} \sum_{\mathbf{k}} e^{-\mathrm{i}\mathbf{k}\cdot\mathbf{R}} \phi_{\mathbf{k},b,\sigma}(\mathbf{r}) \ , \end{aligned} \tag{111}$$

where $\mathbf{k}$ is from the first Brillouin zone and L is the number of lattice sites. The field operators are given by

$$\hat{\Psi}^\dagger_\sigma(\mathbf{r}) = \sum_{\mathbf{k},b} \phi^*_{\mathbf{k},b,\sigma}(\mathbf{r}) \hat{c}^\dagger_{\mathbf{k},b,\sigma} \quad , \quad \hat{\Psi}_\sigma(\mathbf{r}) = \sum_{\mathbf{k},b} \phi_{\mathbf{k},b,\sigma}(\mathbf{r}) \hat{c}_{\mathbf{k},b,\sigma} \ . \tag{112}$$

In the orbital Wannier basis, the Kohn-Sham single-particle Hamiltonian reads

$$\begin{aligned} \hat{H}^{\rm KS} &= \sum_{\mathbf{k},b,b',\sigma} T^{\rm KS}_{b,b';\sigma}(\mathbf{k}) \hat{c}^\dagger_{\mathbf{k},b,\sigma} \hat{c}_{\mathbf{k},b',\sigma} \ , \\ T^{\rm KS}_{b,b';\sigma}(\mathbf{k}) &= \int {\rm d}\mathbf{r}\ \phi^*_{\mathbf{k},b,\sigma}(\mathbf{r}) h^{\rm KS}_\sigma(\mathbf{r}) \phi_{\mathbf{k},b',\sigma}(\mathbf{r}) \ . \end{aligned} \tag{113}$$

A.2 Atomic Hamiltonian in cubic symmetry

We choose the Hubbard parameters $U(u,v), U(\zeta,\zeta), U(\xi,\eta), U(\zeta,u), U(\zeta,v)$, the four Hund's-rule couplings $J(u,v), J(\xi,\eta), J(\zeta,u), J(\zeta,v)$, and the two-particle transfer matrix element $S(\eta,\xi;\zeta,u)$ as our ten independent Coulomb matrix elements in cubic symmetry. The other matrix elements in eq. (83) can be expressed as [26, 17]

$$\begin{aligned} U(u,u) &= U(v,v) &&= U(u,v) + 2J(u,v) \ , \\ U(\xi,u) &= U(\eta,u) &&= (U(\zeta,u) + 3U(\zeta,v))/4 \ , \\ U(\xi,v) &= U(\eta,v) &&= (3U(\zeta,u) + U(\zeta,v))/4 \ , \\ J(\xi,u) &= J(\eta,u) &&= (J(\zeta,u) + 3J(\zeta,v))/4 \ , \\ J(\xi,v) &= J(\eta,v) &&= (3J(\zeta,u) + J(\zeta,v))/4 \ , \\ T(\eta;u,v) &= -T(\xi;u,v) &&= \sqrt{3}(U(\zeta,u) - U(\zeta,v))/4 \ , \\ A(\eta;u,v) &= -A(\xi;u,v) &&= \sqrt{3}(J(\zeta,u) - J(\zeta,v))/4 \ , \\ & S(\xi,\eta;\zeta,u) &&= S(\eta,\xi;\zeta,u) \ , \\ & S(\zeta,\xi;\eta,u) &&= -2S(\eta,\xi;\zeta,u) \ , \\ & S(\xi,\eta;\zeta,v) &&= -\sqrt{3}S(\eta,\xi;\zeta,u) \ , \\ & S(\zeta,\xi;\eta,u) &&= \sqrt{3}S(\eta,\xi;\zeta,u) \ . \end{aligned} \tag{114}$$

If we further assume that the radial part of the t_{2g}-orbitals and the e_g-orbitals are identical ('spherical approximation'), we may express all matrix elements in terms of three parameters, e.g., the Racah parameters A, B, and C that are related to the Slater-Condon parameters by $A = F^{(0)} - F^{(4)}/9$, $B = (F^{(2)} - 5F^{(4)}/9)/49$, and $C = 5F^{(4)}/63$; inversely, $F^{(0)} = A + 7C/5$, $F^{(2)} = 49B + 7C$, $F^{(4)} = 63C/5$. In particular,

$$\begin{aligned} U(u,v) &= A - 4B + C \ , \\ J(u,v) &= 4B + C \ , \\ U(\zeta,\zeta) &= A + 4B + 3C \ , \\ U(\xi,\eta) &= A - 2B + C \ , \\ J(\xi,\eta) &= 3B + C \ , \\ U(\zeta,u) &= A - 4B + C \ , \\ U(\zeta,v) &= A + 4B + C \ , \\ J(\zeta,v) &= C \ , \\ J(\zeta,u) &= 4B + C \ , \\ S(\eta,\xi;\zeta,u) &= -\sqrt{3}B \ . \end{aligned} \tag{115}$$

The average Coulomb interaction between electrons in same orbitals is given by

$$U = \frac{1}{5} \sum_{c=\xi,\eta,\zeta,u,v} U(c,c) = A + 4B + 3C \ , \tag{116}$$

the average Coulomb interaction between electrons in different orbitals is given by

$$U' = \frac{1}{10} \sum_{c,c'=\xi,\eta,\zeta,u,v;c<c'} U(c,c') = A - B + C \ , \tag{117}$$

and the average Hund's-rule coupling becomes

$$J = \frac{1}{10} \sum_{c,c'=\xi,\eta,\zeta,u,v;c<c'} J(c,c') = \frac{5}{2}B + C \ . \tag{118}$$

These three quantities are not independent but related by the symmetry relation $U' = U - 2J$. This means that by choosing two of these parameters (e.g., U and J) the three Racah parameters, and therefore all the parameters in Eq. (80) are not uniquely defined. Hence, we use the additional relation $C/B = 4$ which is a reasonable assumption for transition metals [26]. It corresponds to $F^{(2)}/F^{(4)} = 55/36 = 1.53$, in agreement with the estimate $F^{(2)}/F^{(4)} \approx 1.60 = 8/5$ by de Groot et al. [37].

References

[1] For an overview, see R.M. Martin, *Electronic Structure* (Cambridge Univ. Press, 2004); R.M. Dreizler and E.K.U. Gross, *Density Functional Theory* (Springer, Berlin, 1990)

[2] V.L. Moruzzi, J.F. Janak, and A.R. Williams, *Calculated Electronic Properties of Metals* (Pergamon Press, New York, 1978)

[3] J. Hubbard, Proc. Roy. Soc. A **276**, 238 (1963)

[4] M.C. Gutzwiller, Phys. Rev. Lett. **10**, 159 (1963)

[5] V.I. Anisimov, F. Aryasetiawan, and A.I. Lichtenstein, J. Phys.: Condens. Matter **9**, 767 (1997)

[6] G. Kotliar, S. Savrasov, K. Haule, V. Oudovenko, O. Parcollet, and C. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)

[7] E. Pavarini, E. Koch, D. Vollhardt, and A. Lichtenstein (eds.): *The LDA+DMFT approach to strongly correlated materials*, Modeling and Simulation, Vol. 1 (Forschungszentrum Jülich,2 011) `http://www.cond-mat.de/events/correl11`

[8] J. Bünemann, F. Gebhard, and W. Weber in A. Narlikar (ed.): *Frontiers in Magnetic Materials* (Springer, Berlin, 2005), p. 117.

[9] K.M. Ho, J. Schmalian, and C.Z. Wang, Phys. Rev. B **77**, 073101 (2008); Y.-X. Yao, J. Schmalian, C.Z. Wang, H.M. Ho, and G. Kotliar, Phys. Rev. B **84**, 245112 (2011); N. Lanatà, Y.-X. Yao, C.Z. Wang, K.M. Ho, J. Schmalian, K. Haule, and G. Kotliar, Phys. Rev. Lett. **111**, 196801 (2013)

[10] G.-T. Wang, X. Dai, and Z. Fang, Phys. Rev. Lett. **101**, 066403 (2008); X. Deng, X. Dai, and Z. Fang, Eur. Phys. Lett. **83**, 37008 (2008); X. Deng, L. Wang, X. Dai, and Z. Fang, Phys. Rev. B **79**, 075114 (2009); G.-T. Wang, Y. Qian, G. Xu, X. Dai, and Z. Fang, Phys. Rev. Lett. **104**, 047002 (2010); M.-F. Tian, X. Deng, Z. Fang, and X. Dai, Phys. Rev. B **84**, 205124 (2011)

[11] R. Dong, X. Wan, X. Dai, and S.Y. Savrasov, Phys. Rev. B **89**, 165122 (2014)

[12] T. Schickling, J. Bünemann, F. Gebhard, and W. Weber, New J. Phys. **16**, 93034 (2014)

[13] T. Schickling, J. Bünemann, F. Gebhard, and L. Boeri, Phys. Rev. B **93**, 205151 (2016)

[14] M. Levy, Phys. Rev. A **26**, 1200 (1982); E.H. Lieb, Int. J. Quan. Chem. **24**, 243 (1983)

[15] The condition (22) is a bit too strong. We can equally work with the functional $\widetilde{D}_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right] = D_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right] + \Delta\left[\{n_\sigma(\mathbf{r})\}\right]$ as long as we demand $\Delta\left[\{n^0_\sigma(\mathbf{r})\}\right] = 0$ and $(\partial\Delta\left[\{n_\sigma(\mathbf{r})\}\right])/(\partial n_\sigma(\mathbf{r}))|_{n_\sigma(\mathbf{r})=n^0_\sigma(\mathbf{r})} = 0$. Then, $\widetilde{D}_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right]$ leads to the same ground-state energy and ground-state density as $D_{\rm sp}\left[\{n_\sigma(\mathbf{r})\}\right]$.

[16] J. Bünemann Chap. 5 in E. Pavarini, E. Koch, F. Anders, and M. Jarrell (eds.): *Correlated Electrons: From Models to Materials* Modeling and Simulation, Vol. 2 (Forschungszentrums Jülich, 2012) `http://www.cond-mat.de/events/correl12`.

[17] J. Bünemann and F. Gebhard, J. Phys.: Condens. Matt. **29**, 165601 (2017)

[18] F. Gebhard, Phys. Rev. B **41**, 9452 (1990)

[19] J. Bünemann, T. Schickling, and F. Gebhard, Europhys. Lett. **98**, 27006 (2012)

[20] J. Kaczmarczyk, J. Spałek, T. Schickling, and J. Bünemann, Phys. Rev. B **88**, 115127 (2013)

[21] K. zu Münster and J. Bünemann, Phys. Rev. B **94**, 045135 (2016)

[22] J. Bünemann, T. Linneweber, U. Löw, F.B. Anders, and F. Gebhard, Phys. Rev. B **94**, 035116 (2016)

[23] J. Bünemann, T. Linneweber, and F. Gebhard, phys. stat. sol. (b) **254**, 1600166 (2017)

[24] J. Bünemann, W. Weber, and F. Gebhard, Phys. Rev. B **57**, 6896 (1998)

[25] J. Bünemann, T. Schickling, F. Gebhard, and W. Weber, physica status solidi (b) **249**, 1282 (2012)

[26] S. Sugano, Y. Tanabe, and H. Kamimura, *Multiplets of Transition-Metal Ions in Crystals* Pure and Applied Physics **33** (Academic Press, New York, 1970)

[27] A.S. Belozerov and V.I. Anisimov, J. Phys.: Condens. Matt. **26**, 375601 (2014)

[28] A.A. Katanin, A.I. Poteryaev, A.V. Efremov, A.O. Shorikov, S.L. Skornyakov, M.A. Korotin, and V.I. Anisimov, Phys. Rev. B **81**, 045117 (2010)

[29] I. Leonov, A.I. Poteryaev, V.I. Anisimov, and D. Vollhardt, Phys. Rev. Lett. **106**, 106405 (2011)

[30] I. Leonov, A.I. Poteryaev, Y.N. Gornostyrev, A.I. Lichtenstein, M.I. Katsnelson, V.I. Anisimov, and D. Vollhardt, Scientific Reports **4**, 5585 (2014)

[31] A.I. Lichtenstein, M.I. Katsnelson, and G. Kotliar, Phys. Rev. Lett. **87**, 067205 (2001)

[32] L.V. Pourovskii, J. Mravlje, M. Ferrero, O. Parcollet, and I.A. Abrikosov, Phys. Rev. B **90**, 155120 (2014)

[33] T. Miyake, F. Aryasetiawan, and M. Imada, Phys. Rev. B **80**, 155134 (2009)

[34] T. Schickling, F. Gebhard, and J. Bünemann, Phys. Rev. Lett. **106**, 146402 (2011)

[35] T. Schickling, F. Gebhard, J. Bünemann, L. Boeri, O.K. Andersen, and W. Weber, Phys. Rev. Lett. **108**, 036406 (2012)

[36] F. Aryasetiawan, K. Karlsson, O. Jepsen, and U. Schönberger, Phys. Rev. B **74**, 125106 (2006)

[37] F.M.F. de Groot, J.C. Fuggle, B.T. Thole, and G.A. Sawatzky, Phys. Rev. B **42**, 5459 (1990)

[38] J. Schäfer, M. Hoinkis, E. Rotenberg, P. Blaha, and R. Claessen, Phys. Rev. B **72**, 155115 (2005)

[39] J. Sánchez-Barriga, J. Fink, V. Boni, I. Di Marco, J. Braun, J. Minár, A. Varykhalov, O. Rader, V. Bellini, F. Manghi, H. Ebert, M.I. Katsnelson, A.I. Lichtenstein, O. Eriksson, W. Eberhardt, and H.A. Dürr, Phys. Rev. Lett. **103**, 267203 (2009)

[40] J. Bünemann, F. Gebhard, and R. Thul, Phys. Rev. B **67**, 075103 (2003)

[41] D.-J. Zhao, K. Albe, and H. Hahn, Scripta Materialia **55**, 473 (2006)

[42] J. Bünemann, F. Gebhard, T. Ohm, R. Umstätter, S. Weiser, W. Weber, R. Claessen, D. Ehm, A. Harasawa, A. Kakizaki, A. Kimura, G. Nicolay, S. Shin, and V.N. Strocov, Europhys. Lett. **61**, 667 (2003)

[43] J. Bünemann, F. Gebhard, T. Ohm, S. Weiser, and W. Weber, Phys. Rev. Lett. **101**, 236404 (2008)

[44] A. Hofmann, X. Cui, J. Schäfer, S. Meyer, P. Höpfner, C. Blumenstein, M. Paul, L. Patthey, E. Rotenberg, J. Bünemann, F. Gebhard, T. Ohm, W. Weber, and R. Claessen, Phys. Rev. Lett. **102**, 187204 (2009)

[45] C.S. Wang, B.M. Klein, and H. Krakauer, Phys. Rev. Lett. **54**, 1852 (1985)

[46] S.K. Bose, O.V. Dolgov, J. Kortus, O. Jepsen, and O.K. Andersen, Phys. Rev. B **67**, 214518 (2003)

[47] G. Steinle-Neumann, L. Stixrude, and R.E. Cohen, Phys. Rev. B**60**, 791 (1999)

[48] G.Y. Guo and H.H. Wang, Chin. J. of Phys. **38**, 949 (2000)

[49] H.L. Zhang, S. Lu, M.P.J. Punkkinen, Q.-M. Hu, B. Johansson, and L. Vitos, Phys. Rev. B **82**, 132409 (2010)

[50] H. Danan, A. Herr, and A.J.P. Meyer, J. Appl. Phys. **39**, 669 (1968)

[51] D. Bancroft, E.L. Peterson, and S. Minshall, J. Appl. Phys. **27**, 291 (1956)

[52] A.P. Jephcoat, H.K. Mao, and P.M. Bell, J. Geophys. Res.: Solid Earth **91**, 4677 (1986)

[53] A.M. Turner, A.W. Donoho, and J.L. Erskine, Phys. Rev. B **29**, 2986 (1984)

[54] G. Seibold, F. Becca, and J. Lorenzana, Phys. Rev. B **78**, 045114 (2008)

[55] J. Bünemann, M. Capone, J. Lorenzana, and G. Seibold, New J. Phys. **10**, 053050 (2013)

[56] J. Bünemann, S. Wasner, E. von Oelsen, and G. Seibold, Phil. Mag. **95**, 550 (2015)

9 Mott transition: DFT+U vs DFT+DMFT

Eva Pavarini
Institute for Advanced Simulation
Forschungszentrum Jülich

Contents

1 Introduction **2**

2 The Hubbard model **6**
2.1 Introduction 6
2.2 The Hubbard dimer 10

3 The Anderson model **20**
3.1 Introduction 20
3.2 The Anderson molecule 24
3.3 Anderson molecule vs Hubbard dimer 25

4 DMFT and DFT+DMFT **26**
4.1 Method 26
4.2 Model building in DFT+DMFT 28

5 Metal-insulator transition **30**
5.1 Hartree-Fock method 30
5.2 HF vs DMFT 34
5.3 DFT+U vs DFT+DMFT 35

6 Conclusions **39**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

One of the early successes of quantum mechanics was explaining the difference between metals and insulators. The core of this theory is the *independent-electron picture*. In the latter, the electronic states of a given periodic system, a crystal, are described via a set of bands filled following the Pauli principle. As a result, two cases are possible: in the first, each band is either completely filled or totally empty (*band insulator*), and in the second, some of the bands are only partially filled (*conventional metal*). In a system in which all bands are either full or empty, a finite energy is required to bring one electron from the ground state to the lowest-lying excited state. Indeed, an insulator can be viewed as a system with an *energy gap* in the excitation spectrum. The energy gap is not uniquely defined, since it has a different nature depending on the experimental tool used to measure it. Photoemission and inverse photoemission probe the spectral function. The latter yields the charge gap

$$E^c_{\rm gap} = E_0(N+1) + E_0(N-1) - 2E_0(N),$$

where $E_0(N)$ is the ground-state energy for N electrons. This is the difference between the ionization energy, $I = E_0(N-1) - E_0(N)$ and electronic affinity, $A = E_0(N) - E_0(N+1)$. In the independent-electron picture, at $T = 0$

$$E_0(N) = 2\sum_{m\boldsymbol{k}} \varepsilon_{m\boldsymbol{k}}\, \Theta(-\varepsilon_{m\boldsymbol{k}} + \varepsilon_F),$$

where m is an index labelling different bands, $\varepsilon_{m\mathbf{k}}$ the band dispersion, ε_F the Fermi level and $\Theta(x)$ the Heaviside step function. Thus the charge gap is basically identical to the difference

$$E^o_{\rm gap} = E_1(N) - E_0(N).$$

where $E_1(N)$ is the energy of the N-electron first excited state (Fig. 1). The energy difference $E^o_{\rm gap}$ can be directly probed in experiments which do not change the number of electrons, e.g., absorption spectroscopy. In the presence of a gap, the finite-temperature properties are typically characterized by an activation energy ΔE. For example, the static optical conductivity of a band insulator has, in first approximation, the low-temperature form

$$\sigma(T) \sim \sigma_0(T) e^{-\Delta E/2k_BT},$$

where ΔE is the band gap and $\sigma_0(T)$ a prefactor. The size of the gap varies from system to system, giving rise to different behaviors and appearances. Representative examples of band insulators are two well known materials with the same crystal structure and yet rather different properties, diamond and silicon. Diamond is transparent thanks to its large gap (~ 5.5 eV). Silicon has a smaller gap (~ 1.1 eV), a gray color and could be taken at a first glance for a metal. A conventional metal behaves very differently than a band insulator, however. In a conventional metal, since some bands are partially empty, it is possible to excite electrons with infinitesimal energy. Thus, e.g., the conductivity is finite even at $T = 0$

$$\sigma(0) \sim \frac{ne^2\tau}{m_e},$$

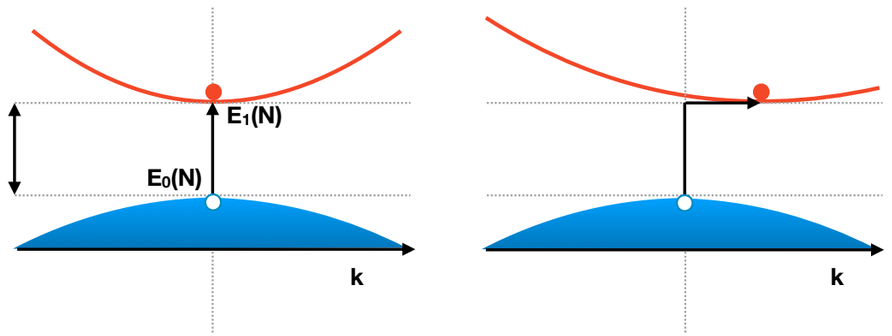


Fig. 1: *The band gap $E_g^o = E_1(N) - E_0(N)$ in the independent-electron picture. Left: direct gap. Right: indirect gap. Silicon and diamond both have an indirect gap. Blue: top of the filled valence band. Red: bottom of the empty conduction band.*

where n is the electron density and τ the average time between two collisions. Classical examples of conventional metals are gold, silver, and copper. They are all characterized by shiny metallic colors. Have we then explained all matter via the rather simple independent-electron band theory? One could naively think that this is, indeed, the case. Reality, however, has always surprises in store. It became clear early on that the independent-electron theory is not the complete story. Some transition-metal oxides, which were supposed to be good metals in the independent-electron picture, turned out to be either insulators or very bad conductors. It was soon understood that a possible cause of the anomalous behavior could be the electron-electron Coulomb repulsion; the latter could localize electrons giving rise to a metal-insulator transition (MIT). That things are very different when the electronic Coulomb interaction is taken into account can be seen already in Fig. 2, which shows the charge gap for an idealized atom made by a single level $\varepsilon_d < 0$ occupied by one electron. If we assume that the electrons do not interact ($U = 0$), the charge gap is zero

$$E^c_{\rm gap} = [E_0(2) - E_0(1)] + [E_0(0) - E_0(1)] = \varepsilon_d - \varepsilon_d = 0.$$

If, however, electrons repel each other ($U \neq 0$), the gap is finite

$$E^c_{\rm gap} = U.$$

Let us define *strongly-correlated systems* the materials whose behavior qualitatively differs from the independent-electron picture because of the electron-electron repulsion. While the theory of conventional metals and band insulators is rather straightforward, the theory of the MIT in strongly-correlated systems has kept theoreticians busy for almost a century, and still the problem is only partially solved. This happens because, when the independent-electron picture fails, we are confronted with the hardness of the quantum many-body problem. The latter can already be grasped by looking at the classical N body problem (Fig. 3), describing masses interacting via gravity. When only one body is present, there is no interaction, and the problem

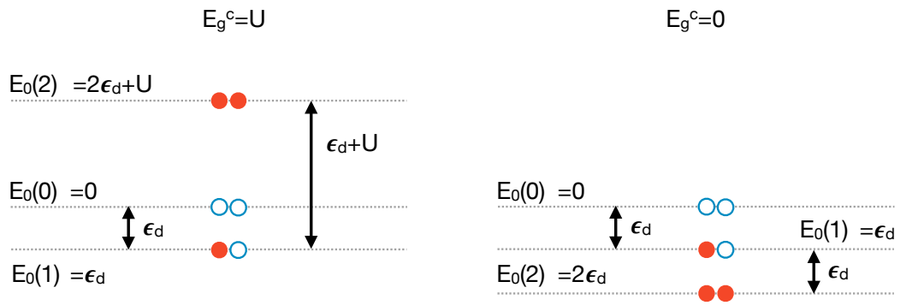


Fig. 2: *The charge gap for an idealized atom described by a level $\varepsilon_d < 0$ occupied by one electron ($N = 1$). Left: When two electrons are on the same level, the system has energy $2\varepsilon_d + U$, where U is the electron-electron repulsion. Right: Non-interacting-electron picture.*

is trivial. For two bodies we can find the analytical solution by working in the center of mass and relative coordinates system. The general three-body problem is a major challenge [1] and it can lead to chaotic behavior; the complexity of the N-body problem grows dramatically with the number of bodies involved [2]. Quantum effects further add to the complications, and the exact solution of the many-body problem is totally out of reach. Even if we cannot count on the exact diagonalization of the full many-body Hamiltonian, however, this does not imply the end of physics. We can still explain the origin of specific co-operative behaviors, such as the nature of the insulating or metallic state; we need, however, to first identify the core nature of the phenomenon along with the relevant effective entities involved, and then build the corresponding effective theory. A natural question arises at this point. Although we know that it eventually fails for strongly-correlated systems, the independent-electron picture is very appealing for its simplicity. Furthermore it works rather well for many systems, at least in first approximation; we have already mentioned among insulators silicon and diamond, and among metals gold, silver or copper. Could we then perhaps explain strongly-correlated insulators without leaving the independent-electron picture, via, e.g, a Coulomb-induced one-electron potential of some type? Or do we really need a more complex theory, in which true many-body effects – those that cannot be reduced to a simple potential – are key? Let us call a system for which the first picture applies *Slater insulator* and one for which the second picture is relevant *Mott insulator*. The answer to the question above is important also in view of the fact that, while solving exactly the many-body problem is basically impossible, we do have very advanced tools to solve material-specific one-electron-like Hamiltonians. These are *ab-initio* methods based on density-functional theory (DFT) in the local-density approximation or its simple extensions. Is it possible to find a *simple* potential that captures the essential nature of the MIT and can be easily embedded in DFT-based codes? The DFT+U method [3] is one of the most important attempts in this direction; in this approach Coulomb repulsion effects are treated at the *static mean-field* level and they are then essentially reduced to a spin-, site- and orbital-

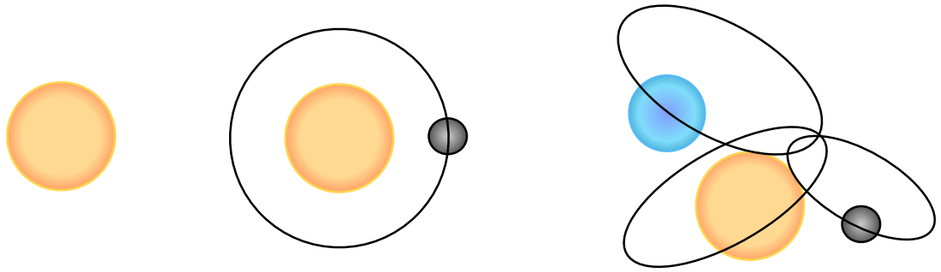


Fig. 3: *The increasing complexity of the classical N-body problem. One body: no interaction. Two bodies: we can find the solution analytically. Shown is a solution describing a lighter body rotating about a heavier body. Three-body: chaotic solutions are possible.*

dependent potential. The resulting MIT is of the Slater type, and it occurs at the onset of long-range magnetic order. Unfortunately, correlated insulators typically do not behave in this way, however. Although most of them have a magnetic ground state, above the magnetic transition temperature T_N they usually remain insulators; furthermore, they typically behave as local-moment paramagnets with Curie-Weiss magnetic susceptibilities. Instead, in the non-magnetic phase DFT+U yields metallic Pauli-like paramagnets. This shows that some crucial aspects are missing in DFT+U. Which ones, however? To answer to this question, one can use an alternative approach. This consist in giving up the band picture and DFT completely, switch to simple representative models and try to solve them beyond the static mean-field level via many-body techniques. Even in simple models, however, truly strongly-correlated phenomena, which escape a static mean-field description, remain a challenge. An example is the Kondo effect, which was solved only after decades of struggle, and the solution lead to the developments of new theoretical approaches such as the numerical renormalization group. In the case of the metal-insulator transition, the breakthrough was the *dynamical mean-field theory* (DMFT) [4–8]. This method was at first designed to solve the one-band *Hubbard model*. It consists in mapping the lattice Hubbard model into a self-consistent quantum-impurity model, described for example by the *Anderson Hamiltonian*. The DMFT technique succeeds in describing the Coulomb-driven transition from paramagnetic metal to local-moment paramagnetic insulator. Furthermore, it can be used for solving material-specific Hubbard models built from DFT-based calculations; this is the DFT+DMFT approach. In this lecture, after an introduction to the Hubbard and the Anderson Hamiltonian, we will discuss some of the basic ideas behind both the DFT+U and the DFT+DMFT method. We will compare the very different pictures of the metal-insulator transition emerging from the two approaches, the first of the Slater and the second of the Mott type. As a concluding remark, it is important to remember that band, Slater, and Mott insulators do not exhaust all possible types of insulators. Electron localization can, e.g., also occur because of disorder alone. This phenomenon is known as Anderson localization [9]. Although the latter is a very important and interesting effect, we will not discuss it in this lecture.

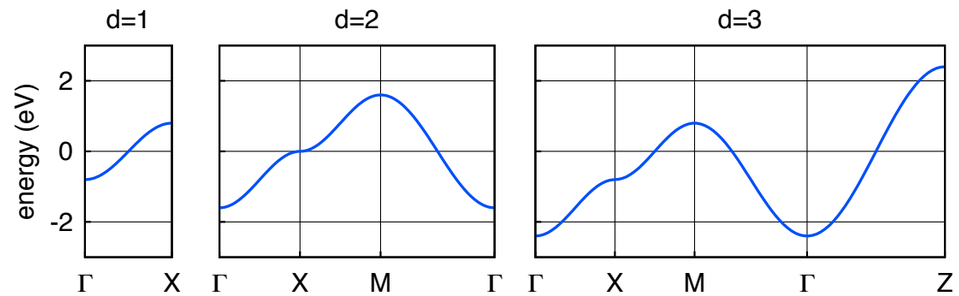


Fig. 4: *Band structure of the one-band tight-binding model (hypercubic lattice). The hopping integral is* $t = 0.4$ *eV. From left to right: one-, two-, and three-dimensional case. At half filling* ($n = 1$) *the Fermi level is at zero energy. The* $\boldsymbol{k}$ *points are* $\Gamma = (0,0,0)$, $X = (\pi/a, 0, 0)$, $M = (\pi/a, \pi/a, 0)$, *and* $Z = (0, 0, \pi/a)$.

2 The Hubbard model

2.1 Introduction

The simplest lattice model describing a correlated system is the one-band Hubbard model

$$\hat{H} = \varepsilon_d \underbrace{\sum_i \sum_\sigma c^\dagger_{i\sigma} c_{i\sigma}}_{\hat{H}_d} \; \underbrace{-t \sum_{\langle ii' \rangle} \sum_\sigma c^\dagger_{i\sigma} c_{i'\sigma}}_{\hat{H}_T} + \underbrace{U \sum_i \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}}_{\hat{H}_U} = \hat{H}_d + \hat{H}_T + \hat{H}_U, \qquad (1)$$

where ε_d is the on-site energy, t is the hopping integral between first-nearest neighbors $\langle ii' \rangle$, and U the on-site Coulomb repulsion; $c^\dagger_{i\sigma}$ creates an electron in a Wannier state with spin σ centered at site i, and $\hat{n}_{i\sigma} = c^\dagger_{i\sigma} c_{i\sigma}$.

In the $U = 0$ limit the Hubbard model describes a system of independent electrons. The Hamiltonian is then diagonal in the Bloch basis

$$\hat{H}_d + \hat{H}_T = \sum_{\boldsymbol{k}\sigma} \Big[\varepsilon_d + \varepsilon_{\boldsymbol{k}} \Big] c^\dagger_{\boldsymbol{k}\sigma} c_{\boldsymbol{k}\sigma}.$$

The energy dispersion $\varepsilon_{\boldsymbol{k}}$ depends on the geometry and dimensionality d of the lattice. For a hypercubic lattice of dimension d

$$\varepsilon_{\boldsymbol{k}} = -2t \sum_{\nu=1}^{d} \cos(k_{r_\nu} a),$$

where a is the lattice constant, and $r_1 = x, r_2 = y, r_3 = z$. The energy $\varepsilon_{\boldsymbol{k}}$ does not depend on the spin. In Fig. 4 we show $\varepsilon_{\boldsymbol{k}}$ in the one-, two- and three-dimensional cases. The corresponding density of states is shown in Fig. 5.

In the opposite limit ($t = 0$) the Hubbard model describes a collection of isolated atoms. Each

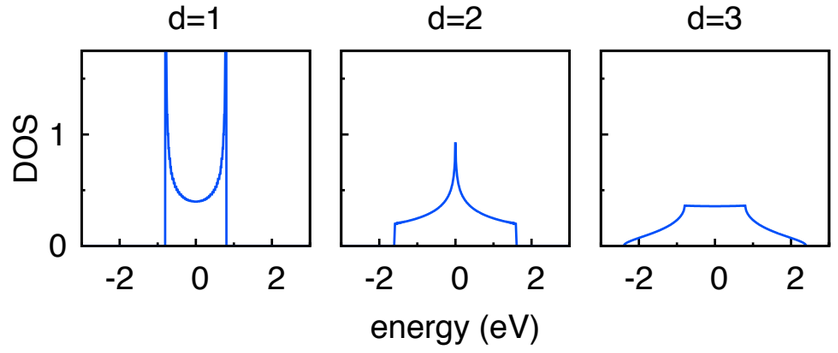


Fig. 5: *Density of states (DOS) per spin, $\rho(\varepsilon)/2$, for a hypercubic lattice in one, two, and three dimensions. The energy dispersion is calculated for $t = 0.4$ eV. The curves exhibit different types of Van-Hove singularities.*

atom has four electronic many-body states

$\|N, S, S_z\rangle$		N	S	$E(N)$
$\|0, 0, 0\rangle$	$= \quad \|0\rangle$	0	0	0
$\|1, \frac{1}{2}, \uparrow\rangle$	$= \quad c^\dagger_{i\uparrow}\|0\rangle$	1	1/2	ε_d
$\|1, \frac{1}{2}, \downarrow\rangle$	$= \quad c^\dagger_{i\downarrow}\|0\rangle$	1	1/2	ε_d
$\|2, 0, 0\rangle$	$= c^\dagger_{i\uparrow}c^\dagger_{i\downarrow}\|0\rangle$	2	0	$2\varepsilon_d + U$

where $E(N)$ is the total energy, N the total number of electrons and S the total spin. We can express the atomic Hamiltonian $\hat{H}_d + \hat{H}_U$ in a form in which the dependence on $\hat{N}_i$, $\hat{S}_i$, and $\hat{S}^i_z$ is explicitly given

$$\hat{H}_d + \hat{H}_U = \varepsilon_d \sum_i \hat{n}_i + U \sum_i \left[-\left(\hat{S}^i_z\right)^2 + \frac{\hat{n}^2_i}{4} \right],$$

where $\hat{S}^i_z = (\hat{n}_{i\uparrow} - \hat{n}_{i\downarrow})/2$ is the z component of the spin operator and $\hat{n}_i = \sum_\sigma \hat{n}_{i\sigma} = \hat{N}_i$.
In the large t/U limit and at half filling we can downfold charge fluctuations and map the Hubbard model into an effective spin model of the form

$$\hat{H}_S = \frac{1}{2}\Gamma \sum_{\langle ii' \rangle} \left[\boldsymbol{S}_i \cdot \boldsymbol{S}_{i'} - \frac{1}{4}\hat{n}_i\hat{n}_{i'} \right]. \tag{2}$$

The coupling Γ can be calculated by using second-order perturbation theory. For a state in which two neighbors have opposite spin, $|\uparrow,\downarrow\rangle = c^\dagger_{i\uparrow}c^\dagger_{i'\downarrow}|0\rangle$, we obtain the energy gain

$$\Delta E_{\uparrow\downarrow} \sim -\sum_I \langle\uparrow,\downarrow|\hat{H}_T|I\rangle\langle I \left| \frac{1}{E(2) + E(0) - 2E(1)} \right| I\rangle\langle I|\hat{H}_T|\uparrow,\downarrow\rangle \sim -\frac{2t^2}{U}.$$

Here $|I\rangle$ ranges over the excited states with one of the two neighboring sites doubly occupied and the other empty, $|\uparrow\downarrow,0\rangle = c^\dagger_{i\uparrow}c^\dagger_{i\downarrow}|0\rangle$, or $|0,\uparrow\downarrow\rangle = c^\dagger_{i'\uparrow}c^\dagger_{i'\downarrow}|0\rangle$; these states can be occupied via virtual hopping processes. For a state in which two neighbors have parallel spins, $|\uparrow,\uparrow\rangle = c^\dagger_{i\uparrow}c^\dagger_{i'\uparrow}|0\rangle$, no virtual hopping is possible because of the Pauli principle, and $\Delta E_{\uparrow\uparrow} = 0$. Thus

$$\frac{1}{2}\Gamma \sim (\Delta E_{\uparrow\uparrow} - \Delta E_{\uparrow\downarrow}) = \frac{1}{2}\frac{4t^2}{U}. \tag{3}$$

The exchange coupling $\Gamma = 4t^2/U$ is positive, i.e., antiferromagnetic.
Canonical transformations [10] provide a scheme for deriving the effective spin model systematically at any perturbation order. Let us consider a unitary transformation of the Hamiltonian

$$\hat{H}_S = e^{i\hat{S}}\hat{H}e^{-i\hat{S}} = \hat{H} + \Big[i\hat{S},\hat{H}\Big] + \frac{1}{2}\Big[i\hat{S},\Big[i\hat{S},\hat{H}\Big]\Big] + \dots.$$

We search for a transformation operator that eliminates, at a given order, hopping integrals between states with a different number of doubly-occupied states. To do this, first we split the kinetic term $\hat{H}_T$ into a component $\hat{H}^0_T$ that does not change the number of doubly-occupied states and two terms that either increase it ($\hat{H}^+_T$) or decrease it ($\hat{H}^-_T$) by one

$$\hat{H}_T = -t\sum_{\langle ii'\rangle}\sum_\sigma c^\dagger_{i\sigma}c_{i'\sigma} = \hat{H}^0_T + \hat{H}^+_T + \hat{H}^-_T,$$

where

$$\hat{H}^0_T = -t\sum_{\langle ii'\rangle}\sum_\sigma \hat{n}_{i-\sigma}\, c^\dagger_{i\sigma}c_{i'\sigma}\, \hat{n}_{i'-\sigma} - t\sum_{\langle ii'\rangle}\sum_\sigma \Big[1-\hat{n}_{i-\sigma}\Big]c^\dagger_{i\sigma}c_{i'\sigma}\Big[1-\hat{n}_{i'-\sigma}\Big],$$

$$\hat{H}^+_T = -t\sum_{\langle ii'\rangle}\sum_\sigma \hat{n}_{i-\sigma}\, c^\dagger_{i\sigma}c_{i'\sigma}\Big[1-\hat{n}_{i'-\sigma}\Big],$$

$$\hat{H}^-_T = \left(\hat{H}^+_T\right)^\dagger.$$

The term $\hat{H}^0_T$ commutes with $\hat{H}_U$. The remaining two terms fulfill the commutation rules

$$[\hat{H}^\pm_T, \hat{H}_U] = \mp U\hat{H}^\pm_T.$$

The operator $\hat{S}$ can be expressed as a linear combination of powers of the three operators $\hat{H}^0_T, \hat{H}^+_T$, and $\hat{H}^-_T$. The actual combination, which gives the effective spin model at a given order, can be found via a recursive procedure [10]. At half filling and second order, however, we can simply guess the form of $\hat{S}$ that leads to the Hamiltonian (2). By defining

$$\hat{S} = -\frac{i}{U}\left(\hat{H}^+_T - \hat{H}^-_T\right)$$

we obtain

$$\hat{H}_S = \hat{H}_U + \hat{H}^0_T + \frac{1}{U}\Big(\Big[\hat{H}^+_T,\hat{H}^-_T\Big] + \Big[\hat{H}^0_T,\hat{H}^-_T\Big] + \Big[\hat{H}^+_T,\hat{H}^0_T\Big]\Big) + \mathcal{O}(U^{-2}).$$

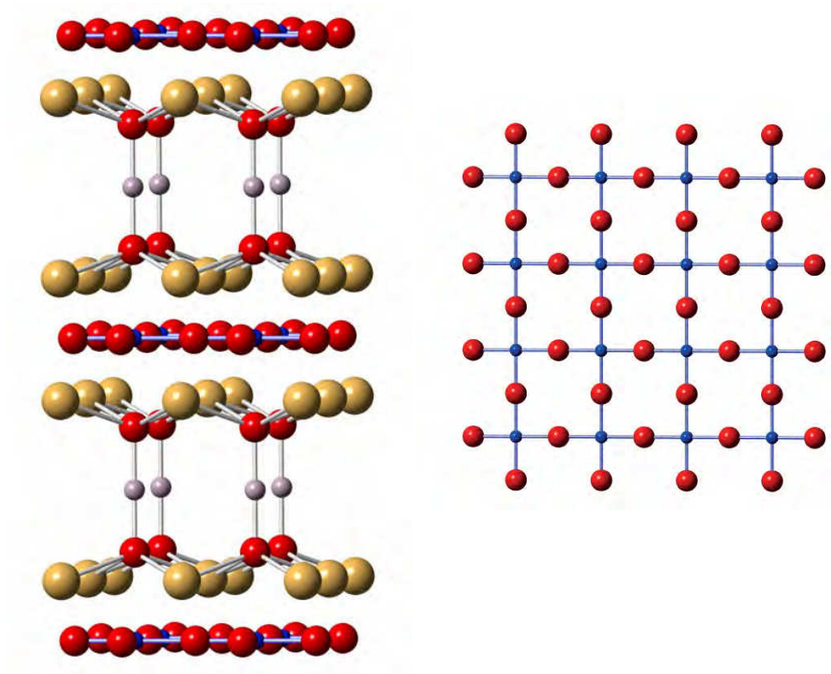


Fig. 6: *Left: Crystal structure of $HgBa_2CuO_4$ showing the two-dimensional CuO_2 layers. Spheres represent atoms of Cu (blue), O (red), Ba (yellow), and Hg (grey). Right: A CuO_2 layer. The hopping integral t between neighboring Cu sites is $t \sim 4t_{pd}^2/\Delta_{dp}$, where t_{pd} is the hopping between Cu d and O p states and $\Delta_{dp} = \varepsilon_d - \varepsilon_p$ their charge-transfer energy.*

If we restrict the Hilbert space of $\hat{H}_S$ to the subspace with one electron per site (half filling), no hopping is possible without increasing the number of doubly-occupied states; hence, only the term $\hat{H}_T^- \hat{H}_T^+$ contributes. After some algebra, we obtain $\hat{H}_S = \hat{H}_S^{(2)} + \mathcal{O}(U^{-2})$ with

$$\hat{H}_S^{(2)} = \frac{1}{2}\frac{4t^2}{U}\sum_{ii'}\left[\boldsymbol{S}_i \cdot \boldsymbol{S}_{i'} - \frac{1}{4}\hat{n}_i\hat{n}_{i'}\right].$$

The Hubbard model (1) is rarely realized in nature in this form. To understand real materials one typically has to take into account orbital degrees of freedom, long-range hopping integrals, and sometimes longer-range Coulomb interactions or perhaps even more complex many-body terms. Nevertheless, there are very interesting systems whose low-energy properties are, to first approximation, described by (1). These are strongly-correlated organic crystals [11] (one-dimensional case) and high-temperature superconducting cuprates [12], in short HTSCs (two-dimensional case). An example of HTSC is $HgBa_2CuO_4$, whose structure is shown in Fig. 6. It is made of CuO_2 planes well divided by BaO-Hg-BaO blocks. The x^2-y^2-like states stemming from the CuO_2 planes can be described via a one-band Hubbard model. The presence of a $x^2 - y^2$-like band at the Fermi level is a common feature of all HTSCs.

2.2 The Hubbard dimer

The Hubbard model cannot be solved exactly. It is thus interesting to consider an even simpler model, for which we can find analytically eigenvectors and eigenvalues. This is the Hubbard dimer, whose Hamiltonian is given by

$$\hat{H} = \varepsilon_d \sum_{i\sigma} n_{i\sigma} - t \sum_{\sigma} \left[c^\dagger_{1\sigma} c_{2\sigma} + c^\dagger_{2\sigma} c_{1\sigma} \right] + U \sum_{i=1,2} \hat{n}_{i\uparrow} \hat{n}_{i\downarrow}. \tag{4}$$

2.2.1 Exact diagonalization

Hamiltonian (4) commutes with the number of electron operator $\hat{N}$, the total spin $\hat{S}$ and $\hat{S}_z$. In the atomic limit, the eigenstates states can be therefore classified as

$\|N,S,S_z\rangle$			N	S	$E(N,S)$
$\|0,0,0\rangle$	$=$	$\|0\rangle$	0	0	0
$\|1,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}\|0\rangle$	1	$1/2$	ε_d
$\|1,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}\|0\rangle$	1	$1/2$	ε_d
$\|2,1,1\rangle$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{1\uparrow}\|0\rangle$	2	1	$2\varepsilon_d$
$\|2,1,-1\rangle$	$=$	$c^\dagger_{2\downarrow}c^\dagger_{1\downarrow}\|0\rangle$	2	1	$2\varepsilon_d$
$\|2,1,0\rangle$	$=$	$\frac{1}{\sqrt{2}}\left[c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} + c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right]\|0\rangle$	2	1	$2\varepsilon_d$
$\|2,0,0\rangle_0$	$=$	$\frac{1}{\sqrt{2}}\left[c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} - c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right]\|0\rangle$	2	0	$2\varepsilon_d$
$\|2,0,0\rangle_1$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	2	0	$2\varepsilon_d + U$
$\|2,0,0\rangle_2$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	2	0	$2\varepsilon_d + U$
$\|3,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	3	$1/2$	$3\varepsilon_d + U$
$\|3,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	3	$1/2$	$3\varepsilon_d + U$
$\|4,0,0\rangle$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	4	0	$4\varepsilon_d + 2U$

Let us order the $N = 1$ states as in the table above, first the spin up and then spin down block. For finite t the Hamiltonian matrix for $N = 1$ electrons takes then the form

$$\hat{H}_1 = \begin{pmatrix} \varepsilon_d & -t & 0 & 0 \\ -t & \varepsilon_d & 0 & 0 \\ 0 & 0 & \varepsilon_d & -t \\ 0 & 0 & -t & \varepsilon_d \end{pmatrix}.$$

This matrix can be easily diagonalized and yields the *bonding* $(-)$ and *antibonding* $(+)$ states

$\lvert 1,S,S_z\rangle_\alpha$		$E_\alpha(1,S)$	$d_\alpha(1,S)$
$\lvert 1,1/2,\sigma\rangle_+$	$= \frac{1}{\sqrt{2}}\left[\lvert 1,1/2,\sigma\rangle_1 - \lvert 1,1/2,\sigma\rangle_2\right]$	$\varepsilon_d + t$	2
$\lvert 1,1/2,\sigma\rangle_-$	$= \frac{1}{\sqrt{2}}\left[\lvert 1,1/2,\sigma\rangle_1 + \lvert 1,1/2,\sigma\rangle_2\right]$	$\varepsilon_d - t$	2

where $d_\alpha(N)$ is the spin degeneracy of the α manifold. Let us now increase the total number of electrons. For $N = 2$ electrons (half filling), the hopping integrals only couple the three $S = 0$ states, and therefore the Hamiltonian matrix is given by

$$\hat{H}_2 = \begin{pmatrix} 2\varepsilon_d & 0 & 0 & 0 & 0 & 0 \\ 0 & 2\varepsilon_d & 0 & 0 & 0 & 0 \\ 0 & 0 & 2\varepsilon_d & 0 & 0 & 0 \\ 0 & 0 & 0 & 2\varepsilon_d & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & 2\varepsilon_d + U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & 2\varepsilon_d + U \end{pmatrix}.$$

The eigenvalues and the corresponding (normalized) eigenvectors are

$\lvert 2,S,S_z\rangle_\alpha$		$E_\alpha(2,S)$	$d_\alpha(2,S)$
$\lvert 2,0,0\rangle_+$	$= a_1\lvert 2,0,0\rangle_0 - \frac{a_2}{\sqrt{2}}\left[\lvert 2,0,0\rangle_1 + \lvert 2,0,0\rangle_2\right]$	$2\varepsilon_d + \frac{1}{2}\left[U + \Delta(t,U)\right]$	1
$\lvert 2,0,0\rangle_o$	$= \frac{1}{\sqrt{2}}\left[\lvert 2,0,0\rangle_1 - \lvert 2,0,0\rangle_2\right]$	$2\varepsilon_d + U$	1
$\lvert 2,1,m\rangle_o$	$= \lvert 2,1,m\rangle$	$2\varepsilon_d$	3
$\lvert 2,0,0\rangle_-$	$= a_2\lvert 2,0,0\rangle_0 + \frac{a_1}{\sqrt{2}}\left[\lvert 2,0,0\rangle_1 + \lvert 2,0,0\rangle_2\right]$	$2\varepsilon_d + \frac{1}{2}\left[U - \Delta(t,U)\right]$	1

where

$$\Delta(t,U) = \sqrt{U^2 + 16t^2},$$

and $a_1 a_2 = 2t/\Delta(t,U)$. For $U = 0$ we have $a_1 = a_2 = 1/\sqrt{2}$, and the two states $|2,0,0\rangle_-$ and $|2,0,0\rangle_+$ become, respectively, the state with two electrons in the bonding orbital and the state with two electrons in the antibonding orbital; they have energy $E_\pm(2,0) = 2\varepsilon_d \pm 2t$; the remaining states have energy $2\varepsilon_d$ and are non-bonding. For $t > 0$, the ground state is unique and it is always the singlet $|2,0,0\rangle_-$; in the large U limit its energy is

$$E_-(2,0) \sim 2\varepsilon_d - 4t^2/U.$$

In this limit the energy difference between the first excited state, a triplet state, and the singlet ground state is thus equal to the Heisenberg antiferromagnetic coupling

$$E_o(2,1) - E_-(2,0) \sim 4t^2/U = \Gamma.$$

Finally, for $N = 3$ electrons, eigenstates and eigenvectors are

$\|3, S, S_z\rangle_\alpha$	$E_\alpha(3)$	$d_\alpha(3, S)$
$\|3, 1/2, \sigma\rangle_+ = \frac{1}{2}\left[\|1, 1/2, \sigma\rangle_1 + \|1, 1/2, \sigma\rangle_2\right]$	$3\varepsilon_d + U + t$	2
$\|3, 1/2, \sigma\rangle_- = \frac{1}{2}\left[\|1, 1/2, \sigma\rangle_1 - \|1, 1/2, \sigma\rangle_2\right]$	$3\varepsilon_d + U - t$	2

If we exchange holes and electrons, the $N = 3$ case is identical to the $N = 1$ electron case. This is due to the particle-hole symmetry of the model.

2.2.2 Local Matsubara Green function

Let us now calculate the local Matsubara Green function for site i, defined as

$$G_{ii,\sigma}(i\nu_n) = -\int_0^\beta d\tau e^{i\nu_n\tau}\langle \mathcal{T} c_{i\sigma}(\tau) c^\dagger_{i\sigma}(0)\rangle,$$

where $\mathcal{T}$ is the time-ordering operator and ν_n a fermionic Matsubara frequency. We use to this end the Lehmann representation

$$G_{ii,\sigma}(i\nu_n) = \frac{1}{Z}\sum_{nn'N} e^{-\beta(E_n(N)-\mu N)}\left[\frac{|\langle n'N-1|c_{i\sigma}|nN\rangle|^2}{i\nu_n - [E_n(N) - E_{n'}(N-1) - \mu]} + \frac{|\langle n'N+1|c^\dagger_{i\sigma}|nN\rangle|^2}{i\nu_n - [E_{n'}(N+1) - E_n(N) - \mu]}\right], \tag{5}$$

where $|nN\rangle$ is the N-electron eigenstate with energy $E_n(N)$, $\beta = 1/k_B T$, μ is the chemical potential, and Z the partition function. In order to calculate the Green function (5) we thus need all eigenstates and their energies; from the eigenstates we have to compute the weights $w_i^\sigma = |\langle n'N'|\hat{o}_{i\sigma}|nN\rangle|^2$, where $\hat{o}_{i\sigma}$ is either $c_{i\sigma}$ or $c^\dagger_{i\sigma}$. The Green function is by symmetry identical for spin up and spin down, and for site 1 and site 2. Thus it is sufficient to perform the calculation for $i = 1$ and $\sigma =\uparrow$. In the atomic limit, the only non-zero terms are collected in the table shown in the next page; in the first half of the table $\hat{o}_{1\uparrow} = c_{1\uparrow}$ and $N' = N$, and in the

second half of the table $\hat{o}_{1\uparrow} = c^\dagger_{1\uparrow}$ and $N' = N + 1$

$\hat{o}_{1\uparrow}\lvert N, S, S_z\rangle$			$w_1^\uparrow$	$E_n(N') - E_{n'}(N'-1)$
$c_{1\uparrow}\lvert 1, 1/2, \sigma\rangle_1$	$=$	$\delta_{\sigma,\uparrow}\lvert 0\rangle$	1	ε_d
$c_{1\uparrow}\lvert 2, 1, 0\rangle$	$=$	$\frac{1}{\sqrt{2}}c^\dagger_{2\downarrow}\lvert 0\rangle$	$\frac{1}{2}$	ε_d
$c_{1\uparrow}\lvert 2, 1, 1\rangle$	$=$	$-c^\dagger_{2\uparrow}\lvert 0\rangle$	1	ε_d
$c_{1\uparrow}\lvert 2, 0, 0\rangle_0$	$=$	$\frac{1}{\sqrt{2}}c^\dagger_{2\downarrow}\lvert 0\rangle$	$\frac{1}{2}$	ε_d
$c_{1\uparrow}\lvert 2, 0, 0\rangle_1$	$=$	$c^\dagger_{1\downarrow}\lvert 0\rangle$	1	$\varepsilon_d + U$
$c_{1\uparrow}\lvert 3, 1/2, \sigma\rangle_1$	$=$	$\delta_{\sigma,\uparrow}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\lvert 0\rangle$	1	ε_d
$c_{1\uparrow}\lvert 3, 1/2, \sigma\rangle_2$	$=$	$-c^\dagger_{2\sigma}c^\dagger_{1\downarrow}\lvert 0\rangle$	1	$\varepsilon_d + U$
$c_{1\uparrow}\lvert 4, 0, 0\rangle$	$=$	$c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\lvert 0\rangle$	1	$\varepsilon_d + U$
$c^\dagger_{1\uparrow}\lvert 0, 0, 0\rangle$	$=$	$c^\dagger_{1\uparrow}\lvert 0\rangle$	1	ε_d
$c^\dagger_{1\uparrow}\lvert 1, 1/2, \sigma\rangle_1$	$=$	$\delta_{\sigma,\downarrow}c^\dagger_{1\uparrow}c^\dagger_{1\sigma}\lvert 0\rangle$	1	$\varepsilon_d + U$
$c^\dagger_{1\uparrow}\lvert 1, 1/2, \sigma\rangle_2$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{2\sigma}\lvert 0\rangle$	1	ε_d
$c^\dagger_{1\uparrow}\lvert 2, 1, 0\rangle$	$=$	$\frac{1}{\sqrt{2}}c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\lvert 0\rangle$	$\frac{1}{2}$	$\varepsilon_d + U$
$c^\dagger_{1\uparrow}\lvert 2, 1, -1\rangle$	$=$	$-c^\dagger_{2\downarrow}c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\lvert 0\rangle$	1	$\varepsilon_d + U$
$c^\dagger_{1\uparrow}\lvert 2, 0, 0\rangle_0$	$=$	$-\frac{1}{\sqrt{2}}c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\lvert 0\rangle$	$\frac{1}{2}$	$\varepsilon_d + U$
$c^\dagger_{1\uparrow}\lvert 2, 0, 0\rangle_2$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\lvert 0\rangle$	1	ε_d
$c^\dagger_{1\uparrow}\lvert 3, 1/2, \sigma\rangle_1$	$=$	$\delta_{\sigma,\downarrow}c^\dagger_{1\uparrow}c^\dagger_{1\sigma}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\lvert 0\rangle$	1	$\varepsilon_d + U$

For $t \neq 0$ we have to recalculate the weights because the eigenstates are different. Let us first exploit the mirror symmetry of the Hamiltonian, however; thanks to it, any hermitian quadratic operator is diagonal in the basis of the bonding and anti-bonding state. Thus the local Green function can be expressed as the average of the bonding and antibonding one

$$G_{11,\sigma}(i\nu_n) = \frac{1}{2}\left[G_{++,\sigma} + G_{--,\sigma}\right]$$

where

$$G_{\pm\pm,\sigma}(i\nu_n) = -\int_0^\beta d\tau e^{i\nu_n\tau}\langle \mathcal{T} c_{\pm\sigma}(\tau)c^\dagger_{\pm\sigma}(0)\rangle,$$

and

$$c_{\pm\sigma} = \frac{1}{\sqrt{2}} \left(c_{1\uparrow} \mp c_{2\uparrow} \right).$$

For $U = 0$, the local Green function is thus simply

$$G^0_{11,\sigma}(i\nu_n) = \frac{1}{2} \sum_{\alpha=\pm} \frac{1}{i\nu_n - (\varepsilon_\alpha - \mu)} = \frac{1}{i\nu_n - (\varepsilon_d + F^0(i\nu_n) - \mu)},$$

where $\varepsilon_\pm = \varepsilon_d \pm t$. The quantity

$$F^0(i\nu_n) = \frac{t^2}{i\nu_n - (\varepsilon_d - \mu)},$$

is the so-called non-interacting hybridization function, and it can be seen as a self-energy for the uncorrelated atomic level ε_d. Let us now suppose that we are in the opposite limit, the one in which $4t \ll U$, and hence $E_-(2,0) \sim E_o(2,1)$. Furthermore, let us assume that $k_B T$ is much lower than the energy difference $E_o(2,0) - E_-(2,0)$; this implies that the two higher-energy states in the 2-electron sector can be neglected in calculating the Green function. In this limit the local Matsubara Green function is given by

$$\begin{aligned} G_{11,\sigma}(i\nu_n) &\sim \frac{1}{4} \sum_{\alpha=\pm} \left[\frac{1}{i\nu_n - (\varepsilon_\alpha - \mu)} + \frac{1}{i\nu_n - (\varepsilon_\alpha + U - \mu)} \right] \\ &= \frac{1}{2} \sum_{\alpha=\pm} \frac{1}{i\nu_n - (\varepsilon_\alpha - \mu + \Sigma_{\alpha\alpha}(i\nu_n))}. \end{aligned}$$

The bonding and antibonding self-energy are

$$\Sigma_{\alpha\alpha}(i\nu_n) = \frac{U}{2} + \frac{U^2}{4} \frac{1}{i\nu_n - (\varepsilon_\alpha + \frac{1}{2}U - \mu)}.$$

In the large frequency limit, as will become clear later, the exact self-energy equals the Hartree-Fock self-energy for zero magnetization, $U/2$. The gap is given by

$$E^c_g = E_0(N+1) + E_0(N-1) - 2E_0(N) \sim U - 2t.$$

The formulas above show that the self-energy is different for the bonding and antibonding state. By making the analogy with an infinite tight-binding chain with dispersion $-2(t/2)\cos ka$, the bonding state corresponds to $k = 0$ and the anti-bonding state to $k = \pi/a$. Thus, in the lattice limit, our result reflects the fact that in general the self-energy depends on $\boldsymbol{k}$. In addition, the gap, which has the value U in the atomic limit, is reduced by the energy difference between antibonding and bonding state, $2t$; in the lattice limit, this difference becomes the band-width, W. The reason of the gap reduction is that, once we remove or add one electron, it does not cost Coulomb energy to move the hole/extra electron from one site to the other in the Hubbard dimer. Finally, we can rewrite the local Green function in a form that will become useful later

$$G_{11,\sigma}(i\nu_n) = \left[\frac{1}{i\nu_n - (\varepsilon_d - \mu + \Sigma_{l\sigma}(i\nu_n) + F_\sigma(i\nu_n))} \right]. \tag{6}$$

In this expression $\Sigma_l(i\nu_n)$ is the local self-energy

$$\begin{aligned}\Sigma_l(i\nu_n) =& \frac{1}{2}(\Sigma_{++}(i\nu_n) + \Sigma_{--}(i\nu_n)),\\ =& \frac{U^2}{4}\,\frac{1}{i\nu_n - (\varepsilon_d + \frac{1}{2}U - \mu + \frac{t^2}{(i\nu_n - (\varepsilon_d + \frac{1}{2}U - \mu))})},\end{aligned}$$

and $F_\sigma(i\nu_n)$ the hybridization function for the correlated dimer

$$F_\sigma(i\nu_n) = \frac{(t + \Delta\Sigma_l(i\nu_n))^2}{i\nu_n - (\varepsilon_d - \mu + \Sigma_{l\sigma}(i\nu_n))}.$$

The difference

$$\begin{aligned}\Delta\Sigma_{l\sigma}(i\nu_n) =& \frac{1}{2}(\Sigma_{++}(i\nu_n) - \Sigma_{--}(i\nu_n))\\ =& \frac{U^2}{4}\,\frac{t}{(i\nu_n - (\varepsilon_d + \frac{1}{2}U - \mu))^2 - t^2},\end{aligned}$$

measures the strength of non-local effects. The sum $F_l(i\nu_n) + \Sigma_l(i\nu_n)$ yields the total modification of the isolated ($t = 0$) and uncorrelated ($U = 0$) level ε_d. Later we will compare expression (6) to its analogous for another simple model, the Anderson molecule.

2.2.3 Long-range Coulomb interaction

A natural question that follows is: what happens if the Coulomb repulsion is longer range? For a dimer, extending the Coulomb interaction to first neighbors leads to the Hamiltonian

$$\begin{aligned}\hat{H} =& \varepsilon_d \sum_{i\sigma} \hat{n}_{i\sigma} - t \sum_{\sigma} \left[c^\dagger_{1\sigma} c_{2\sigma} + c^\dagger_{2\sigma} c_{1\sigma}\right] + U \sum_{i=1,2} \hat{n}_{i\uparrow}\hat{n}_{i\downarrow}\\ &+ \sum_{\sigma\neq\sigma'} (V - 2J_V - J_V\delta_{\sigma\sigma'})\hat{n}_{1\sigma}\hat{n}_{2\sigma'} - J_V \sum_{i\neq i'} \left[c^\dagger_{i\uparrow} c_{i\downarrow} c^\dagger_{i'\downarrow} c_{i'\uparrow} + c^\dagger_{i'\uparrow} c^\dagger_{i'\downarrow} c_{i\uparrow} c_{i\downarrow}\right],\end{aligned}$$

where the parameters in the last two terms are the intersite direct (V) and exchange (J_V) Coulomb interaction. For two electrons the Hamiltonian becomes

$$\hat{H}_2 = \begin{pmatrix} 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 & 0 & 0\\ 0 & 2\varepsilon_d + V - 3J_V & 0 & 0 & 0 & 0\\ 0 & 0 & 2\varepsilon_d + V - 3J_V & 0 & 0 & 0\\ 0 & 0 & 0 & 2\varepsilon_d + V - J_V & -\sqrt{2}t & -\sqrt{2}t\\ 0 & 0 & 0 & -\sqrt{2}t & U & -J_V\\ 0 & 0 & 0 & -\sqrt{2}t & -J_V & U \end{pmatrix}.$$

Thus, if $J_V = 0$, apart from an irrelevant shift, the Hamiltonian at half-filling equals the $\hat{H}_2$ matrix that we obtained for $V = 0$, provided that in the latter U is replaced by $U - V$; hence the

V term effectively reduces the strength of the local Coulomb interaction and at the same time enhances the exchange coupling, which becomes $\Gamma \sim 4t^2/(U-V)$. What about the charge gap? Let us calculate the gap exactly, without assuming $4t \ll U$ as we have done previously. This leads to the formula

$$E_g^c(V) = -2t + V + \sqrt{(U-V)^2 + 16t^2}.$$

Let us consider the case in which $4t/U$ is small. There are two interesting limits. The first, $V/U \to 0$, yield the previous result, $E_g(V) \sim U - 2t$. The second is $V/U \sim 1$, which gives $E_g(V) \sim 2t+V$. In this case the gap equals the one of an uncorrelated dimer with enhanced hopping integrals, $t \to t+V/2$. In this limit, the elements of the matrices $\hat{H}_N$ are basically identical to those we obtained for $U = V = J_V = 0$, apart for a shift on the diagonal; thus also the eigenstates are close to those of the non-interacting dimer. Although for realistic lattices the effect of V is more complex [11], the simple result above explains why actual strong-correlation effects mostly appear when the local Coulomb coupling is large compared to longer-range terms.

2.2.4 Hartree-Fock approximation

Let us now compare the exact solution of the Hubbard dimer with the result of the Hartree-Fock approximation. Here we return for simplicity to the case $V = J_V = 0$. The Hartree-Fock Hamiltonian can be obtained by replacing

$$\hat{H}_U = U \sum_i \hat{n}_{i\uparrow}\hat{n}_{i\downarrow} \to \hat{H}_U^{\rm HF} = U \sum_i [\hat{n}_{i\uparrow}\bar{n}_{i\downarrow} + \hat{n}_{i\downarrow}\bar{n}_{i\uparrow} - \bar{n}_{i\uparrow}\bar{n}_{i\downarrow}], \tag{7}$$

where $\bar{n}_{i\sigma}$ is the HF expectation value of the operator $\hat{n}_{i\sigma}$. Thus we have

$$\hat{H}^{\rm HF} = \varepsilon_d \sum_{i\sigma} \hat{n}_{i\sigma} - t \sum_\sigma \left[c_{1\sigma}^\dagger c_{2\sigma} + c_{2\sigma}^\dagger c_{1\sigma} \right] + U \sum_{\sigma\neq\sigma'} [\hat{n}_{1\sigma}\bar{n}_{1\sigma'} + \hat{n}_{2\sigma}\bar{n}_{2\sigma'}] - U \sum_i \bar{n}_{i\uparrow}\bar{n}_{i\downarrow}.$$

It is convenient to introduce the quantities

$$n_i = \bar{n}_{i\uparrow} + \bar{n}_{i\downarrow} \qquad n = \frac{1}{2}(n_1 + n_2) \qquad \delta n = \frac{1}{2}(n_1 - n_2)$$
$$m_i = \frac{1}{2}(\bar{n}_{i\uparrow} - \bar{n}_{i\downarrow}) \qquad m_+ = \frac{1}{2}(m_1 + m_2) \qquad m_- = \frac{1}{2}(m_1 - m_2)$$

Inverting these relations

$$n_{1\uparrow} = (m_+ + m_-) + (n + \delta n)/2 \qquad n_{1\downarrow} = -(m_+ + m_-) + (n + \delta n)/2$$
$$n_{2\uparrow} = (m_+ - m_-) + (n - \delta n)/2 \qquad n_{2\downarrow} = -(m_+ - m_-) + (n - \delta n)/2$$

The Hartree-Fock version of the Hubbard dimer Hamiltonian equals the non-interacting Hamiltonian plus a shift of the on-site level. This shift depends on the site and the spin

$$\begin{aligned}
\hat{H}_{\mathrm{HF}} &= \sum_{i\sigma} \left(\varepsilon_d + \Delta_{i\sigma}\right) \hat{n}_{i\sigma} - t \sum_{\sigma} \left(c^{\dagger}_{1\sigma} c_{2\sigma} + c^{\dagger}_{2\sigma} c_{1\sigma} \right) - \Delta_0 \\
\Delta_0 &= 2U \left[\frac{n^2 + \delta n^2}{4} - m_+^2 - m_-^2 \right] \\
\Delta_{i\sigma} &= U \left[(-1)^{\sigma} \left(m_+ + (-1)^{i-1} m_- \right) + \frac{1}{2} (n + (-1)^{i-1} \delta n) \right].
\end{aligned}$$

Thus we can write immediately the local Green function matrix for site 1. It is convenient to use this time the site basis, hence, to calculate the matrix $G_{ii',\sigma}(i\nu_n)$. Then we have

$$G_{11,\sigma}(i\nu_n) = \begin{bmatrix} i\nu_n - (\varepsilon_d - \mu + \Sigma_{11,\sigma}(i\nu_n)) & t \\ t & i\nu_n - (\varepsilon_d - \mu + \Sigma_{22,\sigma}(i\nu_n)) \end{bmatrix}^{-1}_{11}$$

where

$$\Sigma_{ii,\sigma}(i\nu_n) = \Delta_{i\sigma}.$$

This shows that the self-energy is not dependent on the frequency, i.e., Hartree-Fock is a *static* mean-field approach. The value of the parameters m_+, m_- and δn have to be found solving the system of self-consistent equations given by

$$\bar{n}_{i\sigma} = \frac{1}{\beta} \sum_n e^{-i\nu_n 0^-} G_{ii,\sigma}(i\nu_n).$$

For ferromagnetic (F) and antiferromagnetic (AF) solutions we have, in the absence of charge disproportionation, the following simplifications

$$\begin{aligned}
\Delta^{\mathrm{F}}_{1\sigma} &= U \left(\frac{n}{2} + \sigma m_+ \right) & \Delta^{\mathrm{AF}}_{1\sigma} &= U \left(\frac{n}{2} + \sigma m_- \right) \\
\Delta^{\mathrm{F}}_{2\sigma} &= U \left(\frac{n}{2} + \sigma m_+ \right) & \Delta^{\mathrm{AF}}_{2\sigma} &= U \left(\frac{n}{2} - \sigma m_- \right)
\end{aligned}$$

In the AF case, the self-energy depends on the site. In the lattice limit, this implies that the interaction couples $\boldsymbol{k}$ states. Indeed, by rewriting the Green-function matrix in the basis of the bonding $(k = 0)$ and anti-bonding $(k = \pi)$ creation/annihilation operators we have

$$G_{\sigma}(i\nu_n) = \frac{1}{2} \begin{bmatrix} i\nu_n - (\varepsilon_d - t - \mu + \frac{1}{2}\sum_i \Sigma_{i\sigma}(i\nu_n)) & \frac{1}{2}\sum_i (-1)^{i-1} \Sigma_{i\sigma}(i\nu_n) \\ \frac{1}{2}\sum_i (-1)^{i-1} \Sigma_{i\sigma}(i\nu_n) & i\nu_n - (\varepsilon_d + t - \mu + \frac{1}{2}\sum_i \Sigma_{i\sigma}(i\nu_n)) \end{bmatrix}^{-1}.$$

The diagonal terms are identical, hence

$$\Sigma_{++,\sigma}(i\nu_n) = \Sigma_{--,\sigma}(i\nu_n) = \Sigma_l(i\nu_n).$$

The off-diagonal terms $\Sigma_{+-}(i\nu_n)$ and $G_{+-}(i\nu_n)$ are not zero, however. This tells us that, by introducing the HF correction, we can lower the symmetry of the system. Let us now calculate

explicitly the eigenstates for different fillings. For this it is sufficient to diagonalize $\hat{H}_1$, the Hamiltonian in the 1-electron sector; the many-electron states can be obtained by filling the one-electron states via the Pauli principle. The Hamiltonian $\hat{H}_1$ can be written as $\hat{H}_1 = \hat{H}_1' + \varepsilon_d \hat{N} - \Delta_0$, and, in the AF case we then have

$$\hat{H}_1' = \begin{pmatrix} U(\frac{1}{2}n - m_-) & -t & 0 & 0 \\ -t & U(\frac{1}{2}n + m_-) & 0 & 0 \\ 0 & 0 & U(\frac{1}{2}n + m_-) & -t \\ 0 & 0 & -t & U(\frac{1}{2}n - m_-) \end{pmatrix}.$$

This leads to the (normalized) states

$\|1\rangle_l$	$E_l(1)$
$\|1\rangle_3 = a_2\|1,1/2,\uparrow\rangle_1 - a_1\|1,1/2,\uparrow\rangle_2$	$\varepsilon_0(1) + \Delta_1(t,U)$
$\|1\rangle_2 = a_1\|1,1/2,\downarrow\rangle_1 - a_2\|1,1/2,\downarrow\rangle_2$	$\varepsilon_0(1) + \Delta_1(t,U)$
$\|1\rangle_1 = a_1\|1,1/2,\uparrow\rangle_1 + a_2\|1,1/2,\uparrow\rangle_2$	$\varepsilon_0(1) - \Delta_1(t,U)$
$\|1\rangle_0 = a_2\|1,1/2,\downarrow\rangle_1 + a_1\|1,1/2,\downarrow\rangle_2$	$\varepsilon_0(1) - \Delta_1(t,U)$

where $\varepsilon_0(1) = \varepsilon_d + U(1/2 + 2m_-^2 - n^2/2)$ and $a_1^2 = \frac{1}{2}\left(1 + \frac{Um_-}{\Delta_1(t,U)}\right)$. The charge gap at half filling is

$$E_g^{\rm HF} = 2\Delta_1(t,U) = 2\sqrt{(m_- U)^2 + t^2}.$$

In general the Hartree-Fock gap tends to be larger than the exact value. If we assume that only the ground state is occupied, solving the self-consistent equations yields the solutions

$$m_- = 0 \quad \text{or} \quad m_- = \frac{1}{2}\sqrt{1 - \frac{4t^2}{U^2}}.$$

Using this result we find $E_g^{\rm HF} = U$. It is useful to look more in detail at $\hat{H}_2'$, with $\hat{H}_2 = \hat{H}_2' + \varepsilon_d \hat{N} - \Delta_0$; in the absence of charge disproportionation, it has the general form

$$\hat{H}_2' = \begin{pmatrix} U & 0 & 0 & -2Um_- & 0 & 0 \\ 0 & U(1-2m_+) & 0 & 0 & 0 & 0 \\ 0 & 0 & U(1+2m_+) & 0 & 0 & 0 \\ -2Um_- & 0 & 0 & U & -\sqrt{2}t & -\sqrt{2}t \\ 0 & 0 & 0 & -\sqrt{2}t & U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t & 0 & U \end{pmatrix}$$

If we search for an AF solution, the normalized Hartree-Fock eigenvalues and eigenvectors are

$\|2\rangle_l$	$E_l(2)$
$\|2\rangle_5 = \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_0 + a_2\|2,1,0\rangle - \frac{a_1}{\sqrt{2}}\left[\|2,0,0\rangle_1 + \|2,0,0\rangle_2\right]\right]$	$\varepsilon_0(2) + 2\Delta_1(t,U)$
$\|2\rangle_4 = \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_1 - \|2,0,0\rangle_2\right]$	$\varepsilon_0(2)$
$\|2\rangle_3 = \|2,1,1\rangle$	$\varepsilon_0(2)$
$\|2\rangle_2 = \|2,1,-1\rangle$	$\varepsilon_0(2)$
$\|2\rangle_1 = a_1\|2,1,0\rangle + a_2\frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_1 + \|2,0,0\rangle_2\right]$	$\varepsilon_0(2)$
$\|2\rangle_0 = \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_0 - a_2\|2,1,0\rangle + \frac{a_1}{\sqrt{2}}\left[\|2,0,0\rangle_1 + \|2,0,0\rangle_2\right]\right]$	$\varepsilon_0(2) - 2\Delta_1(t,U)$

where $\varepsilon_0(2) = 2\varepsilon_d + U(1+2m_-^2 - n^2/2)$, and $a_1^2 = t^2/\Delta_1^2(t,U)$. There are several observations to make. The Hartree-Fock ground state has an overlap with the correct ground state, however incorrectly mixes triplet and singlet states, thus breaking the rotational symmetry of the model. For this reason, its energy, in the large U limit, is $2\varepsilon_d - 2t^2/U$ and not $2\varepsilon_d - 4t^2/U$ as in the exact case. For a F solution, the eigenvalues and eigenvectors are

$\|2\rangle_l$	$E_l(2)$
$\|2\rangle_5 = \|2,1,-1\rangle$	$\varepsilon_0^+(2) + 2Um_+$
$\|2\rangle_4 = \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_0 - \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_1 + \|2,0,0\rangle_2\right]\right]$	$\varepsilon_0^+(2) + 2t$
$\|2\rangle_3 = \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_1 - \|2,0,0\rangle_2\right]$	$\varepsilon_0^+(2)$
$\|2\rangle_2 = \|2,1,\quad 0\rangle$	$\varepsilon_0^+(2)$
$\|2\rangle_1 = \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_0 + \frac{1}{\sqrt{2}}\left[\|2,0,0\rangle_1 + \|2,0,0\rangle_2\right]\right]$	$\varepsilon_0^+(2) - 2t$
$\|2\rangle_0 = \|2,1,\quad 1\rangle$	$\varepsilon_0^+(2) - 2Um_+$

where $\varepsilon_0^+(2) = 2\varepsilon_d + U(1+2m_+^2 - n^2/2)$. The ferromagnetic Hartree-Fock correction thus yields an incorrect sequence of levels; the ground state for large U/t, indicated as $|2\rangle_0$ in the table, has no overlap with the exact ground state of the Hubbard dimer. It is, instead, one of the states of the first excited triplet. The energy difference between F- and AF-magnetic ground state is

$$E_{AF} - E_F \sim -\frac{2t^2}{U},$$

which is indeed the exact energy difference between antiferromagnetic and ferromagnetic state. It does not correspond, however, to the actual singlet-triplet excitation energy, $\Gamma \sim 4t^2/U$.

3 The Anderson model

3.1 Introduction

A magnetic impurity in a metallic host can be described by the Anderson model

$$\hat{H}_{\rm A} = \underbrace{\sum_{\sigma}\sum_{\boldsymbol{k}} \varepsilon_{\boldsymbol{k}} n_{\boldsymbol{k}\sigma} + \sum_{\sigma} \varepsilon_f \, \hat{n}_{f\sigma} + U \hat{n}_{f\uparrow}\hat{n}_{f\downarrow}}_{\hat{H}_0} + \underbrace{\sum_{\sigma}\sum_{\boldsymbol{k}} \left[V_{\boldsymbol{k}} c^{\dagger}_{\boldsymbol{k}\sigma} c_{f\sigma} + h.c. \right]}_{\hat{H}_1},$$

where ε_f is the impurity level (occupied by $n_f \sim 1$ electrons), $\varepsilon_{\boldsymbol{k}}$ is the dispersion of the metallic band, and $V_{\boldsymbol{k}}$ the hybridization. If we assume that the system has particle-hole symmetry with respect to the Fermi level, then $\varepsilon_f - \mu = -U/2$. The Kondo regime is characterized by the parameter values $\varepsilon_f \ll \mu$ and $\varepsilon_f + U \gg \mu$ and by a weak hybridization, i.e., the hybridization width, which is the imaginary part of the hybridization function for the Anderson model,

$$\Delta(\varepsilon) = \pi \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} |V_{\boldsymbol{k}}|^2 \delta(\varepsilon_{\boldsymbol{k}} - \varepsilon)$$

is such that $\Delta(\mu) \ll |\mu - \varepsilon_f|, |\mu - \varepsilon_f - U|$. The Anderson model is important in this lecture because it is used as quantum-impurity model in dynamical mean-field theory. Through the Schrieffer-Wolff canonical transformation [10] one can map the Anderson model onto the Kondo model, in which only the effective spin of the impurity enters

$$\hat{H}_{\rm K} = \hat{H}'_0 + \Gamma \boldsymbol{S}_f \cdot \boldsymbol{s}_c(\boldsymbol{0}) = \hat{H}'_0 + \hat{H}_{\Gamma}, \tag{8}$$

where

$$\Gamma \sim -2|V_{k_F}|^2 \left[\frac{1}{\varepsilon_f} - \frac{1}{\varepsilon_f + U} \right] > 0$$

is the antiferromagnetic coupling arising from the hybridization, $\boldsymbol{S}_f$ the spin of the impurity ($S_f = 1/2$), and $\boldsymbol{s}_c(\boldsymbol{0})$ is the spin-density of the conduction band at the impurity site. For convenience we set the Fermi energy to zero; k_F is a $\boldsymbol{k}$ vector at the Fermi level. The Schrieffer-Wolff canonical transformation works as follows. We introduce the operator $\hat{S}$ that transforms the Hamiltonian $\hat{H}$ into $\hat{H}_S$

$$\hat{H}_S = e^{\hat{S}} \hat{H} e^{-\hat{S}}.$$

We search for an operator $\hat{S}$ such that the transformed Hamiltonian $\hat{H}_S$ has no terms of first order in $V_{\boldsymbol{k}}$. Let us first split the original Hamiltonian $\hat{H}_A$ into two pieces: $\hat{H}_0$, the sum of all terms except the hybridization term, and $\hat{H}_1$, the hybridization term. Let us choose $\hat{S}$ linear in $V_{\boldsymbol{k}}$ and such that

$$[\hat{S}, \hat{H}_0] = -\hat{H}_1. \tag{9}$$

From Eq. (9) one finds that the operator $\hat{S}$ is given by

$$\hat{S} = \sum_{\boldsymbol{k}\sigma} \left[\frac{1-\hat{n}_{f-\sigma}}{\varepsilon_{\boldsymbol{k}} - \varepsilon_f} + \frac{\hat{n}_{f-\sigma}}{\varepsilon_{\boldsymbol{k}} - \varepsilon_f - U} \right] V_{\boldsymbol{k}} c^\dagger_{\boldsymbol{k}\sigma} c_{f\sigma} - \text{h.c.}.$$

The transformed Hamiltonian is complicated, as can be seen from explicitly writing the series for a transformation satisfying Eq. (9)

$$\hat{H}_S = \hat{H}_0 + \frac{1}{2}\left[\hat{S}, \hat{H}_1\right] + \frac{1}{3}\left[\hat{S}, \left[\hat{S}, \hat{H}_1\right]\right] + \dots.$$

In the limit in which the hybridization strength Γ is small this series can, however, be truncated at second order. The resulting Hamiltonian has the form

$$\hat{H}_S = \hat{H}_0 + \hat{H}_2,$$

with

$$\hat{H}_2 = \hat{H}_\Gamma + \hat{H}_{\text{dir}} + \Delta\hat{H}_0 + \hat{H}_{\text{ch}}.$$

The first term is the exchange interaction

$$\hat{H}_\Gamma = \frac{1}{4} \sum_{\boldsymbol{k}\boldsymbol{k}'} \Gamma_{\boldsymbol{k}\boldsymbol{k}'} \left[\sum_{\sigma_1\sigma_2} c^\dagger_{\boldsymbol{k}'\sigma_1} \langle\sigma_1|\hat{\sigma}|\sigma_2\rangle c_{\boldsymbol{k}\sigma_2} \cdot \sum_{\sigma_3\sigma_4} c^\dagger_{f\sigma_3} \langle\sigma_3|\hat{\sigma}|\sigma_4\rangle c_{f\sigma_4} \right]$$

where

$$\Gamma_{\boldsymbol{k}\boldsymbol{k}'} = V^*_{\boldsymbol{k}} V_{\boldsymbol{k}'} \left[\frac{1}{\varepsilon_{\boldsymbol{k}} - \varepsilon_f} + \frac{1}{\varepsilon_{\boldsymbol{k}'} - \varepsilon_f} + \frac{1}{U + \varepsilon_f - \varepsilon_{\boldsymbol{k}}} + \frac{1}{U + \varepsilon_f - \varepsilon_{\boldsymbol{k}'}} \right].$$

Let us assume that the coupling $\Gamma_{\boldsymbol{k}\boldsymbol{k}'}$ is weakly dependent on $\boldsymbol{k}$ and $\boldsymbol{k}'$; then by setting $|\boldsymbol{k}| \sim k_F$, and $|\boldsymbol{k}'| \sim k_F$ we recover the antiferromagnetic contact coupling in Eq. (8).
The second term is a potential-scattering interaction

$$\hat{H}_{\text{dir}} = \sum_{\boldsymbol{k}\boldsymbol{k}'} \left[A_{\boldsymbol{k}\boldsymbol{k}'} - \frac{1}{4}\Gamma_{\boldsymbol{k}\boldsymbol{k}'}\hat{n}_f \right] \sum_\sigma \hat{c}^\dagger_{\boldsymbol{k}'\sigma}\hat{c}_{\boldsymbol{k}\sigma},$$

where

$$A_{\boldsymbol{k}\boldsymbol{k}'} = \frac{1}{2} V^*_{\boldsymbol{k}} V_{\boldsymbol{k}'} \left[\frac{1}{\varepsilon_{\boldsymbol{k}} - \varepsilon_f} + \frac{1}{\varepsilon_{\boldsymbol{k}'} - \varepsilon_f} \right].$$

This term is spin-independent, and thus does not play a relevant role in the Kondo effect. The next term merely modifies the $\hat{H}_0$ term

$$\Delta\hat{H}_0 = -\sum_{\boldsymbol{k}\sigma} \left[A_{\boldsymbol{k}\boldsymbol{k}} - \frac{1}{2}\Gamma_{\boldsymbol{k}\boldsymbol{k}}\, \hat{n}_{f-\sigma} \right] \hat{n}_{f\sigma}.$$

Finally, the last term is a pair-hopping interaction, which changes the charge of the f site by two electrons and thus can be neglected if $n_f \sim 1$

$$\Delta\hat{H}_{\text{ch}} = -\frac{1}{4} \sum_{\boldsymbol{k}\boldsymbol{k}'\sigma} \Gamma_{\boldsymbol{k}\boldsymbol{k}'} c^\dagger_{\boldsymbol{k}'-\sigma} c^\dagger_{\boldsymbol{k}\sigma} c_{f\sigma} c_{f-\sigma} + \text{h.c.}.$$

The essential term in $\hat{H}_2$ is the exchange term $\hat{H}_\Gamma$, which is the one that yields the antiferromagnetic contact interaction in the Kondo Hamiltonian (8).

3.1.1 Poor man's scaling

We can understand the nature of the ground state of the Kondo model by using a simple approach due to Anderson called *poor man's scaling* [13] and an argument due to Nozières. First we divide the Hilbert space into a high- and a low-energy sector. We define as *high-energy* states those with at least one electron or one hole at the top or bottom of the band; the corresponding constraint for the high-energy electronic level $\varepsilon_{\boldsymbol{q}}$ is $D' < \varepsilon_{\boldsymbol{q}} < D$ or $-D < \varepsilon_{\boldsymbol{q}} < -D'$, where $D' = D - \delta D$. Next we introduce the operator $\hat{P}_H$, which projects onto the high-energy states, and the operator $\hat{P}_L = \hat{1} - \hat{P}_H$, which projects onto states with no electrons or holes in the high-energy region. Then we downfold the high-energy sector of the Hilbert space. To do this we rewrite the original Kondo Hamiltonian,

$$\hat{H} \equiv \hat{H}'_0 + \hat{H}_\Gamma,$$

as the energy-dependent operator $\hat{H}'$, which acts only in the low-energy sector

$$\hat{H}' = \hat{P}_L \hat{H} \hat{P}_L + \delta \hat{H}_L = \hat{H}_L + \delta \hat{H}_L,$$
$$\delta \hat{H}_L = \hat{P}_L \hat{H} \hat{P}_H \left(\omega - \hat{P}_H \hat{H} \hat{P}_H\right)^{-1} \hat{P}_H \hat{H} \hat{P}_L.$$

Here $\hat{H}_L$ is the original Hamiltonian, however in the space in which the high-energy states have been downfolded; the term $\delta \hat{H}_L$ is a correction due to the interaction between low-energy and (downfolded) high-energy states. Up to this point, the operator $\hat{H}'$ has the same spectrum as the original Hamiltonian. To make use of this expression, however, we have to introduce approximations. Thus, let us calculate $\delta \hat{H}_L$ using many-body perturbation theory. The first non-zero contribution is of second order in Γ

$$\delta \hat{H}_L^{(2)} \sim \hat{P}_L \hat{H}_\Gamma \hat{P}_H \left(\omega - \hat{P}_H \hat{H}'_0 \hat{P}_H\right)^{-1} \hat{P}_H \hat{H}_\Gamma P_L \,.$$

There are two types of processes that contribute at the second order, an electron and a hole process, depending on whether the downfolded states have (at least) one electron or one hole in the high-energy region. Let us consider the electron process. We set

$$\hat{P}_H \sim \sum_{\boldsymbol{q}\sigma} c^\dagger_{\boldsymbol{q}\sigma} |FS\rangle\langle FS| c_{\boldsymbol{q}\sigma},$$

$$\hat{P}_L \sim \sum_{\boldsymbol{k}\sigma} c^\dagger_{\boldsymbol{k}\sigma} |FS\rangle\langle FS| c_{\boldsymbol{k}\sigma} \,,$$

where $|\varepsilon_{\boldsymbol{k}}| < D'$ and

$$|FS\rangle = \prod_{\boldsymbol{k}\sigma} c^\dagger_{\boldsymbol{k}\sigma} |0\rangle$$

is the Fermi sea, i.e., the many-body state corresponding to the metallic conduction band. Thus

$$\delta H_L^{(2)} = -\frac{1}{2} \Gamma^2 \sum_{\boldsymbol{q}} \frac{1}{\omega - \varepsilon_{\boldsymbol{q}}} \boldsymbol{S}_f \cdot \boldsymbol{s}_c(\boldsymbol{0}) + \dots$$

$$\sim \frac{1}{4} \rho(\varepsilon_F) \Gamma^2 \frac{\delta D}{D} \boldsymbol{S}_f \cdot \boldsymbol{s}_c(\boldsymbol{0}) + \dots .$$

We find an analogous contribution from the hole process. The correction $\delta H_L^{(2)}$ modifies the parameter Γ of the Kondo Hamiltonian as follows

$$\Gamma \to \Gamma' = \Gamma + \delta\Gamma,$$

and

$$\frac{\delta\Gamma}{\delta \ln D} = \frac{1}{2}\rho(\varepsilon_F)\Gamma^2, \tag{10}$$

where

$$\delta \ln D = \delta D/D.$$

It can be seen that equation (10) has two fixed points

(i) $\Gamma = 0$ (*weak coupling*)

(ii) $\Gamma \to \infty$ (*strong coupling*)

By solving the scaling equation we find

$$\Gamma' = \frac{\Gamma}{1 + \frac{1}{2}\rho(\varepsilon_F)\Gamma \ln \frac{D'}{D}} \ .$$

If the original coupling Γ is antiferromagnetic, the renormalized coupling constant Γ' diverges (i.e., it scales to the strong coupling fixed point) for

$$D' = De^{-2/\Gamma\rho(\varepsilon_F)}.$$

We can define this value of D' as the Kondo energy

$$k_B T_{\mathrm{K}} = De^{-2/\Gamma\rho(\varepsilon_F)}. \tag{11}$$

The divergence at $k_B T_{\mathrm{K}}$ indicates that at low energy the interaction between the spins dominates, and therefore the system forms a singlet in which the impurity magnetic moment is screened. The existence of this strong coupling fixed point is confirmed by the numerical renormalization group of Wilson [14]. Nozières [15] has used this conclusion to show that the low-temperature behavior of the system must be of Fermi liquid type. His argument is the following. For infinite coupling Γ' the impurity traps a conduction electron to form a singlet state. For a finite but still very large Γ', any attempt at breaking the singlet will cost a very large energy. Virtual excitations (into the $n_f = 0$ or $n_f = 2$ states and finally the $n_f = 1$ triplet state) are, however, possible and they yield an effective indirect interaction between the remaining conduction electrons surrounding the impurity. This is similar to the phonon-mediated attractive interaction in metals. The indirect electron-electron coupling is weak and can be calculated in perturbation theory ($1/\Gamma$ expansion). Nozières has shown that, to first approximation, the effective interaction is between electrons of opposite spins lying next to the impurity. It is of order D^4/Γ^3 and repulsive, hence it gives rise to a Fermi liquid behavior with enhanced susceptibility [15].

3.2 The Anderson molecule

As in the case of the Hubbard model, it is useful to look at a simpler case, the Anderson molecule. The corresponding Hamiltonian is given by

$$\hat{H} = \varepsilon_f \, \hat{n}_{1\sigma} + \varepsilon_s \hat{n}_{2\sigma} - t_A \sum_\sigma \left[c^\dagger_{1\sigma} c_{2\sigma} + c^\dagger_{2\sigma} c_{1\sigma} \right] + U \hat{n}_{1\uparrow} \hat{n}_{1\downarrow}. \tag{12}$$

Also this Hamiltonian commutes with the number of electron operator $\hat{N}$, with the total spin $\hat{S}$ and with $\hat{S}_z$. Thus we can express the states in the atomic limit as

$\|N,S,S_z\rangle$			N	S	$E(N)$
$\|0,0,0\rangle$	$=$	$\|0\rangle$	0	0	0
$\|1,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}\|0\rangle$	1	1/2	ε_f
$\|1,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}\|0\rangle$	1	1/2	ε_s
$\|2,1,0\rangle$	$=$	$\frac{1}{\sqrt{2}}\left[c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} + c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right]\|0\rangle$	2	1	$\varepsilon_f + \varepsilon_s$
$\|2,1,1\rangle$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{1\uparrow}\|0\rangle$	2	1	$\varepsilon_f + \varepsilon_s$
$\|2,1,-1\rangle$	$=$	$c^\dagger_{2\downarrow}c^\dagger_{1\downarrow}\|0\rangle$	2	1	$\varepsilon_f + \varepsilon_s$
$\|2,0,0\rangle_0$	$=$	$\frac{1}{\sqrt{2}}\left[c^\dagger_{1\uparrow}c^\dagger_{2\downarrow} - c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}\right]\|0\rangle$	2	0	$\varepsilon_f + \varepsilon_s$
$\|2,0,0\rangle_1$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	2	0	$2\varepsilon_f + U$
$\|2,0,0\rangle_2$	$=$	$c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	2	0	$2\varepsilon_s$
$\|3,1/2,\sigma\rangle_1$	$=$	$c^\dagger_{1\sigma}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	3	1/2	$\varepsilon_f + 2\varepsilon_s$
$\|3,1/2,\sigma\rangle_2$	$=$	$c^\dagger_{2\sigma}c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}\|0\rangle$	3	1/2	$2\varepsilon_f + \varepsilon_s + U$
$\|4,0,0\rangle$	$=$	$c^\dagger_{1\uparrow}c^\dagger_{1\downarrow}c^\dagger_{2\uparrow}c^\dagger_{2\downarrow}\|0\rangle$	4	0	$2\varepsilon_f + 2\varepsilon_s + U$

Again, for $N = 2$ electrons, the hopping integrals only couple the $S = 0$ states. The Hamiltonian looks like

$$\hat{H}_2 = \begin{pmatrix} \varepsilon_f + \varepsilon_s & 0 & 0 & 0 & 0 & 0 \\ 0 & \varepsilon_f + \varepsilon_s & 0 & 0 & 0 & 0 \\ 0 & 0 & \varepsilon_f + \varepsilon_s & 0 & 0 & 0 \\ 0 & 0 & 0 & \varepsilon_f + \varepsilon_s & -\sqrt{2}t_A & -\sqrt{2}t_A \\ 0 & 0 & 0 & -\sqrt{2}t_A & 2\varepsilon_f + U & 0 \\ 0 & 0 & 0 & -\sqrt{2}t_A & 0 & 2\varepsilon_s \end{pmatrix}$$

The ground-state is a singlet, as in the Kondo problem. In order to calculate its energy, let us downfold the doubly-occupied states. We find

$$E_0(\omega) = \omega = \varepsilon_f + \varepsilon_s - \frac{2t_A^2}{2\varepsilon_f + U - \omega} - \frac{2t_A^2}{2\varepsilon_s - \omega}.$$

If we set $\omega = \varepsilon_f + \varepsilon_s - \Delta E$, and $\varepsilon_s \sim 0$ we have the solution

$$\Delta E \sim -2t_A^2 \left[\frac{1}{\varepsilon_f} - \frac{1}{\varepsilon_f + U}\right] \equiv \Gamma.$$

We can define ΔE as Kondo energy for the Anderson molecule. There is an important difference with respect to the real Kondo model, namely that in that case the Kondo energy, defined in Eq. (11), decreases exponentially with the inverse of Γ. The non-perturbative nature of the problem is thus not captured by the Anderson dimer.

3.3 Anderson molecule vs Hubbard dimer

Let us now compare the Anderson molecule and the Hubbard dimer. The non-interacting Green function for the Anderson molecule can be obtained directly from the non-interacting part of the Hamiltonian

$$\mathcal{G}_\sigma^{-1}(i\nu_n) = \begin{pmatrix} i\nu_n - \varepsilon_f + \mu & t_A \\ t_A & i\nu_n - \varepsilon_s + \mu \end{pmatrix}^{-1}.$$

By downfolding the s orbital we obtain

$$\mathcal{G}_{ff,\sigma}(i\nu_n) = \frac{1}{i\nu_n - (\varepsilon_f - \mu + \mathcal{F}(i\nu_n))},$$

where $\mathcal{F}(i\nu_n)$ is the non-interacting hybridization function for the Anderson molecule

$$\mathcal{F}(i\nu_n) = \frac{t_A^2}{i\nu_n - (\varepsilon_s - \mu)} = i\nu_n - \varepsilon_f + \mu - \mathcal{G}_{ff,\sigma}^{-1}(i\nu_n).$$

Using the Dyson equation, we can then write the interacting local Green function as

$$G_{ff,\sigma}(i\nu_n) = \frac{1}{i\nu_n - (\varepsilon_f - \mu + \mathcal{F}(i\nu_n) + \Sigma_{ff}(i\nu_n))}. \tag{13}$$

The impurity Green function (13) and the local Green function $G_{ii,\sigma}(i\nu_n)$ of the Hubbard dimer, Eq. (6), have a similar form. In view of this observation, it is legitimate to ask ourselves the following question: Can we reproduce some properties of the Hubbard dimer via an Anderson molecule in which $\varepsilon_f = \varepsilon_d$, while ε_s and t_A are free parameters? In the limit $U = 0$, indeed, setting $\varepsilon_s = \varepsilon_d$ and $t_A = t$ the two models are identical. For finite U, in general, they strongly differ. Let us request first that the occupation numbers is the same for the two models at half filling. This can be achieved with the choice

$$\varepsilon_s = \varepsilon_f + U/2.$$

For this value of ε_s, the eigenstates in the $N = 2$ electron sector are identical for the Hubbard dimer and the Anderson model. We can then in addition demand that at half-filling the gap is the same for the two models. This leads to the condition

$$\frac{1}{2}\sqrt{U^2 + 16t_A^2} = -2t + \sqrt{U^2 + 16t^2}$$

which for small t/U has the solution $t_A \sim \frac{\sqrt{3}}{4}U$. The message is that we could in principle use the Anderson molecule as an approximate version of the Hubbard dimer; choosing the parameters of the first ad hoc, we can reproduce some properties of the second, for example occupation number and gap. Could we go beyond that, and reproduce the full local Green function of the Hubbard dimer via an Anderson-like molecule? Comparing the local Green functions of the two models, we can see that it would be possible under, e.g., the following conditions

- the non-local part of the self-energy of the Hubbard dimer is negligible
- the local self-energy $\Sigma_l(i\nu_n)$ equals $\Sigma_{ff}(i\nu_n)$
- the hybridization function $\mathcal{F}(i\nu_n)$ equals $F(i\nu_n)$

As we have seen, for the Hubbard dimer the non-local part of the self-energy is finite and, in general, non-negligible; thus already the first condition is not fulfilled. For the lattice Hubbard model it can be shown, however, that diagrammatic perturbation theory greatly simplifies in the limit of infinite dimensions, and the self-energy becomes local [4,6]. This important conclusion is exploited in the DMFT approach.

4 DMFT and DFT+DMFT

4.1 Method

Although apparently simple, the Hubbard Hamiltonian (1) cannot be solved exactly except in special cases. For the Hubbard dimer defined via the Hamiltonian (4), we have seen that some properties can be reproduced via the even simpler Anderson molecule, Hamiltonian (12), provided that the parameters of the latter are chosen ad hoc. Can we do the same for the general Hubbard and Anderson model? This idea is at the core of dynamical mean-field theory. DMFT maps the correlated *lattice* problem described by the Hubbard model onto a correlated single-impurity problem [8, 4–6], e.g., an effective Anderson-like model. The latter can be solved exactly, differently than the original Hubbard model; to solve it we have to use numerical techniques, for example quantum Monte Carlo. The Anderson model is defined via either the hybridization function $\mathcal{F}(i\nu_n)$ or the bath Green function $\mathcal{G}(i\nu_n) = (i\nu_n - \varepsilon_d + \mu - \mathcal{F}(i\nu_n))^{-1}$. Solving it yields the impurity Green function $G(i\nu_n)$. From the Dyson equation we can calculate the impurity self-energy

$$\Sigma(i\nu_n) = \mathcal{G}^{-1}(i\nu_n) - G^{-1}(i\nu_n).$$

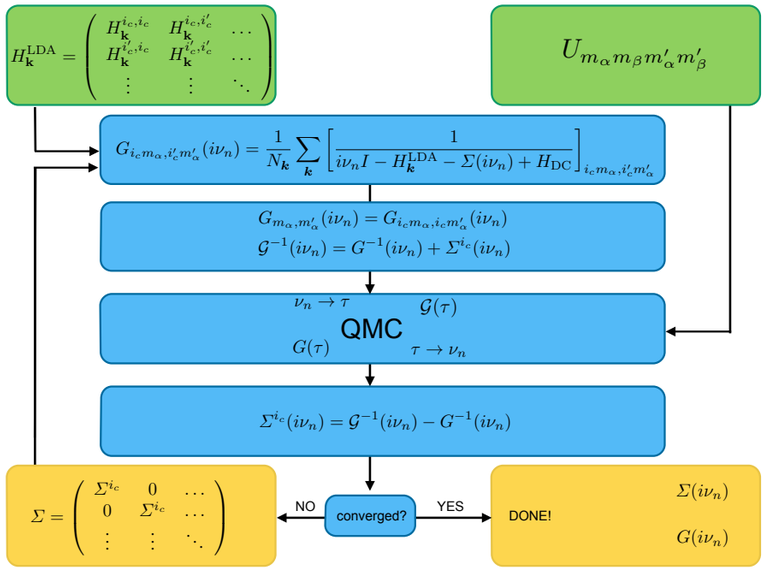


Fig. 7: *DFT+DMFT self-consistency loop. The DFT Hamiltonian is built in the basis of Bloch states obtained from localized Wannier functions, for example in the local-density approximation (LDA); this gives $H_{\boldsymbol{k}}^{\mathrm{LDA}}$. The set $\{i_c\}$ labels the equivalent correlated sites inside the unit cell. The local Green-function matrix is at first calculated using an initial guess for the self-energy matrix. The bath Green-function matrix is then obtained via the Dyson equation and used to construct an effective quantum-impurity model. The latter is solved via a quantum-impurity solver, here quantum Monte Carlo (QMC), yielding the impurity Green-function matrix. Through the Dyson equation the self-energy is then obtained, and the procedure is repeated till self-consistency is reached.*

Next, we assume that non-local contributions to the self-energy of the Hubbard model can be neglected, and that the local self-energy equals the impurity self-energy. Then, the local Green function is given by

$$G_{i_c,i_c}(i\nu_n) = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \left[i\nu_n - \varepsilon_{\boldsymbol{k}} - \Sigma(i\nu_n) \right]^{-1}.$$

Here $N_{\boldsymbol{k}}$ is the number of $\boldsymbol{k}$ points. Self-consistency is reached when the impurity Green function $G(i\nu_n)$ equals the actual local Green function $G_{i_c,i_c}(i\nu_n)$

$$G_{i_c,i_c}(i\nu_n) = G(i\nu_n).$$

The main approximation adopted is that the self-energy of the Hubbard model is local; as already mentioned, it can be shown that the self-energy becomes indeed local in the infinite-coordination-number limit [4, 6]. The DMFT approach can be extended to material-specific multi-orbital Hamiltonians. In this case we replace

$$\varepsilon_{\boldsymbol{k}} \to H_{\boldsymbol{k}}^{0},$$

where $H^0_{\boldsymbol{k}}$ is the non-interacting Hamiltonian. Furthermore, the local Green function and self-energy become matrices in spin-orbital space. Typically, to build minimal material-specific models, we use density-functional theory, for example in the local-density approximation. First we construct a basis of localized Wannier functions that span the relevant bands and then use this basis to build material-specific Hubbard models. The combination of DMFT with density-functional theory, sketched above in short, defines the DFT+DMFT approach. The DMFT self-consistency loop is shown in Fig. 7, where it is assumed that quantum Monte Carlo (QMC) is used as the quantum-impurity solver. It has to be pointed out that the computational time needed to solve a multiband quantum-impurity models with QMC increases rapidly with the number of degrees of freedom. How rapidly depends on the specific QMC flavor used. Thus, in practice, only few correlated orbitals/sites can be treated fully with DMFT. Furthermore, increasing the number of degrees of freedom, eventually leads to the infamous sign problem. It is thus very important to properly build minimal material-specific models.

4.2 Model building in DFT+DMFT

In the Born-Oppenheimer approximation, the non-relativistic electronic Hamiltonian for an ideal crystal, $\hat{H}_e$, can be written as the sum of a one-electron $\hat{H}_0$ and an interaction part $\hat{H}_U$

$$\hat{H}_e = \hat{H}_0 + \hat{H}_U.$$

Let us assume that we have constructed a complete basis of Wannier functions $\psi_{in\sigma}(\boldsymbol{r})$. Then, in this basis, the one-electron term is given by

$$\hat{H}_0 = -\sum_{\sigma}\sum_{ii'}\sum_{nn'} t^{i,i'}_{n,n'} c^{\dagger}_{in\sigma} c_{i'n'\sigma},$$

where $c^{\dagger}_{in\sigma}$ ($c_{in\sigma}$) creates (destroys) an electron with spin σ in orbital n at site i. The on-site ($i = i'$) terms yield the crystal-field matrix while the $i \neq i'$ contributions are the hopping integrals. This part of the Hamiltonian describes the attraction between electrons and nuclei, the latter forming an ideal lattice. The electron-electron repulsion $\hat{H}_U$ is instead given by

$$\hat{H}_U = \frac{1}{2}\sum_{ii'jj'}\sum_{\sigma\sigma'}\sum_{nn'pp'} U^{iji'j'}_{np\,n'p'} c^{\dagger}_{in\sigma} c^{\dagger}_{jp\sigma'} c_{j'p'\sigma'} c_{i'n'\sigma}.$$

For a given system, *material-specific* Wannier functions can be obtained via DFT-based calculations [7, 8]. This immediately gives hopping integrals and crystal-field splittings

$$t^{i,i'}_{n,n'} = -\int d\mathbf{r}\, \overline{\psi_{in\sigma}}(\boldsymbol{r}) \left[-\frac{1}{2}\nabla^2 + v_{\rm R}(\boldsymbol{r})\right] \psi_{i'n'\sigma}(\boldsymbol{r}),$$

where $v_R(\boldsymbol{r})$ is the self-consistent DFT reference potential. The *bare* Coulomb integrals can be expressed in terms of Wannier functions as well

$$U^{iji'j'}_{np\,n'p'} = \int d\mathbf{r}_1 \int d\mathbf{r}_2\, \overline{\psi_{in\sigma}}(\mathbf{r}_1)\overline{\psi_{jp\sigma'}}(\mathbf{r}_2)\frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|}\psi_{j'p'\sigma'}(\mathbf{r}_2)\psi_{i'n'\sigma}(\mathbf{r}_1).$$

Here we have to be careful, however. The DFT potential includes in $v_{\rm R}(\boldsymbol{r})$ also Coulomb effects, via the long-range Hartree term and the exchange-correlation contribution; if we use, e.g., LDA Wannier functions as one-electron basis, to avoid double counting we have to subtract from $\hat{H}_U$ the effects already included in the LDA. This means that we have to replace

$$\hat{H}_U \to \Delta\hat{H}_U = \hat{H}_U - \hat{H}_{\rm DC},$$

where $\hat{H}_{\rm DC}$ is the double-counting correction. Unfortunately we do not know which correlation effects are exactly included in the LDA, and therefore the exact expression of $\hat{H}_{\rm DC}$ is also unknown. The remarkable successes of the LDA suggest, however, that in many materials the LDA is overall a good approximation, and therefore, in those systems at least, the term $\Delta\hat{H}_U$ can be neglected. What about strongly-correlated materials? Even in correlated systems, most likely the LDA works rather well for the delocalized electrons or in describing the average or the long-range Coulomb effects. Thus one can think of separating the electrons into *uncorrelated* and *correlated*; only for the latter we do take the correction $\Delta\hat{H}_U$ into account explicitly, assuming furthermore that $\Delta\hat{H}_U$ is local or almost local [7]. Typically, correlated electrons are those that partially retain their atomic character, e.g., those that originate from localized d and f shells; for convenience here we assume that in a given system they stem from a single atomic shell l (e.g., d for transition-metal oxides or f for heavy-fermion systems) and label their states with the atomic quantum numbers l and $m = -l, \ldots, l$ of that shell. Thus

$$U^{iji'j'}_{np,n'p'} \sim \begin{cases} U^l_{m_\alpha m_\beta m'_\alpha m'_\beta} & iji'j' = iiii \quad npn'p' \in l \\ 0 & iji'j' \neq iiii \quad npn'p' \notin l \end{cases}$$

and $\Delta\hat{H}_U$ is replaced by $\Delta\hat{H}^l_U = \hat{H}^l_U - \hat{H}^l_{\rm DC}$, where $\hat{H}^l_{\rm DC}$ is, e.g., given by the static mean-field contribution of $\hat{H}^l_U$. There is a drawback in this procedure, however. By splitting electrons into correlated and uncorrelated we implicitly assume that the main effect of the latter is the renormalization or *screening* of parameters for the former, in particular of the Coulomb interaction. The calculation of screening effects remains, unfortunately, a challenge to date. Approximate schemes are the constrained LDA and the constrained random-phase approximation (RPA) methods [7, 8]. Nevertheless, we have now identified the general class of models for strongly-correlated systems, namely the generalized Hubbard model

$$\hat{H}_e = \hat{H}^{\rm LDA} + \hat{H}^l_U - \hat{H}^l_{\rm DC}. \tag{14}$$

It is often convenient to integrate out or downfold empty and occupied states and work directly with a set of Wannier functions spanning the correlated bands only. In this case we have

$$\hat{H}^{\rm LDA} = -\sum_{ii'}\sum_{\sigma}\sum_{m_\alpha m'_\alpha} t^{i,i'}_{m_\alpha,m'_\alpha} c^\dagger_{im_\alpha\sigma} c_{i'm'_\alpha\sigma} = \sum_{\boldsymbol{k}}\sum_{\sigma}\sum_{m_\alpha m'_\alpha} \left[H^{\rm LDA}_{\boldsymbol{k}}\right]_{m_\alpha,m'_\alpha} c^\dagger_{\boldsymbol{k}m_\alpha\sigma} c_{\boldsymbol{k}m'_\alpha\sigma},$$

where the right-hand side is rewritten using as a one-electron basis Bloch functions $\psi_{\boldsymbol{k}m_\alpha\sigma}$ constructed from the Wannier functions $\psi_{im_\alpha\sigma}$. The local *screened* Coulomb interaction is instead given by

$$\hat{H}^l_U = \frac{1}{2}\sum_i \sum_{\sigma\sigma'} \sum_{m_\alpha m'_\alpha} \sum_{m_\beta m'_\beta} U_{m_\alpha m_\beta m'_\alpha m'_\beta} c^\dagger_{im_\alpha\sigma} c^\dagger_{im_\beta\sigma'} c_{im'_\beta\sigma'} c_{im'_\alpha\sigma}.$$

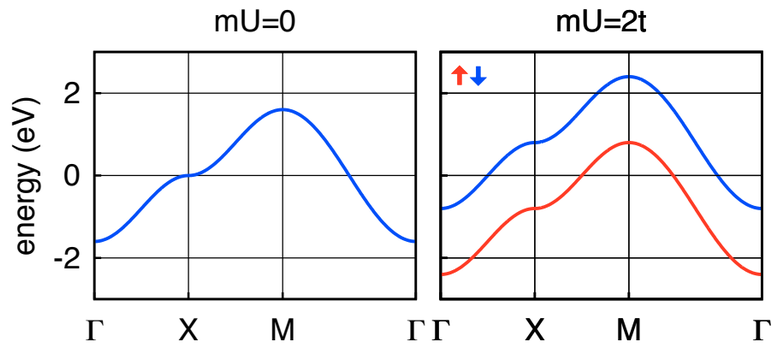


Fig. 8: *Ferromagnetism in Hartree-Fock. The chemical potential is taken as the energy zero.*

5 Metal-insulator transition

5.1 Hartree-Fock method

We have seen in section 2 the Hartree-Fock approximation for the Hubbard dimer. Here we want to extend it to the Hubbard model, and compare the description of the metal-insulator transition from Hartree-Fock to the one that emerges from DMFT. We assume that the system is at half-filling ($n = 1$) and exclude charge-disproportionation phenomena ($n_i = n$). Let us first consider the ferromagnetic HF solution. The HF approximation of the Coulomb term in the Hubbard model, as we have seen, consist in replacing the Coulomb term in the Hamiltonian with the expression given in Eq. (7). For the F solution we rewrite it as

$$\hat{H}_U^{\rm HF} = U \sum_i \left[-2m\hat{S}_z^i + m^2 + \frac{1}{4}n^2 \right],$$

were $m = (\bar{n}_{i\uparrow} - \bar{n}_{i\downarrow})/2 = m_+$. For the Hubbard model, it is convenient to Fourier transform the Hamiltonian to $\boldsymbol{k}$ space. We then adopt as one-electron basis the Bloch states

$$\Psi_{\boldsymbol{k}\sigma}(\boldsymbol{r}) = \frac{1}{\sqrt{N_s}} \sum_i e^{i\boldsymbol{k}\cdot\boldsymbol{T}_i}\, \Psi_{i\sigma}(\boldsymbol{r}),$$

where $\Psi_{i\sigma}(\boldsymbol{r})$ is a Wannier function with spin σ, $\boldsymbol{T}_i$ a lattice vector, and N_s the number of lattice sites. The term $\hat{H}_U^{\rm HF}$ depends on the spin operator $\hat{S}_z^i$, which can be written in $\boldsymbol{k}$ space as

$$\hat{S}_z^i = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{kq}} e^{-i\boldsymbol{q}\cdot\boldsymbol{T}_i} \underbrace{\frac{1}{2}\sum_\sigma \sigma c^\dagger_{\boldsymbol{k}\sigma} c_{\boldsymbol{k}+\boldsymbol{q}\sigma}}_{S_z(\boldsymbol{k},\boldsymbol{k}+\boldsymbol{q})} = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{kq}} e^{-i\boldsymbol{q}\cdot\boldsymbol{T}_i} \hat{S}_z(\boldsymbol{k},\boldsymbol{k}+\boldsymbol{q}).$$

The term $\hat{H}_U^{\rm HF}$ has the same periodicity as the lattice and does not couple states with different $\boldsymbol{k}$ vectors. Thus only $\hat{S}_z(\boldsymbol{0})$ contributes, and the Hamiltonian can be written as

$$\hat{H} = \sum_\sigma \sum_{\boldsymbol{k}} \varepsilon_{\boldsymbol{k}} \hat{n}_{\boldsymbol{k}\sigma} + \underbrace{U \sum_{\boldsymbol{k}} \left[-2m\, \hat{S}_z(\boldsymbol{k},\boldsymbol{k}) + m^2 + \frac{n^2}{4} \right]}_{\hat{H}_U^{\rm HF} = U\sum_i [-2m\hat{S}_z^i + m^2 + \frac{1}{4}n^2]},$$

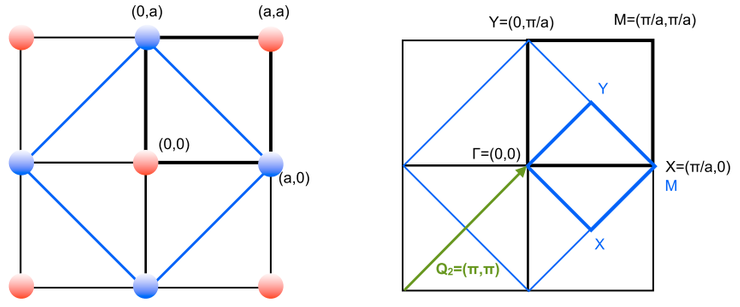


Fig. 9: *Doubling of the cell due to antiferromagnetic order and the corresponding folding of the Brillouin zone (BZ) for a two-dimensional hypercubic lattice. The antiferromagnetic* $\boldsymbol{Q}_2 = (\pi/a, \pi/a, 0)$ *vector is also shown.*

where for simplicity we set $\varepsilon_d = 0$. The HF correction splits the bands with opposite spin, leading to new one-electron eigenvalues, $\varepsilon_{\boldsymbol{k}\sigma} = \varepsilon_{\boldsymbol{k}} + \frac{1}{2}U - \sigma U m$. The separation between $\varepsilon_{\boldsymbol{k}\uparrow} - \mu$ and $\varepsilon_{\boldsymbol{k}\downarrow} - \mu$ is $2mU$, as can be seen in Fig. 8. The system remains metallic for U smaller than the bandwidth W. In the small-t/U limit and at half filling we can assume that the system is a ferromagnetic insulator and $m = 1/2$. The total energy of the ground state is then

$$E_{\mathrm{F}} = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} [\varepsilon_{\boldsymbol{k}\sigma} - \mu] = \frac{1}{N_{\boldsymbol{k}}} \sum_{\boldsymbol{k}} \left[\varepsilon_{\boldsymbol{k}} - \frac{1}{2}U\right] = -\frac{1}{2}U.$$

Let us now describe the same periodic lattice via a supercell which allows for a two-sublattice antiferromagnetic solution; this supercell is shown in Fig. 9. We rewrite the Bloch states of the original lattice as

$$\Psi_{\boldsymbol{k}\sigma}(\boldsymbol{r}) = \frac{1}{\sqrt{2}} \left[\Psi^A_{\boldsymbol{k}\sigma}(\boldsymbol{r}) + \Psi^B_{\boldsymbol{k}\sigma}(\boldsymbol{r})\right], \qquad \Psi^\alpha_{\boldsymbol{k}\sigma}(\boldsymbol{r}) = \frac{1}{\sqrt{N_{s_\alpha}}} \sum_{i_\alpha} e^{i\boldsymbol{T}^\alpha_i \cdot \boldsymbol{k}}\, \Psi_{i_\alpha\sigma}(\boldsymbol{r}).$$

Here A and B are the two sublattices with opposite spins and $\boldsymbol{T}^A_i$ and $\boldsymbol{T}^B_i$ are their lattice vectors; $\alpha = A, B$. We take as one-electron basis the two Bloch functions $\Psi_{\boldsymbol{k}\sigma}$ and $\Psi_{\boldsymbol{k}+\boldsymbol{Q}_2\sigma}$, where $\boldsymbol{Q}_2 = (\pi/a, \pi/a, 0)$ is the vector associated with the antiferromagnetic instability and the corresponding folding of the Brillouin zone, also shown in Fig. 9. Then, in HF approximation, setting $m_- = m$, the Coulomb interaction is given by

$$\hat{H}^{\mathrm{HF}}_U = \sum_{i\in A} \left[-2m\hat{S}^i_z + m^2 + \frac{n^2}{4}\right] + \sum_{i\in B} \left[+2m\hat{S}^i_z + m^2 + \frac{n^2}{4}\right].$$

This interaction couples Bloch states with $\boldsymbol{k}$ vectors made equivalent by the folding of the Brillouin zone. Thus the HF Hamiltonian takes the form

$$\hat{H} = \sum_{\boldsymbol{k}} \sum_{\sigma} \varepsilon_{\boldsymbol{k}} \hat{n}_{\boldsymbol{k}\sigma} + \sum_{\boldsymbol{k}} \sum_{\sigma} \varepsilon_{\boldsymbol{k}+\boldsymbol{Q}_2} \hat{n}_{\boldsymbol{k}+\boldsymbol{Q}_2\sigma} + \underbrace{U \sum_{\boldsymbol{k}} \left[-2m\, \hat{S}_z(\boldsymbol{k}, \boldsymbol{k}+\boldsymbol{Q}_2) + 2m^2 + \frac{n^2}{2}\right]}_{\text{static mean-field correction } \hat{H}^{\mathrm{HF}}_U}.$$

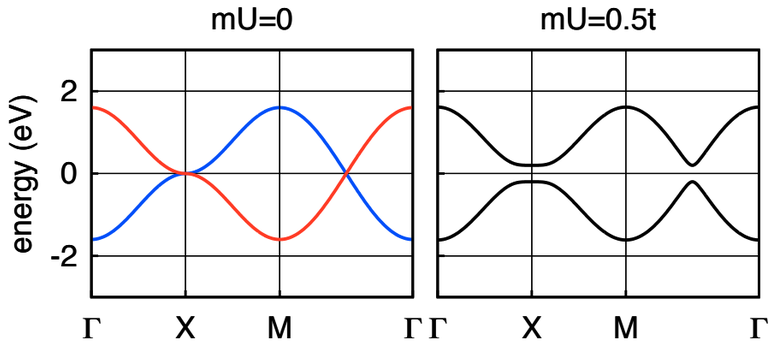


Fig. 10: *Antiferromagnetism in Hartree-Fock. The chemical potential is taken as the energy zero. Blue:* $\varepsilon_{\boldsymbol{k}}$*. Red:* $\varepsilon_{\boldsymbol{k}+\boldsymbol{Q}_2} = -\varepsilon_{\boldsymbol{k}}$*. The high-symmetry lines are those of the large BZ in Fig. 9.*

The sum over $\boldsymbol{k}$ is restricted to the Brillouin zone of the antiferromagnetic lattice. We find the two-fold degenerate eigenvalues

$$\varepsilon_{\boldsymbol{k}\pm} - \mu = \frac{1}{2}(\varepsilon_{\boldsymbol{k}} + \varepsilon_{\boldsymbol{k}+\boldsymbol{Q}_2}) \pm \frac{1}{2}\sqrt{(\varepsilon_{\boldsymbol{k}} - \varepsilon_{\boldsymbol{k}+\boldsymbol{Q}_2})^2 + 4(mU)^2}. \tag{15}$$

A gap opens where the bands $\varepsilon_{\boldsymbol{k}}$ and $\varepsilon_{\boldsymbol{k}+\boldsymbol{Q}_2}$ cross, e.g., at the X point of the original Brillouin zone (Fig. 10). At half filling and for $mU = 0$ the Fermi level crosses the bands at the X point; thus the system is an insulator for any finite value of mU. In the small-t/U limit we can assume that $m = 1/2$ and expand the eigenvalues in powers of $\varepsilon_{\boldsymbol{k}}/U$. For the occupied states we find

$$\varepsilon_{\boldsymbol{k}-} - \mu \sim -\frac{1}{2}U - \frac{\varepsilon_{\boldsymbol{k}}^2}{U} = -\frac{1}{2}U - \frac{4t^2}{U}\left(\frac{\varepsilon_{\boldsymbol{k}}}{2t}\right)^2.$$

The ground-state total energy for the antiferromagnetic supercell is then $2E_{\rm AF}$ with

$$E_{\rm AF} = -\frac{1}{2}U - \frac{4t^2}{U}\frac{1}{N_{\boldsymbol{k}}}\sum_{\boldsymbol{k}}\left(\frac{\varepsilon_{\boldsymbol{k}}}{2t}\right)^2 \sim -\frac{1}{2}U - \frac{4t^2}{U}$$

so that the energy difference per pair of spins between ferro- and antiferro-magnetic state is

$$\Delta E^{\rm HF} = E^{\rm HF}_{\uparrow\uparrow} - E^{\rm HF}_{\uparrow\downarrow} = \frac{2}{n_{\langle ii'\rangle}}\left[E_{\rm F} - E_{\rm AF}\right] \sim \frac{1}{2}\frac{4t^2}{U} \sim \frac{1}{2}\Gamma, \tag{16}$$

which is similar to the result obtained from the Hubbard model in many-body second order perturbation theory, Eq. (3). We notice here the same problems that we already observed for the Hubbard dimer. Despite the similarity with the actual solution, the spectrum of the Hartree-Fock Hamiltonian has very little to do with the spectrum of the Hubbard Hamiltonian at half filling. If we restrict ourselves to the AF solution, the first excited state in HF is at an energy $\propto U$ rather than $\propto \Gamma$; thus, we cannot use a single HF calculation to understand the magnetic excitation spectrum of a given system. It is more meaningful to use HF to compare the total energy of different states and determine in this way, within HF, the ground state. Even in this

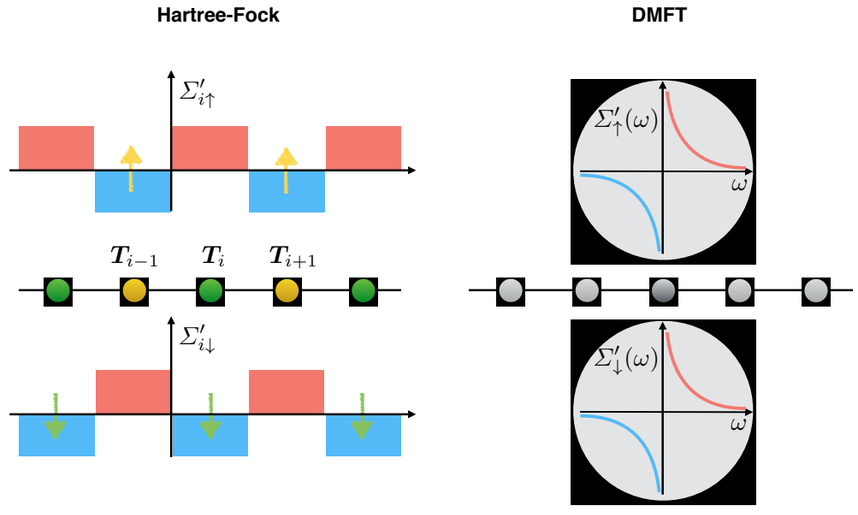


Fig. 11: *Idealized correlated crystal, schematically represented by a half-filled single-band Hubbard chain. Sketch of the real-part of the self-energy in the insulating phase, as described by Hartree-Fock (left-hand side) and DMFT (right-hand side). In HF the self-energy is a spin- and site-dependent potential (Slater insulator). In DMFT it is spin- and site-independent; it is, however, dynamical and its real part diverges at zero frequency (Mott insulator). The imaginary part of the self-energy is always zero in Hartree-Fock (i.e., quasiparticles have infinite lifetimes).*

case, however, as we have seen for the Hubbard dimer, HF suffers from spin contamination, i.e., singlet states and $S_z = 0$ triplet states mix. The energy difference per bond $E^{\mathrm{HF}}_{\uparrow\uparrow} - E^{\mathrm{HF}}_{\uparrow\downarrow}$ in Eq. (16) only resembles the exact result; the exact energy difference between triplet and singlet state in the Hubbard dimer is a factor of two larger

$$\Delta E = E_{S=1} - E_{S=0} = \Gamma.$$

Thus, overall, HF is not the ideal approach to determine the onset of magnetic phase transitions. The major problem of the HF approximation is, however, the description of the metal-insulator transition. In HF the metal-insulator transition is, as we have seen, intimately related to long-range magnetic order, i.e., it is a Slater rather than a Mott transition. If we write the HF correction in the form of a self-energy, the latter is a real, static but spin- and site-dependent potential. More specifically, in the AF case at half filling we have for two neighboring sites i and j

$$\Sigma^{\mathrm{HF}}_{i\sigma}(\omega) = U\left[\frac{1}{2} + m\right], \quad \Sigma^{\mathrm{HF}}_{j\sigma}(\omega) = U\left[\frac{1}{2} - m\right].$$

This spatial structure of the self-energy is what opens the gap shown in Fig. 10. For $m = 0$ the self-energy is a mere energy shift – the same for all sites and spins – and does not change the band structure or the properties of the system, which is then a conventional metal.

5.2 HF vs DMFT

The main difference between DMFT and Hartree-Fock is that in DMFT the self-energy is frequency dependent but local (i.e., site- or $\boldsymbol{k}$-independent), while in HF is static but site dependent. Let us discuss the DMFT description of the metal-insulator transition. The poles of the Green function, i.e., the solutions of the equation

$$\omega - \varepsilon_{\boldsymbol{k}} - \Sigma'(\omega) = 0\,,$$

where $\Sigma'(\omega)$ is the real part of the self-energy, yield the excitations of our system. For small U, in the Fermi liquid regime, the Green function has a pole at zero frequency. Around it, the DMFT self-energy for the Hubbard model has, on the real axis, the following form

$$\Sigma(\omega) \sim \frac{1}{2}\,U + \left(1 - \frac{1}{Z}\right)\omega - \frac{i}{2\tau^{\rm QP}},$$

where the positive dimensionless number Z yields the mass enhancement,

$$\frac{m^*}{m} \sim \frac{1}{Z} = 1 - \left.\frac{d\Sigma'(\omega)}{d\omega}\right|_{\omega\to 0}$$

and the positive parameter $\tau^{\rm QP}$ is the quasiparticles lifetime

$$\frac{1}{\tau^{\rm QP}} \sim -2Z\Sigma''(0) \propto (\pi k_B T)^2 + \omega^2.$$

At higher frequency the self-energy yields additional poles corresponding to the Hubbard bands. For large U, in the insulating regime, the central quasiparticle peak disappears, and only the Hubbard bands remain. The self-energy has approximately the form

$$\Sigma(\omega) \sim \frac{rU^2}{4}\left[\frac{1}{\omega} - i\pi\delta(\omega) - i f_U(\omega)\right],$$

where $f_U(\omega)$ is a positive function that is zero inside the gap and r is a model-specific renormalization factor. Hence, the real-part of the self-energy diverges at zero frequency, and there are no well defined low-energy quasiparticles. Furthermore, since we are assuming that the system is paramagnetic, the self-energy and the Green function are independent of spin. Thus, in DMFT the gap opens via the divergence at zero frequency in the real-part of the self-energy; this happens already in a single-site paramagnetic calculation, i.e., we do not have to assume any long-range magnetic order. What is then the relation between HF and DMFT? As can be understood from the discussion above, HF is not the large-U limit of DMFT. Since the HF self-energy is frequency independent, HF quasi-particles have infinite lifetime and bare masses ($Z = 1$ and $m^* = m$). These quasi-particles exist both in the metallic and in the insulating regime. It can be shown, however, that the DMFT self-energy reduces to the HF self-energy in the large-frequency limit. The main differences between HF and DMFT are pictorially shown in Fig. 11 for an idealized one-dimensional crystal.

5.3 DFT+*U* vs DFT+DMFT

The DFT+*U* method was the first systematic attempt to construct and solve *ab-initio* many-body Hamiltonians [3]. The model building part is very similar in DFT+DMFT and DFT+U, except that the latter was developed already fully embedded in density-functional theory, and therefore it might appear different at a first glance. In DFT+U the Coulomb interaction is treated in static mean-field theory, and therefore, as we can now understand, true many-body effects, such as the frequency dependence of the self-energy, are lost. Let us first assume that Hamiltonian (14) has the simplified form

$$\hat{H}^{\mathrm{LDA}} + \hat{U}^l - \hat{H}^l_{\mathrm{DC}} = \hat{H}^{\mathrm{LDA}} + \frac{1}{2}U \sum_i \sum_{m\sigma \neq m'\sigma'} \hat{n}_{im\sigma}\hat{n}_{im'\sigma'} - \underbrace{\frac{1}{2}U \sum_i \sum_{m\sigma \neq m'\sigma'} \langle \hat{n}_{im\sigma}\rangle \langle \hat{n}_{im'\sigma'}\rangle}_{\text{mean-field energy, } \mathrm{E_{MF}}} .$$

Next, we treat the Coulomb interaction in static mean-field via the HF decoupling; we approximate the mean-field energy in the expression above by the Hartree energy, taking for convenience as the energy zero the atomic chemical potential $\mu_{\mathrm{AT}} = U/2$

$$E_{MF} = E_H - \mu_{\mathrm{AT}} N^l = \frac{1}{2}UN^lN^l - \frac{1}{2}UN^l.$$

Here $N^l = \sum_{m\sigma}\langle \hat{n}_{im\sigma}\rangle$ is the number of heavy electrons per site. The mean-field Hamiltonian takes then the form

$$\hat{H} = \hat{H}^{\mathrm{LDA}} + \sum_{im\sigma} t^\sigma_m \hat{n}_{im\sigma}, \quad \text{with} \quad t^\sigma_m = U\left(\frac{1}{2} - \langle \hat{n}_{im\sigma}\rangle\right).$$

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 LDA orbital energies in this way is

$$E_{\mathrm{LDA+U}}[n] = E_{\mathrm{LDA}}[n] + \sum_i \left[\frac{1}{2}U \sum_{m\sigma \neq m'\sigma'} \langle \hat{n}_{im\sigma}\rangle \langle \hat{n}_{im'\sigma'}\rangle - E_{\mathrm{DC}}\right],$$

where the double-counting term is

$$E_{\mathrm{DC}} = \frac{1}{2}UN^l(N^l - 1)$$

and $E_{LDA}[n]$ is the total energy obtained using the spin-polarized version of the local-density approximation for the exchange-correlation functional. Indeed

$$\varepsilon^{\mathrm{LDA+U}}_{im\sigma} = \frac{\partial E_{\mathrm{LDA+U}}}{\partial \langle \hat{n}_{im\sigma}\rangle} = \varepsilon^{\mathrm{LDA}}_{im\sigma} + U\left(\frac{1}{2} - \langle \hat{n}_{im\sigma}\rangle\right) = \varepsilon^{\mathrm{LDA}}_{im\sigma} + t^\sigma_m.$$

More generally, the DFT+*U* functional is given by a form of the type

$$E_{\mathrm{LDA+U}}[n] = E_{\mathrm{LDA}}[n] + \frac{1}{2}\sum_{i\sigma} \sum_{mm'm''m'''} U_{mm''m'm'''} \langle \hat{n}^\sigma_{imm'}\rangle \langle \hat{n}^{-\sigma}_{im''m'''}\rangle$$

$$+ \frac{1}{2}\sum_{i\sigma} \sum_{mm'm''m'''} \left[U_{mm''m'm'''} - U_{mm''m'''m'}\right] \langle \hat{n}^\sigma_{imm'}\rangle \langle \hat{n}^\sigma_{im''m'''}\rangle - E_{\mathrm{DC}},$$

where $\langle \hat{n}^{\sigma}_{imm'} \rangle = \langle c^{\dagger}_{im\sigma} c_{im'\sigma} \rangle$ is the density matrix, and $\langle \hat{n}_{im\sigma} \rangle = \langle \hat{n}^{\sigma}_{imm} \rangle$. One of the most common recipes for the double-counting correction is the fully-localized limit

$$E_{\rm DC} = \frac{1}{2} U_{\rm avg} N^l (N^l - 1) - \frac{1}{2} J_{\rm avg} \sum_{\sigma} N^l_{\sigma} (N^l_{\sigma} - 1),$$

where

$$U_{\rm avg} = \frac{1}{(2l+1)^2} \sum_{m,m'} U_{mm'mm'},$$

$$U_{\rm avg} - J_{\rm avg} = \frac{1}{2l(2l+1)} \sum_{m,m'} (U_{mm'mm'} - U_{mm'm'm}).$$

The corresponding one-electron DFT+*U* Hamiltonian is

$$\hat{H} = \hat{H}^{\rm LDA} + \sum_{imm'\sigma} t^{\sigma}_{mm'} c^{\dagger}_{im\sigma} c_{im'\sigma}, \tag{17}$$

where

$$t^{\sigma}_{mm'} = \sum_{i\sigma} \sum_{m''m'''} U_{mm''m'm'''} \langle \hat{n}^{-\sigma}_{im''m'''} \rangle + \left[U_{mm''m'm'''} - U_{mm''m'''m'} \right] \langle \hat{n}^{\sigma}_{im''m'''} \rangle$$
$$- \left[U_{\rm avg} \left(N^l - \frac{1}{2} \right) - J_{\rm avg} \left(N^l_{\sigma} - \frac{1}{2} \right) \right] \delta_{m,m'}.$$

The second common recipe for the double-counting correction is the around mean-field limit; here the double-counting energy is the mean-field energy for equally occupied orbitals

$$E_{\rm DC} = U_{\rm avg} N^l_{\uparrow} N^l_{\downarrow} + \frac{2l}{2(2l+1)} \left(U_{\rm avg} - J_{\rm avg} \right) \left(N^2_{\uparrow} + N^2_{\downarrow} \right).$$

The corresponding one-electron LDA+*U* Hamiltonian is (17) with

$$t^{\sigma}_{mm'} = \sum_{i\sigma} \sum_{m''m'''} U_{mm''m'm'''} \langle \hat{n}^{-\sigma}_{im''m'''} \rangle + \left[U_{mm''m'm'''} - U_{mm''m'''m'} \right] \langle \hat{n}^{\sigma}_{im''m'''} \rangle$$
$$- \left[U_{\rm avg} \left(N^l - n_{\sigma} \right) - J_{\rm avg} \left(N^l_{\sigma} - n_{\sigma} \right) \right] \delta_{m,m'},$$

where $n_{\sigma} = N_l/(2(2l+1))$ is the average occupation per spin. In DFT+DMFT the same recipes are used for the double-counting correction; this is reasonable because the source of double-counting is the same in the two methods. In DFT+*U*, differently than in static mean-field for model Hamiltonians, $\hat{H}^{\rm LDA}$ is obtained self-consistently. The DFT+*U* correction in (17) modifies the occupations of the correlated sector 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 (17). Using

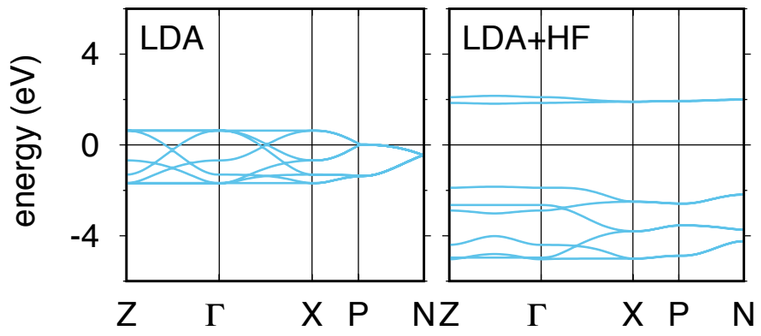


Fig. 12: *Left: LDA e_g band structure of cubic $KCuF_3$ calculated using the experimental magnetic unit cell with four formula units. Right: Static mean-field band structure, calculated for the experimental orbital and spin order. Parameters: $U = 7\ eV$ and $J = 0.9$ eV.*

the same procedure, charge self-consistency can of course also be achieved in DFT+DMFT calculations. A difficulty is, however, the basis. DFT+*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 DFT+*U* correction is determined through projections onto such atomic orbitals. Thus DFT+*U* results are essentially dependent on the choice of the set of correlated electrons and their atomic spheres. If the correlated electrons are well localized, however, they retain to a good extent their atomic character in a solid. Thus, within reasonable sphere choices, the dependence on the sphere size is less crucial than could be expected. Still, from a theoretical point of view, there is an inconsistency in this procedure. Orbitals defined only within the atomic spheres do not form a complete basis (even for the correlated sector), and thus they do not really yield a many-body Hamiltonian of the form (14). One of the successes of DFT+*U* is that it describes well the magnetic ground-state of Mott insulators. The method has however all defects of the HF approximation; it opens a gap by making long-range order, eigenvalues are real, and quasi-particles have an infinite lifetime. One further example of the failure of DFT+U is the description of the superexchange driven orbital-order transition. Let us consider the insulating perovskite $KCuF_3$ as representative material. Instead of the full DFT+*U* calculation, for simplicity we discuss the calculation for the e_g-band Hubbard model describing the low-energy states and do not perform any charge self-consistency. For this Hamiltonian the double-counting correction is a mere shift of the chemical potential and can be neglected. The model has the form

$$\hat{H} = -\sum_{m,m',i,i',\sigma} t^{i,i'}_{mm'} c^{\dagger}_{im\sigma} c_{im'\sigma} + U \sum_{i\ m} \hat{n}_{im\uparrow}\hat{n}_{im\downarrow} + \frac{1}{2} \sum_{\substack{i\sigma\sigma' \\ m\neq m'}} (U - 2J - J\delta_{\sigma,\sigma'}) \hat{n}_{im\sigma}\hat{n}_{im'\sigma'}$$

$$-J \sum_{i\ m\neq m'} \left[c^{\dagger}_{im\uparrow} c^{\dagger}_{im\downarrow} c_{im'\uparrow} c_{im'\downarrow} + c^{\dagger}_{im\uparrow} c_{im\downarrow} c^{\dagger}_{im'\downarrow} c_{im'\uparrow} \right]$$

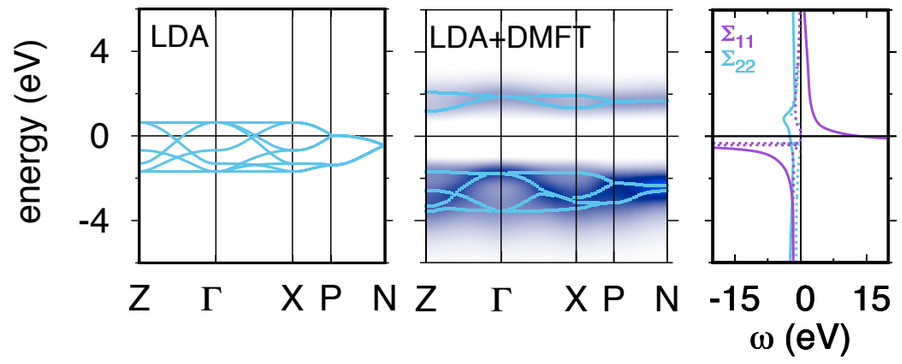


Fig. 13: *Left: LDA band structure of $KCuF_3$, e_g bands. Center: corresponding LDA+DMFT correlated band structure in the orbitally-ordered phase [16]. Dots: poles of the Green function. Right: Self-energy matrix in the basis of the natural orbitals. Full lines: real part. Dotted lines: imaginary part. Parameters: $U = 7$ eV, $J = 0.9$ eV.*

where $m, m' = 3z^2 - r^2, x^2 - y^2$. The last two terms describe the pair-hopping ($U_{mmm'm'} = J_{m,m'}$ for real harmonics, while for spherical harmonics $U_{mmm'm'} = 0$) and spin-flip processes. The question to be answered is: Can orbital order arise spontaneously for the ideal perovskite cubic structure in which the e_g orbitals are degenerate? $KCuF_3$ is, in nature, an insulator, but in LDA it turns out to be metallic and exhibits no orbital order. This can be seen in Fig. 12. In DFT+U, in order to open the gap, we have first to double the cell. The gap opens only in the presence of both spin and orbital order. This means that there is no phase in which the system is non-magnetic but has a gap and exhibits orbital order. We could paraphrase this result by saying that,[1] in DFT+U

$$T_N = T_{OO} = T_{MI}$$

where T_N, T_{OO}, T_{MI} are the critical temperature at which the magnetic, orbital and metal-insulator transition occur. For the magnetic orbitally ordered phase, the resulting electronic structure is shown in Fig. 12. Let us now discuss the solution of the same problem with DFT+DMFT. With this approach we obtain an insulating orbitally ordered solution even in the absence of long-range magnetic order. DMFT describes the correct sequence of phenomena

$$T_N < T_{OO} < T_{MI}.$$

In Fig. 13 we show the DFT+DMFT *paramagnetic* e_g correlated band structure of $KCuF_3$ in the orbitally ordered phase. We can compare these bands with the static mean-field *antiferromagnetic* band structure in Fig. 12. The DFT+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. 13), which in static mean-field theory does not depend on ω, diverges at low frequencies.

[1] This, of course, oversimplifies the discussion, since DFT+U is a $T = 0$ method.

6 Conclusions

In this lecture we have discussed two methods that can be used for describing the metal-insulator transition in Hubbard-like models. The first method is the Hartree-Fock approach. Here the Coulomb interaction part of the Hubbard Hamiltonian is treated at the static mean-field level. The occupation matrix is determined self-consistently. The Hartree-Fock self-energy is equivalent to a site-, orbital-, and spin-dependent potential. The metal-insulator transition occurs via an enlargement of the unit cell and a lowering of the symmetry. In this approach, all states have infinite lifetime and the masses of electrons are not renormalized. This is a Slater-type transition. The Hartree-Fock method is the basis of the DFT+U approach. Numerically, DFT+U calculations are as fast as standard DFT calculations. Furthermore, in DFT+U, the Hartree-Fock correction is embedded in the DFT formalism via a modification of the total-energy functional. One drawback is that typically the correction is for orbitals defined within an atomic sphere, and not Wannier function. This means that, strictly speaking, we could not use them alone to directly construct generalized Hubbard models. The second approach examined in this lecture is DMFT, dynamical mean-field theory. In DMFT the Hubbard model is mapped onto a quantum-impurity model, for example the Anderson Hamiltonian. The latter is solved exactly and yields the impurity self-energy. The hybridization function of the Anderson Hamiltonian is determined self-consistently, requiring the local Green function equals the impurity Green function. The central approximation is that the self-energy of the Hubbard model is assumed to be local. This approximation becomes progressively better with increasing coordination number; in infinite dimensions, indeed, the self-energy is local. In DMFT the metal-insulator transition has a very different nature than in Hartree-Fock. It occurs already above the magnetic transition and it happens via a divergence of the low-frequency self-energy. Switching on the Coulomb interaction leads at first to the formation of heavy quasi-particles with renormalized masses and finite life-times. Eventually, when U is above a critical value, the self-energy and the masses diverge, and the spectral function exhibits a gap. The metal-insulator transition described via DMFT is of Mott type. In DMFT we neglect the momentum-dependence of the self-energy. As we have seen in the case of the Hubbard dimer, such effects can be important; in particular they become important in low dimensions. Straightforward extensions of DMFT to include some non-local effects are cluster approaches, in real and $\mathbf{k}$ space. The bottleneck of DMFT is, however, the quantum impurity solver, typically quantum Monte Carlo. The computational time can increase very rapidly with the number of orbitals and sites, and eventually the infamous sign problem appears. The model and the cluster size has thus to be carefully chosen. The DMFT approach can be used also for realistic Hamiltonians built via density-functional theory. This is the DFT+DMFT approach. In both DFT+U and DMFT+U, a double-counting correction has to be subtracted, since part of the Coulomb effects are already included in the DFT functional, for example the LDA.

References

[1] K.F. Sundman, Acta Mathematica **36**, 105 (1912)

[2] F. Diacu, The Mathematical Intelligencer **18**, 66 (1996)

[3] A.I. Lichtenstein, M.I. Katsnelson, and V.A. Gubanov, J. Phys. F **14**, L125;
Solid State Commun. **54**, 327 (1985);
A.I. Lichtenstein, M.I. Katsnelson, V.P. Antropov, and V.A. Gubanov,
J. Magn. Magn. Mater. **67**, 65 (1987);
M.I. Katsnelson and A.I. Lichtenstein, Phys. Rev. B **61**, 8906 (2000)

[4] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989);
A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)

[5] M. Jarrell, Phys. Rev. Lett. **69**, 168 (1992)

[6] E. Müller-Hartmann, Z. Phys. B **74**, 507 (1989);
Z. Phys. B **76**, 211 (1989); Int. J. Mod. Phys. B **3**, 2169 (1989)

[7] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt:
The LDA+DMFT approach to strongly-correlated materials,
Reihe Modeling and Simulation, Vol. 1 (Forschungszentrum Jülich, 2011)
`http://www.cond-mat.de/events/correl11`

[8] E. Pavarini, E. Koch, A. Lichtenstein, D. Vollhardt: *DMFT at 25: Infinite Dimensions*,
Reihe Modeling and Simulation, Vol. 4 (Forschungszentrum Jülich, 2014)
`http://www.cond-mat.de/events/correl14`

[9] A. Lagendijk, B. van Tiggelen, and D.S. Wiersma, Physics Today **62**, 24 (2009)

[10] J.R. Schrieffer and P.A. Wolff, Phys. Rev. **149**, 491 (1966);
A.H. MacDonald, S.M. Girvin and D. Yoshioka, Phys. Rev. B **37**, 9753 (1988)

[11] L. Cano-Cortés, A. Dolfen, J. Merino, J. Behler, B. Delley, K. Reuter and E. Koch,
Eur. Phys. J. B **56**, 173 (2007)

[12] E. Pavarini, I. Dasgupta, T. Saha-Dasgupta, O. Jepsen and O.K. Andersen,
Phys. Rev. Lett. **87**, 047003 (2001)

[13] P.W. Anderson, J. Phys. C: Solid State Phys. **3**, 2436 (1970)

[14] K. Wilson, Rev. Mod. Phys. **47**, 773 (1975)

[15] P. Nozières, J. Low. Temp. Phys. **17**, 31 (1974)

[16] E. Pavarini, E. Koch, A.I. Lichtenstein, Phys. Rev. Lett. **101**, 266405 (2008)

10 Path Integrals and Dual Fermions

Alexander Lichtenstein
I. Institut für Theoretische Physik
Universität Hamburg, 20355 Hamburg

Contents

1 Introduction: DMFT and beyond **2**

2 Path-integral for fermions **4**

3 Functional approach **8**

4 Dual Fermion approach for non-local correlations **11**

5 Dual Boson approach for non-local interactions **17**

6 Numerical results **20**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction: DMFT and beyond

In this lecture we give an introduction to the theoretical description of strongly correlated materials based on non-local extensions of dynamical mean-field theory (DMFT). This scheme combines the numerically exact DMFT solution of an effective impurity problem with an analytical non-local perturbation scheme. The frequency-dependent effective-impurity DMFT problem is solved within the continuous-time quantum Monte Carlo (CT-QMC) scheme [1]. Therefore the perturbation theory needs to be formulated in the action path integral formalism. We give a brief introduction to the path integral over fermionic Grassmann fields and formulate a general scheme for expansion around the DMFT solution using a special dual space transformation. We discuss a general way to include nonlocal correlations beyond DMFT for generalized Hubbard models, based on the dual-fermion [2] and the dual-boson approach [3].

Consider the noninteracting, "kinetic" part H_t of the Hubbard model first [4]. It is determined by specifying the hopping-matrix elements t_{ij} between sites i and j. In the absence of the local Hubbard-interaction term, H_t is easily diagonalized. For a Hubbard model on a lattice, diagonalization is achieved by Fourier transforming the hopping parameters to $\mathbf{k}$-space, where we obtain the normal "band structure" $\varepsilon_{\mathbf{k}}$ with band width W. If, on the other hand, only the local part of the Hamiltonian is kept, i.e., the Hubbard interaction H_U with interaction strength U and the local term of H_t is given by an on-site energy ε_0 only, the diagonalization of the Hamiltonian is trivial again and reduces to the diagonalization of a single "Hubbard atom".

The great success of the DMFT approach is related to its ability to numerically interpolate between these two limits [5]. For the half-filled Hubbard model on an infinite-dimensional Bethe lattice DMFT gives the exact description of the Mott-transition [6] between the weak-coupling ($U/W \ll 1$) metallic state and the strong-coupling ($U/W \gg 1$) insulating paramagnetic state [7]. In a nutshell, DMFT maps the correlated Hubbard lattice problem onto the self-consistent solution of an effective Anderson impurity problem with a single interacting Hubbard atom (interaction strength U) in a non-interacting fermionic bath, which mimics the rest of the crystal. This impurity model, which is fully determined by the local hybridization function Δ_ν on fermionic Matsubara frequencies $i\nu_n$ is solved by the numerically exact CT-QMC scheme and exact the local Green function g_ν is obtained. The DMFT self-consistency condition for the hybridization function equates the local part of the lattice Green function and the impurity Green function

$$\sum_{\mathbf{k}} \left(g_\nu^{-1} + \Delta_\nu - \varepsilon_{\mathbf{k}} \right)^{-1} = g_\nu \,,$$

which shows that DMFT minimizes, in local sense, the $|\varepsilon_{\mathbf{k}} - \Delta_\nu|$ distance.

We can now think about how to incorporate nonlocal correlations beyond DMFT: since the Hubbard and the Anderson-impurity model share the same interaction part, we can think of the Hubbard model as the impurity model plus a residual term $\propto \varepsilon_{\mathbf{k}} - \Delta_\nu$ and treat this *perturbatively*. Since this term is frequency dependent we need a novel perturbation theory based on the action formalism. One may view this idea as a generalization of the Kohn-Sham idea in density functional theory (DFT) [8] of an optimal reference system, but with a crucial difference. Here,

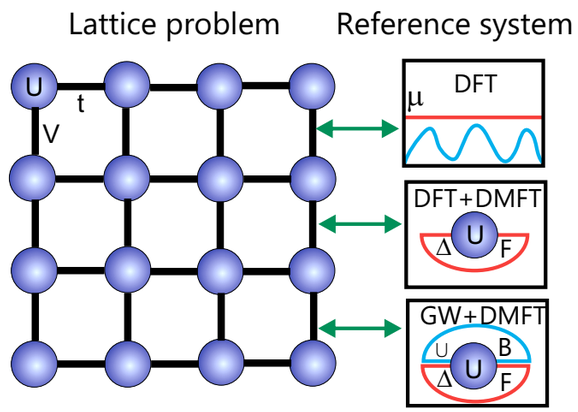


Fig. 1: *Schematic representation of the reference systems in many-body approaches to lattice-fermion models: (i) Density-functional theory (DFT) with the interacting homogeneous electron gas as a reference system, defined by a constant external potential μ. (ii) Dynamical mean-field theory (DMFT) with an effective impurity problem as a reference system, defined by afermionic bath, specified by the hybridization function Δ. (iii) GW+DMFT with a correlated atom in a fermionic (Δ) and a bosonic bath (Λ) due to effects of the frequency-dependent screening of long-range Coulomb (V) interactions.*

not an interacting homogeneous electron gas, but an effective impurity model, tailored to the problem of strong correlations, serves as the reference system, see Fig. 1. Since at zeroth order of this perturbative expansion, i.e., on the level of the DMFT problem, we already have an interacting problem and since the perturbation is momentum and frequency dependent, we are forced to replace the Hamiltonians by actions within the path-integral formalism. Note that the fermion path integral can also be used to formulate the DMFT itself [5, 9]. Now, the separation of local and nonlocal terms is achieved by a Hubbard-Stratonovich transformation applied to the single-particle $(\varepsilon_{\mathbf{k}} - \Delta_\nu)$-term [2]. This provides us with a new action. Moreover, it is formally possible to integrate-out the original local degrees of freedom and in this way generate an effective action in the transformed, so-called dual-fermion representation [2]. Note that integrating out the local degrees of freedom is not just a formal step but can be achieved in practice, by solving the impurity problem within the numerically exact CT-QMC method.

The dual action consists of a bare dual propagator (the non-local part of the DMFT Green function) $\tilde{G}^0_{\mathbf{k},\nu} = [g_\nu^{-1} + \Delta_\nu - \varepsilon_{\mathbf{k}}]^{-1} - g_\nu$, and a local but frequency-dependent effective potential related to scattering processes of two, three, and more dual particles on the impurity site. The simplest two-particle dual potential coincides with the fully connected part of the screened impurity interaction vertex $\gamma^\omega_{\nu\nu'}$, which can be calculated with the impurity CT-QMC solver as a function of bosonic (ω) and fermionic (ν, ν') Matsubara frequencies. Normally, correlations between three particles on the DMFT impurity site are much weaker than two-particle correlations and can be ignored. The same applies to higher-order terms. One can think of the dual-fermion formalism as an expansion in the order of local multi-particle correlation func-

tions. This means that "bare" interactions between dual fermions are related with the connected part of the screened impurity vertex. Standard diagrammatic techniques can be applied for calculations of the bold dual propagator $\tilde{G}_{\mathbf{k},\nu}$, which allows to obtain the nonlocal self-energy for the original fermions [2] and to describe nonlocal correlations beyond the DMFT.
The dual-fermion approach is not necessarily bound to a specific starting point. However, the DMFT starting point is very efficient. Namely, it corresponds to the elimination of all local diagrams for any n-particle correlation of dual fermions when using the DMFT self-consistency equation (1). In the dual space, this simply reduces to $\sum_{\mathbf{k}} \tilde{G}^0_{\mathbf{k},\nu} = 0$ and means that, on average over the whole Brillouin zone, Δ_ν optimally approximates the electron spectrum $\varepsilon_{\mathbf{k}}$, including its local correlation effects. Therefore, the noninteracting dual fermions correspond to strongly correlated DMFT quasiparticles, and the remaining nonlocal effects can be quite small and reasonably described by, e.g., ladder summations of dual diagrams. This also explains the notion "dual fermions".

2 Path-integral for fermions

We first introduce a formalism of path-integration over fermionic fields [10]. Let us consider a simple case of a single quantum state $|i\rangle$ occupied by fermionic particles [11]. Due to the Pauli principle the many-body Hilbert space is spanned only by two orthonormal states $|0\rangle$ and $|1\rangle$. In the second quantization scheme for fermions with annihilation $\hat{c}_i$ and creations $\hat{c}_i^\dagger$ operators with anticommutation relations $\{\hat{c}_i, \hat{c}_j^\dagger\} = \delta_{ij}$ we have the following simple rules

$$\begin{aligned} \hat{c}_i\,|1\rangle = |0\rangle \qquad & \hat{c}_i\,|0\rangle = 0 \\ \hat{c}_i^\dagger\,|0\rangle = |1\rangle \qquad & \hat{c}_i^\dagger\,|1\rangle = 0\,. \end{aligned} \tag{1}$$

Moreover, the density operator and the Pauli principle take the form

$$\hat{c}_i^\dagger \hat{c}_i\,|n\rangle = n_i\,|n\rangle \qquad \text{and} \qquad \hat{c}_i^2 = (\hat{c}_i^\dagger)^2 = 0\,.$$

The central object here are the so-called fermionic coherent states $|c\rangle$, which are eigenstates of annihilation operator $\hat{c}_i$ with eigenvalue c_i:

$$\hat{c}_i\,|c\rangle = c_i\,|c\rangle\,. \tag{2}$$

It is worthwhile to note that such a left-eigenbasis has only annihilation operators, due to the fact that they are bounded from below and one can rewrite one of equation from Eq. (1) in the following "eigenvalue" form

$$\hat{c}_i\,|0\rangle = 0\,|0\rangle\,.$$

Due to the anti-commutation relations for the fermionic operators the eigenvalues of coherent states c_i are so-called Grassmann numbers with the following multiplication rules [12]:

$$c_i c_j = -c_j c_i \qquad \text{and} \qquad c_i^2 = 0\,. \tag{3}$$

It is convenient to assume that the Grassmann numbers also anti-commute with the fermionic operators

$$\{c, \hat{c}\} = \{c, \hat{c}^\dagger\} = 0\,.$$

An arbitrary function of a Grassmann variable can be represented only by its first two Taylor coefficients

$$f(c) = f_0 + f_1 c\,. \tag{4}$$

One can prove the following general many-body representation of coherent states

$$|c\rangle = e^{-\sum_i c_i \hat{c}_i^\dagger}\,|0\rangle\,. \tag{5}$$

Let us show this for the simple case of one fermionic state:

$$\hat{c}\,|c\rangle = \hat{c}(1 - c\hat{c}^\dagger)\,|0\rangle = \hat{c}\,(|0\rangle - c\,|1\rangle) = -\hat{c}c\,|1\rangle = c\,|0\rangle = c\,|c\rangle\,. \tag{6}$$

One can also define a "left" coherent state $\langle c|$ as the left-eigenstates of creations operators $\hat{c}_i^\dagger$

$$\langle c|\,\hat{c}_i^\dagger = \langle c|\,c_i^*\,.$$

Note that new eigenvalues c_i^* is just another Grassmann number and not a complex conjugate of c_i. The left coherent state can be obtained similarly to Eq. (5) as following

$$\langle c| = \langle 0|\,e^{-\sum_i \hat{c}_i c_i^*}.$$

A general function of two Grassmann variables, analogously to Eq. (4), can be represented by only four Taylor coefficients

$$f(c^*, c) = f_{00} + f_{10}c^* + f_{01}c + f_{11}c^*c\,. \tag{7}$$

Using this expansion we can define a derivative of Grassmann variables in the natural way

$$\frac{\partial c_i}{\partial c_j} = \delta_{ij}\,.$$

One needs to be careful with the "right-order" in such a derivative and remember the anti-commutation rules, i.e.

$$\frac{\partial}{\partial c_2} c_1 c_2 = -c_1\,.$$

For the case of a general two-variable function in Eq. (7) we have

$$\frac{\partial}{\partial c^*}\frac{\partial}{\partial c} f(c^*, c) = \frac{\partial}{\partial c^*}(f_{01} - f_{11}c^*) = -f_{11} = -\frac{\partial}{\partial c}\frac{\partial}{\partial c^*} f(c^*, c)\,.$$

One also needs a formal definition of the integration over Grassmann variables, and the natural way consists with the following rules [12]:

$$\int 1\,dc = 0 \qquad \text{and} \qquad \int c\,dc = 1$$

which just show that integration over a Grassmann variable is equivalent to differentiation

$$\int ... \, dc \to \frac{\partial}{\partial c} ...$$

The coherent states are not orthonormal and the overlap of any two coherent fermionic states is

$$\langle c|c\rangle = e^{\sum_i c_i^* c_i},$$

which is easy to see for the case of one state

$$\langle c|c\rangle = \left(\langle 0| - \langle 1| \, c^*\right)\left(|0\rangle - c\,|1\rangle\right) = 1 + c^* c = e^{c^* c}.$$

An important property of coherent states is related to resolution of the unity operator

$$\int dc^* \int dc \; e^{-\sum_i c_i^* c_i} \, |c\rangle \, \langle c| = \hat{1} = \iint dc^* \, dc \; \frac{|c\rangle \, \langle c|}{\langle c|c\rangle} \, .$$

For simplicity we demonstrate this relation only for one fermion state:

$$\iint dc^* \, dc \; e^{-c^* c} \, |c\rangle \, \langle c| = \iint dc^* \, dc \, (1 - c^* c) \left(|0\rangle - c\,|1\rangle\right)\left(\langle 0| - \langle 1| \, c^*\right) =$$
$$- \iint dc^* \, dc \; c^* c \left(|0\rangle \, \langle 0| + |1\rangle \, \langle 1|\right) = \sum_n |n\rangle \, \langle n| = \hat{1} \, .$$

Matrix elements of normally ordered operators are very easy to calculate in a coherent basis by operating of $\hat{c}^\dagger$ to the states on the right and $\hat{c}$ to the left ones

$$\langle c^*| \, \hat{H}(\hat{c}^\dagger, \hat{c}) \, |c\rangle = H(c^*, c) \, \langle c^*|c\rangle = H(c^*, c) \, e^{\sum_i c_i^* c_i} \tag{8}$$

Within the manifold of coherent states we can map the fermionic operators to functions of Grassmann variables $(\hat{c}_i^\dagger, \hat{c}_i) \to (c_i^*, c_i)$.

Finally, we prove the so-called "trance-formula" for arbitrary fermionic operators in normal order (in one fermion notation):

$$\begin{aligned} Tr \, \hat{O} &= \sum_{n=0,1} \langle n| \, \hat{O} \, |n\rangle = \sum_{n=0,1} \iint dc^* \, dc \; e^{-c^* c} \, \langle n| \, c\rangle \, \langle c| \, \hat{O} \, |n\rangle = \\ &= \iint dc^* \, dc \; e^{-c^* c} \sum_{n=0,1} \langle -c| \, \hat{O} \, |n\rangle \, \langle n| \, c\rangle = \iint dc^* \, dc \; e^{-c^* c} \, \langle -c| \, \hat{O} \, |c\rangle \, . \end{aligned}$$

The fermionic "minus" sign in the left coherent states comes from the commutation of the (c^*) and (c) coherent states in such a transformation: $\langle n|c\rangle \, \langle c|n\rangle = \langle -c|n\rangle \, \langle n|c\rangle$. One has to use the standard Grassmann rules: $c_i^* c_j = -c_j c_i^*$ and $|-c\rangle = |0\rangle + c\,|1\rangle$.

We are now ready to write a partition function or the grand-canonical quantum ensemble with $H = \hat{H} - \mu \hat{N}$ and inverse temperature β. We have to use the N-slices Trotter decomposition

for the partition function in the interval $[0,\beta)$ with imaginary times $\tau_n = n\Delta\tau = n\beta/N$ ($n = 1,...,N$), and insert N-times the resolution of unity as follows

$$\begin{aligned} Z &= Tr\, e^{-\beta H} = \iint dc^*\, dc\, e^{-c^* c} \left\langle -c\right| e^{-\beta H} \left|c\right\rangle \\ &= \int \Pi_{n=1}^{N} dc_n^*\, dc_n\, e^{-\sum_n c_n^* c_n} \left\langle c_N\right| e^{-\Delta\tau H} \left|c_{N-1}\right\rangle \left\langle c_{N-1}\right| e^{-\Delta\tau H} \left|c_{N-2}\right\rangle ... \left\langle c_1\right| e^{-\Delta\tau H} \left|c_0\right\rangle \\ &= \int \Pi_{n=1}^{N} dc_n^*\, dc_n\ \exp\left(-\Delta\tau \sum_{n=1}^{N} \left[c_n^* \left(c_n - c_{n-1}\right)/\Delta\tau + H\left(c_n^*, c_{n-1}\right)\right]\right). \end{aligned}$$

In the continuum limit ($N \mapsto \infty$)

$$\begin{aligned} \Delta\tau \sum_{n=1}^{N} ... &\mapsto \int_0^\beta d\tau\, ... \\ \frac{c_n - c_{n-1}}{\Delta\tau} &\mapsto \partial_\tau \\ \Pi_{n=0}^{N-1} dc_n^*\, dc_n &\mapsto D\left[c^*, c\right] \end{aligned}$$

with antiperiodic boundary conditions for the fermionic Grassmann variables on an imaginary time $c(\tau)$ and $c^*(\tau)$

$$c(\beta) = -c(0), \qquad c^*(\beta) = -c^*(0)$$

we end up with the standard path integral formulation of the quantum partition function

$$Z = \int D\left[c^*, c\right] \exp\left(-\int_0^\beta d\tau \left[c^*(\tau)\partial_\tau c(\tau) + H\left(c^*(\tau), c(\tau)\right)\right]\right). \tag{9}$$

For later discussion we mention the Gaussian path integral for a non-interacting "quadratic" fermionic action. For an arbitrary matrix M_{ij} and Grassmann vectors J_i^* and J_i one can calculate analytically the following integral

$$Z_0\left[J^*, J\right] = \int D\left[c^* c\right]\, e^{-\sum_{i,j=1}^{N} c_i^* M_{ij} c_j + \sum_{i=1}^{N}\left(c_i^* J_i + J_i^* c_i\right)} = \det\left[M\right] e^{-\sum_{i,j=1}^{N} J_i^* \left(M^{-1}\right)_{ij} J_j}.$$

To prove this relation one needs first to complete the square in order to eliminate J_i^* and J_i and expand the exponential function (only the N-th oder is non-zero):

$$e^{-\sum_{i,j=1}^{N} c_i^* M_{ij} c_j} = \frac{1}{N!}\left(-\sum_{i,j=1}^{N} c_i^* M_{ij} c_j\right)^N.$$

Finally, different permutations of c_i^* and c_j, and integration over the Grassmann variables gives the $\det M$-answer. As a small exercise we will check such an integral for first two many-particle dimensions. For $N = 1$ it is trivial:

$$\int D\left[c^* c\right]\, e^{-c_1^* M_{11} c_1} = \int D\left[c^* c\right]\, \left(-c_1^* M_{11} c_1\right) = M_{11} = \det M$$

and for $N = 2$ we have

$$\int D\left[c^*c\right]\, e^{-c_1^*M_{11}c_1 - c_1^*M_{12}c_1 - c_2^*M_{21}c_1 - c_2^*M_{22}c_2} =$$
$$\frac{1}{2!}\int D\left[c^*c\right]\, \left(-c_1^*M_{11}c_1 - c_1^*M_{12}c_1 - c_2^*M_{21}c_1 - c_2^*M_{22}c_2\right)^2 = M_{11}M_{22} - M_{12}M_{21} = \det M\,.$$

For the shift (change) of variables in the path integral one uses the following transformation with unit Jacobian: $c \to c - M^{-1}J$ and

$$c^*Mc - c^*J - J^*c = \left(c^* - J^*M^{-1}\right) M \left(c - M^{-1}J\right) - J^*M^{-1}J\,.$$

Using the Gaussian path integral it is very easy to calculate any correlation function of a non-interaction action (Wick-theorem):

$$\begin{aligned}
\left\langle c_i c_j^* \right\rangle_0 &= -\frac{1}{Z_0}\frac{\delta^2 Z_0\left[J^*,J\right]}{\delta J_i^*\,\delta J_j}\bigg|_{J=0} = M_{ij}^{-1} \\
\left\langle c_i c_j c_k^* c_l^* \right\rangle_0 &= \frac{1}{Z_0}\frac{\delta^4 Z_0\left[J^*,J\right]}{\delta J_i^*\delta J_j^*\delta J_l\delta J_k}\bigg|_{J=0} = M_{il}^{-1}M_{jk}^{-1} - M_{ik}^{-1}M_{jl}^{-1}.
\end{aligned}$$

Corresponding bosonic path-integrals can be formulated in a similar way with complex variables and periodic boundary conditions in imaginary time. The Gaussian path integral over bosonic fields is equal to inverse of the M-matrix determinant [10].

3 Functional approach

We introduce a general functional approach which will cover the Density-Functional Theory (DFT), Dynamical Mean-Field Theory (DMFT), and Baym-Kadanoff (BK) theories [9, 13]. Let us start from the full many–body Hamiltonian describing electrons moving in the periodic external potential of ions $V(\mathbf{r})$ with the chemical potential μ and interacting via the Coulomb repulsion $U(\mathbf{r}-\mathbf{r}') = 1/|\mathbf{r}-\mathbf{r}'|$. We use the atomic units $\hbar = m = e = 1$. In the field-operator representation the Hamiltonian has the form

$$\begin{aligned}
H &= \sum_\sigma \int d\mathbf{r}\, \hat{\psi}_\sigma^\dagger(\mathbf{r}) \left(-\frac{1}{2}\bigtriangledown^2 + V(\mathbf{r}) - \mu\right) \hat{\psi}_\sigma(\mathbf{r}) \\
&+\frac{1}{2}\sum_{\sigma\sigma'} \int d\mathbf{r} \int d\mathbf{r}'\, \hat{\psi}_\sigma^\dagger(\mathbf{r})\hat{\psi}_{\sigma'}^\dagger(\mathbf{r}')\, U(\mathbf{r}-\mathbf{r}')\, \hat{\psi}_{\sigma'}(\mathbf{r}')\hat{\psi}_\sigma(\mathbf{r}).
\end{aligned} \tag{10}$$

We can always use the single-particle orthonormal basis set in solids $\phi_n(\mathbf{r})$ for example Wannier orbitals with a full set of quantum numbers, e.g. site, orbital and spin index: $n = (i, m, \sigma)$, and expand the fields in creation and annihilation operators

$$\begin{aligned}
\hat{\psi}(\mathbf{r}) &= \sum_n \phi_n(\mathbf{r})\, \hat{c}_n \\
\hat{\psi}^\dagger(\mathbf{r}) &= \sum_n \phi_n^*(\mathbf{r})\, \hat{c}_n^\dagger\,.
\end{aligned} \tag{11}$$

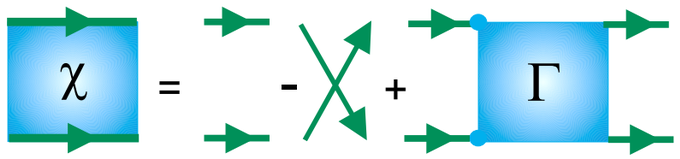


Fig. 2: *Representation of the full two-particle Green function in terms trivial products of single-particle Green functions and the full vertex function Γ.*

Going from fermionic operators to Grassmann variables $\{c_n^*,\, c_n\}$, we can write the functional integral representation of the partition function of the many-body Hamiltonian in the imaginary time domain using the Euclidean action S

$$Z = \int \mathcal{D}\,[c^*, c]\, e^{-S} \tag{12}$$

$$S = \sum_{12} c_1^*(\partial_\tau + t_{12})c_2 + \frac{1}{2}\sum_{1234} c_1^* c_2^*\, U_{1234}\, c_4 c_3\,, \tag{13}$$

where the one- and two-electron matrix elements are defined as

$$t_{12} = \int d\mathbf{r}\, \phi_1^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 + V(\mathbf{r}) - \mu\right) \phi_2(\mathbf{r}) \tag{14}$$

$$U_{1234} = \int d\mathbf{r} \int d\mathbf{r}'\, \phi_1^*(\mathbf{r})\phi_2^*(\mathbf{r}')\, U(\mathbf{r}-\mathbf{r}')\, \phi_3(\mathbf{r})\phi_4(\mathbf{r}')$$

and we use the following short definition of the sum:

$$\sum_1 \ldots \equiv \sum_{im} \int d\tau \ldots \tag{15}$$

The one-electron Green function is defined via a simple non-zero correlation function for fermions

$$G_{12} = -\langle c_1 c_2^* \rangle_S = -\frac{1}{Z}\int \mathcal{D}\,[c^*, c] c_1 c_2^*\, \exp(-S). \tag{16}$$

The main problem of strongly interacting electronic systems is related to the fact that the higher order correlation functions do not separate into a product of lower order correlation functions. For example the two-particle Green function or generalized susceptibility (X) is defined in the following form [14]

$$X_{1234} = \langle c_1 c_2 c_3^* c_4^* \rangle_S = \frac{1}{Z}\int \mathcal{D}[c^*, c]\, c_1 c_2 c_3^* c_4^* \exp(-S), \tag{17}$$

and can be expressed graphically through the Green functions and the full vertex function $\mathit{\Gamma}_{1234}$ [15] (see Fig. (2))

$$X_{1234} = G_{14}G_{23} - G_{13}G_{24} + \sum_{1'2'3'4'} G_{11'}G_{22'}\mathit{\Gamma}_{1'2'3'4'}G_{3'3}G_{4'4}\,. \tag{18}$$

In the case of non-interacting electron systems, the high-order correlations X are reduced to the antisymmetrized product of lower-order correlations G, which would correspond to the first two

terms (Hartree and Fock like), with the vertex function Γ in Eq. (18) equal to zero. In strongly correlated electron systems the last part with the vertex is dominant and even diverges close to electronic phase transitions.

The Baym–Kadanoff functional [13] gives the one–particle Green function and the total free energy at its stationary point. In order to construct the exact functional of the Green function (Baym–Kadanoff) we modify the action by introducing a source term J in the following form

$$S[J] = S + \sum_{12} c_1^* J_{12} c_2 \,. \tag{19}$$

The partition function Z, or equivalently the free energy of the system F, then becomes a functional of the auxiliary source field

$$Z[J] = e^{-F[J]} = \int \mathcal{D}\,[c^*, c] e^{-S'[J]} \,. \tag{20}$$

Variation with respect to this source function gives all correlation functions, for example the Green function

$$G_{12} = \left. \frac{\delta F[J]}{\delta J_{21}} \right|_{J=0} . \tag{21}$$

If we use the definition of the generalized susceptibility as the second variation of the $F[J]$ functional instead of $Z[J]$ we will get only the connected part of correlation the X-function which is represented by the last term in Eq. (18).

The Baym–Kadanoff functional can be obtained by Legendre transform from variable J to G

$$F[G] = F[J] - \mathrm{Tr}\,(JG) \,. \tag{22}$$

We can use the standard decomposition of the free energy F into the single particle part and the correlated part

$$F[G] = \mathrm{Tr} \ln G - \mathrm{Tr}\,(\Sigma G) + \Phi[G], \tag{23}$$

were Σ is single particle self-energy and $\Phi[G]$ is the correlated part of the Baym–Kadanoff functional that is equal to the sum of all two-particle irreducible diagrams. In the stationary point this functional gives the free energy of the system. In practice, $\Phi[G]$ is not known for interacting electron systems, which is similar to the problem in the density functional theory. Moreover, this general functional approach reduces to density-functional theory, if one only uses the diagonal part in space-time of the Green function, which corresponds to the one-electron density

$$n_1 = G_{12}\,\delta_{12} = \langle c_1^* c_1 \rangle_S \,, \tag{24}$$

with the Kohn-Sham potential $V_{KS} = V_{ext} + V_H + V_{xc}$ playing the role of the "constrained field" J. Here V_{ext} is external potential and V_H is a Hartree potential. In principle, the exchange-correlation potential V_{xc} is known only for the homogeneous electron gas, therefore in all practical applications one uses a so-called local density approximation to DFT. In this case the DFT functional defined in the following way

$$F_{DFT}[n] = T_0[n] + V_{ext}[n] + V_H[n] + V_{xc}[n] \,, \tag{25}$$

where T_0 is the kinetic energy of the non-interacting system. Finally, if we define the total electron density as

$$n(\mathbf{r}) = \sum_i \phi_i^*(\mathbf{r})\phi_i(\mathbf{r})$$

the local density approximation to DFT reads

$$T_0[n] + V_{ext}[n] = \sum_i \int d\mathbf{r}\, \phi_i^*(\mathbf{r}) \left(-\frac{1}{2}\nabla^2 + V_{ext}(\mathbf{r}) - \mu \right) \phi_i(\mathbf{r}) \tag{26}$$

$$V_H[n] = \frac{1}{2} \int d\mathbf{r}\, n(\mathbf{r})\, U(\mathbf{r} - \mathbf{r}')\, n(\mathbf{r}') \tag{27}$$

$$V_{xc}[n] = \int d\mathbf{r}\, n(\mathbf{r}) \varepsilon[n(\mathbf{r})]\,, \tag{28}$$

where $\varepsilon(n)$ is the exchange-correlation energy density for homogeneous electron gas, which can be calculated with quantum Monte Carlo [16].
In the DFT scheme we lose information about the non-equal-time Green function, which gives the single particle excitation spectrum as well as the k-dependence of the spectral function, and restrict ourself only to the ground state energy of the many-electron system. Moreover, we also lose also information about all collective excitations in solids, such as plasmons or magnons, which can be obtained from the generalized susceptibility or from the second variation of the free-energy.
One can probably find the Baym–Kadanoff interacting potential $\Phi[G]$ for simple lattice models using the quantum Monte Carlo (QMC). Unfortunately, due to the sign problem in lattice simulations this numerically exact solution of electronic correlations is not possible. On the other hand, one can obtain the solution of local interacting quantum problems in a general fermionic bath, using the QMC scheme, which has no sign problem if it is diagonal in spin and orbital space. Therefore, a reasonable approach to strongly correlated systems is to keep only the local part of the many-body fluctuations. In such a Dynamical Mean-Field Theory (DMFT) one can obtain numerically the correlated part of the local functional. In this scheme we only use the local part of the many electron vertex and obtain in a self-consistent way an effective functional of the local Green function. In the following section we discuss the general dual fermion (DF) transformations [2] which help us to separate the local fluctuations in a many-body system and show a perturbative way to go beyond the DMFT approximations.

4 Dual Fermion approach for non-local correlations

We will consider here the simplest local Hubbard-like interaction vertex U. A generalization to the multi-orbital case is straightforward [17]. All equations will be written in matrix form, giving the idea of how to generalize a DF scheme to the multiorbital case. The general strategy for separating the local and non-local correlation effects is associated with the introduction of auxiliary fermionic fields that couple separated local correlated impurities models back to the lattice [2].

We rewrite corresponding original action from Eq. (12) in Matsubara space as a sum of the non-local one-electron contribution with energy spectrum $\varepsilon_{\mathbf{k}}$ and the local interaction part U:

$$S[c^*, c] = -\sum_{\mathbf{k}\nu\sigma} c^*_{\mathbf{k}\nu\sigma}(i\nu + \mu - \varepsilon_{\mathbf{k}})\, c_{\mathbf{k}\nu\sigma} + \sum_i U\, n^*_{i\uparrow} n_{i\downarrow}\,, \tag{29}$$

where $\nu = (2n+1)\pi/\beta$, $(\omega = 2n\pi/\beta), n = 0, \pm 1, ...$ are the fermionic (bosonic) Matsubara frequencies, β is inverse temperature, and μ is a chemical potential. The index i labels the lattice sites, m refers to different orbitals, σ is the spin projection, and the $\mathbf{k}$-vectors are quasi-momenta. In order to keep the notation simple, it is useful to introduce the combined index $\alpha \equiv \{m, \sigma\}$. In the following, translational invariance is assumed for simplicity, although a real space formulation is straightforward. The local part of the action, $S_{\rm U}$, may contain any type of local multi orbital interaction.

In order to formulate an expansion around the best possible auxiliary local action, a quantum impurity problem is introduced

$$S_{\rm loc}[c^*, c] = -\sum_{\nu,\sigma} c^*_{\nu\sigma}\,(i\nu + \mu - \Delta_\nu)\; c_{\nu\sigma} + U\, n^*_{i\uparrow} n_{i\downarrow}\,, \tag{30}$$

where Δ_ν is the effective hybridization matrix describing the coupling of the impurity to an auxiliary fermionic bath. The main motivation for rewriting the lattice action in terms of a quantum impurity model is that such a reference system can be solved numerically exactly for an arbitrary hybridization function with CT-QMC methods [1]. Using the locality of the hybridization function Δ_ν, the lattice action Eq. (29) can be rewritten exactly in terms of individual impurity models and the effective one-electron coupling $(\varepsilon_{\mathbf{k}} - \Delta_\nu)$ between different impurities

$$S[c^*, c] = \sum_i S_{\rm loc}[c^*_i, c_i] + \sum_{\mathbf{k}\nu\sigma} c^*_{\mathbf{k}\nu\sigma}\,(\varepsilon_{\mathbf{k}} - \Delta_\nu)\, c_{\mathbf{k}\nu\sigma}\,. \tag{31}$$

We will find the condition for the optimal choice of the hybridization function later. Although we can solve the individual impurity model exactly, the effect of spatial correlations due to the second term in Eq. (31) is very hard to treat, even perturbatively, since the impurity action is non-Gaussian and on cannot use of Wick's theorem. The main idea of a dual fermion transformation is the change of variables from (c^*, c) to weakly correlated Grassmann fields (f^*, f) in the path integral representation of the partition function, Eq. (12), followed by a simple perturbation treatment. The new variables are introduced through the following Hubbard-Stratonovich transformation

$$\exp\left(c^*_\alpha b_\alpha (M^{-1})_{\alpha\beta} b_\beta c_\beta\right) = \frac{1}{\det M} \int \mathcal{D}[f^*, f]\, \exp\left(-f^*_\alpha M_{\alpha\beta} f_\beta - c^*_\alpha b_\alpha f_\alpha - f^*_\beta b_\beta c\beta\right). \tag{32}$$

In order to transform the exponential of the bilinear term in Eq. (31), we choose the matrix $M_{\alpha\beta}$ and scaling function b_α in the following way [2]

$$M = g_\nu^{-1}\,(\Delta_\nu - \varepsilon_{\mathbf{k}})^{-1}\, g_\nu^{-1}, \qquad \text{and} \qquad b = g_\nu^{-1}\,, \tag{33}$$

where g_ν is the local, interacting Green function of the impurity problem:

$$g_\nu = -\langle c_\nu c^*_\nu \rangle_{\rm loc} = -\frac{1}{\mathcal{Z}_{\rm loc}} \int \mathcal{D}[c^*, c]\, c\, c^* \exp\Big(-S_{\rm loc}[c^*, c]\Big). \tag{34}$$

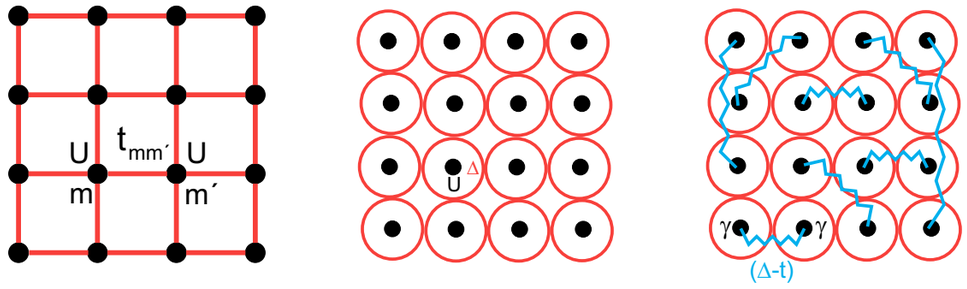


Fig. 3: *From the lattice model (left) to the real-space DMFT (middle) and following up with the non-local DF perturbation (right)*

With this choice, the lattice action transforms to

$$S[c^*, c, f^*, f] = \sum_i S^i_{\rm site} + \sum_{\nu \mathbf{k} \sigma} f^*_{\nu \mathbf{k} \sigma} [g_\nu^{-1} \left(\Delta_\nu - \varepsilon_{\mathbf{k}} \right)^{-1} g_\nu^{-1}] f_{\nu \mathbf{k} \sigma} \, . \tag{35}$$

Hence the coupling between sites is transferred to a local coupling to the auxiliary fermions

$$S^i_{\rm site}[c^*_i, c_i, f^*_i, f_i] = S_{\rm loc}[c^*_i, c_i] + \sum_\sigma f^*_{i\nu\sigma} \, g_\nu^{-1} c_{i\nu\sigma} + c^*_{i\nu\sigma} \, g_\nu^{-1} f_{i\nu\sigma} \, . \tag{36}$$

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 $S_{\rm site}$ decomposes into a sum of local terms. The lattice fermions can therefore be integrated out from $S_{\rm site}$ for each site i separately. This completes the change of variables:

$$\int \mathcal{D}[c^*, c] \, \exp\left(-S_{\rm site}[c^*_i, c_i, f^*_i, f_i]\right) = \mathcal{Z}_{\rm loc} \exp\Big(- \sum_{\nu \, \sigma} f^*_{i\nu\sigma} \, g_\nu^{-1} f_{i\nu\sigma} - V_i[f^*_i, f_i] \Big). \tag{37}$$

The above equation may be viewed as the defining equation for the dual potential $V[f^*, f]$. The choice (33) ensures a particularly simple form of this potential. An explicit expression is found by expanding both sides of Eq. (37) and equating the resulting expressions order 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:

$$V[f^*, f] = \frac{1}{4} \Sigma_{\{\nu\sigma\}} \gamma_{1234} f^*_1 f^*_2 f_4 f_3 \, , \tag{38}$$

where the combined index $1 \equiv \{\nu\sigma\}$ 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

$$\gamma_{1234} = g_1^{-1} g_2^{-1} \left[\chi_{1234} - \chi^0_{1234} \right] g_3^{-1} g_4^{-1}, \tag{39}$$

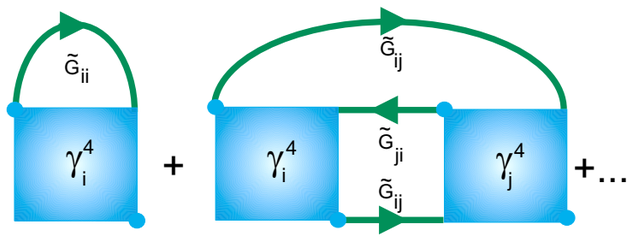


Fig. 4: *Diagrams contributing to the dual self-energy* $\tilde{\Sigma}$

with the two-particle Green function of the impurity being defined as

$$\chi_{1234} = \langle c_1 c_2 c_3^* c_4^* \rangle_{\rm loc} = \frac{1}{\mathcal{Z}_{\rm loc}} \int \mathcal{D}[c^*, c]\, c_1 c_2 c_3^* c_4^* \, \exp\Big(- S_{\rm loc}[c^*, c] \Big). \tag{40}$$

The disconnected part reads

$$\chi^0_{1234} = g_1 g_2 (\delta_{14}\delta_{23} - \delta_{13}\delta_{24}). \tag{41}$$

The single- and two-particle Green functions can be calculated using the CT-QMC algorithms [1]. After integrating out the lattice fermions, the dual action depends on the new variables only

$$\tilde{S}[f^*, f] = -\sum_{\nu \mathbf{k} \sigma} f^*_{\nu \mathbf{k} \sigma} [\tilde{G}^0_\omega(\mathbf{k})]^{-1} f_{\nu \mathbf{k} \sigma} + \sum_i V_i[f_i^*, f_i] \tag{42}$$

and the bare dual Green function is found to be

$$\tilde{G}^0_\nu(\mathbf{k}) = \left(g_\nu^{-1} + \Delta_\nu - \varepsilon_{\mathbf{k}} \right)^{-1} - g_\nu \,, \tag{43}$$

which involves the local Green function g_ν of the impurity model.
Up to now, Eqs. (42) and (43) 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. 4. These are constructed from the impurity vertices and the dual Green functions as lines. The first diagram is purely local, while higher orders contain nonlocal contributions, e.g. the second diagram in Fig. 4. In practice, approximations to the self-energy are constructed in terms of skeleton diagrams. The lines shown in Fig. 4 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 [13], i.e., to fulfill the basic conservation laws for energy, momentum, spin, and particle number. Finally, we can understand the general dual fermion scheme (Fig. 5) as a two-step process for the k-dependent self-energy. First we need to find an optimal hybridization function Δ_ν which defines an effective impurity model. Using the numerically exact Monte-Carlo impurity solver we can obtained the local Green function g_ν which, together with the hopping parameters, defines the non-local dual Green function $\tilde{G}^0_\nu(\mathbf{k})$ and interaction vertex $\gamma^\omega_{\nu,\nu'}$ which can be used in renormalized dual perturbation theory [2]. The hybridization function Δ, which so far has not been specified, allows to optimize the starting point of the perturbation theory and

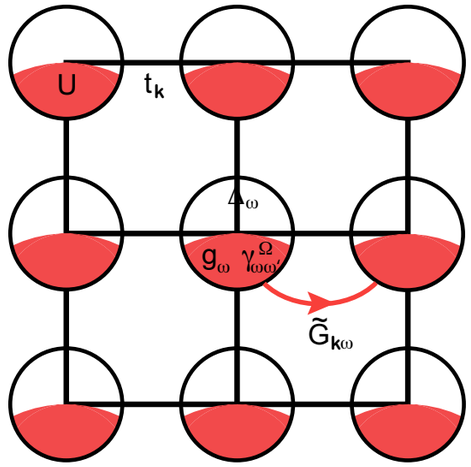


Fig. 5: *General view of the dual-fermion approach: An effective impurity model is defined by the hybridization function Δ_ν. It can be exactly solved within the CT-QMC scheme, resulting in a single particle Green function g_ν and a full connected vertex $\gamma^{\omega}_{\nu,\nu'}$ with two fermionic (ν) and one bosonic (ω) Matsubara frequencies. Based on this local information one can perform an efficient lattice perturbation expansion for the dual Green function $\tilde{G}^0_\nu(\mathbf{k})$.*

should be chosen in an optimal way. The condition of the first diagram (Fig. 4) as well as all local diagrams with higher order correlation functions 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 the γ vertex is local, this condition amounts to demanding that the local part of the dual Green function be zero

$$\sum_{\mathbf{k}} \tilde{G}_\nu(\mathbf{k}) = 0\,. \tag{44}$$

The simplest nontrivial approximation is obtained by taking the leading-order correction, the first diagram in Fig. 4, evaluated with the bare dual propagator (43). Using the expression for the DMFT Green function [5]

$$G^{\mathrm{DMFT}}_\nu(\mathbf{k}) = \left(g_\nu^{-1} + \Delta_\nu - \varepsilon_{\mathbf{k}}\right)^{-1}, \tag{45}$$

it immediately follows that (44) evaluated with the bare dual Green function is exactly equivalent to the DMFT self-consistency condition for Δ_ω

$$\sum_{\mathbf{k}} G^{\mathrm{DMFT}}_\nu(\mathbf{k}) = g_\nu\,. \tag{46}$$

In the limit of infinitely large lattice connectivity the DMFT scheme becomes exact with the local self-energy [5]. The DMFT approximation for real lattice models appears to be one of the most successful many body schemes for realistic multi orbital systems [9]. Since it involves

the exact solution of the many-body multi-orbital impurity model Eq. (34) all local quantum fluctuations of different orbitals, spins, and charges are included in this scheme.

When diagrammatic corrections are taken into account and the first diagram is evaluated with the dressed propagator $\tilde{G}$, the condition (44) will in general be violated. It can be enforced by adjusting the hybridization function iteratively. This corresponds to eliminating an infinite partial series of all local diagrams starting from the first term in Fig. 4. 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. 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 (42) 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 [14]. An 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 in Fig. 4 which contains a single closed loop

$$\tilde{\Sigma}^{(1)}_{12} = -T \sum_{34} \gamma_{1324} \tilde{G}^{\rm loc}_{43} \,, \tag{47}$$

where $\tilde{G}^{\rm loc} = (1/N_k) \sum_{\mathbf{k}} \tilde{G}(\mathbf{k})$ denotes the local part of the dual Green function. The second-order contribution represented in Fig. 4 contains two equivalent lines and one closed loop and hence is k-dependent

$$\tilde{\Sigma}^{(2)}_{12}(\mathbf{k}) = -\frac{1}{2} \left(\frac{T}{N_k} \right)^2 \sum_{\mathbf{k}_1 \mathbf{k}_2} \sum_{345678} \gamma_{1345} \tilde{G}_{57}(\mathbf{k}_1) \tilde{G}_{83}(\mathbf{k}_2) \tilde{G}_{46}(\mathbf{k} + \mathbf{k}_2 - \mathbf{k}_1) \gamma_{6728} \,. \tag{48}$$

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. After calculating the best possible series for the self-energy $\tilde{\Sigma}$ in the dual space one can calculate the renormalized Green function matrix for the original fermions using the following simple transformation [3]

$$G_\nu(\mathbf{k}) = \left[\Big(g_\nu + g_\nu \tilde{\Sigma}_\nu(\mathbf{k}) g_\nu \Big)^{-1} + \Delta_\nu - \varepsilon_{\mathbf{k}} \right]^{-1} , \tag{49}$$

which is a useful generalization of the DMFT Green function (see Eq. (45)) to include the non-local correlation effects. One can see that the dual self-energy plays the role of an effective T-matrix for the exactly solvable local problem.

5 Dual Boson approach for non-local interactions

Many important effects in the physics of correlated systems based on non-local interactions in solids are related with the consistent description of collective excitation (plasmons, magnons, orbitons etc.) which can strongly affect the original electronic degrees of freedom. Using the first-principles constrained-RPA scheme [18] one can obtained non-local interaction parameters for the correlated subspace screened by broad-bands of conducting electrons. The simplest effective Hamiltonian for such an extended Hubbard model reads

$$S = -\sum_{\mathbf{k}\nu\sigma} c^{\dagger}_{\mathbf{k}\nu\sigma}(i\nu + \mu - \varepsilon_{\mathbf{k}})\, c_{\mathbf{k}\nu\sigma} + \frac{1}{2}\sum_{\mathbf{q}\omega} U_{\mathbf{q}} n^{*}_{\mathbf{q}\omega} n_{\mathbf{q}\omega}\,. \tag{50}$$

where the Grassmann variables $c^{\dagger}_{\mathbf{q}\nu}$ ($c_{\mathbf{q}\nu}$) correspond to creation (annihilation) of an electron with momentum $\mathbf{k}$ and fermionic Matsubara frequency ν, and we skip the spin-indices for simplicity. The interaction $U_{\mathbf{q}} = U + V_{\mathbf{q}}$ consists of the on-site (Hubbard term) and non-local long-range Coulomb interactions, respectively. The screened Coulomb interaction can be a frequency dependent $U_{\mathbf{q}\omega}$ as in c-RPA, which does not produce any problems as one can see later. For simplicity we include only charge fluctuations which are given by the complex bosonic variable $n_{\mathbf{q}\omega} = \sum_{\mathbf{k}\nu\sigma}(c^{*}_{\mathbf{k}\nu}c_{\mathbf{k}+\mathbf{q},\nu+\omega} - \langle c^{*}_{\mathbf{k}\nu}c_{\mathbf{k}\nu}\rangle\delta_{\mathbf{q}\omega})$. We do not include exchange interactions in the Hamiltonian as well as local spin degrees of freedom, which can be done with some caution for vector spin boson case [19]. Moreover we will consider only a one-band model but keep the matrix form of all equations for simple generalization to cases of few orbitals (bands). The chemical potential μ defines the average number of electrons per site. Finally, $\varepsilon_{\mathbf{k}}$ is the Fourier transform of the hopping integral t_{ij} between different sites.

The dual boson scheme [3] aims to treat the action (50) in a way, similar to the dual-fermion approach. In addition to the dual fermionic degrees of freedom, the bosonic fields are treated in a similar manner. This allows for the consideration of strongly correlated systems beyond the Hubbard model. Also, it can be employed for an explicit treatment of the collective excitations in the Hubbard model. Here we present the basic idea of this approach (Fig. 6).

First we split the lattice action (50) into a sum of effective single-site local impurity reference actions $S^{(i)}_{\mathrm{ref}}$ defined by hybridization function Δ_{ν} with screened local interaction $\mathcal{U}_{\rightarrow}$ and a non-local remaining part $\tilde{S}$

$$S = \sum_{i} S^{(i)}_{\mathrm{ref}} + \Delta S, \tag{51}$$

which are given by the following explicit relations

$$\begin{aligned} S_{\mathrm{ref}} &= -\sum_{\nu\sigma} c^{\dagger}_{\nu\sigma}(i\nu + \mu - \Delta_{\nu})c_{\nu\sigma} + \frac{1}{2}\sum_{\omega}\mathcal{U}_{\omega}\,\rho^{*}_{\omega}\rho_{\omega}\,, \\ \Delta S &= \sum_{\nu\mathbf{k}\sigma} c^{\dagger}_{\nu\mathbf{k}\sigma}(\varepsilon_{\mathbf{k}} - \Delta_{\nu})c_{\nu\mathbf{k}\sigma} + \frac{1}{2}\sum_{\mathbf{q}\omega}(U_{\mathbf{q}} - \mathcal{U}_{\omega})\,\rho^{*}_{\mathbf{q}\omega}\rho_{\mathbf{q}\omega}\,. \end{aligned} \tag{52}$$

The local bare interaction of the impurity model is then equal to $\mathcal{U}_{\omega} = U_{\omega} + \Lambda_{\omega}$ and it is easy to see that $U_{\mathbf{q}} - \mathcal{U}_{\omega} = V_{\mathbf{q}} - \Lambda_{\omega}$ which makes the method independent of the U–V separation. The

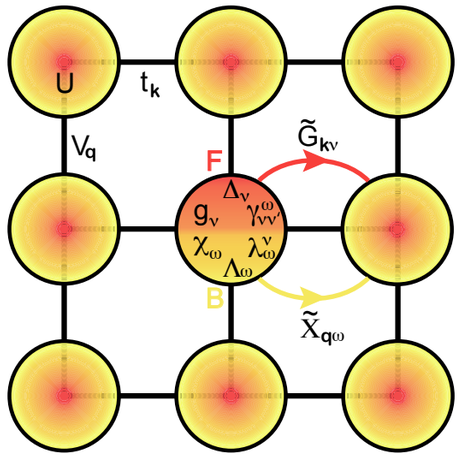


Fig. 6: *General view on dual-fermion approach: effective impurity model defined by fermionic hybridization function Δ_ν and bosonic screened interactions Λ_ω. It can be exactly solved within CT-QMC scheme, resulting in electronic local Green function g_ν and bosonic local susceptibility χ_ω as well as full connected vertex $\gamma^\omega_{\nu,\nu'}$ and electron-boson vertex λ^ν_ω. Based on this local information one can performed an efficient lattice perturbation expansion for the dual Green function $\tilde{G}_{\mathbf{k}\nu}$ and dual boson propagator $\tilde{W}_{\mathbf{q}\omega}$.*

impurity problem with frequency dependent interactions (as well as spin-dependent exchange) can be solved using, e.g., continuous-time quantum Monte Carlo solvers [1], and one can obtain the local impurity Green function g_ν, susceptibility χ_ω, and renormalized interaction $\mathcal{W}_\omega$ as

$$\begin{aligned} g_\nu &= -\langle c_\nu c^*_\nu \rangle_{\text{imp}} \,, \\ \chi_\omega &= -\langle \rho_\omega \rho^*_\omega \rangle_{\text{imp}} \,, \\ \mathcal{W}_\omega &= \mathcal{U}_\omega + \mathcal{U}_\omega \chi_\omega \mathcal{U}_\omega \,, \end{aligned} \tag{53}$$

where the average is taken with respect to the impurity action (52). The strategy here is similar to the dual fermion scheme and consists of an efficient perturbation scheme for ΔS in the action formalism. In addition to the fermionic Hubbard-Stratonovich transformation Eq. (32) on the first term $(\varepsilon_{\mathbf{k}} - \Delta_\nu)\, c^\dagger_{\nu\mathbf{k}\sigma} c_{\nu\mathbf{k}\sigma}$, which give the dual fermion variables $f^\dagger_{\nu\mathbf{k}\sigma}, f_{\nu\mathbf{k}\sigma}$, we perform a bosonic transformation

$$\sqrt{\det[\Lambda_\omega - V_{\mathbf{q}}]}\, e^{\frac{1}{2}\sum\limits_{\mathbf{q}\omega} \rho^*_{\mathbf{q}\omega}[\Lambda_\omega - V_{\mathbf{q}}]\rho_{\mathbf{q}\omega}} = \int D[\phi]\, e^{-\frac{1}{2}\sum\limits_{\mathbf{q}\omega}\left\{\phi^*_{\mathbf{q}\omega}[\Lambda_\omega - V_{\mathbf{q}}]^{-1}\phi_{\mathbf{q}\omega} + \rho^*_\omega \phi_\omega + \phi^*_\omega \rho_\omega\right\}}, \tag{54}$$

and we use that $\mathcal{U}_\omega - U_{\mathbf{q}} = \Lambda_\omega - V_{\mathbf{q}}$. Note that caution should be taken for convergence problem of the integral over the new dual variable $\widetilde{\phi}$ which does not affect the final equations [3]. Rescaling the bosonic fields $\phi_{\mathbf{q}\omega}$ as $\phi_{\mathbf{q}\omega}\alpha^{-1}_\omega$ and integrating out the original degrees of freedom $c^\dagger$ and c we arrive at the dual action

$$\tilde{S} = -\sum_{\mathbf{k}\nu} f^*_{\mathbf{k}\nu} \tilde{G}^{-1}_0 f_{\mathbf{k}\nu} - \frac{1}{2}\sum_{\mathbf{q}\omega} \phi^*_{\mathbf{q}\omega} \tilde{W}^{\nu-1}_0 \phi_{\mathbf{q}\omega} + \tilde{V} \tag{55}$$

with the bare dual fermion-boson propagators

$$\tilde{G}_0 = (G^{-1}_{\mathrm{ref},\nu} + \Delta_\nu - \varepsilon_{\mathbf{k}})^{-1} - g_\nu \qquad = G_{\mathrm{E}} - g_\nu \,, \tag{56}$$

$$\tilde{W}_0 = \alpha^{-1}_\omega \left([U_{\mathbf{q}} - \mathcal{U}_\omega)^{-1} - \chi_\omega \right]^{-1} \alpha^{-1}_\omega = W_{\mathrm{E}} - \mathcal{W}_\omega \,, \tag{57}$$

and the dual interaction term $\tilde{V}$. The explicit form of the dual interaction can be obtained by expanding the $c^\dagger$ and c dependent part of partition function in an infinite series and integrating out these degrees of freedom. The two first terms in $\tilde{V}$ are given by

$$\tilde{V} = \frac{1}{4} \sum_{\nu\nu'\omega} \gamma_{\nu\nu'\omega}\, f^*_\nu f^*_{\nu'} f_{\nu+\omega} f_{\nu'-\omega} + \sum_{\nu\omega} (\lambda_{\nu\omega}\, f^*_\nu f_{\nu+\omega} \phi^*_\omega + h.c.). \tag{58}$$

We define the three-point electron-boson vertex $\lambda_{\nu\omega}$ in the following way

$$\lambda_{\nu\omega} = g^{-1}_\nu g^{-1}_{\nu+\omega} \alpha^{-1}_\omega \langle c_\nu c^*_{\nu+\omega} \rho_\omega \rangle_{\mathrm{loc}} \,, \tag{59}$$

where $\alpha_\omega = \mathcal{W}_\omega / \mathcal{U}_\omega = (1 + \mathcal{U}_\omega \chi_\omega)$ is the local renormalization factor. The four-point vertex function $\gamma_{\nu\nu'\omega}$ can be determined similarly to the dual fermion section. Then, the dual Green function $\tilde{G}_{\mathbf{k}\nu} = -\langle f_{\mathbf{k}\nu} f^*_{\mathbf{k}\nu} \rangle$ and renormalized dual interaction $\tilde{W}_{\mathbf{q}\omega} = -\left\langle \phi_{\mathbf{q}\omega} \phi^*_{\mathbf{q}\omega} \right\rangle$, as well as dual self-energy $\tilde{\Sigma}_{\mathbf{k}\nu}$ and polarization operator $\tilde{\Pi}_{\mathbf{q}\omega}$, can be obtained diagrammatically [20]. These dual Green function have the following relation

$$\tilde{G}^{-1}_{\mathbf{k}\nu} = \tilde{G}^{-1}_0 - \tilde{\Sigma}_{\mathbf{k}\nu} \,, \tag{60}$$

$$\tilde{W}^{-1}_{\mathbf{q}\omega} = \tilde{W}^{-1}_0 - \tilde{\Pi}_{\mathbf{q}\omega} \,. \tag{61}$$

Finally, the Green function $G_{\mathbf{k}\nu}$ and renormalized interaction $W_{\mathbf{q}\omega}$ of the original model can be exactly expressed in terms of dual quantities as

$$G^{-1}_{\mathbf{k}\nu} = G^{-1}_{\mathrm{E}} - \Sigma'_{\mathbf{k}\nu} \,, \tag{62}$$

$$W^{-1}_{\mathbf{q}\omega} = W^{-1}_{\mathrm{E}} - \Pi'_{\mathbf{q}\omega} \,, \tag{63}$$

where the non-local self-energy and polarization operator introduced beyond EDMFT are

$$\Sigma'_{\mathbf{k}\nu} = \frac{\tilde{\Sigma}_{\mathbf{k}\nu}}{1 + g_\nu \tilde{\Sigma}_{\mathbf{k}\nu}} \,, \tag{64}$$

$$\Pi'_{\mathbf{q}\omega} = \frac{\tilde{\Pi}_{\mathbf{q}\omega}}{1 + \mathcal{W}_\omega \tilde{\Pi}_{\mathbf{q}\omega}} \,. \tag{65}$$

The dual-boson self-consistency conditions reads [3]

$$\sum_{\mathbf{k}} G_{\mathbf{k}\nu} = g_\nu \,, \tag{66}$$

$$\sum_{\mathbf{q}} W_{\mathbf{q}\omega} = \mathcal{W}_\omega \,. \tag{67}$$

The DB relations up to this point are exact and derived without any approximations. It is

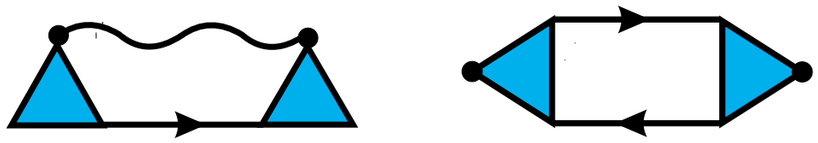


Fig. 7: *Second order approximation for dual-boson scheme with triangular electron-boson vertex ($\lambda_{\nu\omega}$), dual-boson wavy line ($\tilde{W}_{\mathbf{q}\omega}$) and dual-fermion directed line ($\tilde{G}_{\mathbf{k}\nu}$): (Left) Electron self-energy ($\tilde{\Sigma}_{\mathbf{k}\nu}$), (Right) Boson self-energy ($\tilde{\Pi}_{\mathbf{q}\omega}$).*

worth mentioning, that the non-interacting dual theory ($\tilde{V} = 0$) is equivalent to EDMFT. However, even in the weakly-interacting limit of the original model, $U \rightarrow 0$, the fermion-boson vertex $\lambda_{\nu\omega}$ is non-zero. Thus, the Dual Boson formalism explicitly shows, that corrections to EDMFT are not negligible. Therefore, the dynamical mean-field level is insufficient for describing non-local bosonic excitations, because the interactions between the non-local fermionic and bosonic degrees of freedom are always relevant. The simplest approximation for $\tilde{\Sigma}_{\mathbf{k}\nu}$ and $\tilde{\Pi}_{\mathbf{q}\omega}$ related with second oder perturbation theory are presented in (Fig. 7). In principle one can used so-called bold diagrammatic Monte Carlo method to perform the summation of the most important contributions to the dual fermion and boson self-energy, similar to first attempts for dual-fermion action [21].

6 Numerical results

We present now two examples of recent calculations using dual-fermion and dual-boson methods. First, we are able to conduct a realistic dual-fermion study of the interaction of electrons and paramagnons in the strongly correlated sodium cobaltate Na_xCoO_2 close to the in-plane ferromagnetic order at $x = 0.7$ [22]. Sodium cobaltate consists of triangular CoO_2 planes, subject to strong electronic correlations, that are held together by Na ions in between. Using a tailored one-band description of the low-energy cobaltate physics, we derived a realistic DFT-based dispersion, applied a proper Hubbard $U = 5\,\text{eV}$ and solved for the local and nonlocal correlations by the dual-fermion framework (see [22] for details). And indeed, for $x = 0.7$ an additional anti-bound state is detected close to Γ, split off from the renormalized quasiparticle dispersion (Fig. 8a). In terms of physics, it corresponds to the interaction of renormalized electrons with strong ferromagnetic fluctuations. The resulting emerging excitation is a so-called spin polaron [22]. Hence besides refinements of the DMFT-determined electronic spectrum, the inclusion of nonlocal correlations may be important to reveal more complex excitations with possible relevance for future materials science.

The density of states at $x = 0.67$ (Fig. 8b), shows a striking difference between DMFT and DF results. In DMFT, the quasiparticle peak is considerably larger and the QP weight Z is significantly enhanced compared to the case of low doping. This is expected, since $Z \approx 1$ should only hold far away from half-filling towards the opposite (band-insulating) endpoint $x = 1$. We further see that the upper Hubbard band has completely dissolved in the DMFT

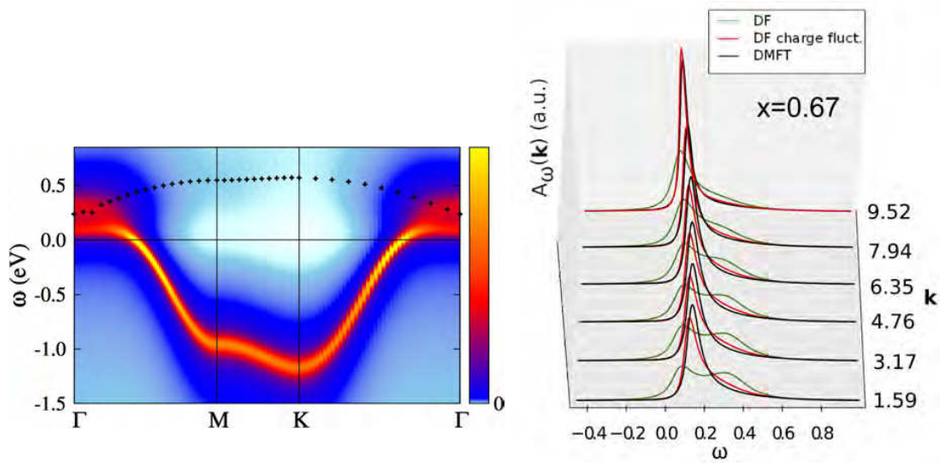


Fig. 8: *The Na_xCoO_2 system close to $x = 0.7$ for electron-paramagnon interactions. (Left) Effective one-band dispersion as obtained by the dual-fermion method. (Right) Spectral function in the DMFT and DB scheme. Crosses mark a model spin-polaron dispersion (from [22]).*

perspective. In DF, on the contrary, the QP peak close to Γ is strongly renormalized. The spectral function additionally exhibits a broad sideband excitation at $\omega \sim 0.3 - 0.4$ eV. By restricting the DF calculation to the charge channel only, this sideband excitation disappears. This is a strong indication that this excitation is of *magnetic* origin.

Despite the successes of GW+EDMFT [23,24], the method does not provide a completely valid description of plasmons. This is due to an inconsistent treatment of the single and two-particle

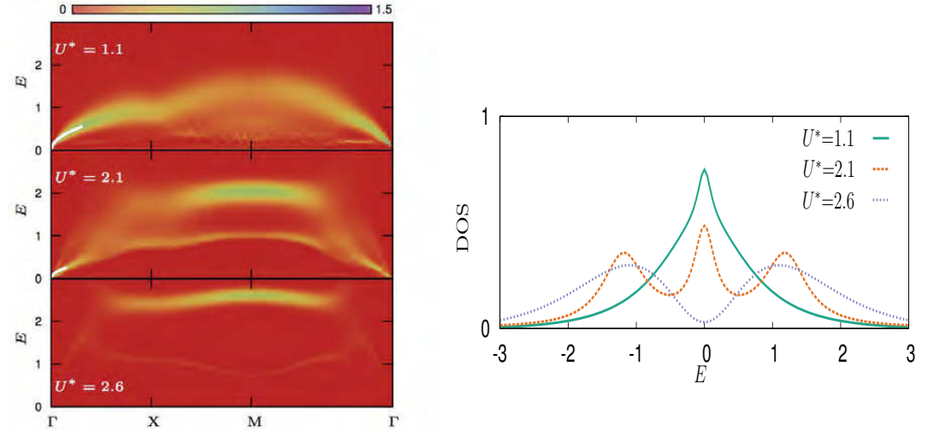


Fig. 9: *Inverse dielectric function (Left) and density of states (Right) of the extended 2d Hubbard model with long-range Coulomb interaction as a function of momentum and energy, for three different values of the effective on-site interaction U^* (from [25]).*

properties, which breaks local charge conservation and gauge invariance. The problem generally occurs when working with renormalized Green functions [13], but does not include the vertex corrections. In the case of the local, frequency-dependent self-energy of EDMFT, vertex corrections from a local but frequency dependent irreducible vertex are necessary to fulfill the Ward identity. In the dual boson approach, they can be included via non-local polarization corrections, which are constructed diagrammatically. The resulting polarization vanishes in the long wavelength limit at finite frequencies, as required by local charge conservation [3, 20]. It therefore becomes possible to study the effect of strong correlations on the plasmon spectra. We showed that the two-particle excitations exhibit both renormalization of the dispersion and spectral weight transfer [25]. Fig. 9 shows the inverse dielectric function $-\mathrm{Im}\,\epsilon^{-1}_{\mathbf{q}}(\omega)$ of 2d surface plasmons in the presence of a long-range interaction $U(q) = U + V_{\mathbf{q}}$. For weak interactions one observes a broad particle-hole continuum and the expected $\sqrt{q}$ dependence of the 2d plasmon dispersion at small q. As the interaction increases, the plasmon dispersion $\omega_p^2(q) \approx \alpha V_0 q$ is renormalized (α is decreased). Two branches are clearly visible in the spectrum. The lower branch can be associated with particle-hole excitations between a Hubbard band and the quasiparticle peak, while the upper branch stems from excitations between the Hubbard bands. Spectral weight is transferred from the lower to the upper branch as the interaction increases. Above a critical effective onsite interaction of $U^* \sim 2.4$ the system is a Mott insulator. In this state a two-particle excitation corresponds to the creation of a doublon and a holon. Such an excitation is highly localized, leading to a weakly dispersing band at energy U^*.

Acknowledgements

We would like to thank A. N. Rubtsov, M. I. Katsnelson, M. Potthoff, F. Lechermann, H. Hafermann and E. G. C. P. van Loon, E. A. Stepanov, D. Hirshmeier, and E. Pavarini for the fruitful collaboration over the years. Financial support of this work by the Deutsche Forschungsgemeinschaft through the Forschergruppe FOR 1346 is gratefully acknowledged.

References

[1] E. Gull, A. Millis, A. Lichtenstein, A. Rubtsov, M. Troyer, and P. Werner, Rev. Mod. Phys. **83**, 349 (2011)

[2] A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **77**, 033101 (2008)

[3] A.N. Rubtsov, M.I. Katsnelson, and A.I. Lichtenstein, Ann. Phys. **327**, 1320 (2012)

[4] F. Lechermann, A.I. Lichtenstein, and M. Potthoff, Eur. Phys. J. Special Topics **226**, 2591 (2017)

[5] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)

[6] N.F. Mott: *Metal-Insulator Transitions*, 2nd ed. (Taylor and Francis, London, 1990)

[7] A. Georges and G. Kotliar, Phys. Rev. B **45**, 6479 (1992)

[8] P. Hohenberg and W. Kohn, Phys. Rev. **136**, 864 (1964)

[9] G. Kotliar, S.Y. Savrasov, K. Haule, V.S. Oudovenko, O. Parcollet, and C.A. Marianetti, Rev. Mod. Phys. **78**, 865 (2006)

[10] J.W. Negele and H. Orland: *Quantum Many-Particle Systems* (Addison-Wesley, New York, 1988)

[11] A. Kamenev: *Field Theory of Non-Equilibrium Systems* (Cambridge University Press, 2011)

[12] F.A. Berezin: *Method of Second Quantization* (Academic Press , New York, 1966)

[13] G. Baym and L.P. Kadanoff, Phys. Rev. **124**, 287 (1961)

[14] A.A. Abrikosow, L.P. Gorkov, and I.E. Dzyaloshinski, *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, New Jersey, 1964)

[15] P. Nozieres: *Theory of Interacting Fermi Systems* (Addison-Wesley, New York, 1964)

[16] D.M. Ceperley and B.J. Alder, Phys. Rev. Lett. 45, 566 (1980)

[17] H. Hafermann, F. Lechermann, A.N. Rubtsov, M.I. Katsnelson, A. Georges, and A.I. Lichtenstein, Lecture Notes in Physics **843**, 145 (2012)

[18] F. Aryasetiawan, M. Imada, A. Georges, G. Kotliar, S. Biermann, and A.I. Lichtenstein, Phys. Rev. B **70**, 195104 (2004)

[19] Th. Ayral, and O. Parcollet, Phys. Rev. B **93**, 235124 (2016)

[20] E.A. Stepanov, E.G.C.P. van Loon, A.A. Katanin, A.I. Lichtenstein, M.I. Katsnelson, and A.N. Rubtsov, Phys. Rev. B **93**, 045107 (2016)

[21] S. Iskakov, A.E. Antipov, and E. Gull, Phys. Rev. B **94**,035102 2016

[22] A. Wilhelm, F. Lechermann, H. Hafermann *et al.*, Phys. Rev. B **91**, 155114 (2015)

[23] S. Biermann, F. Aryasetiawan, and A. Georges, Phys. Rev. Lett. **90**, 086402 (2003)

[24] P. Sun and G. Kotliar, Phys. Rev. B **66**, 085120 (2002)

[25] E.G.C.P. van Loon, H. Hafermann, A.I. Lichtenstein, A.N. Rubtsov, and M.I. Katsnelson, Phys. Rev. Lett. **113**, 246407 (2014)

11 Dynamical Mean-Field Theory of Disordered Electrons: Coherent Potential Approximation and Beyond

Václav Janiš
Institute of Physics, Czech Academy of Sciences
Na Slovance 2, CZ-182 21 Praha, Czech Republic

Contents

1 Introduction **2**

2 Quantum mechanics of a particle in a static random environment **3**
2.1 Lattice model with a random atomic potential 3
2.2 Configurational averaging: coherent potential and T-matrix operator 4

3 Many-body approach to disordered electron systems **6**
3.1 Thermodynamic limit and translational invariance 6
3.2 Green functions and relations between them 7
3.3 Feynman diagrams and multiple-occupancy corrections 8

4 Generating functional for CPA and DMFT **11**
4.1 Functional-integral representation of the thermodynamic potential 11
4.2 The limit of infinite lattice dimensions 13

5 Interacting disordered electrons – Falicov-Kimball model **15**
5.1 Equilibrium thermodynamic properties 15
5.2 Response to external perturbations 16

6 Transport properties within CPA **18**
6.1 Non-local two-particle vertex and electrical conductivity 18
6.2 Gauge invariance and electron-hole symmetry 20

7 Beyond CPA **22**
7.1 Vertex corrections to the electrical conductivity 22
7.2 Making expansions beyond local approximations conserving 23

8 Conclusions **26**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

Impurities are ubiquitous in real materials and are non-negligible in reliable realistic calculations of the low-temperature properties of solids. The impurities are randomly distributed on a macroscopic scale and their impact on thermodynamic, spectral, and transport properties of solids is an important topic of experimental and theoretical research. Advanced experimental techniques allow now for a rather precise determination of the chemical composition of heterogeneous materials, which, on the other hand, increases demands on the precision of the theoretical description of materials with randomness.

Lax [1, 2] was the first to simulate scattering from a random potential by a self-consistently determined homogeneous effective medium. The idea of an effective medium went far beyond the rigid-band or virtual-crystal approximation standardly used at that time. The coherent embedding into a homogeneous environment, when transferred to the context of random alloys, has become the corner-stone of what later has become known as the coherent potential approximation (CPA). The idea of Lax was further extended by Davies and Langer by considering multiple single-site scatterings [3].

The equations for the genuine coherent potential approximation were introduced independently by Soven [4] and Taylor [5]. The multiple-scattering approach was applied in Ref. [4] to electrons on random lattices, while in Ref. [5] it was applied to lattice vibrations of imperfect crystals. The method of coherent potential was extensively studied and applied in various situations during the late sixties and in the seventies of the last century. The progress in the description of random media via multiple single-site scatterings was made possible by two principal theoretical developments. First, sophisticated many-body perturbation techniques using Green functions made it possible to avoid a cumbersome description of random media via an inhomogeneous differential Schrödinger equation. Second, the development of computers capable of determining numerically exactly reasonably large clusters opened the way to the application of the coherent potential methodology to real materials beyond the model level.

The coherent potential approximation has several attractive features. It was shown to be the best single-site approximation [6] and the coherent potential has the proper analytic properties in the complex energy plane, consistent with causality of the averaged Green function [7, 8]. Finally, transport properties of disordered systems can also be determined within this single-site approximation [9]. The early approaches to CPA were reviewed in Ref. [10].

The coherent potential approximation remained for long singled out from other approximations due to its analytic structure and accuracy in the determination of thermodynamic and transport properties. Direct cluster extensions of the single-site multiple-scatterings failed in keeping causality of the Green function [11]. At that time the only causal cluster extension, the so-called traveling cluster approximation [12, 13], was unhandy for applications in realistic settings.

A new impetus in our understanding of CPA in a broader context came in the late eighties and early nineties of the last century with the concept of the Dynamical Mean-Field Theory (DMFT). First, a functional-integral generalization of CPA enabled to understand the concept of coherent potential as a single-site approximation with self-consistently summed single-loop

contributions of the many-body perturbation expansion of the thermodynamic potential [14,15]. Second, these diagrams were then shown to determine the exact solution of models of interacting electrons on a hypercubic lattice in the limit of infinite-dimension [16]. Consequently, CPA then appeared to be an exact solution of models of the disordered Fermi lattice gas in the limit to infinite dimensions within DMFT [15, 17]. Since then, CPA is understood as a special case of DMFT applied to disordered systems. Not only this, DMFT and CPA are interconnected in systematic ways to improve upon these local approximations built on many-body diagrammatic approaches.

2 Quantum mechanics of a particle in a static random environment

2.1 Lattice model with a random atomic potential

Both electron correlations and randomness in configurations of impurities or in the chemical composition are always present to some extent in real materials. It is wise to separate them first to understand their individual impact on the behavior of the electrons. Hence, the easiest model of disorder in metals, crystalline solids with available conduction electrons, is a Fermi gas of moving light particles scattered on heavy, immobile ions, the atomic potential of which fluctuates from site to site. Since there are no electron correlations present, we have a quantum mechanical problem of a test particle scattered on randomly distributed atomic potentials of ions ordered in a regular lattice structure. The generic quantum mechanical Hamiltonian of such an electron can be written as

$$\widehat{H} = \sum_{nm} |m\rangle\, W_{mn}\, \langle n| + \sum_{n} |n\rangle\, V_n\, \langle n| = \widehat{W} + \widehat{V}\,, \tag{1}$$

where $W_{mn} = W(\mathbf{R}_m - \mathbf{R}_n)$ with $W_{nn} = 0$ is the hopping amplitude of the electron between lattice sites $\mathbf{R}_m$ and $\mathbf{R}_n$ and $|n\rangle, |m\rangle$ stand for Wannier states at the respective lattice sites. One usually resorts to hopping only between nearest neighbors. The local potential V_n acquires values due to the atomic occupation of the lattice site $\mathbf{R}_n$. In case of a binary alloy with atoms of type A and B the probability distribution of the atomic potential is

$$g(V) = x_A \delta(V - V_A) + x_B \delta(V - V_B)\,, \tag{2}$$

with $x_A = N_A/N = c$ and $x_B = 1 - c$ are densities of atoms A and B, respectively, and N being the number of the electrons/lattice sites. We assume that the lattice sites are occupied independently according to the distribution $g(V)$ given by eq. (2).

Fluctuations in the values of the atomic potential strongly influence the motion of the electron. Since the operators of the hopping $\widehat{W}$ and the potential $\widehat{V}$ do not commute, the Schrödinger equation for the electron in a randomly distributed scattering potential is not exactly solvable for extended systems. Eigenvalues of the Hamiltonian (1) are random numbers. We hence have

to determine the distribution of the eigenvalues of the random Hamiltonian to draw conclusions about the behavior of the test particle in the random environment.
The fundamental quantity in the description of the quantum particle is the resolvent defined for an arbitrary complex energy z outside the real axis as

$$G_{mn}(z) = \left\langle m \left| \left[z\mathbb{1} - \widehat{W} - \widehat{V} \right]^{-1} \right| n \right\rangle . \tag{3}$$

The distribution of the eigenenergies (density of states) then is

$$\rho(E) = -\frac{1}{\pi V} \sum_n \operatorname{Im} G_{nn}(E + i0^+) , \tag{4}$$

with $V = Nv$, and v is the volume of the elementary cell. This distribution generally depends on the size of the random system as well as on the boundary conditions for solving the Schrödinger equation. This dependence is removed by configurational averaging.

2.2 Configurational averaging: coherent potential and T-matrix operator

Configurational averaging is a tool for restoring translational invariance in random systems. It enables us to develop systematic approximations to the physical quantities of interest. Randomness introduces fluctuations into the physical quantities, since the eigenvalues of the random Hamiltonian are spread over an interval on the real axis. Summing over configurations takes account of the fluctuations only on average and hence not all averaged quantities are relevant. For instance, the averaged Hamiltonian, the energy, is only of a little value. Moreover, products of random variables differ from the product of their averages and vertex corrections for products of random variables must be introduced.
Each averaged quantity is characterized by a translationally invariant function containing, on average, the impact of randomness on this function. The coherent potential for the averaged resolvent is defined from the following equation

$$\langle \mathbb{G} \rangle_{av} = \widehat{G} = \left[z\mathbb{1} - \widehat{\sigma}(z) - \widehat{W} \right]^{-1} . \tag{5}$$

The exact coherent potential $\widehat{\sigma}(z) = \sum_{n,m} |n\rangle \, \sigma_{nm}(z) \, \langle m|$ is, in general, a non-local operator on the lattice, but we resort to single-site approximations with only a diagonal coherent potential $\widehat{\sigma}(z) = \sum_n |n\rangle \, \sigma_n(z) \, \langle n|$.
The coherent potential contains the fluctuations due to the random character of the scattering potential only in an averaged manner. We introduce a configurationally-dependent T-matrix operator $\mathbb{T}(z)$ containing the fluctuations missing in the coherent potential

$$\mathbb{G}(z) = \left\langle \widehat{G}(z) \right\rangle_{av} + \left\langle \widehat{G}(z) \right\rangle_{av} \mathbb{T}(z) \left\langle \widehat{G}(z) \right\rangle_{av} . \tag{6}$$

The T matrix is generally a nonlocal operator and, similarly to the coherent potential, we can introduce local T matrices

$$\mathbb{T}_n(z) = \frac{V_n - \sigma_n(z)}{1 - (V_n - \sigma_n(z)) \, G_{nn}(z)} \tag{7}$$

that depend on the lattice coordinate $\mathbf{R}_n$ only and contain all multiple single-site scatterings on the fluctuations of the atomic potential V_n relatively with respect to the coherent potential of the effective medium $\sigma_n(z)$.
The full T matrix can be represented via the local ones and wave operators $\mathbb{Q}_n(z)$ as

$$\mathbb{T}(z) = \sum_n \mathbb{T}_n(z) \left[\mathbb{1} + \langle \mathbb{G}(z) \rangle_{av} \sum_{m \neq n} \mathbb{Q}_m(z) \right] , \tag{8}$$

$$\mathbb{Q}_n(z) = \mathbb{T}_n(z) + \left[\mathbb{1} + \langle \mathbb{G}(z) \rangle_{av} \sum_{m \neq n} \mathbb{Q}_m(z) \right] . \tag{9}$$

Successive substitution of the wave operators leads to a representation of the T matrix via a multiple-scattering series with the local T matrices connected by the averaged resolvent

$$\begin{aligned} \mathbb{T}(z) = \sum_n \mathbb{T}_n(z) + \sum_{n \neq m} \mathbb{T}_n(z) \langle \mathbb{G}_{nm}(z) \rangle_{av} \mathbb{T}_m(z) \\ + \sum_{n \neq m \neq l} \mathbb{T}_n(z) \langle \mathbb{G}_{nm}(z) \rangle_{av} \mathbb{T}_m(z) \langle \mathbb{G}_{ml}(z) \rangle_{av} \mathbb{T}_l(z) + \dots \quad . \end{aligned} \tag{10}$$

This is a typical excluded-volume problem that is difficult to solve beyond the first few terms of the series.
The averaged T matrix vanishes in the exact solution. This means, that the coherent potential captures all the fluctuations of the random atomic potential. Since we resorted to a single-site coherent potential, we cannot guarantee vanishing of the full T matrix but only of its local part. The vanishing of the local T matrix

$$\langle \mathbb{T}_n(z) \rangle_{av} = \left\langle \frac{V_n - \sigma(z)}{1 - (V_n - \sigma(z)) \langle G_{nn}(z) \rangle_{av}} \right\rangle_{av} = 0 \tag{11}$$

is the Soven equation for the coherent potential $\sigma_n(z) = \sigma(z)$. Averaging restores translational invariance, hence the coherent potential is independent of the lattice coordinate.
The Soven equation for the coherent potential can be solved only iteratively. Due to the correct analytic properties of the coherent potential we can guarantee the convergence of the following iteration procedure for real energies $\lim_{l \to \infty} \sigma^{(l)}(E_+) = \sigma(E_+)$

$$\operatorname{Im} \sigma^{(l+1)}(E_+) = \left[1 - \frac{|G^{(l)}(E_+)|^2}{\langle |G^{(l)}(E_+)|^2 \rangle} \right] \left\langle \frac{1}{|1 + G^{(l)}(E_+) (\sigma^{(l)}(E_+) - V_n)|^2} \right\rangle_{av} \operatorname{Im} \sigma^{(l)}(E_+) , \tag{12}$$

where we denoted $E_+ = E + i0^+$, $\langle |G(z)|^2 \rangle = \operatorname{Im} G(z) / \operatorname{Im} \sigma(z) = \int d\varepsilon \, \rho(\varepsilon) |z - \varepsilon - \sigma(z)|^{-2}$ and $\rho(\varepsilon)$ is the density of states of the electrons on the homogeneous lattice. A negative sign of the imaginary part of the coherent potential is guaranteed during the iterations. The analytic properties of CPA must not be broken during the iterative process of the numerical solution in order to stay within the physical phase space.

3 Many-body approach to disordered electron systems

3.1 Thermodynamic limit and translational invariance

The concept of an effective medium and a coherent potential was derived with quantum mechanics of particles, that is, for Fermi or Bose gases without inter-particle interactions. The construction of the best local approximation is, however, an appealing approach that can be used in a broader context, namely in the statistical mechanics of many-body systems.

The equilibrium properties of macroscopic many-body systems are extracted from the thermodynamic limit. It means that the volume V of the system is sent to infinity. The differences between positions of the individual sites vanish and translational invariance is restored in the thermodynamic limit. We can use the Fourier transform from the direct lattice to momentum space and use the Bloch waves as the elementary quantum states that form a complete orthonormal basis of the states with which we can describe the random system in the thermodynamic limit.

We use second quantization and extend the quantum-mechanical Hamiltonian, eq. (1), to the Anderson disordered model in Fock space by means of fermionic creation and annihilation operators $c_i^\dagger$ and c_i respectively

$$\widehat{H} = -t \sum_{\langle ij \rangle} c_i^\dagger c_j + \sum_i V_i \, c^\dagger c_i = \sum_{\mathbf{k}} \varepsilon(\mathbf{k}) \, c^\dagger(\mathbf{k}) c(\mathbf{k}) + \sum_i V_i c_i^\dagger c_i \,, \tag{13}$$

where $\varepsilon(\mathbf{k}) = \sum_i W_{i0} \, \exp(i\mathbf{R}_i \cdot \mathbf{k})$ is the dispersion relation of the Fermi gas on the lattice, $c^\dagger(\mathbf{k}) = V^{-1} \sum_i c_i^\dagger \, \exp(i\mathbf{R}_i \cdot \mathbf{k})$, and $\langle ij \rangle$ denotes nearest-neighbor lattice sites with coordinates $\mathbf{R}_i$ and $\mathbf{R}_j$.

The existence of the equilibrium state in the thermodynamic limit depends on the validity of the ergodic hypothesis, which means that the particle passes almost everywhere in the phase space after sufficiently long time. Then spatial averaging equals configurational averaging, at least for local quantities that can be proved to possess the so-called self-averaging property. For example,

$$\rho(E) = -\frac{1}{\pi V} \sum_i \operatorname{Im} G_{ii}(E_+) = -\frac{1}{\pi} \left\langle \operatorname{Im} G_{ii}(E_+) \right\rangle_{av} = -\frac{1}{\pi V} \sum_{\mathbf{k}} \operatorname{Im} G(\mathbf{k}, E_+) \tag{14}$$

holds in the thermodynamic limit. The averaged Green function in the thermodynamic limit can then be represented as

$$\left\langle \left\langle \mathbf{k} \left| \frac{1}{z\mathbb{1} - \widehat{H}} \right| \mathbf{k}' \right\rangle \right\rangle_{av} = G(\mathbf{k}, z) \delta(\mathbf{k} - \mathbf{k}') = \frac{\delta(\mathbf{k} - \mathbf{k}')}{z - \varepsilon(\mathbf{k}) - \Sigma(\mathbf{k}, z)} \,, \tag{15}$$

with $|\mathbf{k}\rangle = c^\dagger(\mathbf{k})|\Omega\rangle$ and $|\Omega\rangle$ the vacuum (cyclic) vector in the Fock space. The delta function in the numerator stands for momentum conservation in translationally invariant systems. The self-energy $\Sigma(\mathbf{k}, z)$ contains the entire contribution from the random potential to the one-particle propagator $G(\mathbf{k}, z)$. It is a many-body generalization of the coherent potential. The thermodynamic limit and the ergodic hypothesis not only restore translational invariance in the random

system but they also allow us to use the perturbation expansion in the inhomogeneous/random potential so that configurational averaging can be performed term by term in the perturbation expansion. We can then work only with the averaged Green functions and use the many-body diagrammatic and renormalization techniques of homogeneous systems.

3.2 Green functions and relations between them

The fundamental tool for obtaining quantitative results in disordered systems is a renormalized perturbation theory in the random potential. The perturbation theory works only with translationally invariant averaged Green functions and its basic object is the one-particle Green function of eq. (15). It contains the necessary information about the equilibrium thermodynamic and spectral properties. If we are interested in the response to weak perturbations we need to take into account also the averaged two-particle Green function. If we remove the delta function due to conservation of the total momentum, we can define the two-particle Green function in the basis of Bloch waves as

$$G^{(2)}_{\mathbf{k}\mathbf{k}'}(z_1,z_2;\mathbf{q}) = \left\langle\left\langle \mathbf{q}+\mathbf{k},\mathbf{k}\left|\frac{1}{z_1-\widehat{H}}\otimes\frac{1}{z_2-\widehat{H}}\right|\mathbf{k}',\mathbf{q}+\mathbf{k}'\right\rangle\right\rangle_{av}$$
$$\equiv \left\langle\left\langle \mathbf{k}+\mathbf{q}\left|\frac{1}{z_1-\widehat{H}}\right|\mathbf{k}'+\mathbf{q}\right\rangle\left\langle\mathbf{k}'\left|\frac{1}{z_2-\widehat{H}}\right|\mathbf{k}\right\rangle\right\rangle_{av}, \quad (16)$$

where $\otimes$ denotes the direct product of operators.

The full two-particle Green function can further be represented via a vertex function $\mathit{\Gamma}$

$$G^{(2)}_{\mathbf{k}\mathbf{k}'}(z_1,z_2;\mathbf{q}) = G_{\mathbf{k}+\mathbf{q}}(z_1)G_{\mathbf{k}}(z_2)\times[\delta(\mathbf{k}-\mathbf{k}') + \mathit{\Gamma}_{\mathbf{k}\mathbf{k}'}(z_1,z_2;\mathbf{q})G_{\mathbf{k}'+\mathbf{q}}(z_1)G_{\mathbf{k}'}(z_2)]\ . \quad (17)$$

The two-particle vertex $\mathit{\Gamma}$ represents a disorder-induced correlation between simultaneously propagated pairs of particles. It measures the net impact of the pair scatterings on the random potential.

The vertex $\mathit{\Gamma}$ can further be simplified by introducing an irreducible vertex $\mathit{\Lambda}$ playing the role of a two-particle self-energy. The irreducible and the full vertex are connected by a Bethe-Salpeter equation. Unlike the one-particle irreducibility, the two-particle irreducibility is ambiguous when we go beyond single-site scatterings [18]. Two-particle irreducibilities are characterized by different Bethe-Salpeter equations. Here we introduce only the Bethe-Salpeter equation in the electron-hole scattering channel

$$\mathit{\Gamma}_{\mathbf{k}\mathbf{k}'}(z_1,z_2;\mathbf{q}) = \mathit{\Lambda}_{\mathbf{k}\mathbf{k}'}(z_1,z_2;\mathbf{q}) + \frac{1}{N}\sum_{\mathbf{k}''}\mathit{\Lambda}_{\mathbf{k}\mathbf{k}''}(z_1,z_2;\mathbf{q})G_{\mathbf{k}''+\mathbf{q}}(z_1)G_{\mathbf{k}''}(z_2)\mathit{\Gamma}_{\mathbf{k}''\mathbf{k}'}(z_1,z_2;\mathbf{q}). \quad (18)$$

We use this Bethe-Salpeter equation to introduce the irreducible vertex $\mathit{\Lambda}$ that is important for controlling the consistency of approximations, more precisely, whether they comply with the exact relations between one- and two-particle Green functions.

There are no direct connections between the configurationally dependent one- and two-particle Green functions in disordered systems without inter-particle interactions. This is no longer true

for the averaged Green functions. When constructing approximations one has to comply with the exact relations expressed as Ward identities that are microscopic conditions for macroscopic conservation laws to hold.

The basic conservation law in quantum systems is conservation of probability, that is completeness of the basis formed by the Bloch waves. A first relation between the averaged one- and two-particle Green functions follows from a simple identity for operator (matrix) multiplication

$$\frac{1}{z_1-\widehat{H}}\,\frac{1}{z_2-\widehat{H}}=\frac{1}{z_2-z_1}\left[\frac{1}{z_1-\widehat{H}}-\frac{1}{z_2-\widehat{H}}\right]. \tag{19}$$

If we average both sides of this identity we obtain the Velický-Ward identity [9]

$$\frac{1}{N}\sum_{\mathbf{k}'}G^{(2)}_{\mathbf{k}\mathbf{k}'}(z_1,z_2;\mathbf{0})=\frac{1}{z_2-z_1}\left[G(\mathbf{k},z_1)-\;G(\mathbf{k},z_2)\right]. \tag{20}$$

It holds, provided the Bloch waves form a complete basis in the one-particle representation space. It means that the effect of the random potential is only a rotation in the Hilbert space of states of the homogeneous system.

There is another identity connecting the irreducible one and two-particle functions, Σ and Λ. It is a microscopic condition that guarantees the validity of the macroscopic continuity equation. By analyzing the perturbation contributions to one- and two-particle functions Vollhardt and Wölfle proved the following Vollhardt-Wölfle-Ward identity [19]

$$\Sigma(\mathbf{k}_+,z_+)-\Sigma(\mathbf{k}_-,z_-)=\frac{1}{N}\sum_{\mathbf{k}'}\Lambda_{\mathbf{k}\mathbf{k}'}(z_+,z_-;\mathbf{q})\left[G(\mathbf{k}'_+,z_+)-G(\mathbf{k}'_-,z_-)\right], \tag{21}$$

where we denoted $\mathbf{k}_\pm=\mathbf{k}\pm\mathbf{q}/2$. Identity (21) is related to eq. (20), however, the two Ward identities are identical neither in the derivation nor in the applicability and validity domains. The latter holds for nonzero transfer momentum $\mathbf{q}$, i. e., for an inhomogeneous perturbation, while the former only for $q=0$. On the other hand, the former is nonperturbative while the latter is proved only perturbatively.

3.3 Feynman diagrams and multiple-occupancy corrections

The contribution from the scatterings of the electron on the random potential can be represented diagrammatically in analogy with many-body perturbation theory. There is, however, an important difference between the two perturbation expansions. The former is static, contains only elastic scatterings where energy is conserved. There are no closed loops in its diagrammatic representation. The individual particles are characterized by a fixed energy or Matsubara frequency in the many-body formalism. Since the perturbation theory of random systems is static we have to introduce the so-called multiple-occupancy corrections if we want to keep unrestricted summations over the lattice sites in the representations of physical quantities [20]. We demonstrate this on Green functions.

The standard diagrammatic representation of scatterings of particles on the random potential is an oriented solid line for the particle, a cross (vertex) for the lattice coordinate of the random

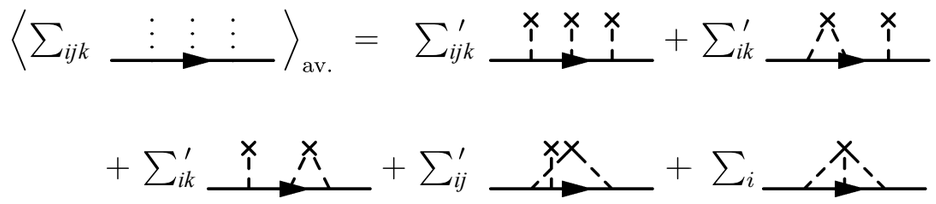


Fig. 1: *Averaging of contributions up to third order in the random potential of the one-particle Green function. A prime means that the multiple sums do not contain any repeating indices, i.e, each vertex is on a different lattice site ($i \neq j \neq k$). Adapted from Ref. [21].*

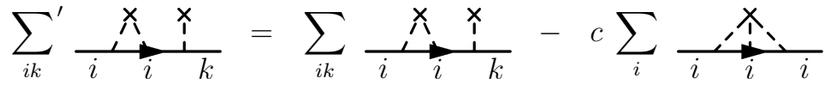
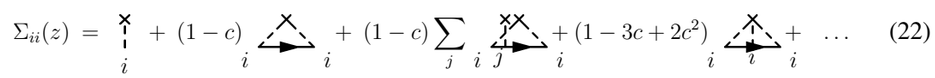


Fig. 2: *Transformation of restricted multiple summations to unrestricted ones with multiple-occupation corrections exemplified in third order of the perturbation expansion. Here c is the concentration of the sites with the random potential. The first diagram on the right-hand side is proportional to c^2 while the second diagram only to c.*

potential and dashed lines connecting the vertex with the solid line. The number of lines starting at the vertex stands for the power of the random potential. Since the random values of the potential are independently distributed at each lattice site, we average separately each vertex of the diagrammatic expansion. When summing over the lattice sites we have to avoid any repetition of lattice indices in the multiple sums as shown in Fig. 1. This restriction in multiple summations makes the perturbation expansion difficult to sum and is equivalent to the excluded volume problem of the T-matrix expansion in (10).

One has to transform the restricted multiple summations to unrestricted ones to be able to reach any nonperturbative results containing multiple scatterings. The transformation from restricted to unrestricted sums is performed by means of multiple-occupancy corrections that subtract events when any two or more lattice sites in the multiple sum are equal, as exemplified in Fig. 2. Counting of the multiple-occupancy corrections becomes more and more cumbersome with increasing order of the perturbation expansion.

Only after we have transformed the restricted sums to unrestricted ones, we can introduce renormalizations of the particle lines in the diagrammatic representation of the perturbation expansion. The renormalization of the one-particle propagator is expressed via the Dyson equation and the self-energy Σ, see Fig. 3. All the multiple-occupancy corrections are contained in the self-energy. Its third order is diagrammatically represented as



(22)

It is, however, impossible to sum up the multiple-occupancy corrections to infinite order.

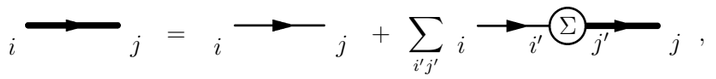
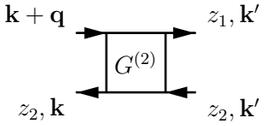


Fig. 3: *Diagrammatic representation of the renormalization of the one-electron propagator via the Dyson equation. Note that the sum over the primed indices is unrestricted.*

The diagrams with multiple-occupancy corrections offer a possibility to directly renormalize all particle lines in the diagram, that is, to replace the bare propagators with the full averaged ones. Inability of finding an analytic expression for the full sum for the self-energy to infinite order reflects the fact that we cannot construct a generating functional consisting of only the renormalized propagators, even in the local mean-field approximation, CPA. It was a breakthrough to find an analytic expression for the Soven equation in terms of the full local averaged propagator G and the self-energy Σ.

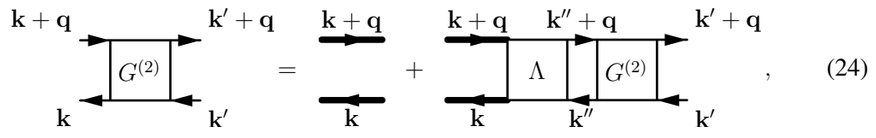
Once we got rid of the restricted summations over lattice sites we can use the Fourier transform to momenta or wave vectors in which the Dyson equation becomes algebraic and easy to solve. The renormalization of the perturbation expansion is not as easy for the two-particle Green function. The averaged two-particle Green function with three independent momenta is diagrammatically represented as

$$G^{(2)}_{\mathbf{k}\mathbf{k}'}(z_1, z_2; \mathbf{q}) =$$



. (23)

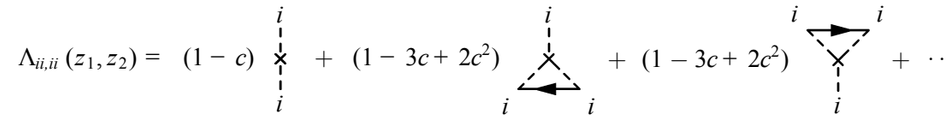
The two-particle renormalization is contained in the two-particle irreducible vertex Λ and a Bethe-Salpeter equation, a two-particle analogy of the Dyson equation. Its diagrammatic representation reads



, (24)

where we sum over the double-primed momentum. We introduced in Sec. 3.2 the full two-particle vertex which obeys an analogous Bethe-Salpeter equation where the absolute term is the irreducible vertex Λ, eq. (18). Its expansion to third order of the perturbation expansion with the multiple-occupancy corrections is

$$\Lambda_{ii,ii}(z_1, z_2) =$$



(25)

The particle lines can again be directly renormalized and replaced by the full averaged one-particle Green function.

4 Generating functional for CPA and DMFT

4.1 Functional-integral representation of the thermodynamic potential

The concept of the best single-site approximation can be generalized beyond the quantum mechanics of disordered systems. The best way to do so is to use a functional integral with which we can describe classical, quantum, disordered, and interacting systems in a unified way. We start with the functional-integral representation of the general partition sum of an interacting and/or disordered system [22]

$$\mathsf{Z}\left[G^{(0)^{-1}}\right] = \int \mathcal{D}\varphi\mathcal{D}\varphi^* \exp\left(-\varphi^*\eta G^{(0)^{-1}}\varphi + h^*\varphi + \varphi^* h + U\left[\varphi,\varphi^*\right]\right) , \tag{26}$$

where φ and φ^* are fluctuating commuting or anticommuting Gaussian fields, $G^{(0)^{-1}}$ is essentially the dispersion relation of the model represented as an inverse of the free propagator of one-body excitations. The sign $\eta = \pm 1$ depends on whether we deal with bosonic (commuting) or fermionic (anticommuting) fluctuating fields, respectively. Further, h is an external source and U is an interaction or a random potential, i.e., a non-quadratic or inhomogeneous function of the fluctuating fields. In eq. (26) we suppressed all internal degrees of freedom of the local fluctuating fields that depend upon particular models under consideration. The thermodynamic potential as a functional of $G^{(0)^{-1}}$ then is

$$\Omega\left[G^{(0)^{-1}}\right] = -\frac{1}{\beta}\ln \mathsf{Z}\left[G^{(0)^{-1}}\right] , \tag{27}$$

where $\beta = 1/k_B T$.

It was the idea of Baym [23] to replace the functional dependence of the thermodynamic potential on the bare propagator by a new representation with a renormalized propagator G. The full propagator G may be defined from the thermodynamic potential itself:

$$G = -\frac{\delta^2\beta\Omega}{\delta h^*\delta h} = \eta\Big(\langle\varphi\varphi^*\rangle - \langle\varphi\rangle\langle\varphi^*\rangle\Big) = \eta\left(\frac{\delta\beta\Omega}{\delta G^{(0)^{-1}}} - \frac{\delta\beta\Omega}{\delta h^*}\frac{\delta\beta\Omega}{\delta h}\right) . \tag{28}$$

We now introduce the full propagator G as a new variable into the thermodynamic potential by a substitution

$$G^{(0)^{-1}} = G^{-1} + \Sigma\,, \tag{29}$$

where Σ is the self-energy. We used here the Dyson equation to relate the bare and renormalized propagators. The self-energy Σ is an accompanying variable that also enters the functional representation of the thermodynamic potential. We can treat the renormalized quantities G and Σ as independent variables in the thermodynamic potential. The new functional must, however, not depend on variations of the new variables G and Σ in order to keep the thermodynamic relations fulfilled. To secure vanishing of variations of the thermodynamic potential with respect to G and Σ we have to modify the functional-integral representation, since the variation with respect to $G^{(0)^{-1}}$ does not vanish. We must add a contribution being a function of only Σ and a contribution being a function of only G. If we denote them Ω_Σ and Ω_G we must fulfill the

following equations to keep variations of the total free energy $\Omega = \Omega_\Sigma + \Omega_G + \Omega\left(G^{(0)^{-1}}\right)$ independent of Σ and G

$$\frac{\delta\beta\Omega_\Sigma}{\delta\Sigma} = \frac{\delta\beta\Omega_G}{\delta G^{-1}} = -\frac{\delta\beta\Omega}{\delta G^{(0)^{-1}}}\,.$$

Using equations (28) and (29) we easily obtain

$$\beta\Omega_\Sigma = \quad \eta\left(\mathrm{tr}\ln\left[G^{(0)^{-1}} - \Sigma\right] + m^*\left[G^{(0)^{-1}} - \Sigma\right]m\right), \tag{30}$$

$$\beta\Omega_G = -\eta\Big(\mathrm{tr}\ln G^{-1} + m^* G^{-1} m\Big)\,, \tag{31}$$

where we have introduced new renormalized variables

$$m = -\frac{\delta\beta\Omega}{\delta h^*} \quad , \qquad m^* = -\frac{\delta\beta\Omega}{\delta h}\,. \tag{32}$$

We use the above definitions and obtain a new representation of the thermodynamic potential

$$\begin{aligned} -\beta\Omega[m; G^{-1}, \Sigma] = -\eta\,\mathrm{tr}\ln\left[G^{(0)^{-1}} - \Sigma\right] + \eta\,\mathrm{tr}\ln G^{-1} - \beta\Omega[h; G^{-1} + \Sigma] \\ - m^*\eta\left[G^{(0)^{-1}} - \Sigma\right]m + m^*\eta\, G^{-1} m\,, \end{aligned} \tag{33}$$

The thermodynamic potential $\Omega[m; G^{-1}, \Sigma]$ in eq. (33) is stationary (extremal) with respect to all its renormalized variables m, Σ, and G. The stationarity with respect to the variables m, m^* leads to trivial equations. We can, however, turn these variables dynamic if we use a substitution in the functional integral (26) $\varphi = \phi + m$, where the new fluctuating field ϕ has vanishing first moment $\langle\phi\rangle = 0$. We then obtain from eqs. (26) and (29)

$$\begin{aligned} -\beta\Omega\left[m, H; G, \Sigma\right] = -\eta\,\mathrm{tr}\ln\left[G^{(0)^{-1}} - \Sigma\right] + \eta\,\mathrm{tr}\ln G^{-1} - \beta F\left[m, H; G^{-1} + \Sigma\right] \\ - m^*\eta\, G^{(0)^{-1}} m + H^* m + m^* H\,. \end{aligned} \tag{34}$$

Now, the new free energy as a functional of m, H and $[G^{-1} + \Sigma]$ reads

$$\begin{aligned} -\beta F\left[m, H; G^{-1} + \Sigma\right] = \\ \ln\int \mathcal{D}\phi\mathcal{D}\phi^* \exp\left(-\phi^*\eta\left[G^{-1} + \Sigma\right]\phi + H^*\phi + \phi^* H + U\left[\phi + m, \phi^* + m^*\right]\right), \end{aligned} \tag{35}$$

where the new external sources H and H^* are new variational parameters, the Legendre conjugates to m^* and m, respectively. The variational parameters and functions m, H, Σ, and G are determined from the saddle-point equations for stationarity of $\beta\Omega$

$$\frac{\delta\beta\Omega}{\delta H} = \frac{\delta\beta\Omega}{\delta m} = \frac{\delta\beta\Omega}{\delta G} = \frac{\delta\beta\Omega}{\delta\Sigma} = 0\,. \tag{36}$$

Expressions (34)-(36) are exact in any spatial dimension for any model, classical or quantum. The thermodynamic potential from eq. (34) is not yet in the Baym form with the Luttinger-Ward functional. Although it is a functional of only renormalized quantities, the diagrammatic representation of $\beta\Omega$ contains the sum of all connected non-renormalized diagrams with the

bare propagator $G^{(0)} = (G^{-1} + \Sigma)^{-1}$. The thermodynamic potential (34) is suitable for an exact solution and for the cases where the result cannot be generated by a sum of simple skeleton diagrams. If we have to rely on sums of classes of particular diagrams it is more practical to define a new functional

$$\Psi\left[m, H; \Sigma\right] = \eta\,\mathrm{tr}\ln G^{-1} - \beta\Omega\left[m, H; G^{-1} + \Sigma\right], \tag{37}$$

that is, due to the stationarity equations (36), independent of the one-body propagator G. We have no diagrammatic representation for the functional Ψ. But if we perform a Legendre transform from Ψ to a functional of the propagator G

$$\Phi\left[m, H; G\right] = \Psi\left[m, H; \Sigma\right] + \eta\,\mathrm{tr}\Sigma G\,, \tag{38}$$

it will be a sum of all connected diagrams free of self-insertions, i.e., skeleton diagrams only. Inserting eqs. (37) and (38) into (34) we obtain the Baym free-energy functional

$$-\beta\widetilde{\Omega}\left[m, H; G\right] = -\eta\,\mathrm{tr}\ln\left[{G^{(0)}}^{-1} - \Sigma\right] - \eta\,\mathrm{tr}\Sigma G + \Phi\left[m, H; G\right] - m^*\eta\,{G^{(0)}}^{-1}m + H^*m + m^*H. \tag{39}$$

Both free-energy functionals (34) and (39) are exact. They are connected by a double Legendre transform (37), (38). While representation (34) is applicable without restrictions, the direct application of the Baym functional is restricted to cases where we are able to find a diagrammatic representation of the functional Φ. We then speak of Φ-derivable approximations. Not all approximations are Φ-derivable. The simplest example for a non-Φ-derivable theory is a 0-dimensional lattice (single site or atomic solution), including CPA and DMFT.

We considered a homogeneous system of interacting particles. It is, however, easy to extend this description to systems with randomness. If the non-quadratic term, becomes random, we simply perform configurational averaging of the free energy $F\left[m, H; G^{-1} + \Sigma\right]$ of eq. (35). The thermodynamic functional is self-averaging and hence it equals its averaged value in the thermodynamic limit.

4.2 The limit of infinite lattice dimensions

We have not yet made any assumption on the form of the interacting term in the functional representation of the thermodynamic potential. A special class of problems are those with a local interaction in the tight-binding representation of statistical systems in crystalline solids for which the limit to high spatial dimensions reduces the lattice to an impurity model.

The fundamental condition in limiting the lattice models to infinite spatial dimensions is the necessity to keep the total energy of the system proportional to volume. This means that we must rescale appropriately the non-local terms in the Hamiltonian. In the case of fermions the fluctuating fields are Grassmannian variables in representation (26) and $\langle\varphi\rangle \equiv 0$. The leading-order contribution of the non-local part of the generic Hamiltonian from eq. (13), the kinetic energy, is

$$E_{kin} = -t\sum_{\langle ij\rangle\sigma}\left\langle c^\dagger_{i\sigma}c_{j\sigma}\right\rangle_{av} = -it\sum_{\langle ij\rangle\sigma}G_{ij,\sigma}(0^+) \propto 2d\mathcal{N}\,t^2\,, \tag{40}$$

where $G_{ij,\sigma}(t) = -i\langle \mathcal{T}[c_{i\sigma}(t)c^\dagger_{j\sigma}(0)]\rangle$ is the time-dependent Green's function, eventually the averaged Green's function [24]. The scaling of the hopping amplitude between the nearest neighbors on a hypercubic lattice with $2d$ nearest neighbors follows from eq. (40)

$$t = t^*/\sqrt{2d}\,. \tag{41}$$

It was derived for the first time by Metzner and Vollhardt in Ref. [16] in the context of the Hubbard model.

The general functional-integral representation of the thermodynamic potential with the renormalized one-electron Green function and the self-energy, eq. (35), offers a direct way to find the generating functional of the solution of a model with local interaction/disorder in infinite dimensions, DMFT. The scaling of the hopping term, eq. (41), leads to

$$G = G^{\rm diag}\left[d^0\right] + G^{\rm off}\left[d^{-1/2}\right], \tag{42}$$

$$\Sigma = \Sigma^{\rm diag}\left[d^0\right] + \Sigma^{\rm off}\left[d^{-3/2}\right]. \tag{43}$$

The functional integral (35) turns local in the limit $d = \infty$. It still may be a functional (infinite-dimensional) integral if the number of the local degrees of freedom is infinite. In the case of quantum itinerant models the local degrees of freedom are spin and Matsubara frequencies. The Matsubara frequencies are coupled in the Hubbard model [24]. They are decoupled in the case of the Fermi gas in a random potential with the Hamiltonian (13). The functional integral (35) then reduces to a product of simple integrals for individual Matsubara frequencies. These integrals can be explicitly performed and inserting the result in eq. (39) we obtain a generating functional for the Coherent Potential Approximation of spinless particles [15]

$$\frac{\Omega\left[G_n,\Sigma_n\right]}{\mathcal{N}} = -\sum_{n=-\infty}^{\infty}\frac{e^{i\omega_n 0^+}}{\beta}\left[\int_{-\infty}^{\infty} d\varepsilon\,\rho_\infty(\varepsilon)\ln\left[i\omega_n+\mu-\Sigma_n-\varepsilon\right] + \langle\ln\left[1+G_n\left(\Sigma_n-V_i\right)\right]\rangle_{av}\right]. \tag{44}$$

This representation of the CPA grand potential for fixed chemical potential μ was derived as the exact grand potential for the model of disordered electrons in $d = \infty$. Apart from the scaling of the nearest-neighbor hopping amplitude (41) we did not use any particular property of the perturbation theory. The limit to infinite dimensions was performed on hypercubic d-dimensional lattices for which the density of states in infinite dimensions reads

$$\rho_\infty(\varepsilon) = \frac{1}{\sqrt{2\pi}t^*}\exp\left(-\varepsilon^2/2t^{*2}\right). \tag{45}$$

The grand potential (44) with the density of states (45) is the exact solution for the Hamiltonian of the Anderson disordered model, eq. (13), in $d = \infty$. It serves as a good local approximation for finite-dimensional systems if the appropriate density of states is used. Notice that CPA makes sense only for lattice models with well separated nearest neighbors. It has no meaning for continuous models where multiple scatterings on continuously spread scatterers cannot be singled out and lose relevance.

5 Interacting disordered electrons – Falicov-Kimball model

5.1 Equilibrium thermodynamic properties

The quantum itinerant models in infinite dimensions (DMFT) can be solved analytically only if the Matsubara frequencies are decoupled and the local functional integral (35) reduces to a product of integrals for individual Matsubara frequencies. This is not the case for the Hubbard model of interacting electrons [24]. But a modification of the Hubbard Hamiltonian, the so-called Falicov-Kimball model, decouples the Matsubara frequencies. Its easiest spinless form is defined by the Hamiltonian [25]

$$\widehat{H}_{FK} = -t \sum_{\langle ij \rangle} c_i^\dagger c_j + \sum_i \varepsilon_i \, f_i^\dagger f_i + \sum_i c_i^\dagger c_i \left(V_i + U f_i^\dagger f_i \right) . \tag{46}$$

This Hamiltonian, in comparison with the Hubbard model, loses some important properties. A great deal of quantum dynamics goes lost in (46), since its equilibrium state is not a Fermi-liquid. The Falicov-Kimball model proved, nevertheless, invaluable in the construction of an analytic mean-field theory in strong coupling of the Hubbard-type models.

The first exact solution of this quantum itinerant model in $d = \infty$ was derived by Brandt and Mielsch [26]. The full grand potential with the renormalized variational parameters was constructed in Ref. [27]. The spinless Falicov-Kimball Hamiltonian (46) does not contain any nonlocal interaction or hybridization and hence no scaling of the coupling term is necessary. The functional $\Omega[G_\alpha^{-1} + \Sigma_\alpha]$ is an atomic solution the partition function of which contains a sum over two possible states of the local electrons. The dynamic electrons have a frequency-dependent local propagator $(G_{\alpha n}^{-1} + \Sigma_{\alpha n})^{-1}$. We obtain the grand potential

$$\frac{2\beta}{\mathcal{N}} \, \Omega \left[G, \Sigma \right] = - \sum_{\alpha = \pm} \left\langle \ln \left[1 + e^{\beta(\mu_\alpha - \varepsilon_i - \mathcal{E}_{i,-\alpha})} \right] \right\rangle_{av} \tag{47}$$

$$- \sum_{n=-\infty}^{\infty} e^{i\omega_n 0^+} \left[\sum_{\alpha=\pm} \langle \ln [1 + G_{n,\alpha} \left(\Sigma_{n,\alpha} - V_i \right)] \rangle_{av} + \int_{-\infty}^{\infty} d\varepsilon \rho_\infty(\varepsilon) \ln \left[(i\omega_n + \mu_+ - \Sigma_{n,+})(i\omega_n + \mu_- - \Sigma_{n,-}) - \varepsilon^2 \right] \right]$$

with

$$\mathcal{E}_{i,\alpha} = - \frac{1}{\beta} \sum_{n=-\infty}^{\infty} e^{i\omega_n 0^+} \ln \left[\frac{1 + G_{n,\alpha} \left(\Sigma_{n,\alpha} - V_i - U \right)}{1 + G_{n,\alpha} \left(\Sigma_{n,\alpha} - V_i \right)} \right] . \tag{48}$$

We used a subscript α to allow for a low-temperature charge order with different sublattices. Conditions on stationarity of the grand potential (47) lead to defining equations for the variational parameters $G_{\alpha n}$ and $\Sigma_{\alpha n}$. We obtain, after a few manipulations,

$$G_{n,\alpha} = \int_{-\infty}^{\infty} d\varepsilon \, \rho_\infty(\varepsilon) \frac{(i\omega_n + \mu_{-\alpha} - \Sigma_{n,-\alpha})}{(i\omega_n + \mu_+ - \Sigma_{n,+}) (i\omega_n + \mu_- - \Sigma_{n,-}) - \varepsilon^2} , \tag{49a}$$

$$1 = \left\langle \frac{n_{i,-\alpha}}{1 + G_{n,\alpha} \left(\Sigma_{n,\alpha} - V_i - U \right)} + \frac{1 - n_{i,-\alpha}}{1 + G_{n,\alpha} \left(\Sigma_{n,\alpha} - V_i \right)} \right\rangle_{av} , \tag{49b}$$

where

$$n_{i,\alpha} = \frac{1}{1+\exp\left(\beta\left(\varepsilon_i + \mathcal{E}_{i,-\alpha} - \mu_\alpha\right)\right)} \tag{49c}$$

is the averaged number of static particles in the Falicov-Kimball model. We see that the variational parameters $G_{\alpha n}$ and $\Sigma_{\alpha n}$ now depend explicitly on the Matsubara frequencies and the thermodynamics of the model contains a portion of quantum many-body fluctuations. The equations of motion are algebraic and the variational variables $G_{\alpha n}$ and $\Sigma_{\alpha n}$ depend only on a single Matsubara frequency ω_n.

It can easily be shown that if we consider a random-alloy with diagonal disorder where the constituent A with the atomic energy U has concentration x and the constituent B with concentration $1-x$ has the atomic energy 0, then the Falicov-Kimball model in $d=\infty$ coincides with CPA of such an alloy. The Falicov-Kimball Hamiltonian defines a semiclassical model with reduced dynamical quantum fluctuations. The dynamical fluctuations are restricted, since we have only one species of dynamical electrons. They interact with static electrons, i.e., they are scattered on static impurity potentials distributed in the lattice. Unlike the static disorder of alloys the localized electrons of the Falicov-Kimball model serving as random scatterers for the mobile electrons are thermally equilibrated, which introduces a nontrivial thermodynamics. The semiclassical character of the Falicov-Kimball model becomes evident from the fact that the partition function of this model can be obtained in any dimension as a static approximation in a special functional-integral representation [28].

Equations (49a) and (49b) coincide with the well-known Hubbard-III approximation [29] if we neglect the static electrons and replace the density of static particles by the density of the dynamic ones. Then equation (49c) must be forgotten and replaced by a sum rule. The analogy between the model of random alloys and the Hubbard-III approximation, discovered in Ref. [6], led at the end of the seventies of the last century to numerous attempts to improve on the weak-coupling Hartree-Fock theory by using the "alloy analogy" reasoning [30]. However, mean-field theories constructed in this way and based on the Hubbard-III approximation become *thermodynamically inconsistent* and lead to unphysical behavior [31]. The Hubbard-III approximation was made thermodynamically consistent by adding a new variational parameter [32].

5.2 Response to external perturbations

The equation for the self-energy of the mobile electrons of the $d=\infty$ Falicov-Kimball model for fixed densities of the local electrons resembles the Soven equation. There is, however, a significant difference when we turn to response functions describing the reaction of the equilibrium state to weak external perturbations. The response functions are derived from the two-particle Green function. The Falicov-Kimball model, unlike the Anderson disordered model, displays a low-temperature critical behavior and a phase transition to a checker-board phase. This difference can be demonstrated on the local two-particle vertex of the two solutions. In both cases, due to conservation of energy, the two-particle vertex contains only two independent variables, Matsubara frequencies. The full local vertex of the Falicov-Kimball model in $d=\infty$ measuring

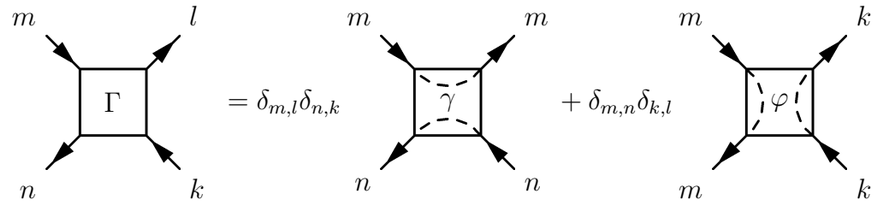


Fig. 4: *Graphical representation of eq.* (50)*. The dashed lines within the boxes indicate charge propagation from the incoming to the corresponding outgoing line. The vertex for the disordered Fermi gas contains only the left diagram.*

correlations between two conduction electrons can be decomposed into two distinct contributions

$$\Gamma^{MF}_{mn,kl} = \delta_{m,l}\,\delta_{n,k}\,\gamma_{m,n} + \delta_{m,n}\,\delta_{k,l}\,\varphi_{m,k}\,, \tag{50}$$

where the integer indices denote fermionic Matsubara frequencies. The full vertex is shown in Fig. 4, where we indicated the way the corners of the vertices are connected by an internal electron line. The CPA vertex is just its first term, $\gamma_{m,n}$, that is relevant for transport properties (electrical conductivity). The second vertex, $\varphi_{m,n}$, determines the thermodynamic response and the low-temperature critical behavior.

The thermodynamic vertex $\varphi_{m,n}$ can be represented via an irreducible one, $\kappa_{m,n}$, from a local Bethe-Salpeter equation

$$\varphi_{m,n} = \kappa_{m,n} + \frac{1}{\beta}\sum_{l}\kappa_{m,l}G_l^2\,\varphi_{l,n}\,. \tag{51}$$

The irreducible vertex satisfies a Ward identity $\kappa_{m,n} = \delta\Sigma_m/\delta G_n$ as is evident from the connection of the corners of vertex $\varphi_{m,n}$ in Fig. 4.

It is more complicated to represent vertex $\gamma_{m,n}$ via an irreducible one. To do so and to derive the corresponding Ward identity we replicate the creation and annihilation operators and introduce external perturbations into the thermodynamic description via a generalized grand potential of a replicated system $\Omega^{\nu}(\mu_1,\mu_2,\ldots\mu_\nu;\Delta)$ with ν chemical potentials $\mu_1,\mu_2,\ldots,\mu_\nu$ [33]. An external perturbation Δ is used to couple different replicas and to break the initial replica independence. We then can write

$$\Omega^{\nu}(\mu_1,\mu_2,\ldots\mu_\nu;\Delta) = -\frac{1}{\beta}\left\langle \ln \mathrm{tr}\exp\left[-\beta\sum_{i,j=1}^{\nu}\left(\widehat{H}^{(i)}_{FKM}\,\delta_{ij} - \mu_i\widehat{N}^{(i)}\delta_{ij} + \Delta\widehat{H}^{(ij)}\right)\right]\right\rangle_{av}, \tag{52}$$

where we assigned to each replica characterized by energy (chemical potential) μ_i a separate Hilbert space and denoted by $\Delta\widehat{H}^{(ij)} = \sum_{kl}\Delta^{(ij)}_{kl}\,\widehat{c}^{(i)\dagger}_k\,\widehat{c}^{(j)}_l$ an external perturbation to be set zero at the end. The thermodynamic potential $\Omega^{\nu}(\mu_1,\nu_2,\ldots\mu_\nu;\Delta)$ is a generating functional for averaged products of Green functions up to ν-th order. In practice, we will use linear-response theory with one- and two-particle Green functions, i. e., $\Omega^{\nu}(\mu_1,\mu_2,\ldots\mu_\nu;\Delta)$ is expanded up to Δ^2. Therefore it is sufficient to introduce only two replicas.

The external disturbance Δ mixes different replicas, and propagators in the replicated space are matrices in the replica indices. Since we are interested only in the averaged two-particle functions, we can represent the propagator by a two-by-two matrix

$$\widehat{G}^{-1}(\mathbf{k}_1, z_1, \mathbf{k}_2, z_2; \Delta) = \begin{pmatrix} z_1 - \varepsilon(\mathbf{k}_1) - \Sigma_{11}(\Delta) & \Delta \;\; - \Sigma_{12}(\Delta) \\ \Delta \;\; - \Sigma_{21}(\Delta) & z_2 - \varepsilon(\mathbf{k}_2) - \Sigma_{22}(\Delta) \end{pmatrix}, \tag{53}$$

where $\varepsilon(\mathbf{k})$ is the lattice dispersion relation and the self-energy elements Σ_{ab} generally depend on both energies z_1, z_2.

The local two-particle vertex is a solution of a Bethe-Salpeter equation with an irreducible two-particle vertex λ and local propagators. We easily find that the Bethe-Salpeter equation in the mean-field approximation reduces to an algebraic one

$$\gamma(z_1, z_2) = \frac{\lambda(z_1, z_2)}{1 - \lambda(z_1, z_2)\, G(z_1)\, G(z_2)}\,. \tag{54}$$

The irreducible vertex λ is determined in equilibrium ($\Delta = 0$) from an equation consistent with the Ward identity, eq. (21) with local propagators,

$$\lambda_{m,n} = \frac{1}{G_m G_n}\left(1 - \left\langle \left[\frac{n_{i,-\alpha}}{1 + G_{n,\alpha}\,(\Sigma_{n,\alpha} - V_i - U)} + \frac{1 - n_{i,-\alpha}}{1 + G_{n,\alpha}\,(\Sigma_{n,\alpha} - V_i)}\right] \right.\right. \\ \left.\left. \times \left[\frac{n_{i,-\alpha}}{1 + G_{n,\alpha}\,(\Sigma_{n,\alpha} - V_i - U)} + \frac{1 - n_{i,-\alpha}}{1 + G_{n,\alpha}\,(\Sigma_{n,\alpha} - V_i)}\right] \right\rangle_{av}^{-1}\right). \tag{55}$$

We can easily verify that this equation coincides with the CPA solution for the irreducible vertex $\lambda(z_1, z_2)$ [9, 35].

6 Transport properties within CPA

6.1 Non-local two-particle vertex and electrical conductivity

There is no ambiguity in the mean-field construction of local one- and two-particle functions. But a mean-field treatment has a physical relevance only if it is able to produce nonlocal correlation functions, the long-range fluctuations of which may significantly influence the thermodynamic and dynamical behavior. There is, however, no unique way to generate the two-particle vertex with non-local contributions within the local (mean-field) approach. The simplest and most straightforward way is to use the Bethe-Salpeter equation with the CPA irreducible vertex λ, eq. (55), and to replace the product of the local propagators with a convolution of the full nonlocal one-electron propagators $G(\mathbf{k}, z)$. Such a Bethe-Salpeter equation then remains algebraic in momentum representation and results in a two-particle vertex with only one transfer momentum. We obtain

$$\Gamma^{\pm}(z_1, z_2; \mathbf{q}^{\pm}) = \frac{\lambda(z_1, z_2)}{1 - \lambda(z_1, z_2)\, \chi^{\pm}(z_1, z_2; \mathbf{q}^{\pm})}\,, \tag{56}$$

where we denoted the two-particle bubble

$$\chi^{\pm}(z_1, z_2; \mathbf{q}) = \frac{1}{N} \sum_{\mathbf{k}} G(\mathbf{k}, z_1)\, G(\mathbf{q} \pm \mathbf{k}, z_2)\,. \tag{57}$$

The ambiguity in this definition of the full mean-field vertex is in the type of nonlocal multiple scatterings we include in the Bethe-Salpeter equation. They are here denoted by the superscript $\pm$. The plus sign corresponds to multiple scatterings of electron-hole pairs, while the minus sign to pairs of electrons. In case of elastic scatterings the electron-hole and electron-electron bubbles produce numerically the same result. However, the difference between the two types of pair scatterings lies in the respective transfer momentum $\mathbf{q}^{\pm}$. Using the notation for momenta in the two-particle Green function from eq. (23) we have $\mathbf{q}^{+} = \mathbf{q}$ for the electron-hole pair scatterings and $\mathbf{q}^{-} = \mathbf{q} + \mathbf{k} + \mathbf{k}'$ for scatterings of two electrons. The nonlocal vertex in CPA is that from eq. (56) and the electron-hole bubble with $\mathbf{q}^{+}$. We discuss this ambiguity more in the next subsection.

We now turn our attention to the electrical conductivity. Using the Kubo formula we obtain a simple representation of the longitudinal conductivity at zero temperature [40]

$$\sigma_{\alpha\alpha} = \frac{e^2}{2\pi N^2} \sum_{\mathbf{k}\mathbf{k}'} v_\alpha(\mathbf{k})\, v_\alpha(\mathbf{k}') \left[G^{AR}_{\mathbf{k}\mathbf{k}'} - \mathrm{Re}\, G^{RR}_{\mathbf{k}\mathbf{k}'} \right], \tag{58}$$

with the values of the two-particle Green function at the Fermi energy. We used an abbreviation $G^{AR}_{\mathbf{k}\mathbf{k}'} = G^{AR}_{\mathbf{k}\mathbf{k}'}(0, 0; \mathbf{0})$ and

$$G^{AR}_{\mathbf{k}\mathbf{k}'}(\omega, \omega'; \mathbf{q}) = G^{\{2\}}_{\mathbf{k}\mathbf{k}'}(\omega - i0^+, \omega' + i0^+;\, \mathbf{q}),$$
$$G^{RR}_{\mathbf{k}\mathbf{k}'}(\omega, \omega'; \mathbf{q}) = G^{\{2\}}_{\mathbf{k}\mathbf{k}'}(\omega + i0^+, \omega' + i0^+;\, \mathbf{q})\,.$$

We decompose the conductivity tensor (58) into two parts by replacing the two-particle Green function by the representation given in eq. (17) with the two-particle vertex Γ. We then have a sum of two terms

$$\sigma_{\alpha\alpha} = \frac{e^2}{\pi N} \sum_{\mathbf{k}} |v_\alpha(\mathbf{k})|^2 \left| \mathrm{Im}\, G^R(\mathbf{k}) \right|^2 + \Delta\sigma_{\alpha\alpha}\,, \tag{59}$$

where the first term is the standard one-electron or Drude conductivity at zero temperature. The second term is the genuine two-particle contribution and is called vertex correction. It is proportional to the appropriate matrix element of the two-particle vertex that, at zero temperature, reads

$$\Delta\sigma_{\alpha\alpha} = \frac{e^2}{2\pi N^2} \sum_{\mathbf{k}\mathbf{k}'} v_\alpha(\mathbf{k})\, v_\alpha(\mathbf{k}') \left(\left| G^R_{\mathbf{k}} \right|^2 \Delta\Gamma^{AR}_{\mathbf{k}\mathbf{k}'} \left| G^R_{\mathbf{k}'} \right|^2 - \mathrm{Re} \left[\left(G^R_{\mathbf{k}} \right)^2 \Delta\Gamma^{RR}_{\mathbf{k}\mathbf{k}'} \left(G^R_{\mathbf{k}'} \right)^2 \right] \right). \tag{60}$$

It is not the full two-particle vertex that is important for the electrical conductivity, but only its odd part $\Delta\Gamma$. That is, only the part of the vertex function depending, on bipartite lattices, on odd powers of the fermionic momenta $\mathbf{k}$ and $\mathbf{k}'$ contributes to the electrical conductivity. Hence, CPA does not contain vertex corrections to the electrical conductivity, since the two-particle vertex $\Gamma^{CPA}_{\mathbf{k}\mathbf{k}'}(\omega, \omega'; \mathbf{q}) = \Gamma(\omega, \omega'; \mathbf{q})$ does not depend on the incoming fermionic momenta $\mathbf{k}, \mathbf{k}'$.

6.2 Gauge invariance and electron-hole symmetry

Electrical conductivity is a form of a response of the charged system to an electromagnetic perturbation. An important feature of the interaction of the charged system with an electromagnetic field is gauge invariance that must be guaranteed in the response functions. There are two fundamental response functions to the electromagnetic field, one based on the current-current and the other on the density-density correlation functions. The former is used in the Kubo formula for the electrical conductivity, eq. (58), and the latter for determination of charge diffusion.
There is a number of more or less heuristic arguments in the literature that relate the density response with the conductivity [34]. They use macroscopic gauge invariance and charge conservation for particles exposed to an electromagnetic field. A microscopic quantum derivation was presented in Ref. [35].
Gauge invariance is used to relate the external scalar potential with the electric field $\mathbf{E} = -\boldsymbol{\nabla}\varphi$. The current-density generated by a harmonic external field then is

$$\mathbf{j}(\mathbf{q},\omega) = \boldsymbol{\sigma}(\mathbf{q},\omega)\cdot\mathbf{E}(\mathbf{q},\omega) = -i\boldsymbol{\sigma}(\mathbf{q},\omega)\cdot\mathbf{q}\,\varphi(\mathbf{q},\omega)\,, \tag{61}$$

where $\boldsymbol{\sigma}(\mathbf{q},\omega)$ denotes the tensor of the electrical conductivity. Charge conservation is expressed by a continuity equation. In equilibrium we can use the operator form of the continuity equation that follows from the Heisenberg equations of motion for the current and density operators. For Hamiltonians with a quadratic dispersion relation we have

$$e\partial_t\widehat{n}(\mathbf{x},t) + \boldsymbol{\nabla}\cdot\widehat{\mathbf{j}}(\mathbf{x},t) = 0\,. \tag{62}$$

The energy-momentum representation of the continuity equation in the ground-state solution is

$$-\,i\omega e\,\delta n(\mathbf{q},\omega) + i\mathbf{q}\cdot\mathbf{j}(\mathbf{q},\omega) = 0\,. \tag{63}$$

We have to use a density variation of the equilibrium density, i. e., the externally induced density $\delta n(\mathbf{q},\omega) = n(\mathbf{q},\omega) - n_0$ in the continuity equation with the averaged values of the operators. From the above equations and for linear response $\delta n(\mathbf{q},\omega) = -e\chi(\mathbf{q},\omega)\varphi(\mathbf{q},\omega)$ we obtain in the isotropic case

$$\sigma = \lim_{\omega\to 0}\lim_{q\to 0}\frac{-ie^2\omega}{q^2}\chi(\mathbf{q},\omega)\,, \tag{64}$$

where at zero temperature

$$\begin{aligned}\chi(\mathbf{q},\omega) = &-\int_{-\omega}^{0}\frac{dx}{2\pi i}\left\langle G^{AR}_{\mathbf{k}\mathbf{k}'}(x,x+\omega;\mathbf{q}) - G^{RR}_{\mathbf{k}\mathbf{k}'}(x,x+\omega;\mathbf{q})\right\rangle_{\mathbf{k},\mathbf{k}'}\\ &+\int_{-\infty}^{0}\frac{dx}{\pi}\,\mathrm{Im}\left\langle G^{RR}_{\mathbf{k}\mathbf{k}'}(x,x+\omega;\mathbf{q})\right\rangle_{\mathbf{k},\mathbf{k}'}\end{aligned} \tag{65}$$

is the density response function. Relation (64) is often taken as granted for the whole range of the disorder strength and used for the definition of the zero-temperature conductivity when describing the Anderson localization transition [36, 37].

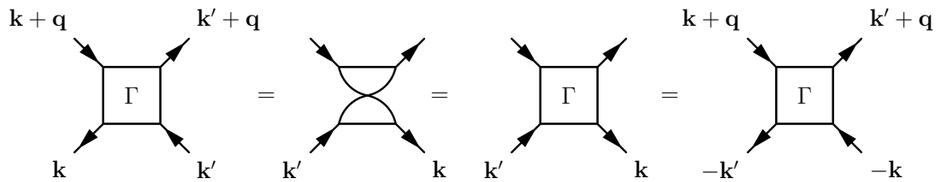


Fig. 5: *Graphical representation of time-reversal symmetry of the two-particle vertex when the lower electron line is reversed.*

Relation (64) between the static optical conductivity and the density response holds if the latter function displays the so-called diffusion pole. One can prove by using the Vollhardt-Wölfle-Ward identity, eq. (21), that in the limit $q \to 0$ and $\omega \to 0$ [35]

$$\chi(\mathbf{q}, \omega) \doteq \frac{D n_F \, q^2}{-i\omega + D \, q^2} \,, \tag{66}$$

where D is the static diffusion constant. Inserting eq. (66) into eq. (64), we end up with the Einstein relation between the diffusion constant and the conductivity $\sigma = e^2 D n_F$, where n_F is the electron density at the Fermi energy. This relation holds in CPA with the Drude conductivity [35].

The coherent potential approximation delivers good results for the one-particle quantities but it fails to take into account back-scatterings that are responsible for the vertex corrections in the electrical conductivity. It also fails to maintain electron-hole symmetry at the level of two-particle functions. Electron-hole symmetry or equivalently time-reversal is an important feature of electron systems without spin- and orbital-dependent scatterings. According to this invariance the physical (measurable) results should not depend on the orientation of the electron propagators. Changing the orientation of the electron line is equivalent to the spatial inversion in the momentum space. The electron-hole symmetry for the one- and two-particle propagators means

$$G(\mathbf{k}, z) = G(-\mathbf{k}, z) \,, \tag{67a}$$

$$\mathit{\Gamma}_{\mathbf{k}\mathbf{k}'}(z_+, z_-; \mathbf{q}) = \mathit{\Gamma}_{\mathbf{k}\mathbf{k}'}(z_+, z_-; -\mathbf{q} - \mathbf{k} - \mathbf{k}') = \mathit{\Gamma}_{-\mathbf{k}'-\mathbf{k}}(z_+, z_-; \mathbf{q} + \mathbf{k} + \mathbf{k}') \,. \tag{67b}$$

The spatial inversion was applied only to one fermion propagator in the two-particle vertex, the upper in the first equality and the lower in the second. The latter transformation is graphically represented in Fig. 5. This is an exact relation which is, however, broken in CPA as discussed in Refs. [38, 33]. One has to go beyond the local mean-field approximation to correct this deficiency.

7 Beyond CPA

7.1 Vertex corrections to the electrical conductivity

Dynamical mean-field theory contains all single-site scatterings. To go beyond, one has to avoid the repetition of lattice indices as in the expansion of the T matrix in eq. (10). The bare expansion parameter is then the off-diagonal one-electron propagator from mean-field theory. It is given by

$$\bar{G}(\mathbf{k},\zeta) = \frac{1}{\zeta - \varepsilon(\mathbf{k})} - \int \frac{d\varepsilon\,\rho(\varepsilon)}{\zeta-\varepsilon}\,, \tag{68}$$

with $\zeta = z - \Sigma(z)$ where the local self-energy $\Sigma(z)$ is that of the mean-field solution. The off-diagonal two-particle bubble describes the simplest non-local contribution. It is the convolution of the off-diagonal one-electron propagators. We have

$$\bar{\chi}(\zeta,\zeta';\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{k}} \bar{G}(\mathbf{k},\zeta)\,\bar{G}(\mathbf{k}+\mathbf{q},\zeta') = \chi(\zeta,\zeta';\mathbf{q}) - G(\zeta)\,G(\zeta')\,, \tag{69}$$

where $\chi(\zeta,\zeta';\mathbf{q})$ is the full two-particle bubble. The frequency indices are external parameters and we suppress them when they are not necessary to specify the particular type of one- or two-electron propagators.

The asymptotic limit of the full two-particle vertex in high spatial dimensions contains beyond the local mean-field vertex γ also non-local contributions from the electron-hole and electron-electron ladders. They are different in their off-diagonal part [38, 33]. The asymptotic two-particle vertex consistent with the electron-hole symmetry can, in leading order, be represented as

$$\mathit{\Gamma}_{\mathbf{k}\mathbf{k}'}(\mathbf{q}) = \gamma\left[1+\gamma\left(\frac{\bar{\chi}(\mathbf{q})}{1-\gamma\,\bar{\chi}(\mathbf{q})}+\frac{\bar{\chi}(\mathbf{Q})}{1-\gamma\,\bar{\chi}(\mathbf{Q})}\right)\right] = \gamma + \mathit{\Delta\Gamma}_{\mathbf{k}\mathbf{k}'}(\mathbf{q})\,, \tag{70}$$

where we denoted by $\mathbf{Q} = \mathbf{q}+\mathbf{k}+\mathbf{k}'$ the momentum conserved in the electron-electron scattering channel. Notice that the contribution from the electron-hole channel with $\bar{\chi}(\mathbf{q})$ is part of the two-particle vertex from CPA and can be derived from the Velický-Ward identity [9]. The two-particle vertex from CPA does not carry the full $1/d$ correction to the local vertex and moreover it is not electron-hole symmetric on the two-particle level [38].

A consistent extension of the local mean-field two-particle vertex must contain both non-local contributions from the electron-hole and electron-electron channels as given in eq. (70). It is not appropriate to calculate the electrical conductivity from the decomposition in eq. (59), since the vertex corrections can outweight the Drude term and the conductivity may get negative [39]. The actual expansion around CPA should be done for the electron-hole irreducible vertex $\mathit{\Lambda}$ from the Bethe-Salpeter equation (18). Actually, only the vertex correction $\overline{\mathit{\Lambda}} = \mathit{\Lambda} - \lambda$ is the object of the perturbation expansion around CPA. The two-particle Green function with only the off-diagonal part of the vertex corrections can be represented as [39]

$$\overline{G}^{ab}_{\mathbf{k}\mathbf{k}'}(\mathbf{q}) = G^{b}_{\mathbf{k}+\mathbf{q}}\left[\delta(\mathbf{k}-\mathbf{k}') + \overline{G}^{a}_{\mathbf{k}}\,\mathit{\Gamma}^{ab}_{\mathbf{k}\mathbf{k}'}(\mathbf{q})\,\overline{G}^{b}_{\mathbf{k}'+\mathbf{q}}\right] G^{a}_{\mathbf{k}'}\,, \tag{71}$$

where superscripts a, b stand for R, A where appropriate. It is a solution of a Bethe-Salpeter equation with the irreducible vertex $\bar{\Lambda}^{ab}$

$$\overline{G}^{ab}_{\mathbf{k}\mathbf{k}'}(\mathbf{q}) = G^{b}_{\mathbf{k}+\mathbf{q}} \left[1 - \widehat{\overline{\Lambda}}^{ab}(\mathbf{q})\star\right]^{-1}_{\mathbf{k}\mathbf{k}'} G^{a}_{\mathbf{k}'} \,. \tag{72}$$

We use the constrained two-electron Green function $\overline{G}^{ab}$ in the calculation of the electrical conductivity. From the Bethe-Salpeter equation (72) we straightforwardly obtain

$$\sigma_{\alpha\beta} = \frac{e^2}{2\pi N^2} \sum_{\mathbf{k}\mathbf{k}'} v_\alpha(\mathbf{k}) \left[G^{A}_{\mathbf{k}} \left[1 - \widehat{\overline{\Lambda}}^{RA}\star\right]^{-1}_{\mathbf{k}\mathbf{k}'} G^{R}_{\mathbf{k}'} - \mathrm{Re}\left(G^{R}_{\mathbf{k}} \left[1 - \widehat{\overline{\Lambda}}^{RR}\star\right]^{-1}_{\mathbf{k}\mathbf{k}'} G^{R}_{\mathbf{k}'} \right) \right] v_\beta(\mathbf{k}'). \tag{73}$$

The expansion around CPA with the off-diagonal propagators will now be applied on the irreducible vertex $\bar{\Lambda}$ in the above equation, accompanied by a non-perturbative matrix inversion as to not break non-negativity of the conductivity [39]. The leading-order contribution to the vertex $\bar{\Lambda}^{ab}$ in high spatial dimensions is

$$\overline{\Lambda}^{ab}_{\mathbf{k}\mathbf{k}'}(\mathbf{q}) = \gamma^{ab} \left[1 + \frac{\gamma^{ab}\overline{\chi}^{ab}(\mathbf{k}+\mathbf{k}'+\mathbf{q})}{1 - \gamma^{ab}\,\overline{\chi}^{ab}(\mathbf{k}+\mathbf{k}'+\mathbf{q})}\right] . \tag{74}$$

The off-diagonal propagator $\overline{G}$ is the fundamental parameter in the expansion around the mean-field limit. It prevents from double counting of multiple local scatterings from the mean-field solution and it also makes the calculation of corrections to the mean-field result numerically more stable. It is preferable to use the full local mean-field vertex γ^{ab} instead of the irreducible one, λ^{ab}, in all formulas of the expansion around mean field, since the latter contains a pole in the RR (AA) channel that is compensated in the perturbation expansion by the former vertex. Notice that the leading-order vertex corrections calculated from the expansion of the right-hand side of eq. (73) coincide with the leading corrections to the mean-field conductivity derived in Ref. [40].

7.2 Making expansions beyond local approximations conserving

The problem of the perturbation expansions is that they break exact relations. Sometimes the deviations from the exact relations are not drastic and do not qualitatively alter the physical behavior. Unfortunately, this is not the case for disordered systems. The Einstein relation (64) between the diffusion constant and the electrical conductivity is of fundamental importance. It holds, however, only if the Vollhardt-Wölfle-Ward identity (21) is obeyed. Each perturbation expansion for two-particle quantities breaks this identity in that a causal self-energy Σ cannot be made compatible with a given approximate two-particle irreducible vertex Λ by fulfilling eq. (21) [41]. Approximations where eq. (21) is broken are no longer conserving and the validity of the continuity equation (62) cannot be guaranteed.

To make the perturbation theory for two-particle functions consistent and conserving we first reconcile the two-particle irreducible vertex in the diagrammatic expansion with the one-electron self-energy via the Ward identity in the best possible way. The full dynamical Ward identity can

neither be used to determine the one-particle self-energy from the two-particle irreducible vertex nor vice versa, since the vertex contains more information than the self-energy. The Ward identity poses a restriction on the form of the two-particle irreducible vertex and generally serves only as a consistency check and a guarantee that the macroscopic conservation laws are obeyed. The Ward identity (21) for $\omega = 0$ and $q = 0$ can nevertheless be used to determine the imaginary part of the self-energy from the electron-hole irreducible vertex via

$$\operatorname{Im}\Sigma^R_{\mathbf{k}}(E) = \frac{1}{N}\sum_{\mathbf{k}'} \Lambda^{RA}_{\mathbf{k}\mathbf{k}'}(E;0,\mathbf{0})\operatorname{Im} G^R_{\mathbf{k}'}(E)\,, \tag{75a}$$

since both sides of this identity contain the same number of degrees of freedom and the equation for the imaginary part of the self-energy can consistently be resolved for each energy E and momentum $\mathbf{k}$. The corresponding real part of the self-energy is then found from the Kramers-Kronig relation

$$\operatorname{Re}\Sigma^R_{\mathbf{k}}(E) = \Sigma_\infty + P\int_{-\infty}^{\infty}\frac{d\omega}{\pi}\frac{\operatorname{Im}\Sigma^R_{\mathbf{k}}(\omega)}{\omega - E} \tag{75b}$$

that ensures analyticity and causality of the self-energy in the plane of complex energies beyond the real axis.

Since we know that the full dynamical Ward identity cannot be fulfilled by the irreducible vertex from the perturbation expansion we introduce a new physical irreducible vertex that we denote by L. It will be connected with vertex Λ from the perturbation theory but will be made to obey the full Ward identity, that is,

$$\Delta\Sigma^{RA}_{\mathbf{k}}(E;\omega,\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{k}'} L^{RA}_{\mathbf{k}_+,\mathbf{k}'_+}(E,\omega;\mathbf{q})\Delta G^{RA}_{\mathbf{k}'}(E;\omega,\mathbf{q}) \tag{76}$$

holds. Here we introduced the discontinuities $\Delta G^{RA}_{\mathbf{k}}(E;\omega,\mathbf{q}) = G^R_{\mathbf{k}_+}(E_+) - G^A_{\mathbf{k}_-}(E_-)$ and $\Delta\Sigma^{RA}_{\mathbf{k}}(E;\omega,\mathbf{q}) = \Sigma^R_{\mathbf{k}_+}(E_+) - \Sigma^A_{\mathbf{k}_-}(E_-)$, and denoted $\mathbf{k}_\pm = \mathbf{k} \pm \mathbf{q}/2$, $E_\pm = E \pm \omega/2$.

The vertex $L^{RA}_{\mathbf{k}_+,\mathbf{k}'_+}(E,\omega;\mathbf{q})$ is not directly accessible in diagrammatic approximations. The two-particle vertex functions in perturbation expansion are represented by classes of diagrams with sums over momenta in the whole two-particle Hilbert space. The output of the diagrammatic expansion is an irreducible vertex $\Lambda_{\mathbf{k}\mathbf{k}'}(E;\omega,\mathbf{q})$ that does not *generically* comply with the Ward identity (76) if the self-energy $\Sigma^{R/A}_{\mathbf{k}}(E)$ is non-local, that is, depends on momentum $\mathbf{k}$. Vertex corrections that take into account the impact of the Ward identity on the irreducible vertex for the given self-energy then must be introduced beyond the standard diagrammatic approach to make the theory conserving. We therefore distinguish the physical vertex Γ^{RA} obeying the Bethe-Salpeter equation with the irreducible vertex L^{RA} from the vertex $\widetilde{\Gamma}^{RA}$ determined from the perturbative vertex Λ^{RA} via the corresponding Bethe-Salpeter equation (18).

We can make the approximations for the two-particle vertex $\Lambda_{\mathbf{k}\mathbf{k}'}(E;\omega,\mathbf{q})$ conserving by appropriately correcting its action in the momentum space. It is namely sufficient to correctly replace the values of vertex Λ on a subspace on which its action is already predefined by the self-energy

via the Ward identity. For this purpose we introduce a new correcting function measuring the deviation of the given vertex Λ^{RA} from the Ward identity [42]

$$R_{\mathbf{k}}(E;\omega,\mathbf{q}) = \frac{1}{N}\sum_{\mathbf{k}'} \Lambda^{RA}_{\mathbf{k}_+\mathbf{k}'_+}(E;\omega,\mathbf{q})\,\Delta G_{\mathbf{k}'}(E;\omega,\mathbf{q}) \;-\; \Delta\Sigma_{\mathbf{k}}(E;\omega,\mathbf{q})\;. \tag{77}$$

This function vanishes in the metallic phase for $\omega = 0$ and $q = 0$ due to the definition of the self-energy (75). It is identically zero if the vertex $\Lambda_{\mathbf{k}\mathbf{k}'}(E;\omega,\mathbf{q})$ obeys the Ward identity. With the aid of function $R_{\mathbf{k}}(E;\omega,\mathbf{q})$ we construct a conserving electron-hole irreducible vertex [42]

$$L^{RA}_{\mathbf{k}_+\mathbf{k}'_+}(E;\omega,\mathbf{q}) = \Lambda^{RA}_{\mathbf{k}_+\mathbf{k}'_+}(E;\omega,\mathbf{q}) - \frac{1}{\langle\Delta G(E;\omega,\mathbf{q})^2\rangle}\Big[\Delta G_{\mathbf{k}}(E;\omega,\mathbf{q})\,R_{\mathbf{k}'}(E;\omega,\mathbf{q})$$
$$+R_{\mathbf{k}}(E;\omega,\mathbf{q})\,\Delta G_{\mathbf{k}'}(E;\omega,\mathbf{q}) \quad - \frac{\Delta G_{\mathbf{k}}(E;\omega,\mathbf{q})\,\Delta G_{\mathbf{k}'}(E;\omega,\mathbf{q})}{\langle\Delta G(E;\omega,\mathbf{q})^2\rangle}\,\langle R(E;\omega,\mathbf{q})\,\Delta G(E;\omega,\mathbf{q})\rangle\Big] \tag{78}$$

that is manifestly compliant with the full Ward identity (76) for arbitrary ω and $\mathbf{q}$. To shorten the expression, we abbreviated the $\mathbf{k}$ sums: $\langle\Delta G(E;\omega,\mathbf{q})^2\rangle = N^{-1}\sum_{\mathbf{k}}\Delta G_{\mathbf{k}}(E;\omega,\mathbf{q})^2$ and $\langle R(E;\omega,\mathbf{q})\,\Delta G(E;\omega,\mathbf{q})\rangle = N^{-1}\sum_{\mathbf{k}} R_{\mathbf{k}}(E;\omega,\mathbf{q})\,\Delta G_{\mathbf{k}}(E;\omega,\mathbf{q})$.

Function L^{RA} is the desired physical irreducible vertex to be used in determining the physical vertex Γ^{RA} from which all relevant macroscopic quantities will be calculated. The conserving vertex $\Gamma_{\mathbf{k}\mathbf{k}'}(E;\omega,\mathbf{q})$, determined from the Bethe-Salpeter equation with the conserving irreducible vertex $L_{\mathbf{k}\mathbf{k}'}(E;\omega,\mathbf{q})$, generally differs from vertex $\widetilde{\Gamma}(E;\omega,\mathbf{q})$ obtained from the perturbative vertex $\Lambda_{\mathbf{k}\mathbf{k}'}(E;\omega,\mathbf{q})$. The two vertices are equal only when the difference function $R_{\mathbf{k}}(E;\omega,\mathbf{q})$ vanishes. Our construction guarantees that this happens for $\omega = 0$ and $q = 0$, i.e.,

$$\widetilde{\Gamma}^{RA}_{\mathbf{k}\mathbf{k}'}(E;0,\mathbf{0}) = \Gamma^{RA}_{\mathbf{k}\mathbf{k}'}(E;0,\mathbf{0})\;, \tag{79}$$

since the self-energy is determined from the two-particle vertex via eqs. (75). This means that the vertex $\Lambda^{RA}_{\mathbf{k}\mathbf{k}'}(E;\omega,\mathbf{q})$ is directly related to the measurable macroscopic quantities only for $\omega = 0$, $\mathbf{q} = \mathbf{0}$.

The physical irreducible vertex $L^{ab}(E;\omega,\mathbf{q})$ can in this way be constructed to any approximate irreducible $\Lambda^{ab}(E;\omega,\mathbf{q})$ from diagrammatic perturbation theory. The continuity equation is then saved and the density response displays a diffusion pole with a diffusion constant. The isotropic diffusion constant is then expressed via a Kubo-like formula with the full two-particle vertex [42]

$$\pi n_F D = \frac{1}{N^2}\sum_{\mathbf{k},\mathbf{k}'}\left[(\hat{\mathbf{q}}\cdot\mathbf{v}_{\mathbf{k}})|G^R_{\mathbf{k}}|^2\left[N\delta_{\mathbf{k},\mathbf{k}'} + \Gamma^{RA}_{\mathbf{k}\mathbf{k}'}|G^R_{\mathbf{k}'}|^2\right]\right.$$
$$\left.\times\left[\mathrm{Im}\,G^R_{\mathbf{k}'}\hat{\mathbf{q}}\cdot\mathbf{v}_{\mathbf{k}'} + \mathrm{Im}\left(G^R_{\mathbf{k}'}\hat{\mathbf{q}}\cdot\nabla_{\mathbf{k}'}\Sigma^R_{\mathbf{k}'}\right)\right]\mathrm{Im}\,\Sigma^R_{\mathbf{k}'}\right], \tag{80}$$

where $\hat{\mathbf{q}}$ is the unit vector pointing in the direction of the drifting electric force. All the frequency variables are set zero, at the Fermi energy. This exact expression is the starting point for the derivation of consistent approximations for the diffusion constant needed to reach quantitative results.

8 Conclusions

The coherent potential approximation was introduced and developed in the late sixties of the last century. Initially it was restricted to the quantum mechanical problem of a particle in a random lattice. The concept of a homogeneous coherent potential simulating the effect of the fluctuating environment so that locally and in the long-time limit there is no difference between the averaged and fluctuating environment proved useful beyond single-particle systems.

The first step in giving the coherent potential a more general meaning was understanding the approximation in terms of Feynman diagrams. Using the functional integral as a generator of Feynman diagrams then allowed for transfering CPA to many-body systems with arbitrary local interaction or site-independent disorder. It appeared that CPA is another form of the cavity field in which all single-loop corrections to tree diagrams are summed. The final framing of CPA was achieved by introducing DMFT as the exact solution in the limit of quantum itinerant models in infinite lattice dimensions.

The coherent potential approximation, its consistency and correct analytic behavior emerged within DMFT in a new light as an exact solution in a special limit. The equations determining the coherent potential deliver also a mean-field solution of interacting models with elastic scattering, i.e., where the energy is conserved during scattering events. Moreover, the thermodynamic formulation of the CPA equations for the disordered Anderson and Falicov-Kimball models revealed differences in the solutions of the two models for their response functions. It also gave a proper understanding of the Hubbard-III approximation, rectifying its inconsistency.

The coherent potential approximation, unlike DMFT for the Hubbard model, is analytically solvable. This is a huge advantage, since it allows for the full analytic control of the mean-field behavior, including quantum criticality. Last but not least, CPA as a simpler DMFT may also serve as a test arena for the reliability of approximations devised for the Hubbard model where it is otherwise uncontrolled. The major restriction in the applicability of CPA is that the equilibrium states are not Fermi liquids and their analytic properties are not directly transferable to heavy-fermion systems.

Local mean-field approximations in whatever formulation are consistent only for the local variables or one-particle properties. Once we need to calculate the response of the extended system to an external perturbation, we must go beyond DMFT. The non-local two-particle functions are not uniquely defined in DMFT, since the limit to infinite dimensions is not interchangeable with functional derivatives [22]. Then either the Ward identity or electron-hole symmetry is broken for non-local response functions.

The ambiguity in the mean-field definition of non-local two-particle Green functions reflects a severe problem of expansions around DMFT. In particular in the case of CPA, the expansion around it fails to reproduce the Ward identity that is needed for the existence of the diffusion pole in the density response function and the validity of macroscopic conservation laws. We proposed a solution to this problem in disordered systems which opens a new and more consistent framework to study systematically the vanishing of diffusion and Anderson localization. A full solution of this problem for strongly correlated Fermi liquids is still to be found.

Acknowledgment

I thank Jindřich Kolorenč for his long-term fruitful collaboration on most of the results presented in this review as well as for his help in the preparation of the diagrammatic representation of the perturbation expansion in disordered systems. I acknowledge support from Grant No. 15-14259S of the Czech Science Foundation.

References

[1] M. Lax, Rev. Mod. Phys. **23**, 287 (1951)

[2] M. Lax, Phys. Rev. **85**, 621 (1952)

[3] R.W. Davies and J.S. Langer, Phys. Rev. **131**, 163 (1963)

[4] P. Soven, Phys. Rev. **156**, 809 (1967)

[5] D.W. Taylor, Phys. Rev. **156**, 1017 (1967)

[6] B. Velický, S. Kirkpatrick, and H. Ehrenreich, Phys. Rev. **175**, 747 (1968)

[7] E. Müller-Hartmann, Solid State Commun. **12**, 1269 (1973)

[8] F. Ducastelle, J. Phys. C: Solid State Physics **7**, 1795 (1973)

[9] B. Velický, Phys. Rev. **184**, 614 (1969)

[10] R.J. Elliott, J.A. Krumhansl, and P.L. Leath, Rev. Mod. Phys. **46**, 465 (1974)

[11] V. Čápek, Phys. Status Solidi B **43**, 61 (1971)

[12] R. Mills, Phys. Rev. B **8**, 3650 (1973)

[13] R. Mills and P. Ratanavararaksa, Phys. Rev. B **18**, 5291 (1978)

[14] V. Janiš, Czechoslovak J. Phys. B **36**, 1107 (1986)

[15] V. Janiš, Phys. Rev. B **40**, 11331 (1989)

[16] W. Metzner and D. Vollhardt, Phys. Rev. Lett. **62**, 324 (1989)

[17] R. Vlaming and D. Vollhardt, Phys. Rev. B **45**, 4637 (1992)

[18] V. Janiš, Phys. Rev. B **64**, 115115 (2001)

[19] D. Vollhardt and P. Wölfle, Phys. Rev. B **22**, 4666 (1980)

[20] F. Yonezawa, Prog. Theor. Phys. **40**, 734 (1968)

[21] J. Kolorenč: *Charge transport in strongly disordered solids*
Ph.D. thesis, Charles University in Prague, Prague, Czech Republic (2004)

[22] V. Janiš, Phys. Rev. Lett. **83**, 2781 (1999)

[23] G. Baym, Phys. Rev. **127**, 1391 (1962)

[24] V. Janiš and D. Vollhardt, Int. J. Mod. Phys. B **6**, 731 (1992)

[25] L.M. Falicov and J.C. Kimball, Phys. Rev. Lett. **22**, 997 (1969)

[26] U. Brandt and C. Mielsch, Z. Phys. B Condens. Matter **75**, 365 (1989)

[27] V. Janiš, Z. Phys. B Condens. Matter **83**, 227 (1991)

[28] V. Janiš, Phys. Rev. B **49**, 1612 (1994)

[29] J. Hubbard, Proc. Roy. Soc. London A **281**, 401 (1964)

[30] G. Czycholl, Phys. Rep. **143**, 277 (1986)

[31] A. Kawabata, Prog. Theor. Phys. **48**, 1793 (1972)

[32] V. Janiš and D. Vollhardt, Z. Phys B Condens. Matter **91**, 317 (1993)

[33] V. Janiš and J. Kolorenč, Phys. Rev. B **71**, 245106 (2005)

[34] J. Rammer: *Quantum Transport Theory*, *Frontiers in Physics*, Vol. 99 (Perseus Books, Reading, MA, 1998)

[35] V. Janiš, J. Kolorenč, and V. Špička, Eur. Phys. J. B **35**, 77 (2003)

[36] D. Vollhardt and P. Wölfle: In W. Hanke and Yu.V. Kopaev (Eds.) *Electronic Phase Transitions* (Elsevier, Amsterdam, 1992) chap. 1, pp. 1–78.

[37] D. Belitz and T.R. Kikpatrick, Rev. Mod. Phys. **66**, 261 (1994)

[38] V. Janiš and J. Kolorenč, Phys. Rev. B **71**, 033103 (2005)

[39] V. Pokorný and V. Janiš, J. Phys.: Condens. Matter **25**, 175502 (2013)

[40] V. Janiš and V. Pokorný, Phys. Rev. B **81**, 165103 (2010)

[41] V. Janiš and J. Kolorenč, Mod. Phys. Lett. B **18**, 1051 (2004)

[42] V. Janiš and J. Kolorenč, Eur. Phys. J. B **89**, 170 (2016)

12 Interplay of Kondo Effect and RKKY Interaction

Johann Kroha
Physikalisches Institut, Universität Bonn
Nussallee 12, 53115 Bonn

Contents

1 Introduction and overview **2**

2 Kondo effect and renormalization group **4**
2.1 Pseudo-fermion representation of spin 5
2.2 Perturbation theory 8
2.3 Renormalization group 10

3 RKKY interaction in paramagnetic and half-metals **12**

4 Interplay of Kondo screening and RKKY interaction **15**
4.1 The concept of a selfconsistent renormalization group 16
4.2 Integration of the RG equation 19
4.3 Universal suppression of the Kondo scale 19

5 Conclusion **21**

A f-spin – conduction-electron vertex $\hat{\Gamma}_{cf}$ **22**
A.1 Spin structure 22
A.2 Energy dependence 23

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction and overview

Magnetic interactions in a metal involving localized magnetic moments give rise to a wealth of phenomena, ranging from the Kondo effect to magnetic ordering and quantum phase transitions. We give a brief overview of such phenomena before, in the main part of these lecture notes, we will focus on a detailed description of the interplay of interactions that tend to quench the local moments or that tend order them.

When a magnetic ion is placed in a metallic host, the Kondo effect [1, 2] occurs: Conduction electrons at the Fermi level, i.e., at zero excitation energy, are in resonance with a flip of the two-fold degenerate spin ground state of the magnetic ion. As the temperature T is lowered, the electrons become confined to the Fermi surface, so that more and more electrons contribute to this resonant quantum spin-flip scattering, leading to a diverging spin scattering amplitude. Hence, when the spin exchange coupling J_0 between the localized moments and the itinerant conduction electrons is antiferromagnetic, a many-body spin-singlet state between the impurity spin and the conduction electron spins is formed below a characteristic temperature, the Kondo temperature T_K. This, however, means that electrons that do not contribute to the singlet bound state, experience merely potential scattering rather than spin scattering, i.e., the impurity spin is effectively removed from the system. This effect is called spin screening. The scattering rate and other physical quantities thus settle smoothly to constant values, leading to Fermi-liquid behavior for $T \ll T_K$ [2]. The Kondo temperature is found to be exponentially small in the exchange coupling, $T_K = D_0 \exp[-1/(2N(0)J_0)]$, with the density of states at the Fermi level $N(0)$ and the conduction band width D_0. The entirety of complex phenomena sketched above, involving the increase of the spin scattering amplitude implying anomalous transport properties, followed by spin screening and the formation of a narrow, but smooth resonance of width T_K in the electronic spectrum at the Fermi energy, comprises the Kondo effect.

When there are several or many localized magnetic moments in a metal, for instance arranged on a lattice, the same spin-exchange coupling J_0 that induces the Kondo effect, induces also a magnetic interaction between the localized spins: The local moments can exchange their spins, mediated by two conduction electrons scattering from and traveling between the impurity sites. Since this effective, long-range spin-exchange coupling K involves two elementary scattering events between electron and und impurity spins, it is of order $K \propto N(0)J_0^2$. It can be ferro- or antiferromagnetic due to the long-range, spatial oscillations of the conduction electron density correlations. This conduction-electron-mediated spin interaction was first considered by Ruderman and Kittel [3], Kasuya [4] and Yosida [5] and is therefore called RKKY interaction. The RKKY interaction usually dominates the magnetic dipole-dipole coupling as well as the direct exchange coupling between neighboring local moments because of the short spatial extent of these couplings or the exponentially small overlap of the local moment wave functions on neighboring lattice sites.

In a Kondo lattice, the local Kondo coupling and the RKKY interaction favor different ground states. The Kondo coupling leads to a paramagnetic Fermi liquid state without local moments. In this state, the local orbitals, whose spectrum has a Kondo resonance at the Fermi energy,

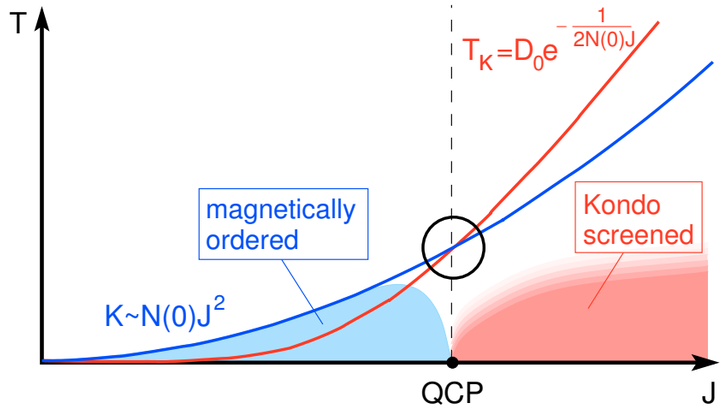


Fig. 1: *Doniach's phenomenological phase diagram for the phase transition between an RKKY-induced, magnetically ordered phase and the Kondo screened, paramagnetic phase. The phase transition occurs when the RKKY coupling K of a local moment to all surrounding moments becomes equal to the Kondo singlet binding energy T_K (black circle). While the RKKY coupling is $K \sim N(0)J_0^2$, the Kondo energy $T_K = D_0 \exp[-1/(2N(0)J_0)]$ is exponentially small in the bare, local spin exchange coupling J. Therefore, the RKKY coupling always dominates for small values of J_0.*

hybridize with each other and eventually become lattice coherent at low temperatures to form Bloch-like quasiparticle states. As a result, a narrow band crossing the Fermi energy is formed. Its bandwidth is controlled by the Kondo resonance width T_K. It thus gives rise to an exponentially strong effective mass enhancement of roughly $m^*/m \approx \exp[1/(2N(0)J_0)]$, which lends the name "heavy Fermi liquid" to this state [6].

By contrast, the RKKY interaction tends to induce magnetic order of the local moments. It was pointed out early on by Doniach [7] that, therefore, the Kondo spin screening of the local moments should eventually break down and give way to magnetic order, when the RKKY coupling energy becomes larger than the characteristic energy scale for Kondo singlet formation, the Kondo temperature T_K, see Fig. 1. Thus, one expects a $T = 0$ quantum phase transition (QPT) to occur [6], with the local spin exchange coupling J_0 as the control parameter. If and how the Kondo breakdown occurs at a magnetic QPT is, however, controversial. In fact, several QPT scenarios in heavy-fermion systems are conceivable.

(1) The heavy Fermi liquid, like any other Fermi liquid, may undergo a spin density-wave (SDW) instability, leading to critical fluctuations of the bosonic magnetic order parameter but leaving the fermionic, heavy quasiparticles intact. This scenario is well described by the pioneering works of Hertz, Moriya, and Millis [8–10].

(2) The local fluctuations of the magnetization, coupling to the nearly localized, heavy quasiparticles, may become critical (divergent) and thereby destroy the heavy Fermi liquid (local quantum criticality) [11, 12].

(3) At the phase transition the Kondo effect and, hence, the heavy-fermion band vanish, which leads to an abrupt change of the Fermi surface (Fermi volume collapse). It has been proposed [13] that the Fermi surface fluctuations associated with this change may self-consistently destroy the Kondo singlet state.
(4) Most recently, a scenario of critical quasiparticles has been put forward, characterized by a diverging effective mass and a singular quasiparticle interaction which is self-consistently generated by the nonlocal order-parameter fluctuations of an impending SDW instability [14–16]. Intriguing in its generality and similar in spirit to Landau's Fermi liquid theory, this scenario does, however, not invoke Kondo physics and, thus, does not address the specific problems associated with the Kondo destruction like Fermi volume collapse or the possibility of small, localized magnetic moments in the magnetically ordered phase.

While the Hertz-Millis-Moriya scenario (1) is described by a critical field theory of the bosonic, magnetic order parameter alone, the complete understanding of the breakdown scenarios (2), (3), and (4) would require a field theory for the fermionic degrees of freedom forming the Kondo effect and the heavy quasiparticles, coupled to the bosonic order parameter field. In lack of such a complete theory, these scenarios presume that specific fluctuations: (2) local fluctuations, (3) Fermi surface fluctuations, or (4) antiferromagnetic fluctuations, become soft for certain values of the system parameters and, thus, dominate the QPT. Therefore, the conditions for these scenarios to be realized are controversial.

In these lecture notes we consider the interplay of Kondo screening and RKKY interaction within the Kondo lattice model. We derive the phenomena of the single-impurity Kondo model in section 2, thereby introducing important concepts and techniques, like the fermionic representation spin, universality, and the analytic (perturbative) renormalization group. Section 3 presents the oscillatory RKKY coupling, calculated as a second-order spin exchange process, mediated by the conduction electrons. In section 4 we show how the Kondo singlet formation as well as the RKKY interaction can be incorporated on the same footing in an analytic renormalization group treatment, leading to a universal Kondo destruction as function of the RKKY coupling parameter. We conclude in section 5 with a discussion how this theory may set the stage for a more complete quantum field theory of heavy-fermion QPTs with Kondo breakdown.

2 Kondo effect and renormalization group

In this section we recollect the essential physics of a single Kondo impurity in a metal and provide the calculational tools for their derivation. We consider the single-impurity Kondo model

$$H = \sum_{\mathbf{k},\sigma} \varepsilon_{\mathbf{k}}\, c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + J_0\, \hat{\mathbf{S}} \cdot \hat{\mathbf{s}} \tag{1}$$

where $c_{\mathbf{k}\sigma}$, $c^\dagger_{\mathbf{k}\sigma}$ denote the conduction (c-) electron operators with momentum $\mathbf{k}$ and dispersion $\varepsilon_{\mathbf{k}}$. $\hat{\mathbf{S}}$ is the impurity spin operator at site $\mathbf{x} = 0$, which is locally coupled to the spins of the

conduction electrons on that site, $\hat{\mathbf{s}}$, via a Heisenberg exchange coupling J_0. We have

$$\hat{\mathbf{s}} = \sum_{\mathbf{k},\mathbf{k}',\sigma,\sigma'} c^\dagger_{\mathbf{k}\sigma}\, \boldsymbol{\sigma}_{\sigma\sigma'}\, c_{\mathbf{k}'\sigma'}, \tag{2}$$

with $\boldsymbol{\sigma} = (\sigma_x,\, \sigma_y,\, \sigma_z)^T$ the vector of Pauli matrices

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \qquad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3}$$

In Eq. (2) the conduction spin eigenvalue $1/2$ has been absorbed in the coupling constant J_0, by convention, and we use units $\hbar = 1$ throughout. The local spins $\hat{\mathbf{S}}$ will henceforth be termed f-spins, as they are typically realized in heavy fermion systems by the rare-earth $4f$ electrons.

2.1 Pseudo-fermion representation of spin

A field theoretical treatment, like the standard functional integral or Wick's theorem and many-body perturbation theory, requires that the corresponding field operators obey canonical commutation rules, i.e., their (anti)commutators must be proportional to the unit operator. However, the spin operators $\hat{\mathbf{S}}$ obey the SU(2) algebra. In order to overcome this difficulty, we use the fermionic representation of spin, first introduced by Abrikosov [17]. For each of the basis states spanning the impurity spin Hilbert space, $|\sigma\rangle$, $\sigma =\uparrow, \downarrow$, fermionic creation and destruction operators $f^\dagger_\sigma$, f_σ are introduced according to $|\sigma\rangle = f^\dagger_\sigma|vac\rangle$, where $|vac\rangle$ denotes the vacuum state (no impurity spin present). The impurity spin operator $\mathbf{S}$ then reads,

$$\hat{\mathbf{S}} = \frac{1}{2}\sum_{\tau,\tau'} f^\dagger_\tau\, \boldsymbol{\sigma}_{\tau\tau'}\, f_{\tau'}. \tag{4}$$

That is, the operator on the right-hand side and $\hat{\mathbf{S}}$ have identical matrix elements in the physical spin Hilbert space. However, repeated action of the fermionic operators would permit unphysical double occupancy or no occupancy of the spin states $|\uparrow\rangle, |\downarrow\rangle$. The dynamics are restricted to the physical spin space by imposing the operator constraint

$$\hat{Q} = \sum_\sigma f^\dagger_{i\tau} f_{i\tau} = \mathbb{1}. \tag{5}$$

Eqs. (4), (5) constitute the exact pseudo-fermion representation of the spin $s = 1/2$.
The impurity-spin operator and, hence, the equation of motion with the Hamiltonian (1) are symmetric under the local U(1) gauge transformation

$$f_\tau \to \mathrm{e}^{-i\phi(t)} f_\tau, \qquad i\frac{d}{dt} \to i\frac{d}{dt} - \frac{\partial\phi(t)}{\partial t}, \tag{6}$$

with an arbitrary, time-dependent phase $\phi(t)$. It is closely related to the conservation of the pseudo-fermion number $\hat{Q}$.

Projection onto the physical Hilbert space. The exact projection of the dynamics onto the physical sector of Fock space with $Q = 1$, is performed by the following procedure. Consider first the grand canonical ensemble with respect to Q, defined by the statistical operator

$$\hat{\rho}_G = \frac{1}{Z_G} \mathrm{e}^{-\beta(H+\lambda Q)}, \tag{7}$$

where $Z_G = \mathrm{tr}[\exp\{-\beta(\hat{H} + \lambda\hat{Q})\}]$ is the grand canonical partition function, $-\lambda$ the associated chemical potential, and $\beta = 1/k_B T$ the inverse temperature. The trace extends over the complete Fock space, including summation over $Q = 0, 1, 2$. The grand canonical expectation value of an observable $\hat{A}$ acting on the impurity spin space is defined as

$$\langle \hat{A} \rangle_G(\lambda) = \mathrm{tr}[\hat{\rho}_G \hat{A}]. \tag{8}$$

The physical expectation value of $\hat{A}$, $\langle \hat{A} \rangle$, must be evaluated in the canonical ensemble with fixed $Q = 1$. It can be obtained from the grand canonical expectation value as [17],

$$\langle \hat{A} \rangle := \frac{\mathrm{tr}_{Q=1}\left[\hat{A} e^{-\beta \hat{H}}\right]}{\mathrm{tr}_{Q=1}\left[e^{-\beta \hat{H}}\right]} = \lim_{\lambda\to\infty} \frac{\mathrm{tr}\left[\hat{A} e^{-\beta[\hat{H}+\lambda(\hat{Q}-1)]}\right]}{\mathrm{tr}\left[\hat{Q} e^{-\beta[\hat{H}+\lambda(\hat{Q}-1)]}\right]} = \lim_{\lambda\to\infty} \frac{\langle \hat{A} \rangle_G(\lambda)}{\langle \hat{Q} \rangle_G(\lambda)} \tag{9}$$

Here, all terms of the grand canonical traces in the numerator and in the denominator with $Q > 1$ are projected away by the limit $\lambda \to \infty$. In the denominator, the operator $\hat{Q}$ makes all terms with $Q = 0$ vanish. In the numerator, the observable $\hat{A}$ acts on the impurity spin space and hence is a power of $\hat{S}$, Eq. (4), which vanishes in the $Q = 0$ subspace. Therefore, in the numerator and in the denominator precisely the canonical traces over the physical sector $Q = 1$ remain, as required. It follows that any impurity-spin correlation function can be evaluated as a pseudo-fermion correlation function in the unrestricted Fock space, where Wick's theorem and the decomposition in terms of Feynman diagrams with pseudo-fermion propagators are valid, and taking the limit $\lambda \to \infty$ at the end of the calculation. Note that for the c electron spin, Eq. (2), the $Q = 1$ projection is not needed, because for the noninteracting c-electrons doubly occupied or empty states are allowed.

Diagrammatic rules. We will now show that the limit $\lambda \to \infty$ translates into simple diagrammatic rules for the evaluation of impurity Green and correlation functions. We denote the local c electron Green function at the impurity site by $G_{c\sigma}(i\omega_n)$ and the bare, grand canonical pseudo-fermion Green function by $G^G_{f\sigma}(i\omega_n)$

$$G_{c\sigma}(i\omega_n) = \sum_{\mathbf{k}} \frac{1}{i\omega_n - \varepsilon_{\mathbf{k}}} \tag{10}$$

$$G^G_{f\sigma}(i\omega_n) = \frac{1}{i\omega_n - \lambda}, \tag{11}$$

with the fermionic Matsubara frequencies $\omega_n = \frac{\pi}{\beta}(2n + 1)$. Consider first $\lim_{\lambda\to 0}\langle \hat{Q} \rangle_G(\lambda)$. Using standard, complex contour integration, we obtain

$$\begin{aligned} \langle \hat{Q} \rangle_G(\lambda) &= \sum_\sigma \frac{1}{\beta} \sum_{\omega_n} G^G_{f\sigma}(i\omega_n) = -\sum_\sigma \oint \frac{dz}{2\pi i} f(z)\, G^G_{f\sigma}(z) \\ &= -\sum_\sigma \int_{-\infty}^{+\infty} \frac{d\varepsilon}{2\pi i} f(\varepsilon) \left[G^G_{f\sigma}(\varepsilon + i0) - G^G_{f\sigma}(\varepsilon - i0)\right], \end{aligned} \tag{12}$$

where $f(\varepsilon) = 1/(e^{\beta\varepsilon}+1)$ is the Fermi function, and the ε-integral extends along the branch cut of $G^G_{f\sigma}(z)$ at the real frequency axis, $\mathrm{Im}\, z = 0$. We can now perform a specific gauge transformation of the operators, $f_\tau \to \mathrm{e}^{-i\lambda t} f_\tau$. It implies, by virtue of Eq. (6), a shift of all pseudo-fermion energies in a diagram by $\varepsilon \to \varepsilon + \lambda$. It eliminates λ from the pseudo-fermion propagator and casts it into the argument of the Fermi function. Thus, we have

$$\begin{aligned} \langle \hat{Q} \rangle_G(\lambda) &= -\sum_\sigma \int_{-\infty}^{+\infty} \frac{d\varepsilon}{\pi}\, f(\varepsilon+\lambda)\, \mathrm{Im}\, G_{f\sigma}(\varepsilon+i0) \\ &\stackrel{\lambda\to\infty}{\longrightarrow} \mathrm{e}^{-\beta\lambda} \sum_\sigma \int_{-\infty}^{+\infty} \frac{d\varepsilon}{\pi}\, \mathrm{e}^{-\beta\varepsilon}\, \mathrm{Im}\, G_{f\sigma}(\varepsilon+i0), \end{aligned} \tag{13}$$

where $G_{f\sigma}(\varepsilon+i0) \equiv G^G_{f\sigma}(\varepsilon+\lambda+i0) = 1/(\varepsilon+i0)$ is independent of λ.

The result Eq. (13) can be generalized by explicit calculation to arbitrary Feynman diagrams involving f- and c-Green functions: (i) Each complex contour integral includes one distribution function $f(z)$. The integral can be written as the sum of integrals along the branch cuts at the real energy axis of all propagators appearing in the diagram. (ii) Consider now one term of this sum. The argument of the distrubution function $f(\varepsilon)$ in that term is real and always equal to the argument ε of that propagator G along whose branch cut the integration extends. (iii) The above energy-shift gauge transformation applies to all pseudo-fermion energies ω in the diagram and, thus, cancels the parameter λ in all pseudo-fermion propagators, $G^G_{f\sigma}(\omega) \to G_{f\sigma}(\omega)$. (iv) If in the considered term the integral is along a pseudo-fermion branch cut, this gauge transformation also shifts the argument of the distribution function, $f(\varepsilon) \to f(\varepsilon+\lambda)$, by virtue of (ii), i.e., the pseudo-fermion branch cut integral vanishes $\sim \mathrm{e}^{-\beta\lambda}$, as in Eq. (13). If the integral is along a c-electron branch cut, the argument of $f(\varepsilon)$ is not affected by the gauge transformation, and the integral does not vanish.

This derivation can be summarized in the following diagrammatic rules for $(Q=1)$-projected expectation values:

(1) In a diagrammatic part that consists of a product of c- and f-Greens's functions, only the integrals along the c-electron branch cuts contribute.
(2) A closed pseudo-fermion loop contains only pseudo-fermion branch cut integrals and thus carries a factor $\mathrm{e}^{-\beta\lambda}$.
(3) Each diagram contributing to the projected expectation value of an impurity spin observable, $\langle \hat{A} \rangle$, contains exactly one closed pseudo-fermion loop per impurity site, because the factor of $\mathrm{e}^{-\beta\lambda}$ cancels in the numerator and denominator of Eq. (9), and higher order loops vanish by virtue of rule (2).

We note in passing that the pseudo-fermion representation can be generalized in a straightforward way to higher local spins than $S = 1/2$ by choosing in Eq. (4) a respective higher-dimensional representation of the spin matrices and defining the constraint $\hat{Q} = \mathbb{1}$ as before, with a summation over all possible spin orientations τ. It can also be extended to include local charge fluctuations by means of the slave boson representation [18–21].

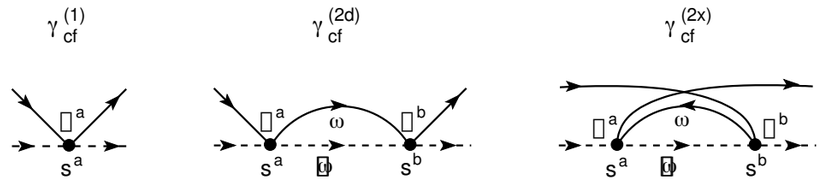


Fig. 2: *Conduction electron-impurity spin vertex $\hat{\gamma}_{cf}$ of the single-impurity Kondo model up to 2nd order in the spin exchange coupling J_0. Conduction electron propagators are denoted by solid, pseudo-fermion propagators by dashed lines. $\hat{\gamma}^{2,d}$ and $\hat{\gamma}^{2,x}$ represent the 2nd-order direct and exchange terms, respectively. The external lines are drawn for clarity and are not part of the vertex.*

2.2 Perturbation theory

It is instructive to analyze the scattering of a conduction electron from a spin impurity in perturbation theory, because this will visualize the physical origin of its singular behavior. The perturbation theory can be efficiently evaluated with the formalism developed in section 2.1.
With the Kondo Hamiltonian Eq. (1) the conduction-electron–impurity-spin vertex $\hat{\gamma}_{cf}$ can be read off from the diagrams in Fig. 2. Denoting the vector of Pauli matrices acting in c-electron spin space by $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)^T$ and the vector of Pauli matrices in f-spin space by $\mathrm{s} = (s^x, s^y, s^z)^T$, $\hat{\gamma}_{cf}$ reads in first and second order of J_0

$$\hat{\gamma}_{cf}^{(1)} = \frac{1}{2} J_0 \, (\mathbf{s} \cdot \boldsymbol{\sigma}) \tag{14}$$

$$\hat{\gamma}_{cf}^{(2,d)} = -\frac{1}{4} J_0^2 \sum_{a,b=x,y,z} \left(s^b \sigma^b\right) \left(s^a \sigma^a\right) \; \frac{1}{\beta} \sum_{\omega_n} G_c(i\omega_n) G_f(i\Omega_m - i\omega_n)|_{\lambda\to\infty} \tag{15}$$

$$\hat{\gamma}_{cf}^{(2,x)} = +\frac{1}{4} J_0^2 \sum_{a,b=x,y,z} \left(s^b \sigma^a\right) \left(s^a \sigma^b\right) \; \frac{1}{\beta} \sum_{\omega_n} G_c(i\omega_n) G_f(i\Omega_m + i\omega_n)|_{\lambda\to\infty}, \tag{16}$$

where matrix multiplications in the f- and c-spin spaces are implied, and the sum $\sum_{a=x,y,z}$ represents the scalar product in position space. The relative minus sign between $\hat{\gamma}_{cf}^{(2,d)}$ and $\hat{\gamma}_{cf}^{(2,x)}$ arises because of the extra fermion loop in the exchange term $\hat{\gamma}_{cf}^{(2,x)}$. Note that the order of the Pauli matrices in Eqs. (15), (16) is crucial. It is determined by their order along the c-electron or f-particle lines running through the diagram. Thus, in $\hat{\gamma}_{cf}^{(2,x)}$ the order of c-electron Pauli matrices is reversed with respect to $\hat{\gamma}_{cf}^{(2,d)}$.
The spin-dependent part of $\hat{\gamma}_{cf}^{(2,d)}$, $\hat{\gamma}_{cf}^{(2,x)}$ can be evaluated using the SU(2) spin algebra, $\sigma^a\sigma^b = \sum_{c=x,y,z} i\varepsilon^{abc}\sigma^c + \delta^{ab}\mathbb{1}$ for a, $b = x$, y, z, where $\mathbb{1}$ is the unit operator in spin space, ε^{abc} the totally antisymmetric unit tensor and δ^{ab} the Kronecker-δ

$$\mathrm{d}: \quad \sum_{a,b=x,y,z} s^b s^a \otimes \sigma^b \sigma^a = -2\,\mathbf{s}\cdot\boldsymbol{\sigma} + 3\,\mathbb{1}\otimes\mathbb{1} \tag{17}$$

$$\mathrm{x}: \quad \sum_{a,b=x,y,z} s^b s^a \otimes \sigma^a \sigma^b = \quad 2\,\mathbf{s}\cdot\boldsymbol{\sigma} + 3\,\mathbb{1}\otimes\mathbb{1} \tag{18}$$

For scattering at the Fermi energy ($\Omega = 0$), the energy-dependent factors in Eqs. (15), (16) are

$$\begin{aligned} \mathrm{d}: \qquad \frac{1}{\beta}\sum_{\omega_n} G_c(i\omega_n)G_f^G(-i\omega_n)|_{\lambda\to\infty} &= \oint \frac{dz}{2\pi i}[1-f(z)]G_c(z)G_f^G(-z)|_{\lambda\to\infty} \\ &= N(0)\int_{-D_0}^{D_0} d\varepsilon\, \frac{1-f(\varepsilon)}{\varepsilon} \qquad (19) \\ \mathrm{x}: \qquad \frac{1}{\beta}\sum_{\omega_n} G_c(i\omega_n)G_f^G(i\omega_n)|_{\lambda\to\infty} &= -\oint \frac{dz}{2\pi i} f(z)G_c(z)G_f^G(z)|_{\lambda\to\infty} \\ &= -N(0)\int_{-D_0}^{D_0} d\varepsilon\, \frac{f(\varepsilon)}{\varepsilon}, \qquad (20) \end{aligned}$$

where we have assumed the Fermi energy in the center of the band of half bandwidth D_0, with a flat conduction electron density of states $N(0) = \mathrm{Im}\, G_c(0-i0)/\pi$. We see (cf. Fig. 2) that in the direct term (d) the intermediate electron must scatter into an unoccupied state, $1-f(\varepsilon)$, while in the exchange term (x) the intermediate electron comes from an occupied state, $f(\varepsilon)$ and then leaves the impurity. Collecting all terms, we obtain $\hat{\gamma}_{cf} = \hat{\gamma}_{cf}^{(1)} + \hat{\gamma}_{cf}^{(2d)} + \hat{\gamma}_{cf}^{(2x)}$ as

$$\begin{aligned} \hat{\gamma}_{cf} &= \frac{1}{2}J_0\,(\mathbf{s}\cdot\boldsymbol{\sigma})\,\left[1 + N(0)J_0\int_{-D_0}^{D_0} d\varepsilon\,\frac{1-2f(\varepsilon)}{\varepsilon} + \mathcal{O}(J_0^2)\right] \\ &\approx \frac{1}{2}J_0\,(\mathbf{s}\cdot\boldsymbol{\sigma})\,\left[1 + 2N(0)J_0\ln\left(\frac{D_0}{T}\right) + \mathcal{O}(J_0^2)\right] \qquad (21) \end{aligned}$$

The calculation clearly shows the physical origin of the logarithmic behavior: the presence of a sharp Fermi edge in the phase space available for scattering, i.e., in the integrals of Eqs. (19), (20), and quantum spin-flip scattering with the nontrivial SU(2) algebra. If the reversed order of Pauli matrices in the exchange term would not introduce a minus sign in the spin channel, Eq. (18), the logarithmic terms would cancel, like in the potential scattering channel, instead of adding up. It is also important that the impurity is localized, because otherwise an integral over the exchanged momentum (recoil) would smear the logarithmic singularity.

Eq. (21) exhibits a logarithmic divergence for low temperatures T. It signals a breakdown of perturbation theory when the 2nd-order contribution to $\hat{\gamma}_{cf}$ becomes equal to the 1st-order contribution. This happens at a characteristic temperature scale, which can be read off from Eq. (21), the Kondo temperature

$$T_K = D_0\, \mathrm{e}^{-1/(2N(0)J_0)}. \qquad (22)$$

Below T_K perturbative calculations about the weak-coupling state break down. To describe the complex physics outlined in the introduction, more sophisticated techniques, predominantly numerical or exact solution methods, are required. The logarithmic behavior of the perturbation expansion, however, sets the stage for the development of the renormalization group method, to be developed in the next section, and which is particularly useful for analytically studying the interplay of Kondo screening and RKKY interaction.

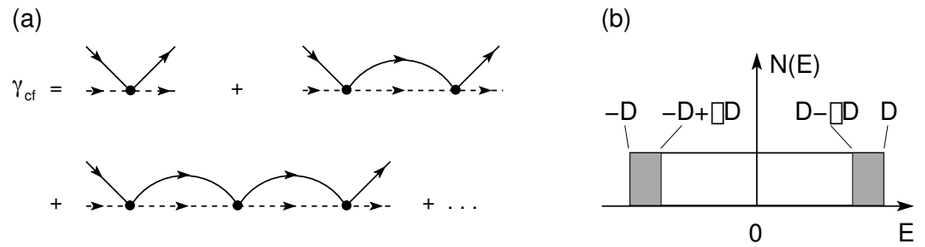


Fig. 3: *Universality and perturbative renormalization group. (a) T-matrix resummation of the c–f vertex. The sum contains, for each conduction electron-pseudo-fermion bubble (direct diagram) shown, the exchange diagram, which is not shown for clarity. (b) Scheme for the cutoff reduction $D \to D - \delta D$.*

2.3 Renormalization group

Since the logarithm is a scale invariant function, there is the possibility that the resummation of a logarithmic perturbation expansion leads to universal behavior in the sense that variables like energy ω, temperature T, etc. can be expressed in units of a single scale, T_K, in such a way that all physical quantities are functions of the dimensionless variables, ω/T_K, T/T_K, etc. only and do not explicitly depend on the microscopic parameters of the Hamiltonian, like J_0, D_0, and $N(0)$. For the Kondo model, this extremely remarkable property can be visualized by a T-matrix-like, partial resummation of the c–f vertex, as sketched in Fig. 3(a). The resummation results in a geometric series for the total c–f vertex or the effective coupling constant $\tilde{J}$

$$N(0)\tilde{J} = 2N(0)J_0\left[1 + 2N(0)J_0 \ln\left(\frac{D_0}{T}\right) + \left(2N(0)J_0 \ln\left(\frac{D_0}{T}\right)\right)^2 + \dots\right] \quad (23)$$

$$= \frac{2N(0)J_0}{1 - 2N(0)J_0 \ln\left(\frac{D_0}{T}\right)} = \frac{1}{\ln\left(\frac{T}{T_K}\right)}, \quad (24)$$

which converges for $T > T_K$. It is seen that, as a consequence of the logarithmic behavior, in the last expression the microscopic parameters J_0, D_0, and $N(0)$ indeed conspire to form the Kondo temperature T_K of Eq. (22) as the only scale in the problem. This universal behavior is inherited by physical quantities, like relaxation rates, transport properties, etc., since they can be expressed in terms of the total c–f vertex. Although the above is only a heuristic argument and other contributions, not contained in the partial summation, could break the universality, it has been shown independently by the Bethe ansatz solution [22] and by numerical renormalization group (NRG) (for a recent review see [23]) that universality in the above sense indeed holds for the Kondo problem.

Universality is the starting point for the renormalization group method whose essence we discuss next. Let all physical quantities $A_n = h_n(\omega/T^*, T/T^*)$ of a system depend on energy ω and temperature T in a universal way, with universal functions h_n and some (yet unknown) characteristic scale T^*, which depends on the microscopic parameters of the Hamiltonian, J_0, D_0, $N(0)$. The fact that the A_n depend on these parameters only implicitly through T^*

implies that different values of this parameter set realize the same physical system (defined by its observables A_n) if only the different parameter set values lead to the same scale T^*. In particular, systems with low and with high values of the conduction bandwidth or cutoff D_0 must be equivalent if the coupling constant J_0 is adjusted appropriately. In the Kondo problem we are mostly interested in the low-energy behavior, where the perturbation theory fails. This regime corresponds to electrons with a low bandwidth, scattering near the Fermi energy. By virtue of the above argument, this low-energy regime is connected with the high bandwidth regime, where perturbative calculations are possible. In the renormalization group method, this relation is established iteratively. Starting from an initial high-energy cutoff D_0, the cutoff is stepwise reduced to low energies, calculating at each step how the coupling constant J of the Hamiltonian must be changed, such that the physical observables A_n remain constant, see Fig. 3(b). This defines a running cutoff D with initial value D_0 and a "renormalized" or "running" coupling constant $J(D)$ with initial value J_0. The running coupling constant, as part of the Hamiltonian, defines a change of the Hamiltonian itself. More generally, the cutoff reduction may even generate new types of interaction operators in the Hamiltonian, implied by the requirement that physical observables be invariant. The repeated operations on the Hamiltonian, defined in this way, form a semigroup (without existence of the inverse operation), the renormalization group (RG). The change of the Hamiltonian by the successive cutoff reduction is called renormalization group flow.

We can now perform the renormalization of the Kondo Hamiltonian (or coupling constant J) explicitly in a perturbative way, following Anderson [24]. To that end, it is convenient to introduce the dimensionless, bare coupling $g_0 = N(0)J_0$ and running coupling $g = N(0)J$. We also define the projector $P_{\delta D}$ of the conduction electron energy onto the intervals $[-D, -D+\delta D] \cup [D-\delta D, D]$ by which the conduction band is reduced in one RG step as well as the projector $(1-P_{\delta D})$ onto the remaining conduction energy interval, cf. Fig. 3(b). To impose the invariance of physical quantities under the RG flow, it is sufficient to keep the total conduction-electron–pseudo-fermion vertex $\hat{\gamma}_{cf}$ invariant, since all physical quantities are derived from it within the Kondo model. $\hat{\gamma}_{cf}$ is defined by the following T-matrix equation

$$\hat{\gamma}_{cf} = \hat{\gamma}_{cf}^{(1)} + \hat{\gamma}_{cf}^{(1)}\, G\, \hat{\gamma}_{cf}. \tag{25}$$

Here, the bare vertex $\hat{\gamma}_{cf}^{(1)}$ is defined as in Eq. (14), G denotes schematically the product of G_c and G_f propagators connecting two bare vertices $\hat{\gamma}_{cf}^{(1)}$ in the direct and exchange diagrams (cf. Fig. 2), and integration over the conduction electron energy in G is implied. Eq. (25) can be rewritten as

$$\begin{aligned}
\hat{\gamma}_{cf} &= \hat{\gamma}_{cf}^{(1)} + \hat{\gamma}_{cf}^{(1)}\,[P_{\delta D}G]\,\hat{\gamma}_{cf} + \hat{\gamma}_{cf}^{(1)}\,[(1-P_{\delta D})G]\,\hat{\gamma}_{cf} \qquad (26)\\
&= \hat{\gamma}_{cf}^{(1)} + \hat{\gamma}_{cf}^{(1)}\,[P_{\delta D}G]\,\left\{\hat{\gamma}_{cf}^{(1)} + \hat{\gamma}_{cf}^{(1)}\,[(P_{\delta D}+(1-P_{\delta D}))\,G]\,\hat{\gamma}_{cf}\right\} + \hat{\gamma}_{cf}^{(1)}\,[(1-P_{\delta D})G]\,\hat{\gamma}_{cf}\\
&= \hat{\gamma}_{cf}^{(1)\prime} + \hat{\gamma}_{cf}^{(1)\prime}\,[(1-P_{\delta D})G]\,\hat{\gamma}_{cf} \quad + \quad \mathcal{O}(P_{\delta D}^2),
\end{aligned}$$

with

$$\hat{\gamma}_{cf}^{(1)\prime} = \hat{\gamma}_{cf}^{(1)} + \hat{\gamma}_{cf}^{(1)}\,[P_{\delta D}G]\,\hat{\gamma}_{cf}^{(1)} =: \hat{\gamma}_{cf}^{(1)} + \delta\hat{\gamma}_{cf}^{(1)}. \tag{27}$$

In the first line of Eq. (26), the integral over the intermediate conduction electron energy has been split into the infinitesimal high-energy part $P_{\delta D}$ and the remaining part $(1-P_{\delta D})$. In the second line, the high-energy part of the equation has been iterated once, and in the third line, only terms up to linear order in $P_{\delta D}$ have been retained, and all terms have been appropriately rearranged. As seen from the third line, the total vertex $\hat{\gamma}_{cf}$ obeys again a T-matrix equation, however with a reduced conduction bandwidth, $(1-P_{\delta D})$. Moreover, $\hat{\gamma}_{cf}$ remains invariant by this procedure, exactly if the bare vertex is changed to $\hat{\gamma}^{(1)}_{cf}$ as defined in Eq. (27). This is the vertex renormalization we are seeking. Note that this expression is a perturbative, because in Eq. (26) we have iterated the T-matrix equation only once (1-loop approximation). Higher-order iterations, leading to higher-order renormalizations in $\hat{\gamma}^{(1)}_{cf}$ are possible. Note that the vertex renormalization $\delta\hat{\gamma}^{(1)}_{cf}$ in Eq. (27) corresponds just to the 2nd-order perturbation theory expression calculated in Eq. (21), see also Fig. 2. Thus, one can read off from these equations the renormalization of the dimensionless coupling constant g under cutoff reduction $-\delta D$ as

$$dg = -\frac{d}{dD}\left[g^2\int_{-D}^{D} d\varepsilon\,\frac{1-2f(\varepsilon)}{\varepsilon}\right]\delta D = -\frac{2g^2}{D}\,\delta D. \tag{28}$$

Usually one takes the logarithmic derivative which ensures that the differential range $-\delta D$ by which the cutoff is reduced is proportional to the cutoff itself: $\delta D = D d(\ln D)$. Thus

$$\frac{dg}{d\ln D} = -2g^2. \tag{29}$$

This is the differential renormalization group equation (of 1-loop order). The function on the right-hand side, $\beta(g) = -2g^2$, which controls the running coupling constant renormalization, is called the β-function of the RG. The RG equation can be integrated in a straightforward way with the initial condition $g(D_0) = g_0$ to give

$$g(D) = \frac{g_0}{1 - 2g_0\ln(D/D_0)}\,. \tag{30}$$

It is seen that this solution becomes again divergent for antiferromagnetic $g_0 > 0$ when the running cutoff reaches the Kondo scale, $D \to T_K = D_0\exp\left[-1/(2g_0)\right]$, a consequence of the perturbative RG treatment above. However, this divergence allows the conclusion that the ground state of the single-impurity Kondo model is a spin-singlet state between the impurity spin and the spin cloud of the surrounding conduction electron spins as outlined in the introduction. Moreover, it allows for a more general definition of the Kondo spin screening scale T_K, namely the value of the running cutoff D where the coupling constant diverges and the singlet starts to be formed. This will be used for the analysis of the Kondo-RKKY interplay in section 4.

3 RKKY interaction in paramagnetic and half-metals

In this section we derive the expressions for the RKKY interaction. To be general, we will allow for an arbitrary spin polarization of the conduction band and then specialize for the paramagnetic case (vanishing magnetization) and the half-metallic case (complete magnetization). Thus,

we consider now the Kondo lattice Hamiltonian of localized spins $\hat{\mathbf{S}}_i$ at the lattice positions $\mathbf{r}_i$

$$H = \sum_{\mathbf{k},\sigma} \varepsilon_{\mathbf{k}}\; c^{\dagger}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + J_0 \sum_i \hat{\mathbf{S}}_i \cdot \hat{\mathbf{s}}_i \,. \tag{31}$$

Usually, the static limit is considered in order to derive a Hamiltonian coupling operator. We will later consider the question of dynamical correlations as well, as it arises in the interplay with the Kondo effect. The interaction Hamiltonian for the conduction electrons and the localized f-spin $\mathbf{S}_j$ at a site $j \neq i$,

$$H_j^{(cf)} = J_0 \hat{\mathbf{S}}_j \cdot \hat{\mathbf{s}}_j \;, \tag{32}$$

acts as a perturbation for the localized f-spin at a site i (and vice versa). Performing standard thermal perturbation theory by expanding the time evolution operator in the interaction picture, $\hat{T}\exp[-\int_0^\beta d\tau H_j^{cf}(\tau)]$ up to linear order in J_0, one obtains for the interaction operator of the f-spin at site i up to $\mathcal{O}(J_0^2)$

$$H_{ij}^{(2)} = J_0\, \hat{\mathbf{S}}_i \cdot \hat{\mathbf{s}}_i \;-\; J_0^2\; \langle (\hat{\mathbf{S}}_i \cdot \hat{\mathbf{s}}_i)(\hat{\mathbf{S}}_j \cdot \hat{\mathbf{s}}_j) \rangle_c \Big|_{\omega=0} \;. \tag{33}$$

Here, $\langle (\dots) \rangle_c := \mathrm{tr}_c\{\mathrm{e}^{-\beta H}(\dots)\}/Z_G$, denotes the thermal trace over the conduction electron Hilbert space, and the static limit, $\omega = 0$, has been taken. Using Wick's theorem with respect to the conduction electron operators, the second term in Eq. (33) can be written as

$$H_{ij}^{RKKY} = -\frac{J_0^2}{4} \sum_{\alpha,\beta=x,y,z} \sum_{\sigma\sigma'} \hat{S}_i^{\alpha}\, \sigma^{\alpha}_{\sigma\sigma'} \sigma^{\beta}_{\sigma'\sigma}\, \hat{S}_j^{\beta}\, \Pi_{ij}^{\sigma\sigma'}(0), \tag{34}$$

where $\hat{S}_i^{\alpha}$, $\alpha = x,\, y,\, z$, are the components of the impurity spin, σ^{α} the Pauli matrices, and $\Pi_{ij}^{\sigma\sigma'}$ the conduction electron density propagator between the sites i and j as depicted diagrammatically in Fig. 4(a). It has the general form,

$$\Pi_{ij}^{\sigma\sigma'}(i\omega) = -\frac{1}{\beta} \sum_{\varepsilon_n} G_{ji\,\sigma}(i\varepsilon_n + i\omega) G_{ij\,\sigma'}(i\varepsilon_n) \,. \tag{35}$$

In the static limit it reads

$$\Pi_{ij}^{\sigma\sigma'}(0) = -\int d\varepsilon\; f(\varepsilon) \left[A_{ij\,\sigma}(\varepsilon) \mathrm{Re} G_{ij\,\sigma'}(\varepsilon) + A_{ij\,\sigma'}(\varepsilon) \mathrm{Re} G_{ij\,\sigma}(\varepsilon) \right] \,,$$

where $A_{ij\,\sigma}(\varepsilon) = -\mathrm{Im} G_{ij\,\sigma}(\varepsilon + i0)/\pi$. Performing the spin contractions in Eq. (34) and defining the longitudinal and the transverse polarization functions, respectively, as

$$\Pi_{ij}^{\parallel}(0) \;=\; \frac{1}{2} \sum_{\sigma} \Pi_{ij}^{\sigma\sigma}(0) \;= -\sum_{\sigma} \int d\varepsilon\; f(\varepsilon)\, A_{ij\,\sigma}(\varepsilon)\, \mathrm{Re}\, G_{ij\,\sigma}(\varepsilon) \tag{36}$$

$$\Pi_{ij}^{\perp}(0) \;=\; \frac{1}{2} \sum_{\sigma} \Pi_{ij}^{\sigma\,-\sigma}(0) = -\sum_{\sigma} \int d\varepsilon\; f(\varepsilon)\, A_{ij\,\sigma}(\varepsilon)\, \mathrm{Re}\, G_{ij\,-\sigma}(\varepsilon) \,, \tag{37}$$

one obtains the RKKY interaction Hamiltonian,

$$H^{RKKY} = \sum_{i \neq j} H_{ij}^{RKKY} = -\sum_{i,j} \left[K_{ij}^{\parallel}\, \hat{S}_i^z \hat{S}_j^z - K_{ij}^{\perp} \left(\hat{S}_i^x \hat{S}_j^x + \hat{S}_i^y \hat{S}_j^y \right) \right]$$

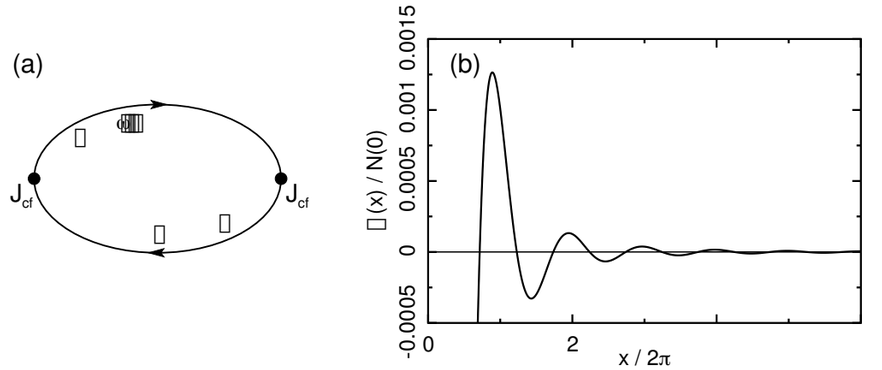


Fig. 4: *(a) Diagram for the spin-dependent conduction electron polarization function $\Pi_{ij}^{\sigma\sigma'}(\omega)$, generating the RKKY interaction. The solid lines represent conduction electron propagators. (b) Oscillatory behavior of $\Pi_{ij}^{\sigma\sigma'}(0)$ in a paramagnetic metal with isotropic dispersion as a function of distance $x = 2k_F|\mathbf{r}_i - \mathbf{r}_j|$, Eq. (45)*

where the sums run over all (arbitrarily distant) lattice sites $i,\ j, i \neq j$ of localized spins $\hat{\mathbf{S}}_i$ and $\hat{\mathbf{S}}_j$, and

$$K_{ij}^{||} = \frac{1}{2} J_0^2 \Pi_{ij}^{||}(0)\ , \qquad K_{ij}^{\perp} = \frac{1}{2} J_0^2 \Pi_{ij}^{\perp}(0)\ , \tag{38}$$

are the longitudinal and transverse RKKY couplings, respectively. The spin being a vector operator, the interaction Hamiltonian H_{ij}^{RKKY} has a tensor structure and is, in general, anisotropic for a magnetized conduction band, as seen from Eq. (38).

We now present explicitly the expressions for the special cases of a paramagnet and of a half-metal. For a paramagnetic conduction band we have $G_{ij\,\sigma} = G_{ij,-\sigma}$, independent of spin. Hence, the RKKY coupling is isotropic, and we have the paramagnetic RKKY Hamiltonian,

$$H_{PM}^{RKKY} = -\sum_{(i,j)} K_{ij}^{PM}\, \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_j\ , \tag{39}$$

with

$$K_{ij}^{PM} = -\frac{J_0^2}{2} \sum_{\sigma} \int d\varepsilon\, f(\varepsilon)\, A_{ij\,\sigma}(\varepsilon)\, \mathrm{Re}\, G_{ij\,\sigma}(\varepsilon)\ . \tag{40}$$

For a half-metal, i.e., for a completely spin-magnetized conduction band with majority spin $\sigma =\uparrow$ we have $A_{ij\downarrow}(\varepsilon) = 0$, Thus, the half-metallic RKKY Hamiltonian reads

$$H_{FM}^{RKKY} = -\sum_{(i,j)} \left[K_{ij}^{FM\,||}\, \hat{S}_i^z \hat{S}_j^z + K_{ij}^{FM\,\perp} \left(\hat{S}_i^x \hat{S}_j^x + \hat{S}_i^y \hat{S}_j^y \right) \right]\ , \tag{41}$$

with

$$K_{ij}^{FM\,||} = -\frac{J_0^2}{2} \int d\varepsilon\, f(\varepsilon)\, A_{ij\uparrow}(\varepsilon)\ \mathrm{Re}\, G_{ij\uparrow}(\varepsilon) \tag{42}$$

$$K_{ij}^{FM\,\perp} = -\frac{J_0^2}{2} \int d\varepsilon\, f(\varepsilon)\, A_{ij\uparrow}(\varepsilon)\ \mathrm{Re}\, G_{ij\downarrow}(\varepsilon)\ . \tag{43}$$

The missing spin summation in Eqs. (42) and (43) as compared to Eq. (40) indicates that in the completely magnetized band only the majority spin species contributes to the coupling. Note, however, that the transverse coupling $K_{ij}^{FM\perp}$ is still non-zero even in the ferromagnetically saturated case because of virtual (off-shell) minority spin contributions represented by the real part $\mathrm{Re}\, G_{ij\downarrow}(\varepsilon)$ in Eq. (43).

The RKKY coupling is long-ranged and has in general complex, oscillatory behavior in space, because it depends on details of the conduction band structure via the position dependent Green functions $G_{ji\,\sigma}(\omega)$ in Eq. (35). For an isotropic system in $d = 3$ dimensions, the retarded/advanced conduction electron Green function $G_{r\sigma}(\varepsilon \pm i0)$ and the paramagnetic polarization $\Pi_r^{\sigma\sigma'}(\omega)$ at temperature $T = 0$ are calculated in position space as,

$$G_{r\sigma}(\varepsilon \pm i0) = -\pi N(\varepsilon)\, \frac{\mathrm{e}^{\pm ik(\varepsilon_F+\varepsilon)r}}{k(\varepsilon_F+\varepsilon)r} \tag{44}$$

$$\begin{aligned} \Pi_r^{\sigma\sigma'}(\omega + i0) = & \left[N(0)\, \frac{\sin(x) - x\cos(x)}{4x^4} + \mathcal{O}\left(\left(\frac{\omega}{\varepsilon_F} \right)^2 \right) \right] \\ & \pm i \left[\frac{1}{\pi} N(0)\, \frac{1-\cos(x)}{x^2}\, \frac{\omega}{\varepsilon_F} + \mathcal{O}\left(\left(\frac{\omega}{\varepsilon_F} \right)^3 \right) \right] \end{aligned} \tag{45}$$

Here, ε_F and k_F are the Fermi energy and Fermi wavenumber, respectively, and $r = |\mathbf{r}_i - \mathbf{r}_j|$, $x = 2k_F r$. For illustration, Fig. 4(b) shows the static polarization $\Pi_r^{\sigma\sigma'}(0)$ as a function of x for the isotropic case.

4 Interplay of Kondo screening and RKKY interaction

We now turn to the interplay of the two interactions on a Kondo lattice, Eq. (31). First, it is crucial to remember that the RKKY interaction between different f-spins is not a direct spin exchange interaction, but mediated by the conduction band [3–5] and generated in second order by the same spin coupling J_0 that is also responsible for the local Kondo spin screening, as shown in the previous section. The essential difference can be seen from the example of a two-impurity Kondo system, $\mathbf{S}_1$, $\mathbf{S}_2$: With a direct impurity-impurity coupling, $K\,\mathbf{S}_1 \cdot \mathbf{S}_2$, this model can exhibit a dimer singlet phase where the dimer is decoupled from the conduction electrons. The dimer singlet and the local Kondo singlet phase are then separated by a quantum critical point (QCP), controlled by K [25, 26]. By contrast, when the inter-impurity coupling is created by the RKKY interaction only, i.e. generated by J_0, a decoupled dimer singlet phase is not possible. Instead, the impurity spins must remain coupled to the conduction sea. We show below that the Kondo singlet formation at $T = 0$ breaks down at a critical strength of the RKKY coupling even if magnetic ordering is suppressed, i.e. without a 2nd-order quantum phase transition and without critical fluctuations. If magnetic ordering occurs, critical ordering fluctuations will be present in addition to, but independently of the RKKY-induced Kondo breakdown.

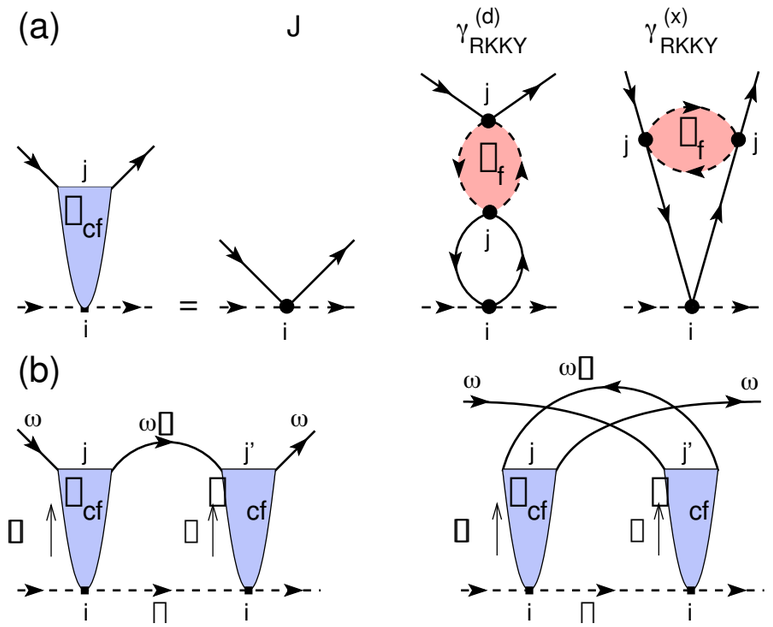


Fig. 5: *(a) f-spin–c-electron vertex $\hat{\Gamma}_{cf}$, composed of the onsite vertex J at site i and the RKKY-induced contributions from surrounding sites $j \neq i$ to leading order in the RKKY coupling, $\gamma^{(d)}_{RKKY}$ (direct term) and $\gamma^{(x)}_{RKKY}$ (exchange term). (b) 1-loop diagrams for the perturbative RG. Solid lines: electron Green functions G_c, dashed lines: pseudo-fermion propagators G_f of the local f-spins. The red bubbles represent the full f-spin susceptibility at sites j.*

4.1 The concept of a selfconsistent renormalization group

The problem of local Kondo screening or breakdown on a Kondo lattice amounts to calculating the vertex for scattering of c-electrons from a local f-spin and analyzing its divergence (Kondo screening of the f-spin, cf. section 2.3) or non-divergence (Kondo breakdown) under RG. In the case of multiple Kondo sites or a Kondo lattice, this vertex $\hat{\Gamma}_{cf}$ acquires nonlocal contributions in addition to the local coupling J_0, because a c-electron can scatter from a distant Kondo site $j \neq i$, and the spin flip at that site is transferred to the f-spin at site i via the RKKY interaction. On the other hand, the RKKY vertex $\hat{\Gamma}_{ff}$ coupling two f-spins has no logarithmic RG flow, since the recoil (momentum integration) of the itinerant conduction electrons prevents an infrared divergence of the RKKY interaction. Thus, $\hat{\Gamma}_{ff}$ remains in the weak coupling regime, and RKKY-induced magnetic ordering must be a secondary effect, not controlled by the RG divergence of a coupling constant.

The diagrams contributing to $\hat{\Gamma}_{cf}$ to leading order in the RKKY coupling are shown in Fig. 5(a). As seen from the figure, a nonlocal scattering process necessarily involves the exact, local dynamical f-spin susceptibility $\chi_f(i\Omega_n)$ on site j. The resulting c–f vertex $\hat{\Gamma}_{cf}$ has the structure of a nonlocal Heisenberg coupling in spin space, see Appendix A.1. The exchange diagram,

$\gamma^{(x)}_{RKKY}$ in Fig. 5(a), contributes only a sub-leading logarithmic term as compared to the direct term $\gamma^{(d)}_{RKKY}$, see Appendix A.2. In particular, it does not alter the universal $T_K(y)$ suppression derived below. It can, therefore, be neglected. To leading (linear) order in the RKKY coupling, $\hat{\Gamma}_{cf}$ thus reads (in Matsubara representation)

$$\begin{aligned} \hat{\Gamma}_{cf} &= \left[J\delta_{i,j} + \gamma^{(d)}_{RKKY}(\mathbf{r}_{ij}, i\Omega_n)\right] \mathbf{S}_i \cdot \mathbf{s}_j \\ &= \left[J\delta_{ij} + 2J J_0^2\ (1-\delta_{ij})\ \Pi(\mathbf{r}_{ij}, i\Omega_n)\, \tilde{\chi}_f(i\Omega_n)\right] \mathbf{S}_i \cdot \mathbf{s}_j\,, \end{aligned} \tag{46}$$

where $\mathbf{r}_{ij} = \mathbf{x}_i - \mathbf{x}_j$ the distance vector between the sites i and j, and Ω is the energy transferred in the scattering process. $\Pi(\mathbf{r}_{ij}, i\Omega_n)$ is the c-electron density correlation function between sites i and j [bubble of solid lines in Fig. 5(a)] and $\tilde{\chi}_f(i\Omega_n) := \chi_f(i\Omega_n)/(g_L\mu_B)^2$, with g_L the Landé factor and μ_B the Bohr magneton. Note that Eq. (46) contains the running coupling J at site i which will be renormalized under RG, while at the site j, where the c-electron scatters, the bare coupling J_0 appears, since all vertex renormalizations on that site are already included in the exact susceptibility χ_f. Higher order terms, as for instance generated by the RG [see below, Fig. 5(b)], lead to nonlocality of the incoming and outgoing coordinates of the scattering c-electrons, $\mathbf{x}_j$, $\mathbf{x}_{j'}$, but the f-spin coordinate $\mathbf{x}_i$ remains strictly local, since the pseudo-fermion propagator $G_f(i\nu_n) = 1/i\nu_n$ is local [20]. For this reason, speaking of Kondo singlet formation on a single Kondo site is well defined even on a Kondo lattice, and so is the local susceptibility χ_f of a single f-spin. The corresponding Kondo scale T_K on a site j is observable, e.g., as the Kondo resonance width measured by STM spectroscopy on one Kondo ion of the Kondo lattice. The temperature dependence of the single-site f-spin susceptibility is known from the Bethe ansatz solution [22] in terms of the Kondo scale T_K. It has a $T = 0$ value $\chi_f(0) \propto 1/T_K$ and crosses over to the $1/T$ behavior of a free spin for $T > T_K$. These features can be modeled in the retarded/advanced, local, dynamical f-spin susceptibility $\chi_f(\Omega \pm i0)$ as

$$\chi_f(\Omega \pm i0) = \frac{(g_L\mu_B)^2 W}{\pi T_K \sqrt{1 + (\Omega/T_K)^2}} \left(1 \pm \frac{2i}{\pi} \mathrm{arsinh} \frac{\Omega}{T_K}\right) \tag{47}$$

where W is the Wilson ratio, and the imaginary part is implied by the Kramers-Kronig relation. Deriving the one-loop RG equation for a multi-impurity or lattice Kondo system proceeds as in section 2.3, however for the c–f vertex $\hat{\Gamma}_{cf}$, including RKKY-induced nonlocal contributions. The one-loop spin vertex function is shown diagrammatically in Fig. 5(b). Using Eq. (46), the sum of these two diagrams is up to linear order in the RKKY coupling,

$$\begin{aligned} Y(\mathbf{r}_{ij}, i\omega_n) = -J\,T \sum_{i\Omega_m} &\left[J\delta_{ij} + \gamma^{(d)}_{RKKY}(\mathbf{r}_{ij}, i\Omega_m) + \gamma^{(d)}_{RKKY}(\mathbf{r}_{ij}, -i\Omega_m)\right] \\ &\times \left[G_c(\mathbf{r}_{ij}, i\omega_n - i\Omega_m) - G_c(\mathbf{r}_{ij}, i\omega_n + i\Omega_m)\right]\, G_f(i\Omega_m). \end{aligned} \tag{48}$$

Here, ω is the energy of incoming conduction electrons and $G_c(\mathbf{r}_{ij}, i\omega_n + i\Omega_m)$ the single-particle c-electron propagator from the incoming to the outgoing site.

For the low-energy physics, the vertex renormalization for c-electrons at the Fermi surface is required. This means setting the energy $i\omega \to \omega = 0 + i0$ and Fourier transforming the total

vertex $Y(\mathbf{r}_{ij}, i\omega)$ with respect to the incoming and outgoing c-electron coordinates, $\mathbf{x}_j$, $\mathbf{x}_i$, and taking its Fourier component for momenta at the Fermi surface, $\mathbf{k}_F$. Note that at the Fermi energy $Y(\mathbf{k}_F, 0)$ is real, even though the RKKY-induced, dynamical vertex $\gamma^{(d)}_{RKKY}(\pm i\Omega_m)$ appearing in Eq. (48) is complex-valued. This ensures the total vertex operator of the renormalized Hamiltonian to be Hermitian. By analytic continuation, the Matsubara summation in Eq. (48) becomes an integration over the intermediate c-electron energy from the lower and upper band cutoff $\mp D$ to the Fermi energy ($\Omega = 0$). The coupling constant renormalization is then obtained by requiring that $Y(\mathbf{k}_F, 0)$ is invariant under an infinitesimal reduction of the running band cutoff D (cf. section 2.3). Note that the band cutoff appears in both, the intermediate electron propagator G_c and in Π. However, differentiation of the latter does not contribute to the logarithmic RG flow. This leads to the 1-loop lattice RG equation for the local coupling [27]

$$\frac{\mathrm{d}g}{\mathrm{d}\ln D} = -2g^2 \left(1 - y\, g_0^2\, \frac{D_0}{T_K}\, \frac{1}{\sqrt{1+(D/T_K)^2}}\right), \tag{49}$$

with the bare band cutoff D_0. The first term in Eq. (49) is the onsite contribution to the β-function, while the second term represents the RKKY contribution. It is seen that χ_f, as in Eq. (47), induces a soft cutoff on the scale T_K and the characteristic $1/T_K$ dependence to the RG flow of this contribution, where T_K is the Kondo scale on the *surrounding* Kondo sites. The dimensionless coefficient

$$y = -\frac{8W}{\pi^2}\mathrm{Im}\sum_{j\neq i} \frac{\mathrm{e}^{-i\mathbf{k}_F\mathbf{r}_{ij}}}{N(0)^2} G_c^R(\mathbf{r}_{ij}, \Omega = 0)\, \Pi(\mathbf{r}_{ij}, \Omega = 0) \tag{50}$$

arises from the Fourier transform $Y(\mathbf{k}_F, 0)$ and parameterizes the RKKY coupling strength. The summation in Eq. (50) runs over all positions $j \neq i$ of Kondo sites in the system. It is important to note that y is generically positive, even though the RKKY correlations $\Pi(\mathbf{r}_{ij}, 0)$ may be ferro- or antiferromagnetic. For instance, for an isotropic and dense system with lattice constant a ($k_F a \ll 1$), the summation in Eq. (50) can be approximated by an integral, and with the substitution $x = 2k_F|\mathbf{r}_{ij}|$, y can be expressed as

$$y \approx \frac{2W}{(k_F a)^3}\int_{k_F a}^{\infty} dx\, (1-\cos x)\, \frac{x\cos x - \sin x}{x^4} > 0\,. \tag{51}$$

As a consequence, the RKKY correlations reduce the g-renormalization in Eq. (49), irrespective of the sign of $\Pi(\mathbf{r}_{ij}, 0)$, as one would physically expect.

The Kondo scale for singlet formation on site i is defined as the running cutoff value where the c–f coupling g diverges. An important feature of the lattice RG equation (49) is that the Kondo screening scale on surrounding sites $j \neq i$ appears as a parameter in the β-function for the renormalization on site i. By equivalence of all Kondo sites, the Kondo scales T_K on all sites i and j must be equal. This leads to the fact that the divergence scale T_K of the lattice RG equation must be determined self-consistently and will imply an implicit equation for the local screening scale $T_K = T_K(y)$ on a Kondo lattice, which will depend on the RKKY parameter y. The equivalence of the c–f vertices on all Kondo sites is reminiscent of a dynamical mean-field theory treatment, however, it goes beyond the latter in taking the long-range RKKY contributions into account.

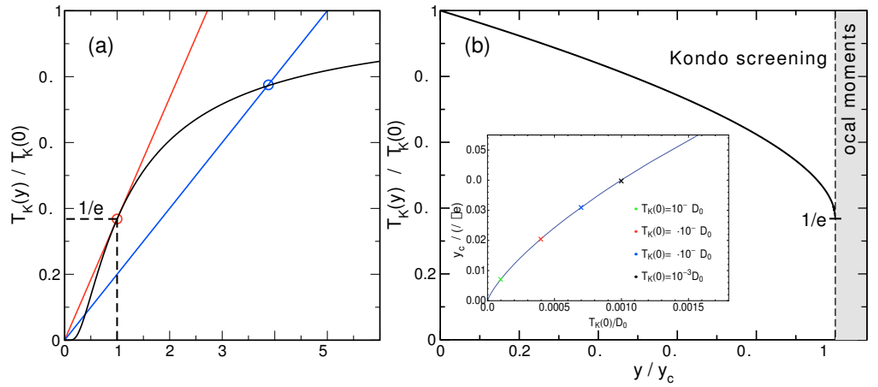


Fig. 6: *(a) Graphical solution of Eq. (55): black, solid curve: right-hand side of Eq. (55), blue line: left-hand side for $y < y_c$, red line: left-hand side for $y = y_c$ (where red line and black curve touch). It proves that there is a critical coupling y_c beyond which Eq. (55) has no solution, and $T_K(y_c)/T_K(0) = 1/e$. (b) Universal dependence of $T_K(y)/T_K(0)$ on the normalized RKKY parameter y/y_c, solution of Eq. (55). The inset shows the critical RKKY parameter y_c for various single-ion Kondo temperatures $T_K(0)$, Eq. (57).*

4.2 Integration of the RG equation

The RG equation Eq. (49) is readily integrated by separation of variables,

$$-\int_{g_0}^{g}\frac{dg}{g^2} \;=\; 2\int_{\ln D_0}^{\ln D} d\ln D' \;-\; 2yg_0^2\frac{D_0}{T_K}\int_{D_0/T_K}^{D/T_K}\frac{dx}{x}\frac{1}{\sqrt{1+x^2}}\,, \tag{52}$$

or

$$\frac{1}{g}-\frac{1}{g_0} \;=\; 2\ln\left(\frac{D}{D_0}\right) \;-\; yg_0^2\frac{D_0}{T_K}\ln\left(\frac{\sqrt{1+(D/T_K)^2}-1}{\sqrt{1+(D/T_K)^2}+1}\right), \tag{53}$$

where we have used $D_0/T_K \gg 1$ in the last expression. The Kondo scale is defined as the value of the running cutoff D where g diverges, i.e., $g \to \infty$ when $D \to T_K$. This yields the defining equation for the Kondo scale $T_K \equiv T_K(y)$,

$$-\frac{1}{g_0} \;=\; 2\ln\left(\frac{T_K(y)}{D_0}\right) \;-\; yg_0^2\frac{D_0}{T_K(y)}\ln\left(\frac{\sqrt{2}-1}{\sqrt{2}+1}\right).$$

Using the definition of the single-impurity Kondo temperature, $-1/g_0 = 2\ln\left(T_K(0)/D_0\right)$, the defining, implicit equation for $T_K(y)$ can finally be written as

$$\frac{T_K(y)}{T_K(0)} = \exp\left(-y\,\alpha\,g_0^2\,\frac{D_0}{T_K(y)}\right)\,, \tag{54}$$

with $\alpha = \ln(\sqrt{2}+1)$.

4.3 Universal suppression of the Kondo scale

By the rescaling, $u = T_K(y)/(y\alpha g_0^2 D_0)$, $y_c = T_K(0)/(\alpha \mathrm{e} g_0^2 D_0)$, Eq. (54) takes the universal form (e is Euler's constant),

$$\frac{y}{\mathrm{e}y_c}\,u = \mathrm{e}^{-1/u}\,. \tag{55}$$

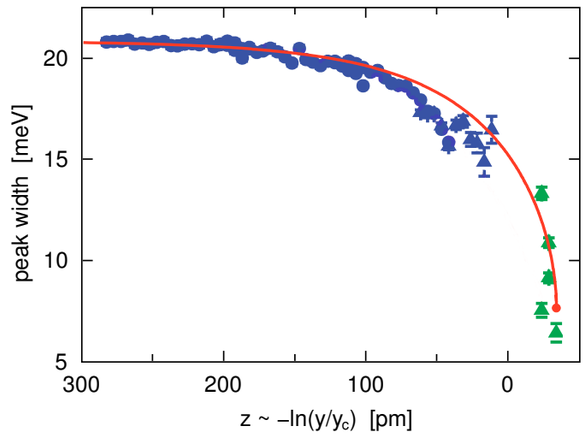


Fig. 7: *Comparison of the theory (red curve), Eq. (55), with STM spectroscopy experiments on a tunable two-impurity Kondo system (data points, Ref. [29]). The data points represent the Kondo scale T_K as extracted from the STM spectra by fitting a split Fano line shape of width T_K to the experimental spectra, see [29] for experimental details.*

Its solution can be expressed in terms of the Lambert W function [28] as $u(y) = -1/W(-y/\mathrm{e}y_c)$. Fig. 6(a) visualizes solving Eq. (55) graphically. It shows that Eq. (55) has solutions only for $y \leq y_c$. This means that y_c marks a Kondo breakdown point beyond which the RG does not scale to strong coupling, i.e., a Kondo singlet is not formed for $y > y_c$ even at the lowest energies. Using the above definitions, the RKKY-induced suppression of the Kondo lattice temperature reads, $T_K(y)/T_K(0) = u(y)y/(\mathrm{e}y_c) = -y/[\mathrm{e}y_c W(-y/\mathrm{e}y_c)]$. It is shown in Fig. 6(b). In particular, at the breakdown point it vanishes *discontinuously* and takes the finite, universal value (see Fig. 6(a)),

$$\frac{T_K(y_c)}{T_K(0)} = \frac{1}{\mathrm{e}} \approx 0.368\,. \tag{56}$$

We emphasize that the RKKY parameter y depends on details of the conduction band structure and of the spatial arrangement of Kondo sites. Sub-leading contributions to Γ_{cf} may modify the form of the cutoff function in the RG Eq. (49) and thus the nonuniversal parameter α. However, all this does not affect the universal dependence $T_K(y)$ on y given by Eq. (55).

The critical RKKY parameter, as defined before Eq. (55), can be expressed solely in terms of the single-ion Kondo scale,

$$y_c = \frac{4}{\alpha \mathrm{e}} \tau_K (\ln \tau_K)^2\,, \tag{57}$$

with $\tau_K = T_K(0)/D_0$. Note that [via $T_K(0) = D_0 \exp(-1/2g_0)$ and $N(0) = 1/(2D_0)$] this is equivalent to Doniach's breakdown criterion [7], $N(0) y_c J_0^2 = T_K(0)$, up to a factor of $\mathcal{O}(1)$. However, the present theory goes beyond the Doniach scenario in that it predicts the behavior of $T_K(y)$.

The present theory applies directly to two-impurity Kondo systems, where magnetic ordering does not play a role, and can be compared to corresponding STM experiments [29, 30]. In Ref. [29], the Kondo scale has been extracted as the line width of the (hybridization-split)

Kondo-Fano resonance. In this experimental setup, the RKKY parameter y is proportional to the overlap of tip and surface c-electron wave functions and, thus, depends exponentially on the tip-surface separation z, $y = y_c \exp(-(z - z_0)/\xi)$. Identifying the experimentally observed breakdown point, $z = z_0$, with the Kondo breakdown point, the only adjustable parameters are a scale factor ξ of the z coordinate and $T_K(0)$, the resonance width at large separation, $z = 300$ pm. The agreement between theory and experiment is striking, as shown in Fig. 7. In particular, at the breakdown point $T_K(y_c)/T_K(0)$ coincides accurately with the prediction, Eq. (56), without any adjustable parameters.

5 Conclusion

We have derived a perturbative renormalization group theory for the interference of Kondo singlet formation and RKKY interaction in Kondo lattice and multi-impurity systems, assuming that magnetic ordering is suppressed, e.g. by frustration. Eqs. (54) or (55) represent a mathematical definition of the energy scale for Kondo singlet formation in a Kondo lattice, i.e., of the Kondo lattice temperature, $T_K(y)$. The theory predicts a universal suppression of $T_K(y)$ and a breakdown of complete Kondo screening at a critical RKKY parameter, $y = y_c$. At the breakdown point, the Kondo scale takes a *finite*, universal value, $T_K(y_c)/T_K(0) = 1/\mathrm{e} \approx 0.368$, and vanishes *discontinuously* for $y > y_c$. In the Anderson lattice, by contrast to the Kondo lattice, the locality of the f-spin does no longer strictly hold, but our approach should still be valid in this case. The parameter-free, quantitative agreement of this behavior with different spectroscopic experiments [29, 30] strongly supports that the present theory captures the essential physics of the Kondo-RKKY interplay.

The results may have profound relevance for heavy-fermion magnetic QPTs. In an unfrustrated lattice, the partially screened local moments existing for $y > y_c$ must undergo a second-order magnetic ordering transition at sufficiently low temperature. This means that the bare c-electron correlation or polarization function Π must be replaced by the full c-correlation function χ_c and will imply a power-law divergence of the latter in Eq. (46). We have checked the effect of such a magnetic instability, induced either by the ordering of remanent local moments or by a c-electron SDW instability: The breakdown ratio $T_K(y_c)/T_K(0)$ will be altered, but must remain nonzero. The reason is that the inflection point of the exponential on the right-hand side of Eq. (55) (see Fig. 6) is not removed by such a divergence and, therefore, the solution ceases to exist at a finite value of $T_K(y_c)$. This points to an important conjecture about a possible, new quantum critical scenario with Kondo destruction: The Kondo spectral weight may vanish continuously at the QCP, while the Kondo energy scale $T_K(y)$ (resonance width) remains finite. Such a scenario may reconcile apparently contradictory experimental results in that it may fulfill dynamical scaling, even though $T_K(y_c)$ is finite at the QCP.

Appendix

A f-spin – conduction-electron vertex $\hat{\Gamma}_{cf}$

Here we present some details on the calculation of the elementary c-electron–f-spin vertex $\hat{\Gamma}_{cf}$ It is defined via the Kondo lattice Hamiltonian,

$$H = \sum_{\mathbf{k},\sigma} \varepsilon_{\mathbf{k}}\, c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + J_0 \sum_i \hat{\mathbf{S}}_i \cdot \hat{\mathbf{s}}_i \;, \tag{58}$$

The direct (d) and exchange (x) parts of the RKKY-induced vertex can be written as the product of a distance and energy dependent function $\Lambda^{(d/x)}_{RKKY}$ and an operator in spin space, $\hat{\Gamma}^{(d/x)}$,

$$\hat{\gamma}^{(d/x)}_{RKKY} = \Lambda^{(d/x)}_{RKKY}(\mathbf{r}_{ij}, i\Omega)\, \hat{\Gamma}^{(d/x)} \tag{59}$$

A.1 Spin structure

Denoting the vector of Pauli matrices acting in c-electron spin space by $\boldsymbol{\sigma} = (\sigma^x, \sigma^y, \sigma^z)^T$ and the vector of Pauli matrices in f-spin space by $\mathrm{s} = (s^x, s^y, s^z)^T$, the RRKY-induced vertex contributions read in spin space,

$$\hat{\Gamma}^{(d)}_{\alpha\beta,\kappa\lambda} = \sum_{a,b,c=x,y,z} \sum_{\gamma,\delta,\mu,\nu=1}^{2} \left(\sigma^a_{\delta\gamma} s^a_{\kappa\lambda}\right) \left(\sigma^b_{\gamma\delta} s^b_{\nu\mu}\right) \left(\sigma^c_{\alpha\beta} s^c_{\mu\nu}\right) \tag{60}$$

$$\hat{\Gamma}^{(x)}_{\alpha\beta,\kappa\lambda} = \sum_{a,b,c=x,y,z} \sum_{\gamma,\delta,\mu,\nu=1}^{2} \left(\sigma^a_{\delta\gamma} s^a_{\kappa\lambda}\right) \left(\sigma^b_{\alpha\delta} s^b_{\nu\mu}\right) \left(\sigma^c_{\gamma\beta} s^c_{\mu\nu}\right) \tag{61}$$

with c-electron spin indices α, β, γ, δ, and f-spin indices κ, λ, μ, ν, as shown in Fig. 8(a). The spin summations can be performed using the spin algebra ($a, b = x, y, z$),

$$\sum_{\gamma=1}^{2} \sigma^a_{\alpha\gamma} \sigma^b_{\gamma\beta} = \sum_{c=x,y,z} i\varepsilon^{abc} \sigma^c_{\alpha\beta} + \delta^{ab} \mathbb{1}_{\alpha\beta} \;, \tag{62}$$

where $\mathbb{1}$ is the unit operator in spin space, ε^{abc} the totally antisymmetric tensor and δ^{ab} the Kronecker-δ. This results in a nonlocal Heisenberg coupling between sites i and j,

$$\hat{\Gamma}^{(d)}_{\alpha\beta,\kappa\lambda} = 4 \sum_{a=x,y,z} \left(\sigma^a_{\alpha\beta} s^a_{\kappa\lambda}\right) \tag{63}$$

$$\hat{\Gamma}^{(x)}_{\alpha\beta,\kappa\lambda} = -2 \sum_{a=x,y,z} \left(\sigma^a_{\alpha\beta} s^a_{\kappa\lambda}\right) \;. \tag{64}$$

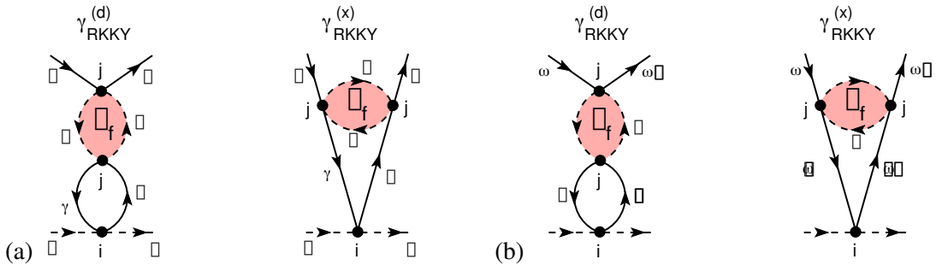


Fig. 8: *Direct (d) and exchange (x) diagrams of the RKKY-induced contributions to the $c-f$ vertex: (a) spin labelling, (b) energy labelling.*

A.2 Energy dependence

With the energy variables as defined in Fig. 8(b), the energy dependent functions in Eq. (59) read in Matsubara representation

$$\Lambda^{(d)}_{RKKY}(\mathbf{r}_{ij}, i\Omega_m) = JJ_0^2 \Pi(\mathbf{r}_{ij}, i\Omega_m)\tilde{\chi}_f(i\Omega_m)$$
$$\Lambda^{(x)}_{RKKY}(\mathbf{r}_{ij}, i\omega_n, i\Omega_m) = -JJ_0^2 T \sum_{i\varepsilon_m} G_c(\mathbf{r}_{ij}, i\omega_n + i\varepsilon_m) G_c(\mathbf{r}_{ij}, i\omega_n + i\Omega_m + i\varepsilon_m)\tilde{\chi}_f(i\varepsilon_m)$$

where

$$\Pi(\mathbf{r}_{ij}, i\Omega_m) \;=\; -T\sum_{\varepsilon_n} G_c(\mathbf{r}_{ij}, i\varepsilon_n) G_c(\mathbf{r}_{ij}, i\varepsilon_n + i\Omega_m) \tag{65}$$

and $\tilde{\chi}_f(i\varepsilon_m) = \chi_f(i\varepsilon_m)/(g_L\mu_B)^2$, with $\chi_f(i\varepsilon_m)$ the full, single-impurity f-spin susceptibility, Eq. (47).

For the renormalization of the total $c-f$ vertex for c-electrons at the Fermi energy, the contributions $\Lambda^{(d)}_{RKKY}$, $\Lambda^{(d)}_{RKKY}$ must be calculated for real frequencies, $i\Omega \to \Omega + i0$, $i\omega \to \omega + i0$, and for electrons at the Fermi energy, i.e., $\omega = 0$. In this limit, only the real parts of $\Lambda^{(d)}_{RKKY}$, $\Lambda^{(d)}_{RKKY}$ contribute to the vertex renormalization, as seen below. In order to analyze their importance for the RG flow, we will expand them in terms of the small parameter T_K/D_0. In the following, the real part of a complex function will be denoted by a prime ' and the imaginary part by a double-prime ".

Direct contribution. Since in $\Lambda^{(d)}_{RKKY}$, $\Pi(i\Omega_m)$ and $\tilde{\chi}_f(i\Omega_m)$ appear as a product and $\tilde{\chi}_f(\Omega)$ cuts off the energy transfer Ω at the scale $T_K \ll \varepsilon_F \approx D_0$, the electron polarization $\Pi(\Omega)$ contributes only in the limit $\Omega \ll \varepsilon_F$ where it is real-valued, as seen in Eq. (45). Using Eq. (45) and Eq. (47), the real part of the direct RKKY-induced vertex contribution reads,

$$\Lambda^{(d)\prime}_{RKKY}(\mathbf{r}_{ij}, \Omega + i0) = JJ_0^2 R(\mathbf{r}_{ij}) AN(0)\,\frac{D_0}{T_K}\,\frac{1}{\sqrt{1+(\Omega/T_K)^2}} + \mathcal{O}\left(\left(\frac{\Omega}{D_0}\right)^2\right), \tag{66}$$

where

$$R(\mathbf{r}_{ij}) \;=\; \frac{\sin(x) - x\cos(x)}{4x^4}, \qquad\qquad x = 2k_F r \tag{67}$$

is a spatially oscillating function.

Exchange contribution. In order to analyze the size of ${\Lambda^{(x)}_{RKKY}}'$ in terms of T_K/D_0, it is sufficient to evaluate it for a particle-hole symmetric conduction band and for $\mathbf{r}_{ij} = 0$, since the T_K/D_0 dependence is induced by the on-site susceptibility $\tilde{\chi}_f(i\Omega)$. The dependence on T_K/D_0 can be changed by the frequency convolution involved in ${\Lambda^{(x)}_{RKKY}}'$, but does not depend on details of the conduction band and distance dependent terms. (The general calculation is possible as well, but considerably more lengthy.) We use the short-hand notation for the momentum-integrated c-electron Green function, $G_c(\mathbf{r} = 0, \omega \pm i0) = G(\omega) = G'(\omega) + iG''(\omega)$, and assume a flat density of states $N(\omega)$, with the upper and lower band cutoff symmetric about ε_F, i.e.,

$$G^{R/A\prime\prime}(\omega) = \mp\frac{\pi}{2D_0}\Theta(D_0 - |\omega|) \tag{68}$$

$$G^{R/A\prime}(\omega) = \frac{1}{2D_0}\ln\left|\frac{D_0+\omega}{D_0-\omega}\right| = \frac{\omega}{D_0^2} + \mathcal{O}\left(\left(\frac{\omega}{D_0}\right)\right) . \tag{69}$$

Furthermore, at $T = 0$ the Fermi and Bose distribution functions are, $f(\varepsilon) = -b(\varepsilon) = \Theta(-\varepsilon)$. ${\Lambda^{(x)}_{RKKY}}'(0, 0, \Omega + i0)$ then reads,

$$\begin{aligned}
&{\Lambda^{(x)}_{RKKY}}'(\mathbf{r}_{ij} = 0, \omega = 0 + i0, \Omega + i0) = \\
&-JJ_0^2\left\{\int\frac{d\varepsilon}{\pi}\left[f(\varepsilon)G^{A\prime\prime}(\varepsilon)G^{R\prime}(\varepsilon+\Omega) + f(\varepsilon+\Omega)\,G^{A\prime}(\varepsilon)G^{A\prime\prime}(\varepsilon+\Omega)\right]\tilde{\chi}_f^{R\prime}(\varepsilon)\right. \\
&\left.-\int\frac{d\varepsilon}{\pi}\left[f(\varepsilon)\,G^{R\prime}(\varepsilon)G^{R\prime}(\varepsilon+\Omega) - f(\varepsilon+\Omega)G^{A\prime\prime}(\varepsilon)G^{A\prime\prime}(\varepsilon+\Omega)\right]\tilde{\chi}_f^{R\prime\prime}(\varepsilon)\right\} .
\end{aligned} \tag{70}$$

With the above definitions, the four terms in this expression are evaluated in an elementary way, using the substitution $\varepsilon_F/T_K = x = \sinh u$,

$$\begin{aligned}
&\int\frac{d\varepsilon}{\pi}f(\varepsilon)G^{A\prime\prime}(\varepsilon)G^{R\prime}(\varepsilon+\Omega)\tilde{\chi}_f^{R\prime}(\varepsilon) \\
&\qquad = AN(0)\frac{T_K}{D_0}\left[1 - \sqrt{1+\left(\frac{D_0}{T_K}\right)^2} + \frac{\Omega}{T_K}\,\mathrm{arsinh}\left(\frac{D_0}{T_K}\right)\right] \\
&\qquad = AN(0)\left[-1 + \frac{\Omega}{D_0}\,\ln\left(\frac{D_0}{T_K}\right) + \mathcal{O}\left(\frac{T_K}{D_0}\right)\right]
\end{aligned} \tag{71}$$

$$\begin{aligned}
&\left|\int\frac{d\varepsilon}{\pi}f(\varepsilon+\Omega)\,G^{A\prime}(\varepsilon)G^{A\prime\prime}(\varepsilon+\Omega)\tilde{\chi}_f^{R\prime}(\varepsilon)\right| \\
&\qquad = AN(0)\frac{T_K}{D_0}\left|\sqrt{1+\left(\frac{\Omega}{T_K}\right)^2} - \sqrt{1+\left(\frac{D_0+\Omega}{T_K}\right)^2}\right| \\
&\qquad \le AN(0) + \mathcal{O}\left(\frac{T_K}{D_0}\right)
\end{aligned} \tag{72}$$

$$\begin{aligned}
&\int\frac{d\varepsilon}{\pi}f(\varepsilon)\,G^{R\prime}(\varepsilon)G^{R\prime}(\varepsilon+\Omega)\tilde{\chi}_f^{R\prime\prime}(\varepsilon) \\
&\qquad = -\frac{4}{\pi^2}AN(0)\left(\frac{1}{2} + \frac{\Omega}{D_0}\right)\ln\left(\frac{D_0}{T_K}\right) + \mathcal{O}\left(\left(\frac{T_K}{D_0}\right)^0\right)
\end{aligned} \tag{73}$$

$$\begin{aligned}\int \frac{d\varepsilon}{\pi} f(\varepsilon+\Omega) G^{A\prime\prime}(\varepsilon) G^{A\prime\prime}(\varepsilon+\Omega) \tilde{\chi}_f^{R\prime\prime}(\varepsilon) \\ &= \frac{\pi}{4} AN(0) \left[-\operatorname{arsinh}\left(\frac{\Omega}{T_K}\right) + \operatorname{arsinh}\left(\min\left(\frac{\Omega}{T_K}, \frac{D_0+\Omega}{T_K}\right)\right)\right] \\ &\leq \frac{\pi}{4} AN(0) + \mathcal{O}\left(\frac{T_K}{D_0}\right) . \end{aligned} \tag{74}$$

Comparing Eqs. (70)–(74) with Eq. (66) shows that all terms of ${\Lambda^{(x)}_{RKKY}}'(\Omega)$ are sub-leading compared to ${\Lambda^{(d)}_{RKKY}}'(\Omega)$ by at least a factor $(T_K/D_0)\ln(T_K/D_0)$ for all transferred energies Ω. Hence, it can be neglected in the RG flow. Combining the results of spin and energy dependence, Eqs. (59), (63), and (66), one obtains the total RKKY-induced $c-f$ vertex as,

$$\hat{\gamma}^{(d)}_{RKKY}(\mathbf{r}_{ij}, i\Omega) \;=\; 2\,(1-\delta_{ij})\,\Pi(\mathbf{r}_{ij}, i\Omega)\chi_f(i\Omega)\,\mathbf{S}_i \cdot \mathbf{s}_j \tag{75}$$

or

$$\mathrm{Re}\hat{\gamma}^{(d)}_{RKKY}(\mathbf{r}_{ij}, \Omega+i0) = 2JJ_0^2 AN(0)\,(1-\delta_{ij}) R(\mathbf{r}_{ij})\,\frac{D_0}{T_K}\,\frac{1}{\sqrt{1+(\Omega/T_K)^2}}\,\mathbf{S}_i \cdot \mathbf{s}_j. \tag{76}$$

References

[1] J. Kondo, Prog. Theor. Phys **32**, 37 (1964)

[2] A.C. Hewson: *The Kondo Problem to Heavy Fermions* (Cambridge University Press, 1993)

[3] M.A. Ruderman and C. Kittel, Phys. Rev. **96**, 99 (1954)

[4] T. Kasuya, Prog. Theor. Phys. **16**, 45 (1956)

[5] K. Yosida, Phys. Rev. **106**, 893 (1957)

[6] H. v. Löhneysen, A. Rosch, M. Vojta, and P. Wölfle, Rev. Mod. Phys. **79**, 1015 (2007)

[7] S. Doniach, Physica B+C **91**, 231 (1977)

[8] J.A. Hertz, Phys. Rev. B **14**, 1165 (1976)

[9] T. Moriya: *Spin fluctuations in itinerant electron magnetism* (Springer, Berlin, 1985)

[10] A.J. Millis, Phys. Rev. B **48**, 7183 (1993)

[11] K.Q. Si, S. Rabello and J.L. Smith, Nature **413**, 804 (2001)

[12] P. Coleman, C. Pépin, Q. Si, and R. Ramazashvili, J. Phys.: Condens, Matter **13**, R723 (2001)

[13] T. Senthil, M. Vojta, and S. Sachdev, Phys. Rev. B **69**, 035111 (2004)

[14] P. Wölfle and E. Abrahams, Phys. Rev. B **84**, 041101 (2011)

[15] E. Abrahams, J. Schmalian, and P. Wölfle, Phys. Rev. B **90**, 045105 (2014)

[16] P. Wölfle and E. Abrahams, Phys. Rev. B **93**, 075128 (2016)

[17] A.A. Abrikosov, Physics **2**, 21 (1965)

[18] S.E. Barnes, J. Phys. F **6**, 1375 (1976)

[19] P. Coleman, Phys. Rev. B **29**, 3035 (1984)

[20] J. Kroha and P. Wölfle, Acta Phys. Pol. B **29**, 3781 (1998)

[21] J. Kroha, P. Wölfle, and T.A. Costi, Phys. Rev. Lett. **79**, 216 (1997)

[22] N. Andrei, K. Furuya, and J.H. Lowenstein, Rev. Mod. Phys. **55**, 331 (1983)

[23] R. Bulla, T.A. Costi, and T. Pruschke, Rev. Mod. Phys. **80**, 395 (2008)

[24] P.W. Anderson, J. Phys. C: Solid State Phys. **3**, 2436 (1970)

[25] B.A. Jones, C.M. Varma, and J.W. Wilkins, Phys. Rev. Lett. **61**, 125 (1988)

[26] I. Affleck, A.W.W. Ludwig, and B.A. Jones, Phys. Rev. B **52**, 9528 (1995)

[27] A. Nejati, K. Ballmann, and J. Kroha, Phys. Rev. Lett. **118**, 117204 (2017)

[28] D. Veberič, Computer Phys. Commun. **183**, 2622 (2012)

[29] J. Bork, Y.-H. Zhang, L. Diekhöner, L. Borda, P. Simon, J. Kroha, P. Wahl, and K. Kern, Nat. Phys. **7**, 901 (2011)

[30] N. Néel, R. Berndt, J. Kröger, T.O. Wehling, A.I. Lichtenstein, and M.I. Katsnelson, Phys. Rev. Lett. **107**, 106804 (2011)

13 Kondo Physics and the Mott Transition

Michele Fabrizio
International School for Advanced Studies, SISSA
Via Bonomea 265, I-34136 Trieste, Italy

Contents

1 A brief recall of Landau-Fermi-liquid theory **2**

2 Ordinary Kondo physics at the Mott transition **4**
2.1 Role of the DMFT self-consistency condition 6

3 Exotic Kondo physics at the Mott transition **8**
3.1 The two-impurity model 8
3.2 The lattice model counterpart of the two-impurity model 15
3.3 Landau-Fermi-liquid picture 18

4 Concluding remarks **21**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 A brief recall of Landau-Fermi-liquid theory

Landau's Fermi-liquid theory [1] explains why interacting fermions, despite repelling each other by Coulomb interaction, almost always display thermodynamic and transport properties similar to those of non-interacting particles, which is, e.g., the reason of success of the Drude-Sommerfeld description of normal metals in terms of free-electrons.

The microscopic justification of Landau's Fermi-liquid theory, see, e.g., Ref. [2], is a beautiful and elegant realization of what we would now denote as a renormalizable field theory. I will not go through all details of such a theory, but just emphasize a few aspects linked to the main subject of the present notes.

Step zero of Landau's Fermi-liquid theory is the assumption[1] that the fully interacting single-particle Green function close to the Fermi level, $||\mathbf{k}-\mathbf{k}_F|| \ll k_F$ and $|\varepsilon| \ll \varepsilon_F$, includes a coherent and an incoherent component, namely

$$G(i\varepsilon,\mathbf{k}) \simeq G_{\text{coh.}}(i\varepsilon,\mathbf{k}) + G_{\text{incoh.}}(i\varepsilon,\mathbf{k}) = \frac{Z_{\mathbf{k}}}{i\varepsilon-\varepsilon_{\mathbf{k}}} + G_{\text{incoh.}}(i\varepsilon,\mathbf{k})\,, \tag{1}$$

where ε are Matsubara frequencies, $\varepsilon_{\mathbf{k}}$ is measured with respect to the Fermi energy ε_F, and $Z_{\mathbf{k}} \leq 1$ is the so-called quasi-particle residue. The Green function continued in the complex frequency plane $i\varepsilon \to z \in \mathbb{C}$ has therefore the simple pole singularity of $G_{\text{coh.}}(z,\mathbf{k})$ plus, generically, a branch cut on the real axis brought by $G_{\text{incoh.}}(z,\mathbf{k})$. By definition the discontinuity on the real axis is $G(z=\varepsilon+i0^+,\mathbf{k}) - G(z=\varepsilon-i0^+,\mathbf{k}) = -2\pi i\,\mathcal{N}(\varepsilon,\mathbf{k})$, where $\mathcal{N}(\varepsilon,\mathbf{k})$ is the single-particle density of states, which therefore reads, according to Eq. (1),

$$\mathcal{N}(\varepsilon,\mathbf{k}) = Z_{\mathbf{k}}\,\delta\big(\varepsilon-\varepsilon_{\mathbf{k}}\big) + \mathcal{N}_{\text{incoh.}}(\varepsilon,\mathbf{k})\,. \tag{2}$$

Since $\mathcal{N}(\varepsilon,\mathbf{k})$ has unit integral over ε, the incoherent component has weight $1-Z_{\mathbf{k}}$. The meaning of Eq. (2) is that an electron added to the system transforms, with weight $Z_{\mathbf{k}}$, into a *quasi-particle* excitation that propagates with dispersion $\varepsilon_{\mathbf{k}}$, but also into a bunch of other excitations that do not propagate coherently. The distinction between coherent and incoherent becomes sharper analyzing the analytic behavior, in the mathematical sense of a distribution, of the product

$$R(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) \equiv G(i\varepsilon+i\omega,\mathbf{k}+\mathbf{q})\,G(i\varepsilon,\mathbf{k})\,, \tag{3}$$

that enters the calculation of low-temperature linear response functions, in the low-frequency, $i\omega = \omega + i0^+$ with $\omega \ll \varepsilon_F$, and long-wavelength, $|\mathbf{q}| \ll k_F$, limit, the measurable quantities which are the ultimate goal of the theory. Indeed, one finds by elementary calculations that

$$\begin{aligned} R(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) &= G_{\text{coh.}}(i\varepsilon+i\omega,\mathbf{k}+\mathbf{q})\,G_{\text{coh.}}(i\varepsilon,\mathbf{k}) + R_{\text{incoh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) \\ &\simeq -\frac{\partial f\,(\varepsilon_{\mathbf{k}})}{\partial\varepsilon_{\mathbf{k}}}\;\delta\,(i\varepsilon)\;Z_{\mathbf{k}}^2\,\frac{\varepsilon_{\mathbf{k}+\mathbf{q}}-\varepsilon_{\mathbf{k}}}{i\omega-\varepsilon_{\mathbf{k}+\mathbf{q}}+\varepsilon_{\mathbf{k}}} + R_{\text{incoh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) \\ &\equiv R_{\text{coh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) + R_{\text{incoh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q})\,. \end{aligned} \tag{4}$$

[1] This assumption can be actually verified order by order in perturbation theory, which however does not guarantee that the perturbation series is convergent.

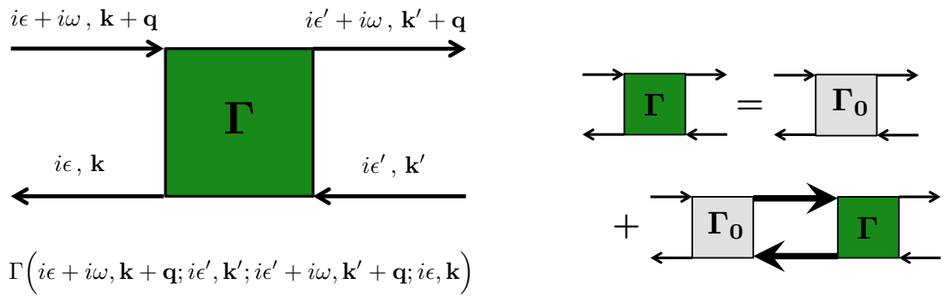


Fig. 1: *Left panel: diagrammatic representation of the interaction vertex Γ in the particle-hole channel with frequency and momentum transferred ω and $\mathbf{q}$, respectively. Right panel: diagrammatic representation of the Bethe-Salpeter equation. Γ_0 is the irreducible vertex, and the product of the two internal Green functions, the two tick lines, is by definition the distribution R.*

The crucial point that distinguishes $R_{\text{coh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q})$ from $R_{\text{incoh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q})$ is that the former is evidently non-analytic in the origin, $\omega=\mathbf{q}=0$, while the latter is assumed to be analytic. In other words, while the limiting value of $R_{\text{incoh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q})$ at $\omega=\mathbf{q}=0$ is unique,

$$\lim_{\omega\to 0}\lim_{\mathbf{q}\to 0} R_{\text{incoh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) = \lim_{\mathbf{q}\to 0}\lim_{\omega\to 0} R_{\text{incoh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) \equiv R_{\text{incoh.}}(i\varepsilon,\mathbf{k})\,, \tag{5}$$

that of $R_{\text{coh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q})$ is instead not unique and depends how the limit is taken:

$$\begin{aligned}
&\lim_{\omega\to 0}\lim_{\mathbf{q}\to 0} R_{\text{coh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) \equiv R^{\omega}_{\text{coh.}}(i\varepsilon,\mathbf{k}) = 0\,,\\
&\lim_{\mathbf{q}\to 0}\lim_{\omega\to 0} R_{\text{coh.}}(i\varepsilon,\mathbf{k};i\omega,\mathbf{q}) \equiv R^{q}_{\text{coh.}}(i\varepsilon,\mathbf{k}) = \frac{\partial f\left(\varepsilon_{\mathbf{k}}\right)}{\partial\varepsilon_{\mathbf{k}}}\,\delta\left(i\varepsilon\right)\,Z^2_{\mathbf{k}}\,,
\end{aligned} \tag{6}$$

where the two different limits are conventionally indicated by the superscripts ω and q. It thus follows that

$$\begin{aligned}
R^{\omega}(i\varepsilon,\mathbf{k}) &= R_{\text{incoh.}}(i\varepsilon,\mathbf{k})\,,\\
R^{q}(i\varepsilon,\mathbf{k}) &= R^{q}_{\text{coh.}}(i\varepsilon,\mathbf{k}) + R_{\text{incoh.}}(i\varepsilon,\mathbf{k})\,.
\end{aligned} \tag{7}$$

The next important step within Landau's Fermi-liquid theory is to absorb the completely unknown $R_{\text{incoh.}}(i\varepsilon,\mathbf{k})$ into few parameters. I will not repeat thoroughly what is well explained in many other places, but just sketch how it works in the case of the Bethe-Salpeter equation that relates the reducible vertex in the particle-hole channel $\mathit{\Gamma}$ to the irreducible one $\mathit{\Gamma}_0$ and to R, see Fig. 1. To simplify the notation, I will not explicitly indicate external and internal variables, frequencies and momenta, and indicate by $\odot$ the summation over the internal ones. With those conventions the Bethe-Salpeter equation reads

$$\mathit{\Gamma} = \mathit{\Gamma}_0 + \mathit{\Gamma}_0 \odot R \odot \mathit{\Gamma}\,, \tag{8}$$

so that $\Gamma^\omega = \Gamma_0 + \Gamma_0 \odot R^\omega \odot \Gamma^\omega$ and $\Gamma^q = \Gamma_0 + \Gamma_0 \odot R^q \odot \Gamma^q$ with the same Γ_0, since by construction Γ_0 is analytic at $\omega = \mathbf{q} = 0$. Solving for Γ_0 one readily finds that

$$\Gamma = \Gamma^\omega + \Gamma^\omega \odot \Big(R - R^\omega\Big) \odot \Gamma = \Gamma^q + \Gamma^q \odot \Big(R - R^q\Big) \odot \Gamma\,, \tag{9}$$

where

$$\begin{aligned} R - R^\omega &\simeq -\frac{\partial f\left(\varepsilon_{\mathbf{k}}\right)}{\partial \varepsilon_{\mathbf{k}}}\, \delta\left(i\varepsilon\right)\, Z^2_{\mathbf{k}}\, \frac{\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}}{i\omega - \varepsilon_{\mathbf{k}+\mathbf{q}} + \varepsilon_{\mathbf{k}}}\,, \\ R - R^q &\simeq -\frac{\partial f\left(\varepsilon_{\mathbf{k}}\right)}{\partial \varepsilon_{\mathbf{k}}}\, \delta\left(i\varepsilon\right)\, Z^2_{\mathbf{k}}\, \frac{i\omega}{i\omega - \varepsilon_{\mathbf{k}+\mathbf{q}} + \varepsilon_{\mathbf{k}}}\,, \end{aligned} \tag{10}$$

do not involve anymore $R_{\text{incoh.}}$, at the expense of introducing two unknown objects, Γ^ω and Γ^q. Those are actually not independent since, e.g.,

$$\Gamma^q = \Gamma^\omega + \Gamma^\omega \odot \Big(R^q - R^\omega\Big) \odot \Gamma^q\,. \tag{11}$$

Conventionally one uses Γ^ω and define the Landau's f-parameters through[2]

$$f_{\mathbf{kp}} = Z_{\mathbf{k}}\, Z_{\mathbf{p}}\, \Gamma^\omega(0, \mathbf{k}; 0, \mathbf{p}; 0, \mathbf{p}; 0, \mathbf{k})\,, \tag{12}$$

where I reintroduced the external variables according to figure 1. Exploiting Ward's identities one can derive the known Fermi-liquid expressions, in terms of the above-defined f-parameters and of the unknown dispersion $\varepsilon_{\mathbf{k}}$, of the linear response functions at small ω and $\mathbf{q}$ for all conserved quantities, for which we refer, e.g., to Ref. [2].

2 Ordinary Kondo physics at the Mott transition

Commonly one cannot model the incoherent background and therefore the best one can do is invoking Landau's Fermi-liquid theory to get rid of $R_{\text{incoh.}}$ and $G_{\text{incoh.}}$. There is however a situation where we can proceed a bit further. Let us imagine to be in a strongly correlated metal phase close to a Mott transition, i.e., a metal-to-insulator transition driven by the electron-electron repulsion. In this circumstance we can grasp what the incoherent background represents. Indeed, in the Mott insulating phase the low-energy coherent component of the Green function, $G_{\text{coh.}}$, has disappeared, while the incoherent $G_{\text{incoh.}}$ must describe the atomic-like excitations of the insulator. I shall assume that the insulator has low-energy degrees of freedom, which cannot involve the charge, since its fluctuations are suppressed, but may involve the spin and/or, if present, the orbital degrees of freedom. We can thus imagine that, in the metal phase contiguous to the Mott insulator, $G_{\text{incoh.}}$ still describes the same atomic-like excitations, though coexisting with low-energy quasiparticle excitations. I shall indicate $G_{\text{coh.}}(i\varepsilon, \varepsilon_{\mathbf{k}})/Z_{\mathbf{k}} = 1/\big(i\varepsilon - \varepsilon_{\mathbf{k}}\big)$ and $G_{\text{incoh.}}/(1-Z)$ with solid and dashed lines, respectively, see Fig. 2. Accordingly, the irreducible vertex becomes $Z(1-Z)\Gamma_0$. Among all irreducible scattering processes that couple among each

[2] Note that the two expressions in Eq. (10) are finite only at $\varepsilon = 0$, so that one only needs the vertex at zero Matsubara frequency in the calculation of linear response functions.

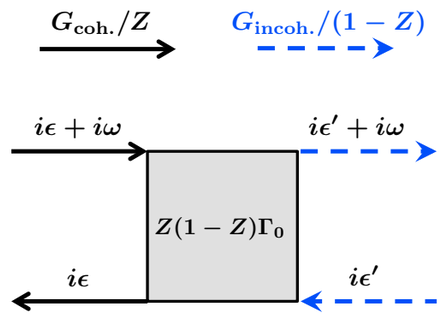


Fig. 2: *Coherent and incoherent components of the Green function, and the irreducible scattering vertex among them that can transfer low frequency ω.*

other coherent and incoherent components, the only one that can transfer low energy is that depicted in Fig. 2. Since charge fluctuations cost energy in the insulator, that scattering vertex $Z(1-Z)\Gamma_0$ acts only in the spin and/or orbital channels. For instance, in the single-band case that vertex should describe a spin exchange between itinerant quasiparticles and localized moments. In other words, the strongly correlated metal close to the Mott transition should behave similarly to a Kondo lattice model, i.e., conduction electrons coupled by a spin exchange $J = Z(1-Z)\Gamma_0$ to local moments, with the major difference that J is not an Hamiltonian parameter but it is self-consistently determined by the fully interacting theory.

The above very crude arguments that suggest a similarity between the physics of the Kondo effect and that of the Mott transition turn into a rigorous proof in lattices with infinite coordination number, $z \to \infty$, a limit in which the so-called dynamical mean-field theory (DMFT) [3] becomes exact. Within DMFT a lattice model is mapped onto an Anderson impurity model coupled to a bath. The mapping is exact for $z \to \infty$ provided a self-consistency condition between the local Green function of the bath and the impurity Greens function is fulfilled. Even though the mapping strictly holds only for $z \to \infty$, the previous heuristic arguments point to a more general validity, with the due differences coming from the fact that spatial fluctuations, which can be neglected in infinitely coordinated lattices, grow in importance as the coordination number decreases.

The obvious step further is therefore how to export the well-established knowledge of the Kondo effect to the physics of the Mott transition. Here one has to face two problems:

1. Even when the two models are rigorously mappable onto each other, i.e., in the limit of infinite coordination number, the mapping holds only under a self-consistency condition. How does such a condition affect the physics?

2. When the lattice has finite coordination, spatial correlations cannot be neglected anymore, e.g., the single-particle self-energy acquires momentum dependence. How does the physics across the Mott transition change?

In what follows I will just touch the first issue, which is also the simpler, assuming a model in an infinitely coordinated lattice.

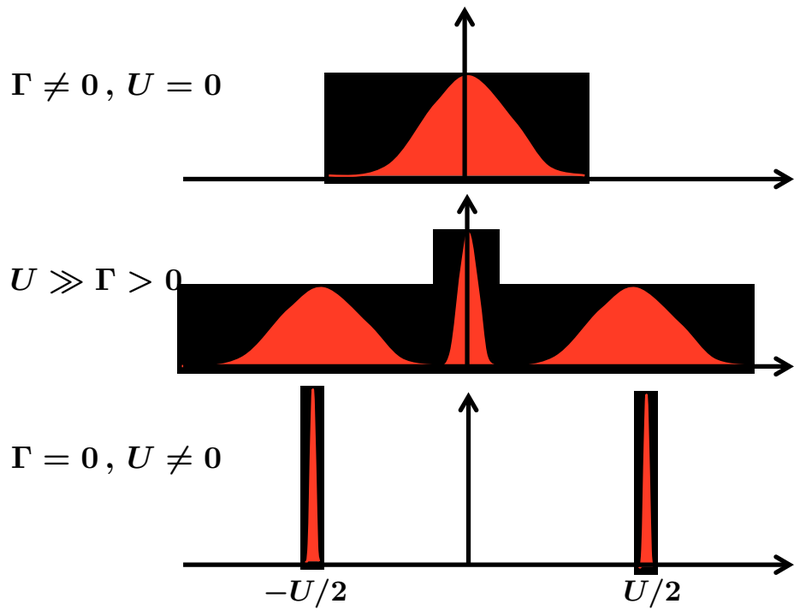


Fig. 3: *Sketch of the impurity density of states in three limiting cases.*

2.1 Role of the DMFT self-consistency condition

Let us start from the simplest case of the single-band Hubbard model at half-filling. Here the mapping is simply that onto a single-orbital Anderson impurity model (AIM) with Hamiltonian

$$\mathcal{H}_{\rm AIM} = \sum_{\mathbf{k}\sigma} \varepsilon_{\mathbf{k}}\, c^\dagger_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \sum_{\mathbf{k}\sigma} V_{\mathbf{k}} \left(c^\dagger_{\mathbf{k}\sigma} d_\sigma + d^\dagger_\sigma c_{\mathbf{k}\sigma} \right) + \frac{U}{2}\big(n-1\big)^2 \,, \tag{13}$$

where $c_{\mathbf{k}\sigma}$ and d_σ are the annihilation operators of the conduction and impurity electrons, respectively, the bath dispersion $\varepsilon_{\mathbf{k}}$ is measured with respect to the chemical potential and finally $n = \sum_\sigma d^\dagger_\sigma d_\sigma$ is the occupation of the impurity level. The model (13) depends actually on two quantities: the Hubbard repulsion U and the so-called hybridization function

$$\Gamma(\varepsilon) = \pi \sum_{\mathbf{k}} |V_{\mathbf{k}}|^2\, \delta\big(\varepsilon - \varepsilon_{\mathbf{k}}\big)\,. \tag{14}$$

When $U = 0$ and $\Gamma(\varepsilon) \neq 0$, the impurity density of states (DOS), $\mathcal{N}(\varepsilon)$, which was a δ-function centered at the chemical potential $\varepsilon = 0$ in the absence of hybridization with the bath, becomes in its presence a Lorentzian of width $\Gamma \simeq \Gamma(0)$, see top panel in Fig. 3. On the contrary, when $U \neq 0$ and $\Gamma(\varepsilon) = 0$, the isolated impurity is singly occupied in its ground state, so that its DOS, which measures at zero temperature the probability of removing, at $\varepsilon < 0$, or adding, at $\varepsilon > 0$, an impurity electron, displays two δ-peaks at $\varepsilon = \pm U/2$, see bottom panel in Fig. 3, where $U/2$ is the energy cost of the empty or doubly occupied impurity states. Those side

peaks are known as the Hubbard bands. When both U and $\Gamma(\varepsilon)$ are non zero, the DOS actually displays both features, namely a roughly Lorentzian peak at $\varepsilon = 0$, whose width is renormalized downwards by U, $\Gamma \to \Gamma_* = Z\,\Gamma$ with $Z < 1$, and two side-peaks centered at $\varepsilon = \pm U/2$ that are broadened by hybridization by an amount $\propto \Gamma$, see middle panel in Fig. 3.

Remarkably, the central peak exists for any value of U, even if bigger than any other energy scale. If U is very large, the impurity is singly occupied by either a spin up or down electron and thus essentially behaves as a spin-1/2 local moment. Nonetheless, the system can still gain hybridization energy by screening the impurity spin through the conduction electrons, what is named the *Kondo effect*. As a result, a tiny fraction of the impurity DOS is promoted at the chemical potential $\varepsilon = 0$ and gives rise to a very narrow peak, the so-called Kondo or Abrikosov-Suhl resonance. Its width $\Gamma_* = Z\,\Gamma \ll \Gamma$ defines the so-called Kondo temperature $T_K = \Gamma_*$, above which screening is not anymore effective. In other words, for temperatures $T > T_K$ the impurity behaves effectively as a *free* spin-1/2 and the Kondo resonance has disappeared.

This is, in brief, the physical behavior of the single-orbital AIM without any DMFT self-consistency. The latter roughly amounts to requiring that the hybridization function, $\Gamma(\varepsilon)$ of Eq. (14), has a similar shape to the impurity DOS. Therefore, once self-consistency is imposed, the effective impurity model is defined by a $\Gamma(\varepsilon)$ that also displays a peak of width Γ_* at the Fermi level, separated from two higher-energy Hubbard side-bands, see middle panel in Fig. 3. As U increases the peak at the Fermi level of $\Gamma(\varepsilon)$ thus becomes narrower and narrower until, at a critical U_c, Kondo screening of the impurity spin by the conduction bath is not anymore sustainable and the Abrikosov-Suhl resonance disappears, i.e., $\Gamma_* \to 0$. Above U_c the impurity DOS, which is also the local Green function of the lattice model, only displays two well separated Hubbard bands; the system is therefore turned into a Mott insulator.

– The first important role of the self-consistency is thus to push down to finite $U = U_c$ what in the impurity model without self-consistency happens only at $U = \infty$, i.e., the disappearance of the Kondo resonance.

In the single-orbital AIM the impurity magnetic susceptibility, $\chi_{\text{imp.}} \sim 1/\Gamma_*$, grows more and more as $\Gamma_* \to 0$. This suggests that the lattice model counterpart should develop some kind of magnetic instability before the Mott transition. Such instability is forbidden in the Anderson impurity model without DMFT self-consistency, since spin $SU(2)$ cannot be locally broken, but it might occur when self-consistency is enforced because a global spontaneous $SU(2)$ symmetry breaking is instead allowed. This is not at all unexpected. Indeed, local moments develop as the metal moves close to the Mott transition; these moments must order one way or another to get free of their $\ln 2$ entropy.

– We can therefore argue that another important effect of self-consistency is to transform the impurity instabilities into genuine bulk instabilities of the corresponding lattice model, which may thus drive a transition a into symmetry broken phases prior or concurrently with the Mott transition.

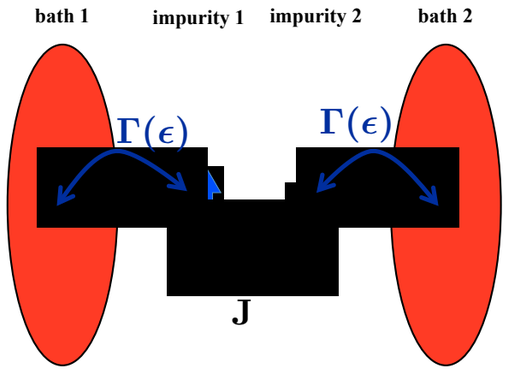


Fig. 4: *The two-impurity Anderson model. Each impurity is coupled by hybridization, parametrized here by the hybridization function $\Gamma(\varepsilon)$, with its own conduction bath. In addition the two impurities are coupled among each other by an antiferromagnetic spin-exchange.*

3 Exotic Kondo physics at the Mott transition

The last conjecture entails appealing scenarios which might be realized in lattice models that map within DMFT onto impurity models with a richer phase diagram than the simple single-orbital one. There is indeed a whole zoo of impurity models with varied physical properties. I note that the metallic phase close to the Mott transition corresponds by DMFT self-consistency to an Anderson impurity model with $U \gg \Gamma_*$ that suppresses valence fluctuations. In this regime the model becomes equivalent to a Kondo model where the impurity effectively behaves as a local moment with spin magnitude S, generically greater that 1/2, and eventually endowed with additional internal degrees of freedom brought, e.g., by orbital degeneracy as in the case of partially filled d or f shells.

Kondo models describing a spin-S impurity, with no other internal degrees of freedom, coupled to k-channels of spin-1/2 conduction electrons are divided into: (1) $k > 2S$ overscreened Kondo models; (2) $k = 2S$ screened Kondo models; and (3) $k < 2S$ underscreened Kondo models. Overscreened Kondo models are potentially interesting since they display instabilities in several channels [4], yet they will never appear in DMFT since by construction a lattice model in infinitely coordinated lattices maps unavoidably onto an impurity model in which the number of degrees of freedom of the impurity is the same as that of the conduction bath, i.e., $k = 2S$ in the above example.

3.1 The two-impurity model

There is however another much more promising class of impurity models characterized by the existence of internal degrees of freedom of the impurity besides the spin, and, more importantly, by an additional Hamiltonian parameter J that is able to quench those degrees of freedom and

thus competes against Kondo screening. Out of this competition a rich phase diagram emerges, which generally includes a quantum critical point or a narrow crossover region that separates the phase in which the impurity degrees of freedom are quenched by Kondo screening from that in which quenching is due to J. The best-known representative of this class is the two-impurity Anderson model [5], which I shall now discuss as the prototypical example. This model is depicted in Fig. 4; it consists of two equivalent single-orbital AIMs in which the two impurities are not only hybridized each to its own bath, but also coupled among each other by an antiferromagnetic spin-exchange J. The Hamiltonian reads

$$\mathcal{H}_{\text{2AIM}} = \mathcal{H}_{\text{AIM-1}} + \mathcal{H}_{\text{AIM-2}} + J\,\mathbf{S}_1 \cdot \mathbf{S}_2\,, \tag{15}$$

where $\mathcal{H}_{\text{AIM-a}}$ is the AIM Hamiltonian Eq. (13) of the impurity $a = 1, 2$, and

$$\mathbf{S}_a = \frac{1}{2} \sum_{\alpha\beta} d^\dagger_{a\alpha}\, \boldsymbol{\sigma}\, d_{b\beta}\,,$$

its spin operator with $\boldsymbol{\sigma} = \left(\sigma_x, \sigma_y, \sigma_z\right)$ the Pauli matrices. The Hamiltonian (15) has three relevant parameters, U, the hybridization function $\Gamma(\varepsilon)$, by definition equal for each impurity, and the exchange J. If $J = 0$, each impurity is Kondo screened by its bath on the energy scale given by the Kondo temperature T_K. If, on the contrary, $\Gamma(\varepsilon) = 0$ but $J \neq 0$, the impurities are decoupled from the baths but coupled among each other into a spin-singlet configuration. Both cases are stable in the sense that no degeneracy is left to be lifted. If all parameters are finite, the Kondo screening, with scale T_K, competes against the direct exchange J. Therefore, if $T_K \gg J$, the system prefers to Kondo screen each impurity with its bath. On the contrary, if $J \gg T_K$, the two impurities lock into a singlet state that is transparent to the conduction electrons. These two limiting cases, which I shall denote as screened, $T_K \gg J$, and unscreened, $J \gg T_K$, correspond to two different phases separated by a genuine quantum critical point (QCP) at $T_K \sim J$. Its critical properties have been uncovered in great detail [6]. Specifically, at the QCP the model display logarithmically singular susceptibilities in several channels:

(1) the "antiferromagnetic" channel defined by the operators

$$\boldsymbol{\Delta}_{\text{AFM}} = \mathbf{S}_1 - \mathbf{S}_2\,, \tag{16}$$

(2) the "hybridization" channels

$$\Delta_x = \sum_\sigma \left(d^\dagger_{1\sigma} d_{2\sigma} + H.c.\right), \qquad \Delta_y = i \sum_\sigma \left(d^\dagger_{1\sigma} d_{2\sigma} - H.c.\right), \tag{17}$$

(3) the spin-singlet Cooper channel

$$\Delta = d^\dagger_{1\uparrow} d^\dagger_{2\downarrow} + d^\dagger_{2\uparrow} d^\dagger_{1\downarrow}\,, \tag{18}$$

and its hermitian conjugate $\Delta^\dagger$.

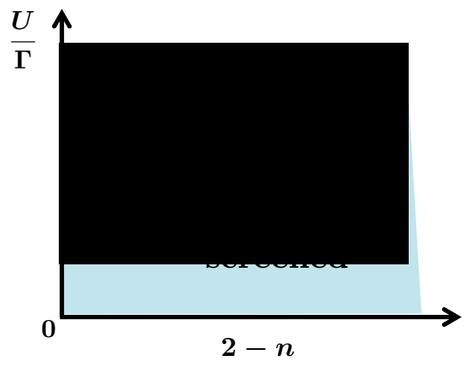


Fig. 5: *Sketch of phase diagram at fixed J/U as function of U/Γ and doping $2-n$ of the two impurities away from half-filling.*

On the contrary, the impurity charge susceptibility is not singular since charge fluctuations are suppressed by the large U. As a consequence, the QCP is stable upon *doping* the impurity site, which corresponds to changing the position of the impurity level so that $n \equiv \langle n_1+n_2\rangle \neq 2$. The phase diagram is schematically shown in Fig. 5. One can observe that the QCP at half-filling, $n=2$, is actually the endpoint of a whole critical line that moves upwards in U/Γ at fixed J/U away from half-filling. In other words, if one starts from the unscreened phase at half-filling and dopes the impurity, at some doping the critical line will be crossed.

The dynamical behavior of the impurity DOS across the QCP has been uncovered quite in detail [7, 8]. The vicinity of the QCP is controlled by two energy scales. One is smooth across the transition; it was denoted as T_+ in Ref. [7] and was found to be of the order $\max\big(T_K, J\big)$, where T_K is the Kondo temperature at $J=0$. The other energy scale, T_-, measures the deviation from the QCP. I recall that the Kondo temperature T_K at $J=0$ is defined by $\Gamma \equiv \Gamma(0)$ and U according to

$$T_K(\Gamma,U) = U\sqrt{\frac{\Gamma}{2U}}\,\exp\left(-\frac{\pi U}{8\Gamma}-\frac{\pi\Gamma}{2U}\right)\,,$$

and decreases with decreasing Γ or increasing U. Let us for instance assume that J and U are fixed while Γ varies. In this case the QCP is identified by $\Gamma=\Gamma_c$ such that $T_K(\Gamma_c,U)\simeq J$, and $T_- \propto \big(\Gamma-\Gamma_c\big)^2$, vanishing quadratically at the transition. It was found [7] that the impurity DOS as obtained by numerical renormalization group is well fit at low energy by the expression

$$\mathcal{N}_\pm(\varepsilon) = \frac{1}{2\pi\Gamma}\left(\frac{T_+^2}{\varepsilon^2+T_+^2} \pm \frac{T_-^2}{\varepsilon^2+T_-^2}\right), \tag{19}$$

where the $+$ refers to the Kondo screened phase, $\Gamma>\Gamma_c$, and the $-$ to the unscreened one, $\Gamma<\Gamma_c$. Right at the QCP

$$\mathcal{N}_c(\varepsilon) = \frac{1}{2\pi\Gamma_c}\,\frac{T_+^2}{\varepsilon^2+T_+^2}\,. \tag{20}$$

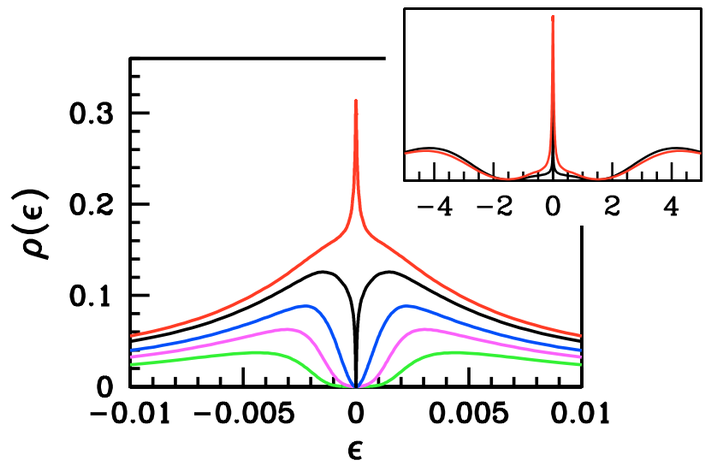


Fig. 6: *Low-energy DOS of the two-impurity Anderson model across the phase transition. The calculation are done at* $U = 8$*,* $J = 0.00125$ *and the curves, from top to bottom, corresponds to* $\Gamma = 0.44, 0.42, 0.4, 0.35, 0.3$ *in units the conduction bandwidth. The critical point is at* $\Gamma = \Gamma_c \simeq 0.43956$*. In the screened phase a narrow Kondo resonance is present, red curve. In the unscreened phase instead the DOS has a pseudo-gap. In the inset the DOS on a larger scale is shown, where the Hubbard bands are visible [From Ref. [8]]. The colors of the curves correspond to those in the main panel.*

I note that the DOS in the screened phase is the sum of two Lorentzians, one of width T_+ and a much narrower one of width T_- that vanishes at the QCP. Here only the broader peak remains. On the unscreened side of the transition, the DOS is the difference of two Lorentzians, and its value at the chemical potential vanishes – a pseudo-gap emerges by the disappearance of Kondo screening. I further note that such a pseudo-gap is not to be confused with the much larger one that separates the lower from the upper Hubbard band, see the inset of Fig. 6. The two are indeed controlled by different energy scales, U the latter and $T_- \ll U$ the former.

When the impurity is doped, i.e., its occupation number n deviates from half-filling $n = 2$, the low-energy DOS was found [7] to be still of the form Eq. (19),

$$\mathcal{N}_\pm(\varepsilon) = \frac{\cos^2\nu}{2\pi\Gamma}\left(\frac{T_+^2+\mu_\pm^2}{\left(\varepsilon+\mu_\pm\right)^2+T_+^2} \pm \cos 2\nu\,\frac{T_-^2}{\varepsilon^2+T_-^2}\right), \tag{21}$$

where $\mu_\pm = \pm T_+ \sin 2\nu$ measures the deviation away from half-filling. Remarkably, the narrower Lorentzian remains peaked at the chemical potential, so that the pseudo-gap in the unscreened phase is pinned at the Fermi level.

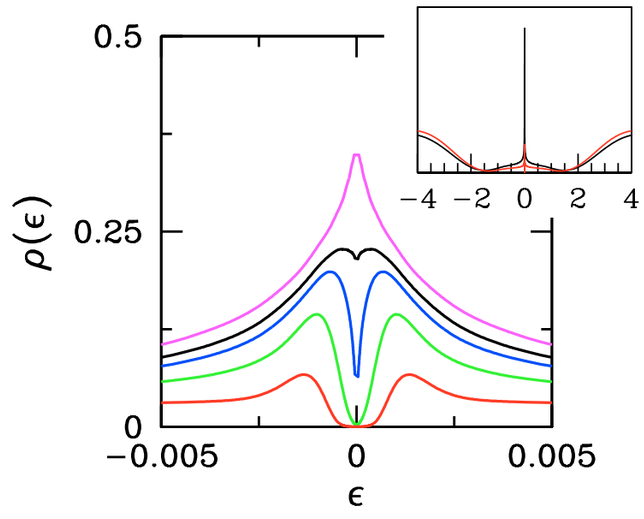


Fig. 7: *Low-energy DOS of the two-impurity Anderson model Eq. (23) for* $U = 8$, $t_\perp = 0.05$, *and, from top to bottom,* $\Gamma = 0.5, 0.47, 0.45, 0.4, 0.3$ *in units the conduction bandwidth. In the inset the DOS on a larger scale is shown, where the Hubbard bands are visible [From Ref. [8]]. The colors of the curves correspond to those in the main panel.*

3.1.1 Explicit symmetry breaking

Let us now discuss more in detail the role of the operators in Eq. (17), focusing in particular on Δ_x that describes a direct real hybridization among the impurities. This operator breaks the $U(1)$ orbital symmetry

$$\begin{aligned} d_{1\sigma} &\to \mathrm{e}^{+i\varphi}\, d_{1\sigma}\,, \qquad & c_{1\mathbf{k}\sigma} &\to \mathrm{e}^{+i\varphi}\, c_{1\mathbf{k}\sigma}\,, \\ d_{2\sigma} &\to \mathrm{e}^{-i\varphi}\, d_{2\sigma}\,, \qquad & c_{2\mathbf{k}\sigma} &\to \mathrm{e}^{-i\varphi}\, c_{2\mathbf{k}\sigma}\,, \end{aligned} \tag{22}$$

of the original Hamiltonian (15), where $d_{a\sigma}$ and $c_{a\mathbf{k}\sigma}$ are the annihilation operators of the $a = 1, 2$ impurity and conduction electrons, respectively. In the language of critical phenomena, such $U(1)$ symmetry breaking is therefore a relevant perturbation that spoils the QCP. In other words, if the Hamiltonian were not invariant under that symmetry, there would not be anymore a quantum phase transition but just a crossover between the screened and unscreened phases. Suppose we consider the following Hamiltonian instead of that in Eq. (15)

$$\mathcal{H}'_{\text{2AIM}} = \mathcal{H}_{\text{AIM-1}} + \mathcal{H}_{\text{AIM-2}} - t_\perp\, \Delta_x\,. \tag{23}$$

This Hamiltonian is not invariant under the $U(1)$ symmetry in (22), therefore should not possess the above QCP. On the other hand, if $U \gg t_\perp$, $\mathcal{H}'_{\text{2AIM}}$ of Eq. (23) can be mapped onto $\mathcal{H}_{\text{2AIM}}$ with $J = 4t_\perp^2/U$, which instead has the QCP. How can we reconcile this apparent paradox? The answer is quite instructive. Indeed, $\mathcal{H}'_{\text{2AIM}}$ of Eq. (23) does map onto $\mathcal{H}_{\text{2AIM}}$ of Eq. (15) with

$J = 4t_\perp^2/U$, but just at leading order in $1/U$. What really prevents the system from encountering the QCP are symmetry variant sub-leading terms, with coupling constant $h_x \propto \Gamma^2\, t_\perp/U^2 \ll J$, which actually correspond to a direct hybridization among the two conduction baths. In other words, a hierarchy of energy scales emerges naturally at large U from the single $t_\perp$: J, which alone would drive the system across the phase transition, and a much smaller scale $h_x \ll J$ that allows the system crossing from the screened phase to the unscreened one without eventually passing through the QCP. In the language of critical phenomena, we could state that even though the system does not cross the QCP, it gets very close to it. Practically, this implies that the quantum phase transition turns into a very sharp crossover between screened and unscreened phases that, for many purposes, is indistinguishable from a phase transition. In Fig. 7 the impurity DOS of the Hamiltonian (23) is shown, with parameters $U = 8$ and $t_\perp = 0.05$, in units of the conduction bandwidth, such that $4t_\perp^2/U$ is equal to the value of J in Fig. 6. We first observe that in this case the DOS is always finite at $\varepsilon = 0$, though very small in the unscreened phase. In addition we can note that, despite the opening of the incomplete pseudo-gap does not occur through a phase transition, nonetheless it is extremely sharp.
In conclusion, we can thus interpret the above results as those of the model Eq. (15) at $J = 4t_\perp^2/U$ in the presence of a small $h_x \ll J$ symmetry breaking field. From this perspective, inside the unscreened phase J is responsible for the pseudo-gap opening, while the much smaller h_x for the partial filling of that same gap. A question immediately arises. How is it possible that the unscreened phase, even though pseudo-gapped, namely despite the impurities have a vanishing quasiparticle residue $Z = 0$, is able to respond so efficiently to the small symmetry breaking field $h_x \ll J$?

3.1.2 How can a pseudo-gap symmetry invariant phase develop a symmetry variant order parameter?

This question has been addressed in Ref. [9] recognizing a curious analogy between this problem and that of disordered s-wave superconductors. I will briefly sketch this relationship since I believe it reveals a basic feature that can be used in many other contexts.
The expression Eq. (19) of the low-energy DOS corresponds to the impurity Green function in Matsubara frequencies

$$\mathcal{G}(i\varepsilon) = \frac{1}{2\Gamma}\left(\frac{T_+}{i\varepsilon + iT_+\,\mathrm{sign}(\varepsilon)} \pm \frac{T_-}{i\varepsilon + iT_-\,\mathrm{sign}(\varepsilon)}\right) + \mathcal{G}_{\text{incoh.}}(i\varepsilon)\,,\,, \tag{24}$$

where $\pm$ refers, as before, to the screened and unscreened phases, respectively, and $\mathcal{G}_{\text{incoh.}}(i\varepsilon)$ is the high-energy contribution from the Hubbard sidebands. In turn the Green function satisfies the Dyson equation

$$\mathcal{G}(i\varepsilon)^{-1} = \mathcal{G}_0(i\varepsilon)^{-1} - \Sigma(i\varepsilon) = i\varepsilon + i\Gamma\,\mathrm{sign}(\varepsilon) - \Sigma(i\varepsilon)\,, \tag{25}$$

where $\mathcal{G}_0(i\varepsilon)$ is the non-interacting Green function and $\Sigma(i\varepsilon)$ the impurity self-energy. We thus find that the impurity self-energy at low-energy and in the unscreened phase has the following

expression

$$\begin{aligned}\Sigma(i\varepsilon) &\simeq i\varepsilon - \frac{i}{4\varepsilon}\,\frac{T_+T_-}{T_+-T_-} - \frac{i}{4}\,\frac{T_++T_-}{T_+-T_-}\,\text{sign}(\varepsilon) - i\,\frac{\varepsilon}{T_+-T_-}\\ &\equiv i\varepsilon - i\,\frac{\varepsilon}{Z(i\varepsilon)}\,,\end{aligned} \tag{26}$$

and diverges at $\varepsilon \to 0$. The quasiparticle residue $Z(i\varepsilon)$ thus vanishes at $\varepsilon = 0$.
Let us consider instead a disordered metal in the normal phase, whose self-energy is

$$\Sigma(i\varepsilon) = \frac{i}{2\tau}\,\text{sign}(\varepsilon) \equiv i\varepsilon - i\varepsilon\,\eta(i\varepsilon)\,, \tag{27}$$

where τ is the relaxation time, and $\eta(i\varepsilon)$ diverges at $\varepsilon \to 0$, which, in analogy with Eq. (26), can be interpreted as a vanishing quasiparticle residue. In the superconducting phase the self-energy acquires anomalous components and must be written as a two by two matrix

$$\hat{\Sigma}(i\varepsilon) = \begin{pmatrix}\Sigma_{11}(i\varepsilon) & \Sigma_{12}(i\varepsilon)\\ \Sigma_{21}(i\varepsilon) & \Sigma_{22}(i\varepsilon)\end{pmatrix}, \tag{28}$$

where $\Sigma_{22}(i\varepsilon) = -\Sigma_{11}(-i\varepsilon)$ and $\Sigma_{21}(i\varepsilon) = \Sigma_{12}(-i\varepsilon)^*$. Because of the perfect cancellation of the disorder-induced corrections to the self-energy and to the vertex in the s-wave Cooper channel, superconductivity regularizes the singularities brought by disorder below some low-energy scale Δ, the superconducting gap, leading to the following expressions of the self-energy matrix elements

$$\begin{aligned}\Sigma_{11}(i\varepsilon) &= i\varepsilon - i\varepsilon\,\eta\Big(i\,\sqrt{\varepsilon^2+\Delta^2}\Big)\,,\\ \Sigma_{12}(i\varepsilon) &= \qquad \Delta\,\eta\Big(i\,\sqrt{\varepsilon^2+\Delta^2}\Big)\,.\end{aligned} \tag{29}$$

A famous consequence of Eq. (29), known as *Anderson theorem*, is that T_c is independent of disorder strength, which readily follows from the BCS gap equation in the presence of an attraction λ

$$1 = \lambda\,\frac{T}{V}\sum_{i\varepsilon}\sum_{\mathbf{k}}\frac{\eta\Big(i\,\sqrt{\varepsilon^2+\Delta^2}\Big)}{\big(\varepsilon^2+\Delta^2\big)\,\eta\Big(i\,\sqrt{\varepsilon^2+\Delta^2}\Big)^2+\varepsilon_{\mathbf{k}}^2}\,. \tag{30}$$

The authors of [9] argued, in analogy with disordered s-wave superconductors, that the corrections brought by J to the self-energy and to the vertex in the Δ_x-channel of Eq. (17) cancel each other also in the impurity model. If one then considers a model with Hamiltonian

$$\mathcal{H} = \mathcal{H}_{\text{2AIM}} - h_x\,\Delta_x\,, \tag{31}$$

see equations (15) and (17), with a symmetry breaking term $h_x \ll J$, the impurity self-energy also becomes a two by two matrix with elements $\Sigma_{ab}(i\varepsilon)$, with $a, b = 1, 2$ labelling the impurities. Following the above arguments one should expect that h_x brings about a low energy scale Δ that cutoffs the singularities of $\Sigma(i\varepsilon)$ in Eq. (26) so that

$$\begin{aligned}\Sigma_{11}(i\varepsilon) &= i\varepsilon - i\varepsilon\,Z\Big(i\,\sqrt{\varepsilon^2+\Delta^2}\Big)^{-1} = \Sigma_{22}(i\varepsilon)\,,\\ \Sigma_{12}(i\varepsilon) &= \qquad \Delta\,Z\Big(i\,\sqrt{\varepsilon^2+\Delta^2}\Big)^{-1} = \Sigma_{21}(i\varepsilon)\,.\end{aligned} \tag{32}$$

This ansatz was shown to fit extremely well the numerical data obtained in Ref. [7] by directly solving the model in Eq. (31) via the numerical renormalization group. This result demonstrates, from a quite general perspective, how a pseudo-gapped symmetry invariant phase can nonetheless develop a sizeable symmetry breaking order parameter, which was used by the authors of Ref. [9] to interpret the phase diagram of a model that maps by DMFT onto the two-impurity model Eq. (15), which I describe below.

3.2 The lattice model counterpart of the two-impurity model

Let us consider the two band Hubbard model in a infinitely coordinated Bethe lattice with Hamiltonian

$$\mathcal{H} = -\frac{t}{\sqrt{z}} \sum_{a=1}^{2} \sum_{\langle i,j \rangle \, \sigma} \left(c^{\dagger}_{ai\sigma} c_{aj\sigma} + H.c. \right) + \frac{U}{2} \sum_{i} \left(n_i - 2 \right)^2 - 2J \sum_{i} \left(T^2_{i\,x} + T^2_{i\,y} \right), \quad (33)$$

where $z \to \infty$ is the coordination number, $n_i = n_{1i} + n_{2i} = \sum_{a=1}^{2} \sum_{\sigma} c^{\dagger}_{ai\sigma} c_{ai\sigma}$ is the charge density at site i and $T_{i\,\alpha}$, $\alpha = x, y, z$, are the components of the orbital pseudo-spin $\mathbf{T}_i$ defined through

$$\mathbf{T}_i = \frac{1}{2} \sum_{a,b=1}^{2} \sum_{\sigma} c^{\dagger}_{ai\sigma} \, \boldsymbol{\sigma}_{ab} \, c^{\dagger}_{bi\sigma} \, , \quad (34)$$

with $\boldsymbol{\sigma}$ the Pauli matrices. This model describes an $e \times E$ Jahn-Teller effect within the antiadiabatic approximation [10]. Alternatively, one may rewrite Eq. (33) as

$$\mathcal{H} = -\frac{t}{\sqrt{z}} \sum_{a=1}^{2} \sum_{\langle i,j \rangle \, \sigma} \left(c^{\dagger}_{ai\sigma} c_{aj\sigma} + H.c. \right) + \frac{U}{2} \sum_{i} \sum_{a=1}^{2} \left(n_{ai} - 1 \right)^2 + \sum_{i} \left(4J \mathbf{S}_{1i} \cdot \mathbf{S}_{2i} + V n_{1i} n_{2i} \right), \quad (35)$$

where $V = U - J$, which represents two Hubbard models coupled by an antiferromagnetic exchange and by a charge repulsion.

Finally, if we interchange spin with orbital indices, the Hamiltonian (33) can also be written as

$$\mathcal{H} = -\frac{t}{\sqrt{z}} \sum_{a=1}^{2} \sum_{\langle i,j \rangle \, \sigma} \left(c^{\dagger}_{ai\sigma} c_{aj\sigma} + H.c. \right) + \frac{U}{2} \sum_{i} \left(n_i - 2 \right)^2 - 2J \sum_{i} \left(S^2_{i\,x} + S^2_{i\,y} \right), \quad (36)$$

where now $S_{i\,\alpha}$ are the components of the total spin at site i, which is a two-band Hubbard model with a single-ion anisotropy that favors the spin to lie in the x-y plane.

Let us for simplicity focus just on the Hamiltonian (33), or its equivalent representation (35). The Hamiltonian contains three parameters, the conduction bandwidth $W = 4t$, the Hubbard U and the exchange J. The latter mediates pairing in the s-wave channel of Eq. (18) that is however contrasted by U. The net effect is a *bare* scattering amplitude in that pairing channel $A_0 = \rho_0 \big(U - 2J \big)$, where ρ_0 is the non-interacting DOS at the Fermi energy. At fixed $J \ll W$, as we shall assume hereafter, mean-field theory predicts upon increasing U a BCS superconducting domain that extends from $U = 0$, where pairing is maximum, to $U = 2J$, where pairing

disappears. Above $2J$ the ground state of the model (33) should turn into a normal metal. This prediction is actually independent of the coordination number $z > 2$ and the electron density.

When $z \to \infty$ this lattice model maps by DMFT onto the two-impurity model Eq. (15) with the addition of a charge repulsion V among the impurities. This repulsion is irrelevant and does not change the phase diagram of the impurity model, which thus remains similar to that in Fig. 5. If DMFT self-consistency is imposed, as U increases at fixed $J \ll W$ and at half-filling, the lattice model is pushed towards a Mott transition, exactly like the single-band Hubbard model of section 2. In the impurity language, this transition corresponds to $T_K \to 0$. However, still in the metal phase before that happens, T_K will become comparable to J, even though the bare conduction bandwidth $W \gg J$. This is right the point where the impurity model crosses its QCP. Already before that happens, the impurity susceptibilities in the channels of equations (16), (17) and (18) will be strongly enhanced and thus may drive, through the self-consistency, a true bulk instability. An instability in the first channel Eq. (16) translates in the lattice model into an instability towards Néel antiferromagnetic order, with the two orbitals occupied by opposite spin electrons. The instability in channel (17) corresponds instead to a cooperative Jahn-Teller effect. However both (16) and (17) are particle-hole channels and thus they require nesting of the Fermi surface to drive a bulk instability. On the contrary, the particle-particle channel in Eq. (18) does not require any particular property of the Fermi surface to drive a superconducting instability, but just a finite DOS at the chemical potential. It is thus tempting to conclude that generically, i.e., in the absence of nesting and with finite DOS at the Fermi level, there must exist another superconducting dome besides the weak coupling $U < 2J \ll W$ BCS one, right next to the Mott transition. This expectation was confirmed by a full DMFT calculation in Ref. [10]. In Fig. 8 the superconducting gap of the model (33) is plotted at electron density $n = 2$ as a function of U and for different values of J. I note that for the largest $J = 0.15$, in units of the conduction bandwidth W, the gap is monotonically decreasing with U and disappears at $U \simeq 2J$ where the superconductor turns by a weakly first order transition into a Mott insulator. This insulating phase is non-magnetic with all sites occupied by two electrons, one on each orbital, coupled into a spin-singlet configuration; a local version of a valence-bond crystal. Already for a weaker $J = 0.1$ the gap becomes non monotonous; it first decreases then increases again before the first order Mott transition. For smaller $J = 0.05$ and $J = 0.02$, the superconducting phase splits, as anticipated, into two well separated domains. One appears at weak coupling and has a tiny BCS-like gap exponentially small in $1/J$, see the inset of Fig. 8. However, another bell-shaped superconducting dome emerges at strong coupling next to the Mott transition and displays a huge gap if compared with the weak coupling BCS one; a striking example of superconductivity boosted by strong correlations [11].

The physics of the impurity model (15) allows anticipating not only the phase diagram but also the dynamical properties of the lattice model. Within DMFT one can prevent superconducting symmetry breaking and thus access the unstable zero-temperature normal phase, whose single-particle self-energy was found [10] to be well fitted by that of the impurity model, see Eq. (26). Fig. 9 shows the impurity scales T_+ and T_- extracted by fitting the DMFT self-energy within the unstable normal phase through Eq. (26). We can observe that the impurity critical point

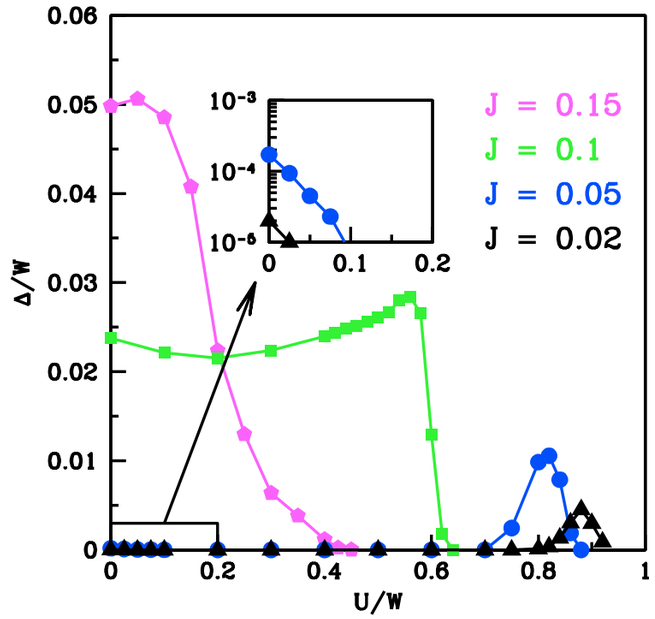


Fig. 8: *Superconducting gap of the model (33) at electron density $n = 2$ as function of U and for different values of J, in units of the conduction bandwidth W. In the inset the weak coupling gap is zoomed. [From Ref. [10].]*

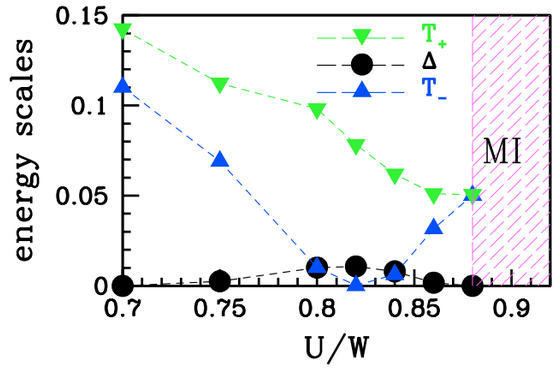


Fig. 9: *Impurity scales T_+ and T_-, see Eq. (26), extracted by the DMFT self-energy within the unstable normal phase, in comparison with the superconducting gap that is obtained once symmetry breaking is permitted. MI indicate the Mott insulating phase. [From Ref. [10].]*

is encountered before the Mott transition and, once one permits superconducting symmetry breaking, it corresponds to the maximum of the superconducting gap Δ. I remark that the

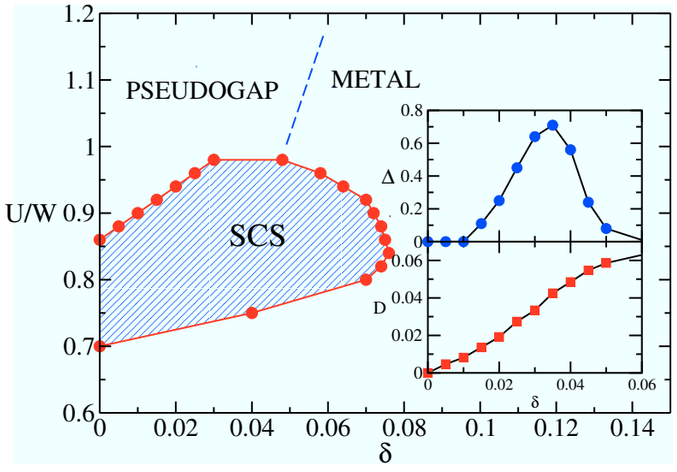


Fig. 10: *Phase diagram of the model Eq. (33) at $J = 0.05$ as function of U/W and doping δ away from half-filling. The insets show the dependence upon doping of the superconducting gap Δ and of the Drude weight D at $U = 0.92W$. [From Ref. [10].]*

Mott insulator appears now when $T_+ = T_-$ on the unscreened side of the impurity model, at which point the two Lorentzians that define the low-energy DOS $\mathcal{N}_-(\varepsilon)$, see Eq. (19), cancel each other. This is evidently different from the single-band case, where, as I mentioned, the transition occurs by the gradual disappearance of a Kondo-like resonance.

One can push the interpretation via the impurity model even further. Suppose we are in the Mott insulator at half-filling $n = 2$, which, as mentioned, corresponds in the impurity model to the unscreened phase with $T_+ = T_-$, and dope it, $n \to 2 - \delta$. According to Eq. (21) we should expect that the insulator turns into a pseudo-gapped normal metal as soon as $\delta > 0$. Moreover, upon increasing further the doping δ, the impurity model should cross its quantum critical point, see Fig. 5, which should correspond in the lattice model after DMFT self-consistency to reappearance of a superconducting dome. Once again this expectation was indeed confirmed. Fig. 10 shows the final phase diagram at $J = 0.05$ as function of U/W and doping $\delta > 0$. The superconducting dome extends at finite doping into a whole region such that, starting from the Mott insulator and doping it, one first finds a pseudo gapped normal phase that, upon further doping, turns into a superconductor that eventually disappears at higher doping into a well-behaved normal metal, i.e., not anymore pseudo gapped, see the insets of Fig. 10.

3.3 Landau-Fermi-liquid picture

It is impossible not to see striking similarities between the phase diagram in Fig. 10 and the phenomenology of high-T_c copper oxides. This is even more evident by the behavior of the Drude weight D, shown in the bottom inset of Fig. 10, which grows linearly in δ upon doping the Mott insulator because of the linear-in-δ filling of the pseudo-gap.

One may thus wonder whether and in which terms the predictive power of the impurity physics transferred to lattice models could be extended even beyond the limits of applicability of DMFT, namely even in realistic lattices with finite coordination number. I already mentioned how the emergence of a Kondo-like physics close to the Mott transition can be inferred quite generally from Landau-Fermi-liquid theory. It is therefore worth addressing how one can translate the physics of the two-impurity model in the language of Landau's Fermi-liquids. Since a local Fermi-liquid description can be defined also for impurity models [12], it is convenient to start with that and eventually extend it to bulk systems.
Concerning the impurity model Eq. (15), both screened and unscreened phases are Fermi-liquid like. While this is evidently the case in the screened phase, where ordinary Kondo effect takes place, it is by far less obvious in the unscreened one that is characterized by a singular impurity self-energy. In Ref. [7] it was shown that the conventional definition [13]

$$\rho_{\rm qp} = \frac{\mathcal{N}(0)}{Z} \,,$$

of the quasiparticle DOS at the impurity site and at the Fermi energy, where $\mathcal{N}(0)$ is the particle DOS at Fermi and Z the quasiparticle residue, must be generalized to account for unscreened Fermi-liquid phases into

$$\rho_{\rm qp} = \int \frac{d\varepsilon}{\pi} \, \frac{\partial f(\varepsilon)}{\partial \varepsilon} \, {\rm Im} \left(\mathcal{G}(\varepsilon + i0^+) \left[1 - \left. \frac{\partial \Delta(\zeta)}{\partial \zeta} \right|_{\zeta=\varepsilon+i0^+} - \left. \frac{\partial \Sigma(\zeta)}{\partial \zeta} \right|_{\zeta=\varepsilon+i0^+} \right] \right) , \quad (37)$$

where

$$\Delta(\zeta) = \int \frac{d\varepsilon}{\pi} \, \frac{\Gamma(\varepsilon)}{\zeta - \varepsilon} \simeq -i \, \Gamma(0) \, {\rm sign} \, ({\rm Im} \, \zeta) \;\; ,$$

is the Hilbert transform of the hybridization function, $\mathcal{G}(\zeta)$ and $\Sigma(\zeta)$ the impurity Green function and self-energy, respectively, continued to the complex frequency plane. In the screened phase of the model Eq. (15) and for negligible $\partial\Delta(\zeta)/\partial\zeta$ one gets the conventional result

$$\rho_{\rm qp+} = \frac{\mathcal{N}_+(0)}{Z} = \frac{1}{2\pi} \, \frac{T_+ + T_-}{T_+ \, T_-} \;\; .$$

In the pseudo-gap phase, even though Z vanishes, still $\rho_{\rm qp}$ in Eq. (37) has a well defined value since the singularity in the self-energy is cancelled by the vanishing DOS

$$\mathcal{N}_-(\varepsilon) = -\frac{1}{\pi} \, {\rm Im} \, \mathcal{G}(\varepsilon + i0^+) \, ,$$

at $\varepsilon = 0$, leading to a finite *quasiparticle* DOS at the Fermi level

$$\rho_{\rm qp-} = \frac{1}{\pi} \, \frac{T_+ + T_-}{T_+ \, T_-} \, , \quad (38)$$

despite the vanishing *particle* DOS. This is remarkable, since common wisdom would rather suggest that a singular self-energy is incompatible with Landau's Fermi-liquid theory.

In addition, even though a local Fermi-liquid description does not apply right at the critical point $T_- = 0$, still one can approach it from either Fermi-liquid side of the transition. In particular, it was explicitly verified [7] that the quasiparticle scattering amplitudes, defined in the screened phase through

$$A_i \equiv Z^2\,\rho_{\rm qp}\,\Gamma_i\,, \tag{39}$$

where Γ_i is the interaction vertex in channel i, tend to finite values approaching the critical point $T_- \to 0$, or equivalently $Z \to 0$, in all three relevant channels of equations (16)-(18), specifically $A \to 1$ in channels (16) and (17), and $A \to -2$ in channel (18). This confirms the expectation that vertex corrections cancel exactly self-energy ones, as assumed in the previously discussed Ref. [9].

Suppose we could export the above local Fermi-liquid results to the lattice model (33). As I mentioned, the bare scattering amplitude in the s-wave Cooper channel Eq. (18) is

$$A_0 = \rho_0\left(U - 2J\right),$$

where ρ_0 is the non-interacting DOS at the Fermi energy. For $J\rho_0 \ll 1$, as we assumed, and close to the Mott transition, the amplitude $A_0 > 0$ and thus one should not expect any superconductivity. In reality, the quasiparticle amplitude in that channel reads

$$A = \rho_{\rm qp}\left(Z^2\,\Gamma_U - 2Z^2\,\Gamma_J\right). \tag{40}$$

The contribution from the charge channel $Z^2\Gamma_U \sim ZU$ becomes negligible approaching the Mott transition; quasiparticles slow down and at the same time they undress from the strong repulsion. Moreover, just because they spend more time on each site before hopping to neighboring ones, the quasiparticles can take more advantage of the local J-term, so that it is well conceivable that $Z^2\Gamma_J \sim J$ is to a large extent unrenormalized by the proximity to a Mott transition, which once again entails cancellation of vertex and self-energy corrections. The outcome is that approaching the Mott transition

$$A \sim \rho_{\rm qp}\,(ZU - 2J) \simeq \rho_0\,\left(U - 2\frac{J}{Z}\right), \tag{41}$$

changes sign from positive to negative, thus permitting a superconducting instability to set in despite the bare value A_0 does not, as indeed found by DMFT [9, 10]. Moreover, A may now become of order $O(1)$ when $\rho_{\rm qp}\,J \simeq \rho_0 J/Z \sim 1$, despite $\rho_0 J \ll 1$, suggesting that superconductivity is effectively pushed to the maximum $T_c \sim 0.055\,g$ attainable at a given pairing strength g, again consistent with DMFT results.

In the impurity model the maximum of A occurs right at the critical point, beyond which, in the unscreened phase, A diminishes again [7]. Here, as I mentioned, one should not use anymore Eq. (39) to define the scattering amplitudes [7]. This suggests that the simple expression Eq. (41) is only valid in the counterpart of the screened phase and cannot be pushed till the Mott transition, which would otherwise imply the unphysical result $A \sim 1/Z \to \infty$; some readjustment must intervene before, which in the impurity model is the pseudo-gap opening that was also observed as a stable phase in the lattice model Eq. (33) away from half-filling.

4 Concluding remarks

In conclusion, the Landau-Fermi-liquid theory, in its original bulk formulation [1] as well as in its local version [12, 13, 7], seems to be the natural framework for building a bridge between the Kondo physics of impurity models and the Mott physics of lattice models, which connects to each other not just gross spectral features, like the Kondo resonance to the quasiparticle peak, but also more subtle properties of even greater impact, like the channels in which the impurity shows enhanced susceptibility to those in which the lattice model develops a true bulk instability. I have shown how this task can be explicitly accomplished in the case of the two-impurity model Eq. (15) in connection with the lattice model Eq. (33) treated by DMFT, i.e., in the limit of infinite coordination number $z \to \infty$. The same approach has also been used to gain insights from the physics of a C_{60}^{n-} impurity model [14] into a model for alkali doped fullerides A_3C_{60} that was studied by DMFT still in the $z \to \infty$ limit [15], but whose results reproduce quite well the physical properties of those molecular conductors. I further mention that the phase diagram of the lattice model Eq. (33) in one dimension, i.e., the opposite extreme to infinitely coordinated lattices, also recalls [16] the same impurity phase diagram of Fig. 5, apart from some obvious differences.

It is therefore quite tempting to speculate that the relationship between impurity and lattice models close to a Mott transition remains even beyond the limit of infinitely coordinated lattices, with the due caution about spatial correlations. As previously discussed, this connection can turn extremely fruitful if the impurity model upon increasing U crosses a critical point, or gets very close to it, i.e., goes through a genuine phase transition or just a very sharp crossover. Any critical point is generically unstable in several symmetry-lowering channels, which share the property of being *orthogonal* to charge that is instead severely suppressed by U. We could then argue that, in the corresponding lattice model upon approaching the Mott transition from the metallic side, a spontaneous symmetry breaking should intervene in one of the impurity instability channels. Which one dominates is going to be dictated by the spatial correlations that it entails with respect to lattice structure and Hamiltonian parameters, besides its relevance relative to all other instability channels of the impurity. For instance, the instability in a particle-particle channel leading to superconductivity is less sensitive to the lattice structure than, e.g., a magnetic instability. However the coupling constant in the magnetic channel is inevitably stronger than that in the pairing channel, especially as the system gets closer to the Mott transition. The outcome of such competing effects might be that superconductivity appears first giving way to magnetism sooner or later depending on the degree of magnetic frustration, as shown, e.g., for doped fullerides [15], or it may be defeated by magnetism and not appear at all. Even in that case, superconductivity can re-emerge upon doping the magnetic insulator.

There is actually a plethora of impurity models whose rich phase diagrams could translate into equally rich phase diagrams of corresponding lattice models. The issue is whether those lattice models are realistic and can describe physical systems. For instance, clusters of Anderson impurities [8] could be used to interpret the results of cluster DMFT calculations [17], even though the N-site extension of DMFT is only exact for $N \to \infty$ and therefore finite N calculations could be biased by the small cluster size.

References

[1] L.D. Landau, Sov. Phys. JETP **3**, 920 (1957);
L.D. Landau, Sov. Phys. JETP **5**, 101 (1957)

[2] P. Nozières and J.M. Luttinger, Phys. Rev. **127**, 1423 (1962);
J.M. Luttinger and P. Nozières, Phys. Rev. **127**, 1431 (1962)

[3] A. Georges, G. Kotliar, W. Krauth, and M.J. Rozenberg, Rev. Mod. Phys. **68**, 13 (1996)

[4] I. Affleck and A.W. Ludwig, Nuclear Physics B **360**, 641 (1991)

[5] B.A. Jones and C.M. Varma, Phys. Rev. Lett. **58**, 843 (1987);
B.A. Jones, C.M. Varma, and J.W. Wilkins, Phys. Rev. Lett. **61**, 125 (1988);
B.A. Jones and C.M. Varma, Phys. Rev. B **40**, 324 (1989)

[6] I. Affleck and A.W.W. Ludwig, Phys. Rev. Lett. **68**, 1046 (1992);
I. Affleck, A.W.W. Ludwig, and B.A. Jones, Phys. Rev. B **52**, 9528 (1995)

[7] L.D. Leo and M. Fabrizio, Phys. Rev. B **69**, 245114 (2004)

[8] M. Ferrero, L.D. Leo, P. Lecheminant, and M. Fabrizio,
Journal of Physics: Condensed Matter **19**, 433201 (2007)

[9] M. Schiró, M. Capone, M. Fabrizio, and C. Castellani, Phys. Rev. B **77**, 104522 (2008)

[10] M. Capone, M. Fabrizio, C. Castellani, and E. Tosatti, Phys. Rev. Lett. **93**, 047001 (2004)

[11] M. Capone, M. Fabrizio, C. Castellani, and E. Tosatti, Science **296**, 2364 (2002)

[12] P. Nozières, Journal of Low Temperature Physics **17**, 31 (1974)

[13] L. Mihály and A. Zawadowski, J. Physique Lett. **39**, 483 (1978)

[14] L. De Leo and M. Fabrizio, Phys. Rev. Lett. **94**, 236401 (2005)

[15] M. Capone, M. Fabrizio, C. Castellani, and E. Tosatti, Rev. Mod. Phys. **81**, 943 (2009)

[16] M. Fabrizio and E. Tosatti, Phys. Rev. Lett. **94**, 106403 (2005)

[17] M. Harland, M.I. Katsnelson, and A.I. Lichtenstein, Phys. Rev. B **94**, 125133 (2016)

14 Hund's Metals Explained

Luca de' Medici
Ecole Supérieure de Physique et Chimie Industrielles
de la Ville de Paris
10 rue Vauquelin, 75005 Paris, France

Contents

1 Introduction **2**

2 Hund's metals in Fe-based superconductors: experimental evidence and *ab-initio* studies **3**

3 Model studies: generality of Hund's metals main features **10**

4 Analytical insights into Hund's metal mechanisms **14**

5 Compressibility enhancement and quasiparticle interactions **17**

6 Conclusions **21**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

The expression "Hund's metal," to the best of the author's knowledge, was introduced in Ref. [1], in the context of an ab-initio study of Fe-based superconductors (FeSC). It has met, since then, quite some success [2,3] and today is commonly used, without a complete consensus on its precise meaning. A phase in which Hund's coupling, the intra-atomic exchange energy, influences crucially the metallic properties is probably a definition that encompasses the different current uses of this buzzword. In this chapter we will use a more specific, and hopefully precise, definition. Indeed we will show that, due to Hund's coupling, a metallic phase with specific features arises consistently in realistic simulations of paramagnetic FeSC, in accord with experiments where this has been tested (see Section 2). This phase is found beyond a crossover line in the doping/interaction strength plane in the space of parameters, and is characterized by three main features, compared to the "normal" metal realized before the same frontier:

1. enhanced electron correlations and masses,
2. high local spin configurations dominating the paramagnetic fluctuations,
3. selectivity of the electron correlation strength depending on the orbital character.

These features grow both with increasing further the interaction strength and with proximity to the half-filling of the conduction bands (thus with hole-doping in FeSC), where a Hund's coupling favored Mott insulator is realized. We will call this phase a Hund's metal.
In Section 3 we will show that these features are not specific of the band structure of FeSC, because they arise consistently in models with featureless simplified densities of states. They are, in fact, outcomes of the many-body physics dictated by the local electronic configurations of the atomic orbitals that give rise to the conduction bands (the five orbitals of mainly Fe $3d$ character, in the case of FeSC). These configurations are shaped by the local coulomb interaction, and in particular by the Hund's coupling [4], as it is known from atomic physics.
The role in this physics of the fermiology, and more generally of the specific features of the bare band structure in each particular case is comparatively lesser, and typically lumped in a few local parameters: the orbital energies, the total and orbitally-resolved kinetic energy, etc. This reduced influence implies the generality of this physics and its relevance for many materials besides FeSC, which can be called Hund's metals. A recently much investigated example is that of the Ruthenates [3, 5].
In Section 4 we will then provide, through analytical arguments, some insights into the basic mechanisms by which Hund's coupling induces the above mentioned features.
Finally in Section 5 we will outline a recent additional feature of Hund's metals, stemming plausibly from the same mechanisms outlined in the previous section: the enhancement (culminating in a divergence) of the electronic compressibility in proximity to the crossover between the Hund's metal and the normal metallic phase. The divergence of the compressibility signals an instability towards phase separation/charge-density waves. Its enhancement signals enhanced quasiparticle interactions that can also favor instabilities. Both possibly link the Hund's metal crossover to high-Tc superconductivity.

2 Hund's metals in Fe-based superconductors: experimental evidence and *ab-initio* studies

Many different families of iron-based superconductors have been synthesized [6], all bearing as a central unit a buckled plane of Fe atoms disposed in a square array with ligands (As, P, Se, or Te) in the middle of each square, alternatively slightly above or below the plane. The so-called "122" family of FeSC (where FeAs layers are interleaved with buffer layers of Ba or other alkaline earth or alkali elements) is particularly well suited to highlight the Hund's metal phenomenology, as defined above. Indeed these compounds can be synthesized in high-quality single-crystals and cleaved easily to yield clean surfaces, all of which facilitates several experimental techniques, such as angular-resolved photoemission spectroscopy (ARPES) for example. As importantly, the mother compound $BaFe_2As_2$ (the chemical formula giving the name to the family) can be engineered through many chemical substitutions, that allow to tune continuously both the structural and the electronic properties so to map out finely a complex phase diagram. There, the gross features are: a high-temperature metallic paramagnetic phase that has tetragonal symmetry, becoming below some temperature either a tetragonal superconducting phase or an orthorhombically distorted magnetic phase, depending on the exact composition.[1]

Pure $BaFe_2As_2$ becomes distorted and magnetic below $\sim$ 140 K, whereas both electron doping (most commonly by partially substituting Fe with Co) and hole doping (substituting Ba with K) lead to the suppression of this phase and the rise of superconductivity (reaching a maximum $T_c \sim$ 23 K for $Ba(Fe_{0.93}Co_{0.07})_2As_2$ and $T_c \sim$ 38 K for $Ba_{0.6}K_{0.4}Fe_2As_2$, respectively). The hole doping can be continued until reaching another stoichiometric compound KFe_2As_2 (where $T_c \sim$ 3 K). This is the most extended doping range that can be continuously obtained in a single family of FeSC to date (the phase diagram on this range is schematically reproduced as a background in Fig. 3).

Both the stoichiometric end members of the family have also been explored with isovalent chemical substitutions that act on the structure at fixed doping. In particular the substitution K $\rightarrow$ Rb, Cs acts as a negative chemical pressure and lengthen the Fe–Fe distance.

Electronic structure, typically density-functional theory (DFT) calculations, show a complex of 5 conduction bands dispersing roughly W $\sim$ 4 eV. These bands are mainly of character coming from all five Fe $3d$ orbitals, with some character of the ligand p-orbitals, and are populated by 6 electrons/Fe, in the stoichiometric parent compounds such as $BaFe_2As_2$. The same calculations predict the dominant magnetic order (collinear antiferromagnetic in most of the compounds) and show, in the paramagnetic phase a semi-metallic bandstructure where the Fermi surface is made up of hole and electron pockets (respectively in the center and at the border of the Brillouin zone), as is indeed verified in ARPES measurements. The nesting of such pockets is responsible, in the mainstream view, for the low-temperature instabilities of the phase diagram, i.e., magnetism and superconductivity [6]. This view based on itinerant electrons is indeed qualitatively quite successful, but substantial discrepancies between calculated and measured band

[1] There are some exceptions in some very small areas of the phase diagram: e.g. the two low-temperature phases can actually coexist, or a non-magnetic distorted (nematic) phase can be realized.

structures and magnetic moments, plus the difficulties arising in explaining materials trends as far as the superconducting properties are concerned point in the direction of electronic many-body correlations (basically neglected in DFT) playing a substantial role.

The most striking feature of these is the quasiparticle mass renormalization. It can be addressed by several experimental probes: low-T specific heat, optical conductivity, ARPES, and quantum oscillations, to name just the main ones. A collection of mass enhancement estimates by all these probes in the 122-family from the literature is shown in Fig. 3. Let us, however, focus first on the very fine measurement of the low-temperature slope (Sommerfeld coefficient) of the temperature dependence of the specific heat in the normal phase, performed on the whole 122-family from Ref. [7],[2] reported in Fig. 1. This coefficient reads $\gamma_b = \pi^2 k_B^2 / 3N^*(\epsilon_F)$ and is directly proportional to the (quasi particle) density of states at the Fermi energy $N^*(\epsilon_F)$ [9], which itself is enhanced for enhanced masses (i.e., for reduced dispersion of the bands: in a single band for instance one has $N^*(\epsilon_F) = (m^*/m_b)N(\epsilon_F)$, where m^* and m_b are the renormalized electron mass and the bare band mass, respectively). From the figure it is clear (blue squares) that the Sommerfeld coefficient grows monotonically with diminishing electron density all across the family. The value reached for the end member KFe_2As_2 with density of 5.5 electrons/Fe is $\sim$ 100 mJ/mol K^2, which approaches the ranges of heavy fermionic compounds. Remarkably, a further raise happens upon isovalent substitution, in the series K, Rb, Cs. These substitutions stretch the lattice parameters, thus reducing the bare electron hopping amplitudes, enhancing further the effect of electron-electron interactions. This is however a subleading effect, compared to the enhancement of correlations due to the doping. Indeed in the progressive substitution Ba $\rightarrow$ K there is actually a contraction of the Fe–Fe distance twice as large as the aforementioned stretch. Still, the correlations increase enormously. Moreover a different hole-doping substitution, Fe $\rightarrow$ Cr, induces negligible changes in the Fe–Fe distance and has the same trend of steadily increasing correlations (see, e.g., Ref. [10] and references therein). DFT calculations are unable to reproduce this trend. Indeed the calculated value of the Sommerfeld coefficient is of the order 10 mJ/mol K^2 or slightly above, throughout the family.

Dynamical many-body correlations can be included with several methods. One of the most popular is dynamical mean-field theory (DMFT). Here we report calculations within a method similar in spirit (i.e., it is a local-mean field capturing the low-frequency part of a frequency-dependent self-energy), but much cheaper in terms of computational resources, the Slave-Spin Mean-Field (SSMF) [8]. This is particularly suited to address the quasiparticle properties and their renormalizations in terms of local interactions. The advantage of a simplified method is to be able to explore thoroughly the space of parameters (both compound-wise and interaction-wise) to highlight the main trends.

Details about this method can be found, e.g., in Ref. [8], let us here just specify that it lumps the effect of the many-body local interactions into a renormalization of the hopping probabilities, in an orbitally resolved way. Indeed the starting point is a bare Hamiltonian, typically a tight-

[2]The extra points compared to the published plot, and relevant to the paramagnetic high-T phase in the zone of the phase diagram where the low-T phase is magnetic, appear already as unpublished material in Ref. [8] and are the result of a private communication with F. Hardy.

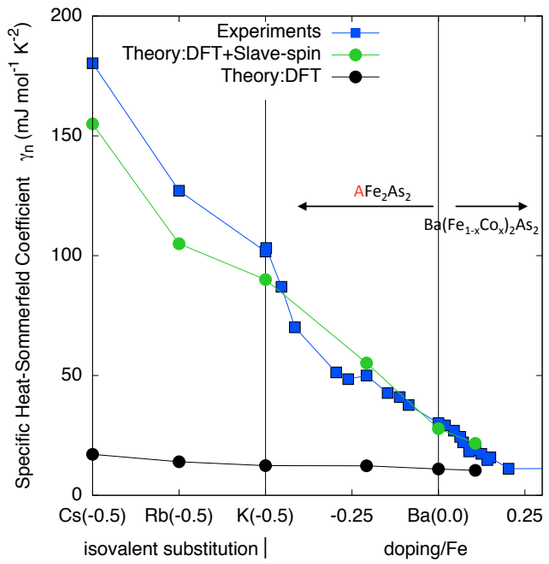


Fig. 1: *(Sommerfeld) coefficient of the linear-in-temperature contribution to the specific heat at low temperature in the normal phase for the 122-family (under both electron and hole doping in $BaFe_2As_2$ and isovalent substitution in KFe_2As_2). The calculations including local electronic correlations (DFT+slave-spin mean-field, green circles) are done with a unique choice of interaction parameters ($U = 2.7$ eV, $J/U = 0.25$) for the whole family. The discrepancy between these calculations, that capture well the experimental data (blue squares), and those for uncorrelated electrons (black dots) shows that these compounds are indeed strongly correlated. Moreover, correlations increase monotonically with hole doping, supporting the fact that a Mott insulating state would be realized for half-filling (doping of 1 hole/Fe). Isovalent substitution in KFe_2As_2 further strengthen this behavior due to an expansion of the in-plane lattice parameters in the series K, Rb, Cs (adapted from Ref. [7], originally appearing in Ref. [8]).*

binding fit of a DFT bandstructure, parametrized by the hopping amplitudes t_{ij}^{lm} (where i and j are the sites of the ionic lattice where the basis functions (typically localized Wannier-like functions) are centered, and l and m are the orbitals giving rise to the treated conduction bands), in which only the electrostatic effects of the electron-electron interactions are included. The dynamical part of these interactions (where we specify the intra-orbital interaction U, the inter-orbital interaction for anti-parallel spins $U - 2J$ and the Hund's exchange energy J, further gained when electrons have parallel spins) treated in slave-spins yields renormalization factors $Z_l < 1$ that reduce the hopping amplitudes, leading to a new (quasiparticle) band structure analogously parametrized[3] by $\tilde{t}_{ij}^{lm} = \sqrt{Z_l Z_m} t_{ij}^{lm}$.

The Sommerfeld coefficient can thus be directly evaluated from the renormalized quasi particle density of states and, as can be seen in Fig. 1, the DFT+SSMF calculations (obtained with a single set of interaction parameters, and varying only the ab-initio structure and the total

[3] This formula holds when $i \neq j$. Local orbital energies for $i = j$ are shifted by other effective orbital-dependent parameters λ_l.

electron density) capture the trend throughout the 122-family. This agreement with experiments together with the clear failure of the bare DFT calculations shows unambiguously the degree of electronic correlation of these compounds.

The same calculations explain easily the reason of this trend. Indeed the correlations increase monotonically with reducing the electron density throughout the range 6.2 to 5.5 electrons/Fe. The simulations show that when reaching 5 electrons/Fe (d^5 configuration) a Mott insulator is obtained. Thus in this picture electron masses increase until diverging for half-filled conduction bands. This Mott insulator, as we will see, is strongly favored by Hund's coupling and influences a large part of the phase diagram, even for densities as large as 1 electron/Fe away from it (the d^6 configuration of the parent FeSC) as it is clear from the above data, and beyond. Moreover $BaMn_2As_2$ (d^5) and $BaCo_2As_2$ (d^7) can also be synthesized and they are respectively a Mott insulator and an uncorrelated metal, clearly in line with the present picture.

Another recent experimental study [10] confirms this Mott-Hund scenario through analysing the formation of the local magnetic moment that fluctuate in the paramagnetic metal. Indeed, the Mott insulating state is the loss of metallicity due to the local configurations becoming energetically very unfavorable to the charge fluctuation necessary for electrons in a metal to flow. In a Mott state where Hund's coupling dominates this reduction of charge fluctuations together with the tendency of electron spins to align favors the highest possible atomic spin configurations. Thus upon approaching the half-filled Mott-Hund insulating state we expect to see these high-spin configurations gradually prevailing and building up a large local moment. In a non-magnetic phase these local moments do not form a static long range order and thus a fast spectroscopic technique sensitive to the size and not the direction of these fluctuating moments is needed to characterize them. X-ray emission spectroscopy (XES) is such a technique. Indeed it probes the energy of a photon emitted in a deexcitation from a valence state (in the present case the Fe $3p$) into a core hole previously created by incident radiation. The decaying electron can have spin up or down and its energy is different in these two configurations when a net magnetic moment is present in the near Fe $3d$ open shell. Thus the spectroscopical line due to the photons emitted in the deexcitation splits, in a proportional way to the magnitude of the local moment, which can be thus characterized.[4] A measure of this splitting (the so called IAD value [11, 10]) for hole-doped $BaFe_2As_2$ series is reported in Fig. 2 and compared with that of the electron doped compound. Indeed a monotonic increase of the Fe $3d$ local moment is observed with reducing density, throughout the phase diagram.

Calculations within DFT+SSMF (inset of Fig. 2) again show that this is a clear indication of FeSC being in a zone of influence of the half-filled Mott insulating state. Indeed the estimate of the local moment from this method shows, both in $BaFe_2As_2$ (d^6 configuration) and in KFe_2As_2 ($d^{5.5}$ configuration), a clear crossover as a function of the interaction strength U (at fixed J/U) between a low-moment (at small U) and a high-moment (at large U) region. The frontier departs from the critical U for the Mott transition at half-filling and moves very slightly to larger values

[4] Absolute measurements of the moment are made uneasy by the line shapes and intensities and by the proportionality factor which is a screened exchange constant not readily obtained (see and [11, 10] and references therein). This constant is however believed to vary slowly within the series of doped $BaFe_2As_2$ reported here.

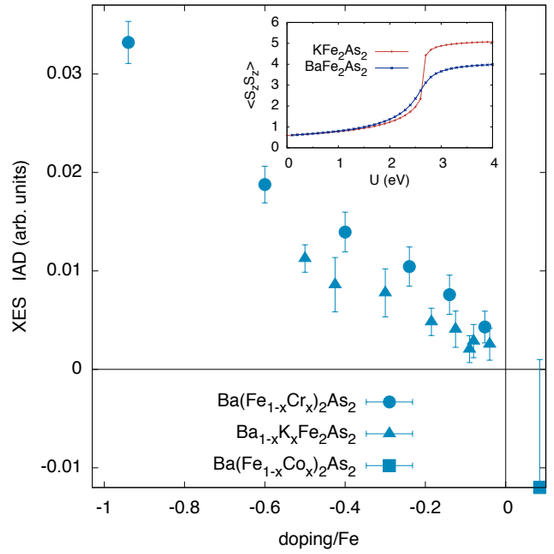


Fig. 2: *Evolution of the Fe* 3d *local magnetic moment in the paramagnetic normal phase for doped $BaFe_2As_2$ estimated with X-ray Emission Spectroscopy at room-temperature. IAD is the result of a deconvolution method, yielding a value that scales with the Fe* 3d *local moment [11, 10]. The local moment increases monotonically with hole doping (which is obtained with two different chemical substitutions yielding very similar results) and in comparison with the electron doped material, in line with the Mott-Hund scenario. Inset: DFT+Slave-Spin estimate of the local moment in $BaFe_2As_2$ and KFe_2As_2 as a function of the interaction strength U: a considerable increase is realized only when U is large enough to be in the Hund's metal regime. In the DFT limit (U=0) the moment is basically the same for the two compounds. (adapted from Refs. [10] and [11])*

with doping, the crossover becoming progressively smoother. From the data it is clear that an increase in the local moment in going from $BaFe_2As_2$ to KFe_2As_2 as seen in the experiments is only possible if the materials are in the large-moment zone.

The third experimental feature well captured by realistic theoretical calculations we want to highlight is the orbital selectivity of the electron correlation strength. This is illustrated, across the 122-family in the upper panel of Fig. 3. Indeed, several experimental probes are sensitive to the mass enhancement and by collecting the various estimates (from specific heat, optical conductivity, ARPES and quantum oscillations, and their comparison to DFT estimates – details are given in Ref. [12]) one finds a substantial agreement on values between 2 and 3 for the electron-doped side, and an increasing disagreement with hole doping.

This can be clearly interpreted in terms of increasingly differentiated correlation strengths leading, at strong hole doping, to the coexistence between strongly correlated and weakly correlated electrons. These electrons enter differently in the different experimental quantities and this explains the diverging estimates. For example the increase of the Sommerfeld coefficient γ we have already seen in this section is caused by an increase of the density of states, which is

Fig. 3: *Upper panel: experimental estimates of the mass enhancement in the 122-family from the various probes indicated in the legend. As detailed in the main text the divergence of the estimates between the various band-integrated techniques can be interpreted in terms of coexisting weakly and strongly correlated electrons, which is confirmed by the spread of the band-resolved values of ARPES and quantum oscillations. Lower panel: orbitally-resolved mass enhancement calculated within DFT+SSMF capturing the above mentioned behavior (adapted from Ref. [12]).*

a sum over the orbital index l of contributions $\sim (m^*/m_b)_l$. The estimates of correlations in optical conductivity instead is done comparing the measured low-frequency (Drude) spectral weight D with the theoretical one calculated in DFT (which in a single band case is inversely proportional to the band mass). In the multi-band case the renormalized Drude weight is then a sum of terms $\sim (m_b/m^*)_l$. If the mass renormalization were the same for all electrons then the common renormalization factor would factorize and the two estimates of correlation strength $\gamma(\text{measured})/\gamma_{DFT}$ and $(D(\text{measured})/D_{DFT})^{-1}$ would coincide. Instead if the various $(m^*/m_b)_l$ differ, as it happens for series or parallels of resistances, γ will be dominated by the largest mass enhancement, and D by the smallest.

Analogously, the renormalization of the whole bandstructure (labeled "ARPES whole" in Fig. 3), this being an entangled complex where all orbital contributions mix, will be renormalized as the least renormalized of the contributions.

On the other hand ARPES can measure the renormalization in a band/orbital resolved fashion, by comparing the slope of the dispersion (the Fermi velocity, $v_F^* = v_F(m_b/m^*)$ for a single band model) measured in different points of the Fermi surface of different orbital character to that coming from band structure calculations. The estimates from these different measures are reported as orange crosses in Fig. 3 and it is clear that they spread increasingly with hole-doping, confirming the increasing differentiation of the correlation strength. This is further confirmed by analogous band-resolved estimates from quantum oscillations (yellow crosses).

This differentiation unambiguously comes out of the same theoretical calculations that reproduce the values of the Sommerfeld coefficient in Fig. 1 and the increase of the local moment in Fig. 2. Indeed in the lower panel of Fig. 3 the orbitally-resolved renormalization factors $(m^*/m_b)_l = Z_l^{-1}$ calculated in DFT+SSMF for doped $BaFe_2As_2$ (lines) and for KFe_2As_2 (squares)[5] are shown. Clearly there is a crossover, roughly around the electron density of the parent compound, between a region at electron doping in which the renormalization is similar for all the electrons, and another at hole doping where the differentiation is strong. The differentiation increases the closer the density approaches half-filling, where the Mott insulating state is obtained.

Summarizing, the three features mentioned in the introduction, namely: i) strong correlations due to a Mott insulating state realized for the half-filled system and increasing when approaching this filling, ii) a correspondingly increasing local moment in the metallic paramagnetic phase, and iii) a strong differentiation of the correlation strength among electrons of different orbital character, are all realized and clearly seen experimentally in the 122-family of FeSC. They happen after a crossover roughly located around the filling of the parent compound $BaFe_2As_2$, so that while the hole-doped side clearly shows these features, the electron-doped side has them much less pronounced and fading into a more common uncorrelated metallic phase.

Calculations within DFT+SSMF correctly describe all this physics within an unbiased unique choice of interaction parameters for the whole family, highlighting the paramount role played by the filling of the conduction bands, and how the distance from the half-filled Hund's induced Mott insulator is a key quantity dominating the many-body physics in these compounds. The same calculations show that the crossover between the normal and the Hund's metal is a frontier in the U/density plane departing from the Mott transition and moving slightly to higher U values with doping, such that it can be crossed both by acting on the doping and on the interaction strength U.

We take these features as defining the Hund's metal, so that it appears, based on the above analysis, that the parent compound with density of 6 electrons/Fe of the 122-FeSC family is located in the proximity of the crossover. Isovalent doping [6] can be performed on $BaFe_2As_2$ too with the substitution As $\rightarrow$ P which reduces the in-plane Fe–Fe distance. Albeit rich of further phenomenology that we will not detail here, there is evidence that this substitution, acting as positive pressure (equivalent to reducing U in our calculations), brings the metal towards a weakly correlated phase, as again predicted by DFT+SSMF calculations.

[5] Calculations are done for a given DFT structure. The lines are obtained varying the electron filling for the DFT structure of $BaFe_2As_2$. The squares are values calculated for the KFe_2As_2 structure.

3 Model studies: generality of Hund's metals main features

The goal of this section is to make a parallel analysis to the one performed in Sec. 2 on the main features of Hund's metals, which are apparent in the experiments on FeSC and well captured by DFT+slave-spin calculations. The aim is to show that these features appear identically in models with featureless bare Hamiltonians in the proximity of a Hund's-induced half-filled Mott insulator, and are thus the outcome of the local many-body physics dominated by Hund's coupling and relatively independent from the fermiology and the specifics of the band structures. Let us thus focus on the multi-orbital Hubbard model of Hamiltonian

$$\hat{H} = \sum_{i \neq jlm\sigma} t_{ij}^{lm} d_{il\sigma}^{\dagger} d_{jm\sigma} + \sum_{il\sigma} (\epsilon_l - \mu) n_{il\sigma}^{d} + \hat{H}_{int} \,, \tag{1}$$

where $d_{il\sigma}^{\dagger}$ creates an electron in orbital l with spin σ on site i and we take $t_{ij}^{lm} = \delta_{lm} t$ equal for all bands as well as all orbitals degenerate $\epsilon_l = 0, \ \forall l$ (the total filling being tuned by the chemical potential μ). The lattice geometry is such to have a semicircular bare density of states (DOS) of half-width $D = 2t$ for each band (Bethe lattice). The interaction Hamiltonian reads[6]

$$\hat{H}_{int} = U \sum_{l} n_{l\uparrow} n_{l\downarrow} + U' \sum_{l \neq m} n_{l\uparrow} n_{m\downarrow} + (U' - J) \sum_{l < m, \sigma} n_{l\sigma} n_{m\sigma} \,, \tag{2}$$

where we make the with typical choice $U' = U - 2J$ (a discussion on this prescription can be found in [4]) and the three contributions mentioned in the previous section are easily read out.[7] We solve this model within the slave-spin mean-field approach [8].

The important point we want to highlight here is that the main features of the Hund's metal are present already in a simple 2-orbital Hubbard model in the presence of Hund's coupling.

Indeed as seen in the previous section, Hund's coupling favors a Mott insulator at half-filling (2 electrons in 2 orbitals, in this case), while the Mott transitions for the other possible integer fillings (1 and 3 electrons in 2 orbitals – which are physically identical due to the particle-hole symmetry of the semi-circular DOS) are sent to very high interaction strength. In Fig. 4 we report the critical interaction strength for the Mott transition calculated within SSMF[8] and indeed it is clear that for the customary value[9] $J/U = 0.25$ the Mott transition at half-filling is brought to much lower values of U compared to the $J = 0$ case, while the opposite happens at the other integer fillings.

[6] The interaction Hamiltonian eq. (2) is an approximation of the more rigorous "Kanamori" Hamiltonian in which, besides the density-density terms here considered, off-diagonal "spin-flip" and "pair-hopping" terms are present. The approximation of dropping them is however quite customary for computational reasons.

[7] This is the same Hamiltonian form that is used in the DFT+SSMF simulations discussed in the previous section. The only difference is in the choice of the t_{ij}^{lm} which are fitted on a DFT bandstructure, yielding a 5-orbital model in the case of typical FeSC.

[8] These calculations were done including spin-flip and pair-hopping terms, but are very close (and qualitatively identical) to the result of the present model in which we neglect these terms.

[9] The typical ab-initio estimates for $3d$ transition metal compounds is rather $J/U \simeq 0.12 \div 0.15$, however it was shown [8] that in the present approximation $J/U = 0.25$ is a suitable choice to reproduce DMFT results with spin-flip and pair hopping at $J/U \simeq 0.15$, in the typical fillings of interest not far from half.

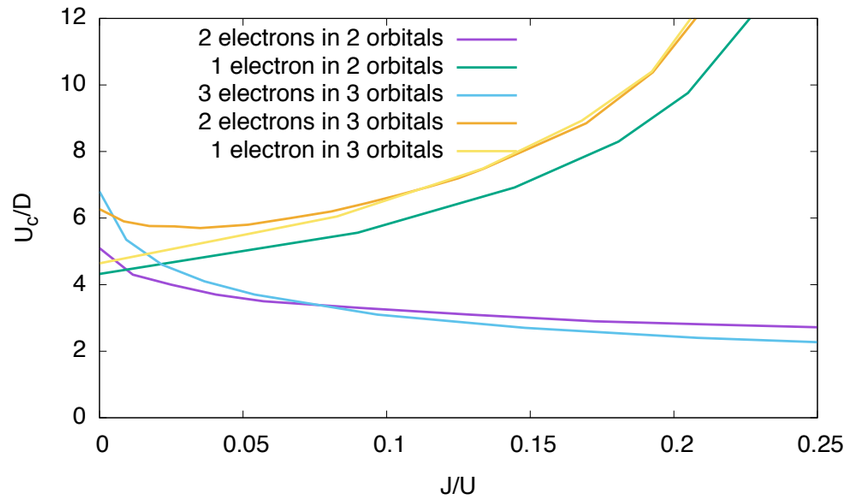


Fig. 4: *Critical interaction strength U_c for the Mott transition in the 2-orbital and 3-orbital Hubbard models with semi-circular DOS of bandwidth $W = 2D$, for all distinct integer fillings, as a function of Hund's coupling J/U (adapted from Ref. [13]).*

Identical trends are found for Hubbard models with 3 orbitals [4] or more [8]: at half-filling the critical interaction strength U_c in the presence of a sizable Hund's coupling (say $J/U = 0.25$) is a fraction of the bandwidth, whereas for all other fillings it is several times the bandwidth.

Moreover the half-filled Mott insulator dominates the phase diagram, for $U \gtrsim U_c$ for an extended range of filling even quite far from half. As an example we report in Fig. 5 several quantities calculated at fixed density that undergo a crossover as a function of U on a frontier departing from the U_c at half-filling. These quantities are the mass enhancement (bottom panel), the inter-orbital charge correlations $\langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle$ (where $n_l = \sum_\sigma n^d_{il\sigma}$, lower middle) and the local moment (upper middle). The top panel reports the inverse of the electronic compressibility that we will discuss later in the chapter.

Obviously the mass enhancement and the local moment undergo the same kind of crossover described in the previous section. The mass enhancement clearly goes from a hardly changing value near 1 at low U to a much larger value increasing with U. Depending on the proximity to half-filling ($n_{tot} = 2.0$) the effect is more or less pronounced. The local moment undergoes a rapid change of behavior from a low value increasing from the uncorrelated one at small U to a saturated high-value at large U, here again the closer to the saturated value for the Mott insulator the nearer the density is to half-filling. The analogy between the local moment behavior in the model plotted in the upper-middle panels in Fig. 5 and the behavior calculated in the ab-initio simulations for $BaFe_2As_2$ and KFe_2As_2 in Fig. 2 is obvious.

Analyzing the third feature, orbital selectivity, is more subtle here, since the model is perfectly degenerate between orbitals 1 and 2, and thus the mass enhancement will be the same for the two orbitals, by symmetry. The tendency towards orbital selective correlations can be however

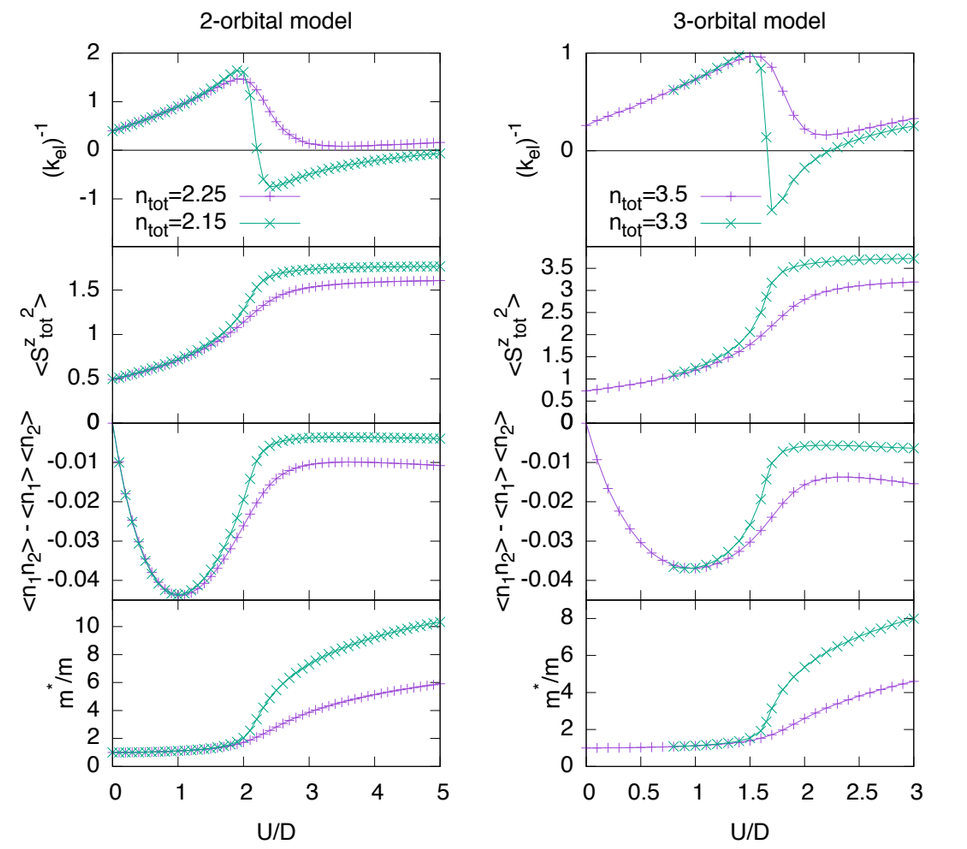


Fig. 5: *Main quantities highlighting the crossover from normal (small U/D) to Hund's metal (large U/D) in the 2-orbital (left) and 3-orbital (right) Hubbard model with Hund's coupling $J/U = 0.25$, solved within slave-spin mean-field for several dopings in proximity of half-filling: (inverse) electronic compressibility, total local moment, inter-orbital charge correlation function, mass enhancement (from Ref. [14]). The analogy with the analogous quantities calculated in realistic simulations validated by experiments shows the robustness of the Hund's metal physics with respect to materials details.*

highlighted by looking at the inter-orbital charge-charge correlations. This correlation function is obviously zero in the uncorrelated limit $U = 0$ and grows initially with the interaction strength. However, in the proximity of the crossover it undergoes a quick suppression, testifying the independence of the charge fluctuations between the two orbitals in the large U phase. Charge excitations, which are ultimately responsible for the metallicity and the suppression of which leads towards the Mott insulating state, being independent in each orbital allow for a different individual proximity of each orbital to the Mott state. This has been termed "orbital decoupling" [13, 4, 12, 15, 8]. Indeed the same correlation function is reported for $BaFe_2As_2$ in Refs. [12, 8] (obviously in a realistic case all pairs of orbitals will give rise to a different correlation function) and the behavior over the whole phase diagram is completely analogous.

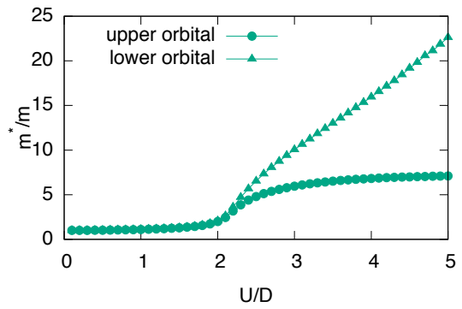


Fig. 6: *As in the left panel of Fig. 5 (2-orbital Hubbard model, $J/U = 0.25$) with total density $n_{tot} = 2.15$, but with a small crystal field splitting $\epsilon_1 - \epsilon_2 = 0.05D$, in order to show the tendency to orbital selectivity upon entering the large U region beyond the Hund's metal crossover.*

A more direct check of the link between the orbitally-decoupled charge excitations and the orbital-selectivity of the correlation strength in the present model case can be obtained by introducing a small crystal field splitting, e.g., $\epsilon_1 - \epsilon_2 = 0.05D$, i.e., 1/40th of the bandwidth. As can be seen from Fig. 6 this results in a clear orbital selectivity of the mass enhancements, starting at the crossover and growing quickly with U. The orbital closer to individual half-filling (orbital 2, which is lowest in energy for a total filling n_{tot}=1.85) is the most correlated of the two, following the orbital-decoupling physics.

All the same crossovers can be observed with similar doping-dependent plots at constant U. Moreover a completely analogous behavior is found in the 3-orbital Hubbard model (right panels in Fig. 5) and for a larger number of orbitals [15]).

Thus in conclusion we have shown that the three features that we have taken as a definition of a Hund's metal and that are found in realistic simulations of FeSC and confirmed by experiments, are also identically found in degenerate models with featureless semi-circular densities of states. Irrespectively of the number of orbitals a Mott insulator is favored at half-filling by Hund's coupling (and is found for U_c of the order of the bandwidth or smaller for realistic Hund's coupling) and dominates a large range of the U–doping phase diagram for $U \gtrsim U_c$. Identically too, the crossover into the Hund's metal phase happens on a frontier stemming from the Mott transition point at half-filling and extending at finite doping for a large range of dopings even far from half-filling.

This robustness is due to the local many body physics being the cause of these distinctive features, and to the fact that the details of the bare band structure enter through few local parameters: crystal-field splitting, kinetic energy (i.e., the first moment of the density of states), possibly orbitally resolved, etc. This is the reason why such a phenomenology can be common to many different materials irrespectively of the details of the band structure (and of the Fermi surface most notably), and can be righteously labeled a general behavior, the Hund's metal behavior.

4 Analytical insights into Hund's metal mechanisms

In this Section we will give analytical arguments that provide some insight into the Hund's metal phenomenology outlined thus far. These arguments are based on an analysis of the spectrum of the half-filled Mott insulator that was seen to influence a large zone of the U-doping parameter space in the previous sections, and is here taken as responsible for the Hund's metal phenomenology.

The spectrum of a Mott insulator can be analyzed in terms of the excitations of the system in the atomic limit, i.e., with all hoppings in eq. (1) set to zero, $t_{ij}^{lm} = 0$. Indeed for the 2-orbital model with $\epsilon_1 = \epsilon_2 = 0$ and $\mu = (3U - 5J)/2$, which ensure half-filling for a particle-hole symmetric DOS [8], the spectrum is (the zero of energy is arbitrarily fixed at the ground state energy):

$$\begin{cases} |\uparrow\downarrow,\uparrow\downarrow\rangle|0,0\rangle & E = 2U - 2J \\ |\uparrow\downarrow,\uparrow\rangle|\uparrow\downarrow,\downarrow\rangle|0,\uparrow\rangle|0,\downarrow\rangle \\ |\uparrow,\uparrow\downarrow\rangle|\downarrow,\uparrow\downarrow\rangle|\uparrow,0\rangle|\downarrow,0\rangle & E = \frac{U+J}{2} \\ |\uparrow\downarrow,0\rangle \quad |0,\uparrow\downarrow\rangle & E = 3J \\ |\uparrow,\downarrow\rangle \quad |\downarrow,\uparrow\rangle & E = J \\ |\uparrow,\uparrow\rangle \quad |\downarrow,\downarrow\rangle & E = 0 \end{cases} \tag{3}$$

The Coulomb repulsion U splits the sectors with the same total charge while J splits the half-filled sector depending on the spin alignment of the two electrons and of them paying inter- or intra-orbital repulsion. As a result the ground state is 6 times degenerate at $J = 0$, while the high-spin doublet is selected for nonzero J.[10]

The ground state of the lattice system in the atomic limit will then be that in which every site hosts two electrons in one of the high-spin configurations. The gap to overcome to establish conduction in such a system amounts to the energy needed to move an electron from one site to another and is thus the sum of the energies necessary for adding a particle on an atom and for subtracting one on another , that is:

$$[E(n+1) - E(n)] + [E(n-1) - E(n)] = E(n+1) + E(n-1) - 2E(n), \tag{4}$$

where $E(n)$ is the atomic ground state with n particles. It can be directly read from the above scheme, where both energy differences read $(U + J)/2$ so that the spectral function has two delta-like features at $\pm(U + J)/2$, and the total atomic gap reads $\Delta_{at} = U + J$.

Upon reintroduction of the hopping, these features in the spectrum broaden in two "Hubbard" bands due to the delocalization of the charge excitations. Indeed at zero hopping there are many degenerate excited states, one for each lattice site, since an extra electron (or an extra hole) with a given orbital and spin flavor can be added on any site. The hopping connects them and

[10] The present discussion, besides a modification of the spectrum that has no impact, is equally valid for the Kanamori Hamiltonian for which the ground state is a high-spin triplet.

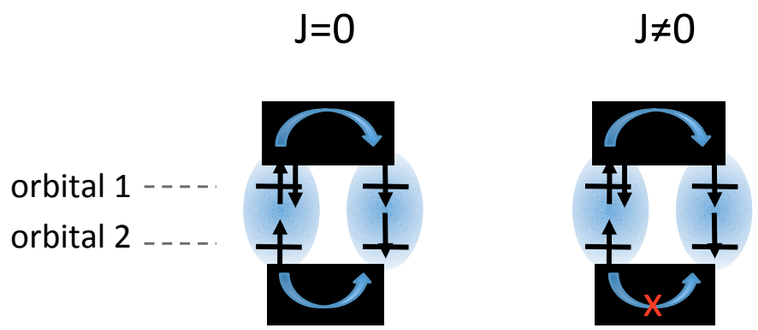


Fig. 7: *Delocalization of charge excitations. In the absence of Hund's coupling ($J = 0$) all 6 configurations with 2 electrons on a given site are degenerate (see eq. (3)). Thus both hopping channels between two sites (light blue arrows) are allowed without an extra energy cost. For finite J instead the only hopping allowed is in the channel where the extra electron was created, since hopping in the other channel would produce a two-electron configuration with a doubly occupied orbital, which has now a higher energy.*

removes the degeneracy, spreading the states over a range roughly the bandwidth W, in analogy with the case of a non-interacting electron.

It should be noticed that at $J = 0$ in a multi-orbital model (take for simplicity only diagonal hopping in the orbital index) the spread is actually larger [16], of order $\sim \sqrt{M}W$, where M is the number of orbitals. Indeed (see Fig. 7) from the site where the extra electron is created, say in orbital 1 – and that thus hosts 3 electrons – not only hopping from orbital 1 but also hopping from orbital 2 connects this state with another state of the same energy. This second hopping process leaves behind a site with a doubly occupied orbital and the other empty, which at $J = 0$ is degenerate with all the other configurations of two electrons on a site.

However when J is nonzero this extra degeneracy is removed. Indeed the two-electron configuration with orbital 1 doubly occupied left behind by the second hopping process at finite J is no longer degenerate with the high-spin configurations of the atomic ground state. Thus an extra electron created in orbital 1 can only delocalize through the hopping process in its own channel, i.e., the charge excitation cannot take advantage of the multi-orbital nature of the system to delocalize and the width of the Hubbard bands becomes of order $\sim W$ again. This shrinking was also verified within dynamical mean-field theory, where the spectral function for a Mott insulator can be directly calculated and the width of the Hubbard bands measured [8].

All in all this means that the gap between the two Hubbard bands, which are Δ_{at} apart is $\Delta \simeq \Delta_{at} - W = U + J - W$. This gap will then close at interaction strength $U_c \simeq W - J$, which is of size of order of the bandwidth or less, and decreasing with increasing J, perfectly in line with what is found numerically [8].

This argument can be generalized to any number of orbitals and explains why the half-filled Mott insulating state is favored by Hund's coupling. Indeed the half-filled sector is the one with a larger number of spare spins to align in order to gain exchange energy. Its distance in energy

from all other sectors will grow with J and so will the Mott gap, needing thus a smaller U to close.

The same kind of arguments can be applied to the Mott transition at other integer fillings [13], where however the effect is opposite. For instance in the 2-orbital model for filling of 1 electrons/site, the atomic ground state will be in the $n = 1$ sector. This sector is unaffected by J, while the excited state with $n + 1$ particles will be the ground state of the atomic $n = 2$ sector. The energy of this state lowers with J, and thus J helps closing the gap in this case, thus disfavoring the Mott transition.

The general outcome [13] is that the critical interaction strength for the Mott transition in a system with M orbitals at large J goes like

$$U_c(n) \propto \begin{cases} 3J, & \forall n \neq N \quad \text{(off half-filling)} \\ -(M-1)J, & n = M \quad \text{(half-filling)}\,, \end{cases} \tag{5}$$

which explains why for values of $U \sim W$ as it happens in FeSC and in the related models with sizable J/U, the Mott insulating state is only realized at half-filling.

This analytic arguments justifies the first two features of the Hund's metals. Indeed the electronic correlation strength naturally grows with reducing doping from the Mott insulator. Also the local moment is maximized in the half-filled Mott insulator, in which the ground state lies in the sector with the highest possible spin configuration and charge fluctuations are minimized. Upon doping this sector will mix increasingly with the other charge sectors, where lower spin configurations are realized, and the resulting local moment will gradually decrease.

The argument above on J decoupling the hopping channels for the charge excitations in the various orbitals is a support to the third Hund's metal feature we have outlined, i.e., the role of Hund's coupling as an orbital decoupler in general, favoring orbital-selectivity in the proximity of the half-filled Mott insulator. Indeed, in a system where the two orbitals differ, be it for the energy $\epsilon_1 \neq \epsilon_2$ or for the hopping integrals (or both), the Hubbard bands in the spectral function will differ for the two orbitals [17]. This implies different gaps and can lead to orbitally-selective Mott transitions, if U is such that the gap is open only for one of the orbitals, and closed for the other. Analogously in the doped case the chemical potential can fall in the gap for one orbital and in the Hubbard band for the other causing again selective localization. In a more realistic case where off-diagonal hoppings are present and thus the character of the orbital mixes, one can expect that orbital selective Mott phases turn into metallic phases with different correlation strength, and this is indeed what is observed in simulations.

These arguments for independent gaps are just indicative for the electron correlations in the metallic phase. There, indeed, more rigorous arguments for the low-energy long lived quasiparticle excitations should be used. SSMF offers a framework where this can be done and a low-energy analysis supporting the role of Hund's coupling as an orbital decoupler was performed in Ref. [8].

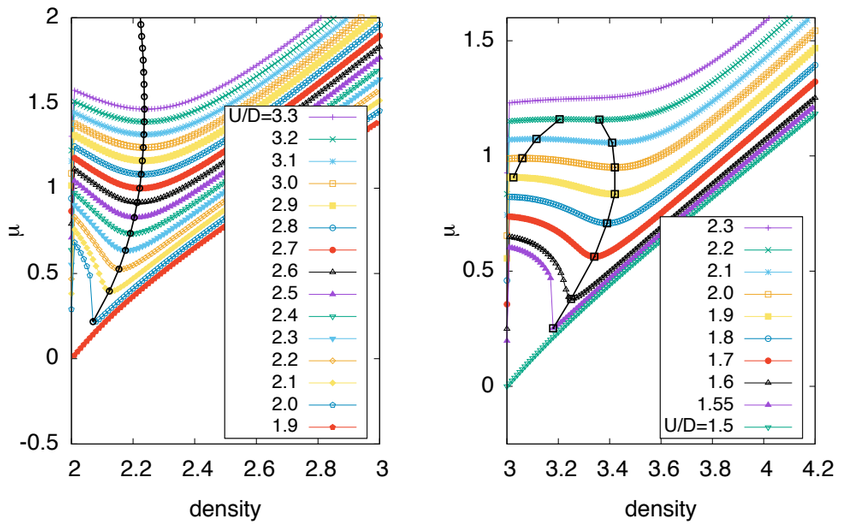
5 Compressibility enhancement and quasiparticle interactions

In this last section it is worth mentioning some more recent work highlighting another feature connected with the normal to Hund's metal crossover. This feature emerges clearly from theoretical calculations: an enhancement (culminating in a divergence) of the electronic compressibility near the crossover.

Indeed when investigating within SSMF the proximity of the half-filled Mott insulator in the presence of Hund's coupling, irrespectively of the number of orbitals, one encounters a zone of negative compressibility [14]. This is illustrated in the upper panels of Fig. 8 where the chemical potential μ as a function of the total density n is plotted, for the 2-orbital and the 3-orbital Hubbard model on the Bethe lattice with $J/U = 0.25$. Curves for various values of U are plotted. The lowest value illustrates a case in which at half filling a metallic solution is obtained, just below U_c for the Mott transition. All other values of U lead to a Mott insulator at half-filling. A clear change of behavior of the μ vs. n curve appears for $U > U_c$. Indeed upon approaching the Mott insulator the electronic compressibility $\kappa_{el} = dn/d\mu$ diverges (i.e. $\mu(n)$ has a flat slope). At large doping the slope is positive, signaling a stable phase, while in a zone near the Mott insulator the negative slope signals negative compressibility and thus an unstable electronic fluid. In the 2-orbital model for all dopings below the one where the divergence happens the system is unstable, giving rise to the purple instability zone in the phase diagram depicted in the lower panel of Fig. 8. In the 3-orbital (Fig. 8, right panel) and in the 5-orbital model (not shown, see supplementary material in Ref. [14]) instead the compressibility becomes positive again before the filling reaches half. This gives rise to a different shape of the instability zone (respectively in green and light blue for the 3- and 5-orbital model in the lower panel of Fig. 8) that appears more like a "moustache."

The striking feature is that in all cases the instability zone departs from the Mott transition point at half-filling (symbols in the lower panel of Fig. 8). It is in fact easy to verify that the lower frontier between the stable and unstable metal approximatively coincides with the crossover into the Hund's metal that we have described in this chapter. Indeed this is illustrated in the upper panels in Fig. 5, which plot quantities at constant density as a function of U along the scans of the phase diagram signaled in the lower panel of Fig. 8 by dashed lines. The (inverse) electronic compressibility κ_{el}^{-1} is plotted and the instability is signaled by its vanishing. It is easy to see that for the scans crossing the frontier the compressibility divergence happens in correspondence (or immediately beyond) the crossover into the Hund's metal phase. It is also worth mentioning that in the proximity of the instability zone the compressibility of the stable Hund's metal remains enhanced. In particular this is shown from the scans reported in Fig. 5 at larger dopings in both models: indeed even without diverging, the compressibility is strongly enhanced (i.e. κ_{el}^{-1} is very small) in a zone starting with the Hund's crossover.

The interest of this finding lies in the connection between enhanced or diverging compressibility and superconductivity. Indeed the compressibility being the uniform and static charge-charge response function, its divergence signals an instability of the system towards phase separation (or more physically towards a charge density wave, when taking into account the long range



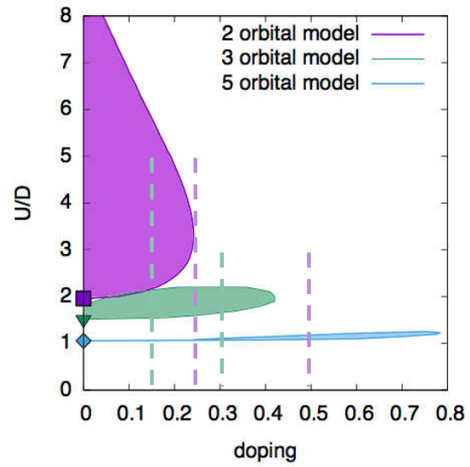


Fig. 8: *Upper panels: chemical potential vs. density curves for the 2-orbital (left) and 3-orbital (right) Hubbard models with $J/U = 0.25$ calculated within DFT+SSMF. For $U > U_c$ (the critical interaction value for the Mott transition at half-filling, with $U_c/D = 1.96$ in the 2-orbital model and $U_c/D = 1.515$ in the 3-orbital model) the curves show a negative slope, indicating an unstable system for a range of doping inside a spinodal line (black dots) which is the locus where the compressibility $\kappa_{el} = dn/d\mu$ diverges. In the 3-orbital model (and in the 5-orbital model – not shown) a second spinodal line where the compressibility becomes positive again delimits a region in the shape of a "moustache." The instability regions for all these models are reported (colored areas) in the lower panel, in the interaction-density plane. The dashed lines represent the scans in the space of parameters corresponding to the calculations reported in Fig. 5 (from Ref. [14]).*

Coulomb interaction neglected in the Hubbard model). A second order transition due to a phase separation instability is a possible cause of Cooper pairing through quantum critical fluctuations in its proximity. Moreover in a Fermi-liquid phase such as the one described in the SSMF the compressibility reads [18][11]

$$\kappa_{el} = \frac{N(\epsilon_F)/Z}{1+F_0^s}\,, \tag{6}$$

where $N(\epsilon_F)$, the total bare density of states at the Fermi energy, coincides with the non-interacting compressibility. F_0^s is the isotropic, spin-symmetric Landau parameter. It is seen that when Z behaves smoothly as in our case, the divergence (or the strong enhancement) is due to a negative Landau parameter approaching the value $F_0^s = -1$. The Landau parameter embodies the effect of quasiparticle interactions and a negative F_0^s signals an attraction in the particle-hole channel, which is known to favor superconductivity.

Moreover the enhanced compressibility signals also the enhancement of some quasiparticle interaction vertices with bosonic excitations (like phonons for instance). Indeed for example for the density-vertex $\Lambda(q,\omega)$, which plays a role in the interaction between electrons and phonons, the following Ward identity holds [19]

$$\Lambda(q\to 0,\omega=0) = \frac{1/Z}{1+F_0^s}\,. \tag{7}$$

One can see that this vertex is enhanced in the same way as the compressibility. This enhancement of quasiparticle-boson interactions can be a source of instabilities and in particular of an instability in the Cooper channel, i.e., superconductivity.

One word is in order on the possible mechanism causing such a compressibility enhancement in correspondence of the Hund's metal crossover. We will only give a plausibility argument, that needs to be verified in further studies: A doped Mott insulator typically has a shifted spectrum, compared to the insulating case, in which the chemical potential has "jumped into" one of the Hubbard bands. At low electron doping it lies typically at the bottom (and symmetrically, in case of hole-doping) of this band. Now we have mentioned that the width of the Hubbard bands, albeit in an M-orbital system orbital fluctuations would favor a value $\sim\sqrt{M}W$, is brought back to values of order W by Hund's coupling, near half-filling. However, the mechanism behind this shrinking that we have described in the previous section holds only near half-filling, and it is natural to expect that with doping it becomes gradually ineffective, and the width of the Hubbard bands is brought back to $\sim\sqrt{M}W$. Now if the Hubbard band swells with increasing doping, the situation in which the chemical potential at a higher density has a lower value than at a smaller density can naturally happen. This indeed coincides with a negative electronic compressibility.

Finally it is worth going back to realistic calculations for FeSC to see if the compressibility enhancement too survives in an ab-initio framework. In Fig. 9 we report the calculation done

[11] The formula reported here (where for instance the renormalized and bare densities of states are simply related by $N^*(\epsilon_F) = N(\epsilon_F)/Z$, with ϵ_F the bare Fermi energy) holds only in simple cases (single band, degenerate identical bands). In the general case, however, proper generalizations can be made, and the considerations made in this section hold valid.

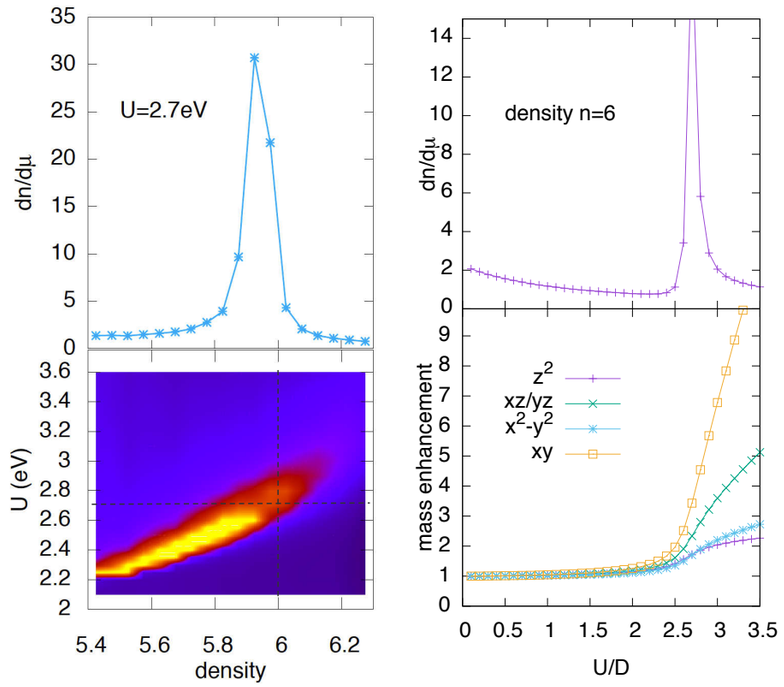


Fig. 9: *Lower left panel: compressibility (color scale) in $BaFe_2As_2$ calculated within DFT+SSMF (J/U=0.25) in the U-density plane. The saturated yellow color corresponds to the unstable region (the pixelation is due to numerical discretization and is unphysical), surrounded by an area of enhanced compressibility (red). Upper panels: compressibility plotted along the cuts (dashed lines) for constant $U = 2.7$ eV and constant density n=6, relevant values for $BaFe_2As_2$. Lower right panel: orbitally resolved mass enhancements, showing that the compressibility enhancement happens near the Hund's metal crossover, signaled by the correlations become orbitally selective.*

within DFT+SSMF for $BaFe_2As_2$. The "moustache" expected in the 5-orbital model is well visible and actually goes through the realistic values for the interaction expected in this material ($U = 2.7$ eV and $J/U = 0.25$) and used in all the calculations reported in this chapter that successfully reproduce the phenomenology of the 122-family. The moustache (or its prolongation where the divergence becomes a strong enhancement) goes through the right interaction strength exactly for the zone of densities where the material shows both the superconductive and the magnetic instabilities. From the plotted panels it is obvious that the enhancement once again happens concomitantly with the crossover in the Hund's metal regime, signaled by the onset of orbital selectivity of the masses in the figure.

This is a possible confirmation of the role played by the enhancement of the quasiparticle interaction signaled by the compressibility peak in promoting the instabilities in general and maybe high-Tc superconductivity in particular.

6 Conclusions

In this chapter we have adopted a clear-cut definition of a Hund's metal, as a metallic phase emerging clearly in the phase diagram of Fe-based superconductors, signaled by three main features: enhanced electron masses, local magnetic moments and orbital-selectivity, and all growing with hole doping (in the case of these materials where the conduction bands are filled more than half).
We have shown that experiments and theoretical simulations within density-functional theory + slave-spin mean-field go hand in hand, pointing to the local electronic correlations triggered by Hund's coupling as the origin of this phenomenology.
Then we have shown that all the features of the Hund's metal are found also in Hubbard models with featureless band structures, thus proving the generality of this physics and its robustness with respect to details of the materials. The only necessary condition is to be under the influence of a half-filled Mott insulator in the presence of sizable Hund's coupling. A crossover line between the normal and the Hund's metal exist, where all the aforementioned features become enhanced, it departs from the Mott transition point at half-filling and extends in the interaction/doping plane.
We have then given some analytical arguments in order to gain insight into this phenomenology, showing how Hund's coupling favors the realization of a half-filled Mott insulator at values of the interaction less than the bandwidth, and a mechanism lying at the basis of the "orbital decoupling" causing orbital-selective correlation strength. More on this can be found in Refs. [4, 8].
We have finally pointed out a relatively recent extra feature of the normal to Hund's metal crossover, which is the enhancement of the electronic compressibility. This culminates in a zone of instability towards phase separation, departing from the Mott transition point at half-filling and following the Hund's metal frontier. This phenomenology is common, once again, to models and realistic simulations, and it stems from the outlined Hund's local correlation physics. A Fermi-liquid analysis highlights the connection with an enhancement of quasiparticle interactions, tracing a possible, hitherto unsuspected link between Hund's physics and high-Tc superconductivity. More on this can be found in Ref. [14] and in its supplementary material.

Acknowledgment

The author is supported by the European Commission through the ERC-StG2016, StrongCoPhy4Energy, GA No724177.

References

[1] Z.P. Yin, K. Haule, and G. Kotliar, Nat. Mater. **10**, 932 (2011)

[2] C.F. Hirjibehedin, Nat. Nano. **10**, 914 (2015)

[3] T. Kondo, M. Ochi, M. Nakayama, H. Taniguchi, S. Akebi, K. Kuroda, M. Arita, S. Sakai, H. Namatame, M. Taniguchi, Y. Maeno, R. Arita, and S. Shin, Phys. Rev. Lett. **117**, 247001 (2016)

[4] A. Georges, L. de' Medici, and J. Mravlje, Annual Review of Condensed Matter Physics **4**, 137 (2013)

[5] J. Mravlje, M. Aichhorn, T. Miyake, K. Haule, G. Kotliar, and A. Georges, Phys. Rev. Lett. **106**, 096401 (2011)

[6] J. Paglione and R.L. Greene, Nat. Phys. **6**, 645 (2010)

[7] F. Hardy, A.E. Böhmer, L. de' Medici, M. Capone, G. Giovannetti, R. Eder, L. Wang, M. He, T. Wolf, P. Schweiss, R. Heid, A. Herbig, P. Adelmann, R.A. Fisher, and C. Meingast, Phys. Rev. B **94**, 205113 (2016)

[8] L. de' Medici and M. Capone: *Modeling Many-Body Physics with Slave-Spin Mean-Field: Mott and Hund's Physics in Fe-Superconductors* (Springer, 2017), pp. 115–185

[9] N. Ashcroft and N. Mermin: *Solid State Physics* (Saunders College, Philadelphia, 1976)

[10] S. Lafuerza, H. Gretarsson, F. Hardy, T. Wolf, C. Meingast, G. Giovannetti, M. Capone, A.S. Sefat, Y.-J. Kim, G.P., and L. de' Medici, ArXiv e-print 1607.07417 (2016)

[11] H. Gretarsson, A. Lupascu, J. Kim, D. Casa, T. Gog, W. Wu, S.R. Julian, Z.J. Xu, J.S. Wen, G.D. Gu, R.H. Yuan, Z.G. Chen, N.-L. Wang, S. Khim, K.H. Kim, M. Ishikado, I. Jarrige, S. Shamoto, J.-H. Chu, I.R. Fisher, and Y.-J. Kim, Phys. Rev. B **84**, 100509 (2011)

[12] L. de' Medici, G. Giovannetti, and M. Capone, Phys. Rev. Lett. **112**, 177001 (2014)

[13] L. de' Medici, Phys. Rev. B **83**, 205112 (2011)

[14] L. de' Medici, Phys. Rev. Lett. **118**, 167003 (2017)

[15] L. de' Medici: *Weak and strong correlations in Iron pnictides* (Springer Series in Materials Science, Vol. 211, 2015) pp. 409–441

[16] O. Gunnarsson, E. Koch, and R.M. Martin, Phys. Rev. B **54**, R11026 (1996)

[17] A. Koga, N.Kawakami, T. Rice, and M. Sigrist, Phys. Rev. Lett. **92**, 216402 (2004)

[18] P. Nozières: *Theory of Interacting Fermi systems* (Westview Press, 1997)

[19] M. Grilli and C. Castellani, Phys. Rev. B **50**, 16880 (1994)

15 Electron-Phonon Coupling

Rolf Heid
Institute for Solid State Physics
Karlsruhe Institute of Technology

Contents

1 Introduction **2**

2 Electron-phonon Hamiltonian **2**
2.1 Electron-phonon vertex 2
2.2 Fröhlich Hamiltonian 3

3 Normal-state effects **5**
3.1 Green functions and perturbation 5
3.2 Electron self-energy 6
3.3 Migdal's theorem 10
3.4 Phonon self-energy and linewidth 11

4 Phonon-mediated superconductivity **12**
4.1 Effective electron-electron interaction 13
4.2 Nambu formalism 15
4.3 Eliashberg theory 17
4.4 Isotropic gap equations 18

5 Density functional theory approach **22**

6 Summary **25**

A Phonon quantization **26**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

The electron-phonon interaction is, besides the Coulomb interaction, one of the fundamental interactions of quasiparticles in solids. It plays an important role for a variety of physical phenomena. In particular in metals, low-energy electronic excitations are strongly modified by the coupling to lattice vibrations, which influences, e.g., their transport and thermodynamic properties. Electron-phonon coupling (EPC) also provides in a fundamental way an attractive electron-electron interaction, which is always present and, in many metals, is the origin of the electron pairing underlying the macroscopic quantum phenomenon of superconductivity.
This lecture addresses the consequences of electron-phonon coupling in both the normal and the superconducting state of metals. In Section 2, the basic Hamiltonian describing the coupled electron-phonon system is introduced. In Section 3, a closer look onto normal state effects in a metal is taken, focusing on the renormalization of quasiparticles, which allows to experimentally quantify the strength of the interaction. Section 4 is devoted to phonon-mediated superconductivity. First a derivation of the effective attractive interaction among electrons mediated by phonon exchange is given. Then we analyze the role of electron-phonon coupling for superconductivity in the context of the strong-coupling Migdal-Eliashberg theory in some detail. In Section 5, we discuss the density-functional based technique to calculate electron-phonon coupling quantities and present two examples to illustrate its predictive power. Throughout this Chapter, only nonmagnetic states are considered and atomic units $\hbar = 2m_e = e^2/2 = 1$ as well as $k_B = 1$ are used.

2 Electron-phonon Hamiltonian

2.1 Electron-phonon vertex

The lowest-order process involving the electron-phonon interaction is the scattering of a single electron by a simultaneous creation or annihilation of a single phonon, as diagrammatically shown in Fig. 1. The probability for the scattering process is called the electron-phonon vertex g. We will briefly sketch its derivation starting from rather general grounds. For more details one can refer to the book of Grimvall [1].
Due to the large ratio of the ionic and electronic mass, the dynamics of the ions and the electrons can be systematically expanded in terms of the small parameter $\kappa = (m/M)^{1/4}$, which results in a partial decoupling [2, 3]. To lowest order in κ, called the adiabatic or Born-Oppenheimer approximation, the total wavefunction of the coupled electron-ion system can be written as a product $\Psi(\underline{\mathbf{r}}, \underline{\mathbf{R}}) = \chi(\underline{\mathbf{R}})\psi(\underline{\mathbf{r}}; \underline{\mathbf{R}})$, where $\underline{\mathbf{r}}$ and $\underline{\mathbf{R}}$ denote the sets of electron and ion coordinates, respectively. The electronic wavefunction obeys the equation

$$[T_e + V_{ee} + H_{e-i}(\underline{\mathbf{R}})]\psi_n(\underline{\mathbf{r}}; \underline{\mathbf{R}}) = E_n(\underline{\mathbf{R}})\psi_n(\underline{\mathbf{r}}; \underline{\mathbf{R}}) \,, \tag{1}$$

where T_e and V_{ee} denote the kinetic energy and Coulomb interaction of the electron system, respectively. Eq. (1) depends parametrically on the ionic positions $\underline{\mathbf{R}}$ via the electron-ion interaction H_{e-i}.

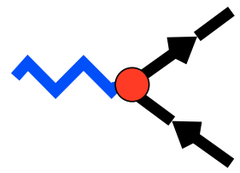


Fig. 1: *Diagrammatic representation of the basic electron-phonon scattering process. Black lines represent electrons, the blue zigzag line a phonon, and the red circle the screened vertex.*

The electron-phonon vertex appears in first order beyond the adiabatic approximation. One can show that it induces off-diagonal matrix elements among the electronic eigenstates ψ_n and has the form

$$\langle n|\delta_{\mathbf{R}} V|n'\rangle \,. \tag{2}$$

The operator $\delta_{\mathbf{R}} V$ stands for the linear change of the potential felt by the electrons under a displacement of an atom from its rest position: $\mathbf{R} = \mathbf{R}_0 + \mathbf{u}$. If the potential V is the bare electron-ion potential V^0, then $\delta_{\mathbf{R}} V = \mathbf{u}\cdot\nabla V^0|_{\mathbf{R}_0}$. Eq. (2) represents the bare vertex. However, in solids, and in particular in metals, the bare electron-ion potential is screened by the other electrons. Screening also alters the vertex significantly. Within linear response theory this operator takes the form

$$\delta_{\mathbf{R}} V = \mathbf{u}\cdot\epsilon^{-1}\nabla V^0|_{\mathbf{R}_0} \,, \tag{3}$$

where ϵ^{-1} is the inverse dielectric matrix [4], which is a measure of the screening. Note that in Eq. (3), the screening operator does not commute with the gradient operation, and thus can not be written in terms of the gradient of a screened potential.

2.2 Fröhlich Hamiltonian

We now aim to develop a systematic perturbative treatment of the mutual influence of the electronic and phononic subsystems in a solid. Thereby the question arises, what are the proper noninteracting quasiparticles to start with. The correct answer requires to know the solution to some extent. As we will see, electronic states are significantly influenced by lattice vibrations mostly in close vicinity of the Fermi energy. It is therefore appropriate to start with electrons moving in the static potential of a rigid ion lattice, without any renormalization by the lattice vibrations. On contrast, the bare vibrations of the ion lattice would be a bad starting point, because they are strongly altered by the screening of the electrons. This screening must be built into the description of the harmonic lattice vibrations which defines the noninteracting phonons. For the discussion of electron-phonon coupling effects in periodic solids, a good starting point is the *Fröhlich Hamiltonian*, which reads in second quantization

$$H = H_e + H_{ph} + H_{e-ph} \,. \tag{4}$$

Here the electron system is described by noninteracting quasi-particles with dispersion ε_k. These quasiparticles are considered to be the stationary solutions of band electrons in a perfect periodic lattice, and include already the renormalization from Coulomb interaction

$$H_e = \sum_{\mathbf{k}\nu\sigma} \varepsilon_{\mathbf{k}\nu}\, c^\dagger_{\mathbf{k}\nu\sigma} c_{\mathbf{k}\nu\sigma}\,. \tag{5}$$

Here $c_{\mathbf{k}\nu\sigma}$ ($c^\dagger_{\mathbf{k}\nu\sigma}$) are the annihilation (creation) operators for an electronic state with momentum $\mathbf{k}$, band index ν, spin σ, and band energy $\varepsilon_{\mathbf{k}\nu}$.
The lattice Hamiltonian is expressed in terms of quantized harmonic vibrations, and represents noninteracting phonons

$$H_{ph} = \sum_{\mathbf{q}j} \omega_{\mathbf{q}j} \left(b^\dagger_{\mathbf{q}j} b_{\mathbf{q}j} + \frac{1}{2} \right), \tag{6}$$

where $b_{\mathbf{q}j}$ ($b^\dagger_{\mathbf{q}j}$) are the annihilation (creation) operators for a phonon with momentum $\mathbf{q}$, branch index j, and energy $\omega_{\mathbf{q}j}$. Phonons are the quanta of the normal mode vibrations (for more details see Appendix A). The operator of atom displacements is expressed in terms of the phonon operators by

$$u_{ls\alpha} = e^{i\mathbf{q}\mathbf{R}^0_{ls}} \frac{1}{\sqrt{N_q}} \sum_{\mathbf{q}j} A^{\mathbf{q}j}_{s\alpha} \left(b_{\mathbf{q}j} + b^\dagger_{-\mathbf{q}j} \right) \qquad \text{with} \qquad A^{\mathbf{q}j}_{s\alpha} = \frac{\eta_{s\alpha}(\mathbf{q}j)}{\sqrt{2M_s\omega_{\mathbf{q}j}}}\,. \tag{7}$$

Atoms are characterized by two indices denoting the unit cell (l) and the atoms inside a unit cell (s), respectively, with M_s the corresponding atom mass. α denotes Cartesian indices, and $\eta_{s\alpha}(\mathbf{q}j)$ is the eigenvector of the normal mode $\mathbf{q}j$. The number of points in the summation over $\mathbf{q}$ is N_q.
The third term describes the lowest-order coupling between electrons and phonons derived from Eq. (3). Using the relationship Eq. (7) it has the form

$$H_{e-ph} = \sum_{\mathbf{k}\nu\nu'\sigma} \sum_{\mathbf{q}j} g^{\mathbf{q}j}_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu} c^\dagger_{\mathbf{k}+\mathbf{q}\nu'\sigma} c_{\mathbf{k}\nu\sigma} \left(b_{\mathbf{q}j} + b^\dagger_{-\mathbf{q}j} \right)\,. \tag{8}$$

$g^{\mathbf{q}j}_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}$ is the electron-phonon matrix element and describes the probability amplitude for scattering an electron with momentum $\mathbf{k}$ from band ν to a state with momentum $\mathbf{k}+\mathbf{q}$ in band ν' under the simultaneous absorption (emission) of a phonon with momentum $\mathbf{q}$ ($-\mathbf{q}$) and branch index j

$$g^{\mathbf{q}j}_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu} = \sum_{s\alpha} A^{\mathbf{q}j}_{s\alpha} \langle \mathbf{k}+\mathbf{q}\nu'\sigma | \delta^{\mathbf{q}}_{s\alpha} V | \mathbf{k}\nu\sigma \rangle\,. \tag{9}$$

Here again the screened first-order variation enters the matrix elements. They are independent of spin for nonmagnetic ground states.
This general form of the Fröhlich Hamiltonian will be the starting point for the many-body perturbation outlined in the next Sections. To simplify the treatment, we will use a compact notation combining momentum and band/branch index into a single symbol: $k = (\mathbf{k}\nu)$, $k' = (\mathbf{k}'\nu')$, and $q = (\mathbf{q}j)$. The EPC matrix elements are then denoted as

$$g^q_{k',k} = g^{\mathbf{q}j}_{\mathbf{k}'\nu',\mathbf{k}\nu} \delta_{\mathbf{k}',\mathbf{k}+\mathbf{q}}\,, \tag{10}$$

which implicitly takes into account momentum conservation.

3 Normal-state effects

3.1 Green functions and perturbation

In this section we will discuss the effects of electron-phonon interaction in the normal state of a metal. This will be done using many-body perturbation techniques [5–7]. The focus will be on the renormalization of electronic and phononic quasiparticles, which provides ways to experimentally gain information about the coupling strength. This will set the stage for the discussion of phonon-mediated superconductivity in the next section.

The following treatment is based on the Fröhlich Hamiltonian Eq. (4), where $H_0 = H_e + H_{ph}$ denotes the Hamiltonian of the unperturbed quasiparticles and H_{e-ph} is perturbation linear in the electron-phonon coupling. We will work with the imaginary-time Green functions

$$G(k,\tau) = -\langle T_\tau c_{k\sigma}(\tau) c^\dagger_{k\sigma}(0)\rangle \tag{11}$$

for the fermionic quasiparticles, where the field operators are given in a Heisenberg picture using an imaginary time $-i\tau$, $c_{k\sigma}(\tau) = e^{H\tau} c_{k\sigma} e^{-H\tau}$ with $-\beta < \tau < \beta$, $\beta = 1/T$. The Wick operator T_τ reorders operators to increasing τ from right to left.

For the bosonic quasiparticles, the Green function of the displacement operators is defined as

$$U_{s\alpha,s'\alpha'}(\mathbf{q},\tau) = -\langle T_\tau u_{\mathbf{q}s\alpha}(\tau) u_{-\mathbf{q}s'\alpha'}(0)\rangle = \sum_j A^{\mathbf{q}j}_{s\alpha} A^{-\mathbf{q}j}_{s'\alpha'} D(\mathbf{q}j,\tau)\,, \tag{12}$$

where D denotes the phonon Green function ($q = (\mathbf{q}j)$)

$$D(q,\tau) = -\langle T_\tau (b_q(\tau) + b^\dagger_{-q}(\tau))(b_{-q}(0) + b^\dagger_q(0))\rangle \tag{13}$$

$G(k,\tau)$ and $D(q,\tau)$ can be defined as periodic functions in τ with symmetry properties $G(k,\tau+\beta) = -G(k,\tau)$ and $D(k,\tau+\beta) = D(k,\tau)$, respectively. Their Fourier transforms are given by

$$G(k,i\omega_n) = \frac{1}{2}\int_{-\beta}^{\beta} d\tau\, e^{i\omega_n\tau} G(k,\tau) \tag{14}$$

$$D(q,i\nu_m) = \frac{1}{2}\int_{-\beta}^{\beta} d\tau\, e^{i\nu_m\tau} D(q,\tau)\,, \tag{15}$$

where $\omega_n = (2n+1)\pi T$ and $\nu_m = 2m\pi T$, with n, m integer values, denote fermionic and bosonic Matsubara frequencies, respectively.

Two further simplifications have been assumed: (i) because we are dealing with nonmagnetic states only, the spin index in the electronic Green function can be suppressed; (ii) the perturbation H_{e-ph} does not mix different electronic bands or phononic modes, such that the interacting Green functions can still be represented by a single band/mode index.

The bare Green functions of the unperturbed Hamiltonian $H_0 = H_e + H_{ph}$ are

$$G_0(k,i\omega_n) = \frac{1}{i\omega_n - \varepsilon_k} \tag{16}$$

$$D_0(q,i\nu_m) = \frac{1}{i\nu_m - \omega_q} - \frac{1}{i\nu_m + \omega_q}\,. \tag{17}$$

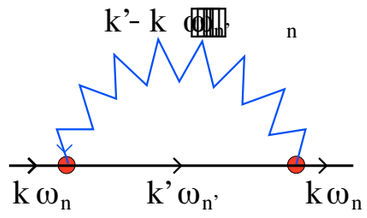


Fig. 2: *Diagrammatic representation of the lowest-order contribution to the electron self-energy from the electron-phonon coupling. Blue zigzag and black lines represent phonon and electron propagators, respectively.*

Electronic energies are measured with respect to the chemical potential. By applying many-body perturbation theory to the Fröhlich Hamiltonian, the interacting Green functions are expressed by an infinite series of Feynman diagrams containing the bare Green functions and an increasing number of electron-phonon vertices.
Partial resummation leads to the Dyson equations

$$G(k, i\omega_n)^{-1} = G_0(k, i\omega_n)^{-1} - \Sigma(k, i\omega_n) \tag{18}$$

$$D(q, i\nu_m)^{-1} = G_0(q, i\nu_m)^{-1} - \Pi(q, i\nu_m)\,, \tag{19}$$

which connects bare and renormalized Green functions via the electron and phonon self-energy, Σ and Π, respectively. The self-energies are defined as the sum of all one-particle irreducible Feynman diagrams, i.e., as the sum of all Feynman diagrams, which cannot be separated into two distinct graphs by cutting a single electron or phonon line.
In the following we will discuss the most important contributions to the self-energies in more detail.

3.2 Electron self-energy

The lowest-order diagram of the electron self-energy represents the virtual exchange of a phonon as shown in Fig. 2

$$\Sigma_{ep}(k, i\omega_n) = -\frac{1}{\beta}\sum_{n'}\frac{1}{N_q}\sum_{k',q} g^q_{k',k} G_0(k', i\omega_{n'})(g^q_{k',k})^* D_0(q, i\omega_{n'} - i\omega_n)\,. \tag{20}$$

After performing the Matsubara sum over $\omega_{n'}$ one obtains

$$\Sigma_{ep}(k, i\omega_n) = \frac{1}{N_q}\sum_{k',q} |g^q_{k',k}|^2 \left(\frac{b(\omega_q) + f(\varepsilon_{k'})}{i\omega_n + \omega_q - \varepsilon_{k'}} + \frac{b(\omega_q) + 1 - f(\varepsilon_{k'})}{i\omega_n - \omega_q - \varepsilon_{k'}}\right). \tag{21}$$

Σ_{ep} depends on temperature T via the Fermi and Bose distribution functions, $f(\varepsilon) = (e^{\varepsilon/T} + 1)^{-1}$ and $b(\omega) = (e^{\omega/T} - 1)^{-1}$, respectively.

To discuss the quasiparticle renormalization, we consider the retarded Green function, which is obtained by analytic continuation to the real axis via $i\omega_n \to \varepsilon + i\delta$ with an infinitesimal positive δ. It is connected to the analytic continuation of the self-energy via the Dyson equation

$$G(k,\varepsilon) = (\varepsilon - \varepsilon_k - \Sigma(k,\varepsilon))^{-1}\,. \tag{22}$$

If the self-energy is small enough, the spectral function $A_k(\varepsilon) = -\mathrm{Im}\, G(k, \varepsilon + i\delta)$ consists of a well defined peak at a shifted quasiparticle energy determined by the real part of Σ

$$\bar{\varepsilon}_k = \varepsilon_k + \mathrm{Re}\,\Sigma(k,\bar{\varepsilon}_k)\,. \tag{23}$$

The quasiparticle acquires a finite lifetime leading to a linewidth (full width at half maximum)

$$\Gamma_k = -2\mathrm{Im}\,\Sigma(k,\bar{\varepsilon}_k)\,, \tag{24}$$

which is determined by the imaginary part.
It is straightforward to perform the analytic continuation of $\Sigma_{ep}(k, i\omega_n \to \varepsilon + i\delta)$ in the form given in Eq. (21) and to derive the expression for the imaginary part

$$\begin{aligned}\mathrm{Im}\,\Sigma_{ep}(k,\varepsilon) = -\pi\frac{1}{N_q}\sum_{k',q}|g^q_{k',k}|^2\Big(\quad &\delta(\varepsilon-\varepsilon_{k'}+\omega_q)(b(\omega_q)+\qquad f(\varepsilon_{k'}))\\ &+\delta(\varepsilon-\varepsilon_{k'}-\omega_q)(b(\omega_q)+1-f(\varepsilon_{k'}))\Big).\end{aligned} \tag{25}$$

This can be rewritten by introducing two spectral functions

$$\alpha^2F^{\pm}_k(\varepsilon,\omega) = \frac{1}{N_q}\sum_q \delta(\omega-\omega_q)\sum_{k'}|g^q_{k',k}|^2\delta(\varepsilon-\varepsilon_{k'}\pm\omega)\,. \tag{26}$$

They depend on the electronic state k via the EPC vertex. The imaginary part can then be cast in the form

$$\mathrm{Im}\,\Sigma_{ep}(k,\varepsilon) = -\pi\int_0^\infty d\omega\Big(\alpha^2F^+_k(\varepsilon,\omega)[b(\omega)+f(\omega+\varepsilon)]+\alpha^2F^-_k(\varepsilon,\omega)[b(\omega)+f(\omega-\varepsilon)]\Big). \tag{27}$$

The physical interpretation of this expression is as follows. When a quasiparticle hole is created at a state k ($\varepsilon < \varepsilon_F$), electrons can scatter from states with higher or lower energies, respectively (see Fig. 3). By conservation of energy, the first process involves a simultaneous emission of a phonon, while the second one is related to the absorption of a phonon. The probability is described by $\alpha^2F^-_k$ and $\alpha^2F^+_k$, respectively, weighted with the appropriate bosonic and fermionic distribution functions. Both processes provide decay channels contributing additively to the linewidth (inverse lifetime) of the quasiparticle. A similar description holds when a quasiparticle (electron) is created at energies above the Fermi level.
Very often, a simplification is made which is called the quasielastic approximation. Because the electronic energy scale is typically much larger than the phonon energies, differences between emission and absorption spectra are rather small, and it is well justified to ignore the phonon energy ω_q in the δ-function of (26), such that $\alpha^2F^{\pm}_k \approx \alpha^2F_k$ with

$$\alpha^2F_k(\varepsilon,\omega) = \frac{1}{N_q}\sum_q \delta(\omega-\omega_q)\sum_{k'}|g^q_{k',k}|^2\delta(\varepsilon-\varepsilon_{k'})\,. \tag{28}$$

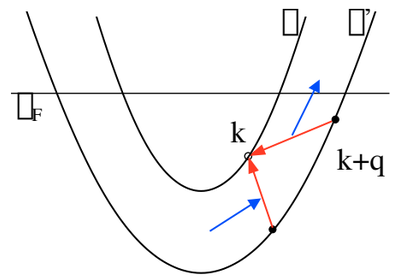


Fig. 3: *Illustration of the scattering processes contributing to the self-energy of a hole quasiparticle with momentum* $\mathbf{k}$ *and band index* ν*. Electrons (red lines) can scatter virtually from states with higher or lower energies under simultaneous emission or absorption of a phonon (blue lines), respectively.*

The self-energy then simplifies to

$$\mathrm{Im}\,\Sigma_{ep}(k,\varepsilon) = -\pi \int_0^\infty d\omega \Big(\alpha^2 F_k(\varepsilon,\omega)[2b(\omega) + f(\omega+\varepsilon) + f(\omega-\varepsilon)] \Big). \tag{29}$$

It is instructive to evaluate this expression for the simple Einstein model, where only a single dispersionless phonon mode with energy Ω couples to the electrons. In the limit $T \to 0$ one finds

$$\mathrm{Im}\,\Sigma_{ep}(k,\varepsilon) \to -\pi A(\varepsilon)\big(2 - \Theta(\Omega-\varepsilon) - \Theta(\Omega+\varepsilon)\big)\,, \tag{30}$$

where $\Theta(x)$ denotes the Heaviside step function, and $A(\varepsilon) = 1/N_k \sum_{k',q} |g^q_{k',k}|^2 \delta(\varepsilon - \varepsilon_{k'})$ represents the density of states at energy ε weighted by scattering matrix elements. Typically $A(\varepsilon)$ is slowly varying on the scale of phonon energies. On contrast, $\Sigma_{ep}(\varepsilon)$ vanishes for energies $|\varepsilon| < \Omega$ and shows a step at Ω, because of the presence of the step functions. This reflects the fact that no phonon modes are available for decay when $|\varepsilon| < \Omega$. $\mathrm{Re}\,\Sigma_{ep}$ can be obtained via the Kramers-Kronig relation

$$\mathrm{Re}\,\Sigma_{ep}(k,\varepsilon) = \frac{1}{\pi} \int d\varepsilon' \, \frac{\mathrm{Im}\,\Sigma_{ep}(k,\varepsilon')}{\varepsilon - \varepsilon'}\,. \tag{31}$$

As shown in Fig. 4(a) it contains a maximum at $\varepsilon = \Omega$ and has a finite slope at $\varepsilon \to 0$. The resulting dispersion for the renormalized quasiparticle is sketched in Fig. 4(b). It shows two characteristics: (i) the dispersion is strongly modified in the vicinity of ε_F in the range of phonon energies, altering the Fermi velocity related to the slope of $\mathrm{Re}\,\Sigma_{ep}(\varepsilon \to 0)$. (ii) A cusp appears at $\varepsilon = \pm\Omega$.

For a more realistic phonon spectrum which covers continuously an energy range $0 \le \omega \le \omega_{\max}$, the step-like feature in $\mathrm{Im}\,\Sigma_{ep}(\varepsilon)$ is washed out, but $\Sigma_{ep}(\varepsilon)$ still varies rapidly in the energy range of the phonons. The cut in the renormalized dispersion is then replaced by a kink. An example of an experimentally determined self-energy is given in Fig. 4(c) and (d).

The spectral function $\alpha^2 F_k$ contains the essential information related to the electron-phonon coupling of the specific electronic state $k = (\mathbf{k}\nu)$. A convenient measure for the strength of the

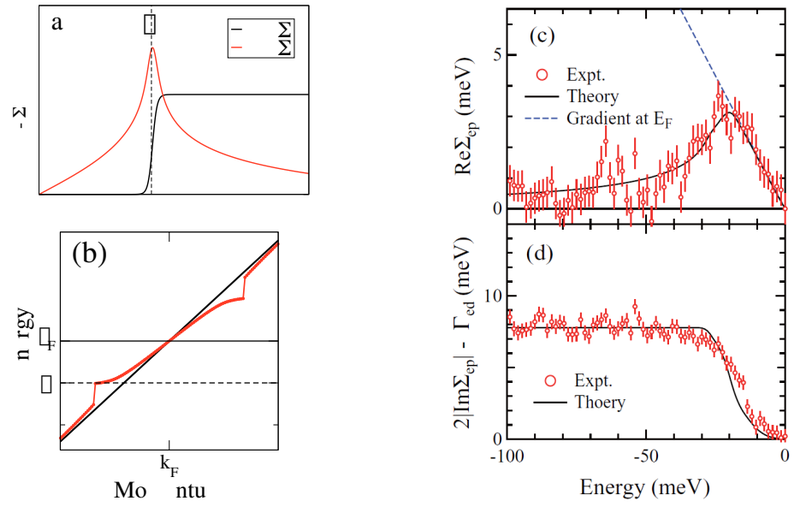


Fig. 4: *Illustration of the renormalization of an electronic band coupling to an Einstein-type phonon branch with energy Ω. (a) Real and imaginary part of the electron self-energy. (b) Renormalized quasiparticle dispersion, showing a kink at the phonon frequency. (c) Real and (d) imaginary part of the electron self-energy extracted from angle-resolved photoemission spectroscopy measurements taken for an electronic surface band of the Cu(110) surface. After Jiang* et al. *[8]*

EPC is the dimensionless coupling parameter

$$\lambda_k = 2 \int d\omega \, \frac{\alpha^2 F_k(\bar{\varepsilon}_k, \omega)}{\omega} \, . \tag{32}$$

It characterizes the strength of the coupling of a specific electronic state to the whole phonon spectrum, and depends both on the momentum and band character of the electronic state.
There are two relations which connect this parameter to experimentally accessible quantities. The first is related to the real part of the self-energy for an electronic band crossing the Fermi level:

$$\lambda_k = \left. \frac{\partial \mathrm{Re}\, \Sigma_{ep}(k, \varepsilon)}{\partial \varepsilon} \right|_{\varepsilon=0, T=0} \, . \tag{33}$$

Thus the coupling constant is given by the slope of $\mathrm{Re}\Sigma_{ep}$ right at the Fermi energy in the limit $T \to 0$. λ_k is also called the mass-enhancement parameter, because the quasiparticle velocity is changed to $v_F^* = v_F/(1+\lambda_k)$ and can be interpreted as an enhanced effective mass $m_k^* = m_k(1+\lambda_k)$, where m_k denotes the unrenormalized mass. Eq. (33) is often utilized in ARPES measurements of bands crossing the Fermi level, which attempt to extract the energy dependence of the real part of the self-energy.
A second route to determine the coupling constant of an electronic state is via the temperature dependence of the linewidth

$$\Gamma_k = \pi \int_0^\infty d\omega \Big(\alpha^2 F_k(\bar{\varepsilon}_k, \omega)[2b(\omega) + f(\omega + \bar{\varepsilon}_k) + f(\omega - \bar{\varepsilon}_k)] \Big) \, . \tag{34}$$

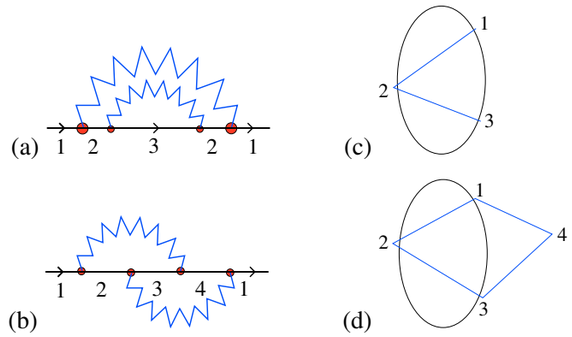


Fig. 5: *(a) and (b) Diagrammatic representation of the two second-order contributions to the electron self-energy. Blue zigzag lines represent phonon and black lines electron propagators. (c) and (d) Schematic drawing of the Fermi surface and the states contributing to the graphs (a) and (b), respectively.*

In Eq. (34), the T-dependence it contained solely in the Bose and Fermi distribution functions. For $T \to 0$, it approaches a finite value given by

$$\Gamma_k \to 2\pi \int_0^{\omega_{\max}} d\omega \, \alpha^2 F_k(\bar{\varepsilon}_k, \omega) \, . \tag{35}$$

With increasing T, the linewidth increases for all energies. For temperatures larger than the maximum phonon frequencies, this T-dependence becomes almost linear, and its slope is determined by the average coupling parameter defined above

$$\Gamma_k \approx 2\pi \lambda_k T \, . \tag{36}$$

This relationship has been widely used to extract λ_k from measurements of $\Gamma_k(T)$, in particular for surface electronic states.

3.3 Migdal's theorem

So far we have discussed the influence of phonons on the electronic properties in lowest order of the electron-phonon coupling. What about higher-order corrections? A very important answer is given by the *Migdal's theorem*, which is relevant for both the normal-state properties discussed here and the Eliashberg theory of superconductivity presented in the next section. We give only a very brief qualitative discussion here, more details can be found in literature [9,10,7]. Fig. 5 (a) and (b) show two next-order corrections to Σ_{ep}. The first is a self-energy contribution to an inner line and can be taken into account by using the full Green function G for the intermediate state instead of G_0. In contrast, the graph in Fig. 5(b) is a *vertex correction*. Migdal's theorem now states that vertex corrections are small compared to self-energy graphs and can be neglected. More precisely, this is true for those parts of the renormalized Green function which are sensitive to the phonons. Such contributions involve intermediate states whose energies are close to each other.



Fig. 6: *(a) Diagrammatic representation of the imaginary part of the phonon self-energy up to second order in the electron-phonon vertex. Blue zigzag lines represent phonon, black lines electron propagators.*

Fig. 5(c) and (d) show schematically the Fermi surfaces and the states which make a contribution to the graphs in Fig. 5(a) and (b), respectively. The first case contains only small energy differences $\varepsilon_{12} = \varepsilon_1 - \varepsilon_2$ and ε_{13}. In the second case, momentum conservation leads to a large difference ε_{14}. This unfavorable situation can only be avoided when one of the intermediate phonon momenta becomes small. Migdal showed, that for normal metals the phase space for such processes is very small, and the contribution from graph Fig. 5(d) is by a factor ω_D/ε_F smaller than that from graph (a), where ω_D denotes the Debye frequency, ε_F the Fermi energy. Thus ω_D/ε_F represents a typical ratio of phononic and electronic energy scales which is of the order of 0.1.

The phase space argument of Migdal's theorem breaks down in two circumstances: (i) For a significant part of processes both phonons have small q (both $k_2 - k_1$ and $k_3 - k_2$ are small). This can happen in metals with very small Fermi surfaces, for example in low-doped semiconductors. (ii) Metals with a one-dimensional Fermi surface topology (quasi-1D). In addition, Migdal's theorem becomes questionable in the case of metals with very small band widths, where the ratio ω_D/ε_F is not small any more.

According to Migdal's theorem, Σ_{ep} is well represented by the single graph shown in Fig. 2, except that G_0 is replaced by G, and other contributions can be neglected. According to arguments given by Migdal and Holstein, this replacement again gives small corrections of the order of ω_D/ε_F [9, 11]. Thus to a good approximation it is justified to use G_0, so the previous formulas still hold. This approximation does not work anymore for the superconducting state, as discussed in the next section.

However, the analysis given above rests on a simplified solution of the Dyson equation which can break down for larger coupling. Then the renormalization becomes much more involved and requires the solution of the Dyson equation in the complex plane [12]. The spectral function develops a complex structure, which indicates the break-down of the quasiparticle picture.

3.4 Phonon self-energy and linewidth

The EPC also renormalizes the phononic quasiparticles. The measurement of the phonon linewidth actually provides another way to gain experimental information about the coupling strength. We will briefly sketch this approach here.

The finite linewidth or inverse lifetime of a phonon mode is connected to the imaginary part of the phonon self-energy by $\gamma_q = -2\mathrm{Im}\Pi_q(\omega)$. The lowest-order diagram contributing to $\Pi_q(\omega)$

is shown in Fig. 6

$$\begin{aligned} \Pi_q(i\nu_m) &= \frac{1}{\beta}\sum_n \frac{1}{N_k}\sum_{k,k'} |g^q_{k',k}|^2 \; G_0(k, i\omega_n)\, G_0(k'\nu', i\omega_n + i\nu_m) \\ &= \frac{1}{N_k}\sum_{k',k} |g^q_{k',k}|^2 \; \frac{f(\varepsilon_k) - f(\varepsilon_{k'})}{i\nu_m + \varepsilon_k - \varepsilon_{k'}}\,, \end{aligned} \tag{37}$$

leading after analytic continuation to the following expression for the linewidth (half-width at half maximum)

$$\gamma_q = -2\mathrm{Im}\Pi_q(\omega_q) = 2\pi\frac{1}{N_k}\sum_{k',k} |g^q_{k',k}|^2 \big(f(\varepsilon_k) - f(\varepsilon_{k'})\big)\delta(\omega_q + (\varepsilon_k - \varepsilon_{k'}))\,. \tag{38}$$

This expression contains the T-dependence via the Fermi distribution functions f. Because phonon energies are typically small compared to electronic energies, the energy difference $\varepsilon_k - \varepsilon_{k'}$ is also small, and one can approximate

$$f(\varepsilon_k) - f(\varepsilon_{k'}) \approx f'(\varepsilon_k)(\varepsilon_k - \varepsilon_{k'}) \to -f'(\varepsilon_k)\omega_q \tag{39}$$

with $f' = df/d\varepsilon$. For $T \to 0$, $f'(\varepsilon_k) \to -\delta(\varepsilon_k)$, and by neglecting ω_q inside the δ-function, the expression further simplifies to

$$\gamma_q = 2\pi\omega_q\frac{1}{N_k}\sum_{k',k} |g^q_{k',k}|^2 \; \delta(\varepsilon_k)\,\delta(\varepsilon_{k'})\,. \tag{40}$$

This approximate expression for the linewidth, first derived by Allen [13], is widely used in numerical calculations. As will be discussed in the next section, γ_q in the form of Eq. (40) enters directly the expression for the coupling strength of a phonon mode relevant for superconductivity. Thus measurements of the phonon linewidths, for example by inelastic neutron or X-ray scattering experiments, provide information about the importance of a phonon mode for the pairing. One has to keep in mind, however, that γ_q only represents the contribution from EPC, while the experimental linewidth also contains other contributions like those from anharmonic decay processes. Furthermore, approximation (40) does not hold in the limit $\mathbf{q} \to 0$ for metals, because the phonon frequency in Eq. (38) cannot be neglected anymore for intraband contributions, which involve arbitrarily small energy differences $\varepsilon_k - \varepsilon_{k'}$.

4 Phonon-mediated superconductivity

Superconductivity is a macroscopic quantum phenomenon of the electron system. Its origin lies in an instability of the Fermi liquid state that leads to a new ground state of correlated paired electrons (Cooper pairs). In their seminal paper, Bardeen, Cooper, and Schrieffer (BCS) [14] have shown that this state is stabilized, whenever there exists an attractive interaction among two electrons. Such an attractive interaction is always provided by the electron-phonon coupling, which thus represents a natural source for pairing in any metal. EPC is known to be the pairing

mechanism in most superconductors, which are commonly termed classical superconductors to distinguish them from more exotic materials where other types of pairing mechanism are suspected.
The BCS theory treated the EPC only in a simplified form appropriate for the weak coupling limit. A more complete theory has been soon after worked out applying many-body techniques (for reviews see, e.g., Refs. [15, 10, 16, 17]) . The resulting Eliashberg theory [18] extends the framework of BCS into the strong coupling regime and allows a quantitative prediction of many properties of the superconducting state. An important property of the superconducting state is that the quasiparticle spectrum is gaped. The size of the gap plays the role of an order parameter. In the following, we discuss the essential ingredients of the theory of strong-coupling phonon-mediated superconductivity, also known as the Migdal-Eliashberg theory. First, we give a simple derivation of an effective electron-electron interaction mediated by phonons. Using many-body techniques we then derive the superconducting gap equations and identify the important quantities related to the electron-phonon coupling, which determine the superconducting properties.

4.1 Effective electron-electron interaction

The coupling of the electrons to the phonon system does introduce an effective electron-electron interaction, which can act as a pairing interaction evoking the superconducting state. The general approach using many-body techniques will be discussed below. Here a simple but instructive derivation of the effective interaction is given with the help of a properly chosen canonical transformation. To simplify the discussion, we will consider the case of a single, spinless quasiparticle band coupled to a single phonon (boson) mode. The Fröhlich Hamiltonian then reads ($g_{\mathbf{k},\mathbf{q}} \equiv g^{\mathbf{q}}_{\mathbf{k}+\mathbf{q},\mathbf{k}}$)

$$H = \sum_{\mathbf{k}} \varepsilon_{\mathbf{k}}\, c^{\dagger}_{\mathbf{k}} c_{\mathbf{k}} + \sum_{\mathbf{q}} \omega_{\mathbf{q}} \left(b^{\dagger}_{\mathbf{q}} b_{\mathbf{q}} + \frac{1}{2} \right) + \sum_{\mathbf{kq}} g_{\mathbf{k},\mathbf{q}}\, c^{\dagger}_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} \left(b_{\mathbf{q}} + b^{\dagger}_{-\mathbf{q}} \right) . \tag{41}$$

Let us consider the Hamiltonian

$$H = H_0 + \eta H_1 \,, \tag{42}$$

where H_0 is the unperturbed Hamiltonian, H_1 the perturbation, and η represents an expansion coefficient, which is considered to be small. The idea is to perform a canonical transformation

$$H' = e^{-\eta S} H e^{\eta S} \tag{43}$$

and eliminate the first-order term in η by choosing the operator S appropriately. Expanding Eq. (43) in a power series in η gives

$$H' = H + \eta[H, S] + \frac{\eta^2}{2}[[H, S], S] + O(\eta^3) \tag{44}$$

$$= H_0 + \eta(H_1 + [H_0, S]) + \eta^2[H_1, S] + \frac{\eta^2}{2}[[H_0, S], S] + O(\eta^3) \,. \tag{45}$$

To eliminate the term linear in η one has to find an S which fulfills the condition

$$H_1 + [H_0, S] = 0\,. \tag{46}$$

Then the transformed Hamiltonian can be written as

$$H' = H_0 + H_{\text{eff}} + O(\eta^3) \qquad \text{with} \quad H_{\text{eff}} = \frac{\eta^2}{2}[H_1, S]\,. \tag{47}$$

This general approach is now applied to the Fröhlich Hamiltonian (41) with $H_0 = H_e + H_{ph}$ and $\eta H_1 = H_{e-ph}$. For the canonical operator we make the ansatz

$$S = \sum_{\mathbf{kq}} g_{\mathbf{k},\mathbf{q}}\, c^\dagger_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} \left(x_{\mathbf{k},\mathbf{q}} b_{\mathbf{q}} + y_{\mathbf{k},\mathbf{q}} b^\dagger_{-\mathbf{q}} \right)\,. \tag{48}$$

The parameters $x_{\mathbf{k},\mathbf{q}}$ and $y_{\mathbf{k},\mathbf{q}}$ will be determined in order to fulfill Eq. (46). Evaluating the commutators gives

$$[H_e, S] = \sum_{\mathbf{kq}} g_{\mathbf{k},\mathbf{q}}\, (\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}}) c^\dagger_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} \left(x_{\mathbf{k},\mathbf{q}} b_{\mathbf{q}} + y_{\mathbf{k},\mathbf{q}} b^\dagger_{-\mathbf{q}} \right) \tag{49}$$

$$[H_{ph}, S] = \sum_{\mathbf{kq}} g_{\mathbf{k},\mathbf{q}}\, c^\dagger_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} \left(-x_{\mathbf{k},\mathbf{q}} \omega_{\mathbf{q}} b_{\mathbf{q}} + y_{\mathbf{k},\mathbf{q}} \omega_{-\mathbf{q}} b^\dagger_{-\mathbf{q}} \right)\,. \tag{50}$$

Using the relation $\omega_{\mathbf{q}} = \omega_{-\mathbf{q}}$ this combines to

$$H_1 + [H_0, S] = \sum g_{\mathbf{k},\mathbf{q}} c^\dagger_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} \Big(\quad \big(1 + (\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}} - \omega_{\mathbf{q}})\, x_{\mathbf{k},\mathbf{q}}\big) b_{\mathbf{q}} \tag{51}$$

$$+ \big(1 + (\varepsilon_{\mathbf{k}+\mathbf{q}} - \varepsilon_{\mathbf{k}} + \omega_{\mathbf{q}})\, y_{\mathbf{k},\mathbf{q}}\big) b^\dagger_{-\mathbf{q}} \Big)\,. \tag{52}$$

This expression vanishes when

$$x_{\mathbf{k},\mathbf{q}} = (\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}} + \omega_{\mathbf{q}})^{-1} \quad \text{and} \quad y_{\mathbf{k},\mathbf{q}} = (\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}} - \omega_{\mathbf{q}})^{-1}\,. \tag{53}$$

The last step is to evaluate the effective interaction Eq. (47). The commutator $[H_1, S]$ has the form $[Aa, Bb]$ with $A, B \propto c^\dagger c$ containing products of fermion operators, and $a, b \propto xb + yb^\dagger$ containing sums of boson operators. From the general relationship $[Aa, Bb] = AB[a, b] + [A, B]ab - [A, B][a, b]$ it is easy to see that there are three types of contributions. Keeping in mind that $[A, B]$ is again a product of the form $c^\dagger c$ and $[a, b]$ a c-number, the last term represents a one-body electron operator, which actually can be shown to vanish. The second term describes an effective coupling of an electron to two phonons, also called a non-linear coupling term.
We are interested in the first term, which is proportional to the product of two fermionic creation and two annihilation operators, $c^\dagger c c^\dagger c$, and thus represents an effective electron-electron interaction. Explicitly it has the form

$$H_{\text{eff}} = \frac{\eta^2}{2} \sum_{\mathbf{kk'q}} g_{\mathbf{k},\mathbf{q}}\, g_{\mathbf{k'},-\mathbf{q}}\, (y_{\mathbf{k'},-\mathbf{q}} - x_{\mathbf{k'},-\mathbf{q}})\, c^\dagger_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} c^\dagger_{\mathbf{k'}-\mathbf{q}} c_{\mathbf{k'}} \tag{54}$$

$$= \eta^2 \sum_{\mathbf{kk'q}} V_{\text{eff}}(\mathbf{k}, \mathbf{k'}, \mathbf{q})\, c^\dagger_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} c^\dagger_{\mathbf{k'}-\mathbf{q}} c_{\mathbf{k'}} \tag{55}$$

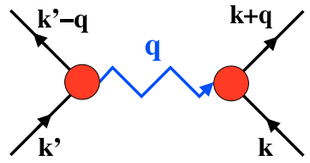


Fig. 7: *Diagrammatic representation of the effective electron-electron interaction mediated by the exchange of a phonon (blue zigzag line). Black lines indicate electronic states.*

with

$$V_{\text{eff}}(\mathbf{k},\mathbf{k}',\mathbf{q}) = g_{\mathbf{k},\mathbf{q}}\, g_{\mathbf{k}',-\mathbf{q}}\, \frac{\omega_{\mathbf{q}}}{(\varepsilon_{\mathbf{k}'} - \varepsilon_{\mathbf{k}'-\mathbf{q}})^2 - \omega_{\mathbf{q}}^2}\,. \tag{56}$$

H_{eff} describes the scattering of two electrons with momenta $\mathbf{k}$ and $\mathbf{k}'$ into states with momenta $\mathbf{k}+\mathbf{q}$ and $\mathbf{k}'-\mathbf{q}$ by the exchange of a virtual boson with momentum $\mathbf{q}$. This process is sketched in Fig. 7.

In the context of pairing in superconductors, the effective interaction between electrons with momenta $\mathbf{k}$ and $-\mathbf{k}$ is of special importance. Using $\varepsilon_{-\mathbf{k}} = \varepsilon_{\mathbf{k}}$ and $g_{-\mathbf{k},-\mathbf{q}} = g^*_{\mathbf{k},\mathbf{q}}$ one obtains

$$V_{\text{eff}}(\mathbf{k},-\mathbf{k},\mathbf{q}) = |g_{\mathbf{k},\mathbf{q}}|^2 \frac{\omega_{\mathbf{q}}}{(\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}})^2 - \omega_{\mathbf{q}}^2}\,. \tag{57}$$

This effective interaction is attractive (negative) for $|\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}}| < \omega_{\mathbf{q}}$ and repulsive (positive) for $|\varepsilon_{\mathbf{k}} - \varepsilon_{\mathbf{k}+\mathbf{q}}| > \omega_{\mathbf{q}}$. Eq. (57) shows that the electron-phonon coupling always introduces an attractive interaction for electronic scattering processes involving small energies of the order of phonon energies.

4.2 Nambu formalism

The superconducting state is a macroscopic quantum state, which is characterized by a coherent occupation of Cooper pairs, i.e., states with $(k\uparrow, -k\downarrow)$. In a many-body description, it is related to the appearance of anomalous Green functions

$$F(k,\tau) = -\langle T_\tau c_{k\uparrow}(\tau) c_{-k\downarrow}(0)\rangle \quad F^*(k,\tau) = -\langle T_\tau c^\dagger_{-k\downarrow}(\tau) c^\dagger_{k\uparrow}(0)\rangle \tag{58}$$

originally introduced by Gor'kov [19]. In the normal state these anomalous Green functions vanish. Starting from the Fröhlich Hamiltonian, one can set up a systematic perturbation expansion of the normal and anomalous Green functions, with the goal of obtaining a set of self-consistent equations. A necessary step is a partial resummation of an infinite number of diagrams, because the superconducting state can not be reached in any finite order of the perturbation.

A very convenient way to organize this algebra of diagrams has been introduced by Nambu [20]. One starts by defining the two-component operators

$$\Psi_k = \begin{pmatrix} c_{k\uparrow} \\ c^\dagger_{-k\downarrow} \end{pmatrix} \qquad \Psi^\dagger_k = \begin{pmatrix} c^\dagger_{k\uparrow}\,, & c_{-k\downarrow} \end{pmatrix} \tag{59}$$

and a 2×2 Green function

$$\underline{G}(k,\tau) = -\langle T_\tau \Psi_k(\tau)\Psi^\dagger_{-k}(0)\rangle = -\begin{pmatrix} \langle T_\tau c_{k\uparrow}(\tau)c^\dagger_{k\uparrow}(0)\rangle & \langle T_\tau c_{k\uparrow}(\tau)c_{-k\downarrow}(0)\rangle \\ \langle T_\tau c^\dagger_{-k\downarrow}(\tau)c^\dagger_{k\uparrow}(0)\rangle & \langle T_\tau c^\dagger_{-k\downarrow}(\tau)c_{-k\downarrow}(0)\rangle \end{pmatrix}$$
$$= \begin{pmatrix} G(k,\tau) & F(k,\tau) \\ F^*(k,\tau) & G(-k,-\tau) \end{pmatrix}. \tag{60}$$

In the following, underlined symbols indicate 2×2 matrices in spin space. Switching to the Fourier transform gives

$$\underline{G}(k,i\omega_n) = \frac{1}{2}\int_{-\beta}^{\beta} d\tau\, e^{i\omega_n\tau}\underline{G}(k,\tau) = \begin{pmatrix} G(k,i\omega_n) & F(k,i\omega_n) \\ F^*(k,i\omega_n) & -G(-k,-i\omega_n) \end{pmatrix}. \tag{61}$$

The next step is to rewrite the Fröhlich Hamiltonian in terms of $\Psi, \Psi^\dagger$. This is most easily done by using the Pauli matrices

$$\underline{\tau}_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},\ \underline{\tau}_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \underline{\tau}_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \underline{\tau}_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{62}$$

The non-interacting electronic part is rewritten as

$$H_e = \sum_{k\sigma} \varepsilon_k\, c^\dagger_{k\sigma}c_{k\sigma} \to \sum_k \varepsilon_k\, \Psi^\dagger_k\, \underline{\tau}_3\, \Psi_k \tag{63}$$

and the interaction part as

$$H_{e-ph} = \sum_{k\sigma}\sum_{\mathbf{q}j} g^q_{k'k}\, c^\dagger_{k'\sigma}c_{k\sigma}\left(b_q + b^\dagger_{-q}\right) \to \sum_k g^q_{k'k}\, \Psi^\dagger_{k'}\, \underline{\tau}_3\, \Psi_k \left(b_q + b^\dagger_{-q}\right). \tag{64}$$

The bare Green function (related to H_e) takes the form

$$\underline{G}_0(k,i\omega_n) = \begin{pmatrix} G_0(k,i\omega_n) & 0 \\ 0 & -G_0(-k,-i\omega_n) \end{pmatrix} = \begin{pmatrix} (i\omega_n - \varepsilon_k)^{-1} & 0 \\ 0 & (i\omega_n + \varepsilon_k)^{-1} \end{pmatrix}$$
$$= \left(i\omega_n\, \underline{\tau}_0 - \varepsilon_k\, \underline{\tau}_3\right)^{-1}. \tag{65}$$

One can show that the Dyson equation retains its usual form

$$\underline{G}^{-1}(k,i\omega_n) = \underline{G}_0^{-1}(k,i\omega_n) - \underline{\Sigma}(k,i\omega_n) \tag{66}$$

with the inversion performed in the 2-dimensional spin space, where the self-energy $\underline{\Sigma}$ is now a 2×2 matrix.

The diagrammatic expansion of the self-energy contains the same diagrams as in the normal state, with the difference that Green functions and vertices are now represented by 2×2 matrices. In particular $g^q_{k'k}$ is replaced by $g^q_{k'k}\,\underline{\tau}_3$.

4.3 Eliashberg theory

The Eliashberg theory is in essence the extension of the normal-state Migdal theory to the superconducting state. Using Migdal's theorem, the only important self-energy diagram is again given by Fig. 2. Within the Nambu formulation this gives

$$\underline{\Sigma}(k,i\omega_n) = -\frac{1}{\beta}\sum_{n'}\frac{1}{N_q}\sum_{k',q} g^q_{k'k}\,\underline{\tau}_3\,\underline{G}(k',i\omega_{n'})\,\underline{\tau}_3\,g^{-q}_{kk'}\,D(q,i\omega_{n'}-i\omega_n)\,. \tag{67}$$

Using the Pauli matrices, $\underline{\Sigma}$ can be written in the general form

$$\underline{\Sigma}(k,i\omega_n) = i\omega_n[1-Z(k,i\omega_n)]\,\underline{\tau}_0 + \chi(k,i\omega_n)\,\underline{\tau}_3 + \Phi(k,i\omega_n)\,\underline{\tau}_1 + \overline{\Phi}(k,i\omega_n)\,\underline{\tau}_2 \tag{68}$$

with as yet unknown and independent real functions Z, χ, Φ, and $\overline{\Phi}$. From the Dyson equation one finds

$$\underline{G}^{-1}(k,i\omega_n) = i\omega_n Z(k,i\omega_n)\,\underline{\tau}_0 - (\varepsilon_k+\chi(k,i\omega_n))\,\underline{\tau}_3 - \Phi(k,i\omega_n)\,\underline{\tau}_1 - \overline{\Phi}(k,i\omega_n)\,\underline{\tau}_2\,. \tag{69}$$

The inverted Green function is then, using $(a_0\underline{\tau}_0+\vec{a}\cdot\vec{\underline{\tau}})(a_0\underline{\tau}_0-\vec{a}\cdot\vec{\underline{\tau}}) = (a_0^2-\vec{a}^2)\underline{\tau}_0$,

$$\underline{G}(k,i\omega_n) = \Big(i\omega_n Z(k,i\omega_n)\,\underline{\tau}_0 + (\varepsilon_k+\chi(k,i\omega_n))\,\underline{\tau}_3 + \Phi(k,i\omega_n)\,\underline{\tau}_1 + \overline{\Phi}(k,i\omega_n)\,\underline{\tau}_2\Big)/\mathcal{D} \tag{70}$$

with $\mathcal{D} := \det\underline{G}^{-1} = (i\omega_n Z)^2 - (\varepsilon_k+\chi)^2 - \Phi^2 - \overline{\Phi}^2$. If one uses this expression for Eq. (67) and separates it into the τ-components, one arrives at four self-consistent equations for the four unknown functions Z, χ, Φ, and $\overline{\Phi}$

$$i\omega_n(1-Z(k,i\omega_n)) = -\frac{1}{\beta}\sum_{n'}\frac{1}{N_q}\sum_{k',q}|g^q_{k'k}|^2 D(q,i\omega_{n'}-i\omega_n)\,\frac{i\omega_{n'}Z(k',i\omega_{n'})}{\mathcal{D}(k',i\omega_{n'})} \tag{71}$$

$$\chi(k,i\omega_n) = -\frac{1}{\beta}\sum_{n'}\frac{1}{N_q}\sum_{k',q}|g^q_{k'k}|^2 D(q,i\omega_{n'}-i\omega_n)\,\frac{\varepsilon_{k'}+\chi(k',i\omega_{n'})}{\mathcal{D}(k',i\omega_{n'})} \tag{72}$$

$$\Phi(k,i\omega_n) = \frac{1}{\beta}\sum_{n'}\frac{1}{N_q}\sum_{k',q}|g^q_{k'k}|^2 D(q,i\omega_{n'}-i\omega_n)\,\frac{\Phi(k',i\omega_{n'})}{\mathcal{D}(k',i\omega_{n'})} \tag{73}$$

$$\overline{\Phi}(k,i\omega_n) = \frac{1}{\beta}\sum_{n'}\frac{1}{N_q}\sum_{k',q}|g^q_{k'k}|^2 D(q,i\omega_{n'}-i\omega_n)\,\frac{\overline{\Phi}(k',i\omega_{n'})}{\mathcal{D}(k',i\omega_{n'})}\,. \tag{74}$$

We note that because momentum conservation determines the phonon momentum, $\mathbf{q}=\mathbf{k}'-\mathbf{k}$, the sum over q is actually only a sum over different phonon branches (j).
Quasiparticle properties are determined by the poles of the Green function after analytic continuation, i.e., from $\mathcal{D}(k,i\omega_n\to\varepsilon+i\delta)=0$. This gives

$$E_k = \sqrt{\frac{(\varepsilon_k+\chi)^2}{Z^2}+\frac{\Phi^2+\overline{\Phi}^2}{Z^2}}\,. \tag{75}$$

The normal state corresponds to a solution $\Phi=\overline{\Phi}=0$. Z is the quasiparticle renormalization factor, and χ describes shifts in the electron energies. The superconducting state is characterized by a non-zero Φ or $\overline{\Phi}$. From Eq. (75) one can see that the gap function is given by

$$\Delta(k,i\omega_n) = \frac{\Phi(k,i\omega_n)-i\overline{\Phi}(k,i\omega_n)}{Z(k,i\omega_n)} \tag{76}$$

and describes the energy gap in the quasiparticle spectrum. Φ and $\overline{\Phi}$ obey the same equations and are expected have the same functional form up to a common phase factor. This phase factor becomes important in the description of Josephson junctions, but is irrelevant for the thermodynamic properties of a homogeneous superconductor. In the following, we choose the simple gauge $\overline{\Phi} = 0$.

4.4 Isotropic gap equations

The Eliashberg equations (74) represent a complicated non-linear set of equations which couple all momenta k with each other. We will now simplify them and derive the so-called isotropic equations where only the frequency dependence remains. A very detailed derivation was given by Allen and Mitrović [10]. Here we only briefly sketch the main steps. (i) We ignore changes of the phonon quasiparticles and replace D by the unrenormalized Green function

$$D(q, i\nu_m) \to D_0(q, i\nu_m) = \int d\omega\, \delta(\omega - \omega_q)\, \frac{2\omega}{(i\nu_m)^2 - \omega^2}\,. \tag{77}$$

It is then convenient to define the coupling function

$$\alpha^2 F(k, k', \omega) = N(0) \sum_q |g^q_{k'k}|^2\, \delta(\omega - \omega_q)\,. \tag{78}$$

Again, the sum extends only over the phonon branches j. $N(0) = \frac{1}{N_k}\sum_k \delta(\varepsilon_k)$ denotes the electronic density of states per spin at the Fermi energy. (ii) Similar to the normal state, the electron-phonon self-energy evokes a significant renormalization of quasiparticles only in an energy range $\pm\omega_D$ around the Fermi energy. It is therefore appropriate to consider the quantities Z and ϕ only at the Fermi energy. (iii) We consider only Fermi-surface averages of these quantities. The justification comes from the observation that the superconducting gaps are often very isotropic. Moreover, in real materials, defects are always present which tend to average anisotropic gaps [21]. Under these conditions we can replace the quantities Z and ϕ by their Fermi surface averages, e.g.

$$Z(i\omega_n) = \frac{1}{N_k} \sum_k w_k Z(k, i\omega_n) \tag{79}$$

with weights $w_k = \delta(\varepsilon_k)/N(0)$. Similarly one replaces the coupling function $\alpha^2 F$ by its value at the Fermi surface and averaged over both electron momenta

$$\begin{aligned} \alpha^2 F(\omega) = & \quad \frac{1}{N_k^2} \sum_{kk'} w_k\, w_{k'}\, \alpha^2 F(k, k', \omega) \\ = & \frac{1}{N(0)} \frac{1}{N_k^2} \sum_{kk'} |g^q_{k'k}|^2\, \delta(\varepsilon_k)\, \delta(\varepsilon_{k'})\, \delta(\omega - \omega_q)\,, \end{aligned} \tag{80}$$

which defines the *isotropic Eliashberg function*.

The only significant energy dependence comes from $\varepsilon_{k'}$ in the determinant $\mathcal{D}$. Putting everything together gives, for example,

$$\Phi(i\omega_n) = -\frac{1}{\beta}\sum_{n'}\int d\omega\, \frac{2\omega\alpha^2F(\omega)}{(\omega_n-\omega_{n'})^2+\omega^2}\,\Phi(i\omega_{n'})\frac{1}{N_q}\sum_{k'}\,\frac{1}{\mathcal{D}(\varepsilon_{k'},i\omega_{n'})} \tag{81}$$

with $\mathcal{D}(\varepsilon_{k'},i\omega_{n'}) = -[(\omega_{n'}Z(i\omega_{n'})]^2+\Phi(i\omega_{n'})^2+\varepsilon_{k'}^2]$ The final k sum is converted into an integral

$$\frac{1}{N_q}\sum_{k'}\frac{1}{\mathcal{D}(\varepsilon_{k'},i\omega_{n'})} = \int d\varepsilon\, N(\varepsilon)\frac{1}{\mathcal{D}(\varepsilon,i\omega_{n'}))} \approx \frac{-\pi N(0)}{\sqrt{[(\omega_{n'}Z(i\omega_{n'})]^2+\Phi(i\omega_{n'})^2}}\,. \tag{82}$$

In the last step, it was assumed that the electronic density of states $N(\varepsilon)$ does not have a pronounced ε dependence and can be replaced by its value at the Fermi energy. To simplify the following discussion, we will drop the equation for χ thus ignoring the related, often small, shift in the electronic energies. Indeed $\chi = 0$ holds exactly in the limit of infinite band width [17]. Using $\Delta(i\omega_n) = \Phi(i\omega_n)/Z(i\omega_n)$, this finally results in the *isotropic gap equations*

$$\begin{aligned}\omega_n(1-Z(i\omega_n)) &= -\pi\frac{1}{\beta}\sum_{n'}\Lambda(\omega_n-\omega_{n'})\,\frac{\omega_{n'}}{\sqrt{\omega_{n'}^2+\Delta(i\omega_{n'})^2}}\\ \Delta(i\omega_n)Z(i\omega_n) &= \pi\frac{1}{\beta}\sum_{n'}\Lambda(\omega_n-\omega_{n'})\,\frac{\Delta(i\omega_{n'})}{\sqrt{\omega_{n'}^2+\Delta(i\omega_{n'})^2}}\end{aligned} \tag{83}$$

with the interaction kernel

$$\Lambda(\nu_m) = \int d\omega\,\frac{2\omega\alpha^2F(\omega)}{\nu_m^2+\omega^2}\,. \tag{84}$$

The set of non-linear equations (83) must be solved self-consistently for a given temperature T and pairing function α^2F. The kernel entering both equations is an even function of ν_m. It takes its largest value at $\nu_m = 0$

$$\lambda = \Lambda(0) = 2\int d\omega\,\frac{\alpha^2F(\omega)}{\omega}\,. \tag{85}$$

λ is called the (isotropic) coupling constant and is a dimensionless measure of the average strength of the electron-phonon coupling. Depending on its value, materials are characterized as strong ($\lambda > 1$) or weak coupling ($\lambda < 1$) . Due to the factor $1/\omega$ in the integral low-energy modes contribute more to the coupling strength than high-energy modes.

The superconducting state is characterized by a solution with $\Delta(i\omega_n) \neq 0$. The largest T which still allows such a solution defines the critical temperature T_c. Because $\alpha^2F(\omega)$ as defined in Eq. (80) is a positive function, (83) always possess such a superconducting solution for low enough temperatures, i.e., a finite T_c.

An important feature of the Eliashberg gap equations is that they only depend on normal-state properties, which specify a particular material. These comprise the electronic band structure, phonons, and the EPC vertex, quantities which are accessible to *first principles* techniques as discussed in the next section.

At this stage it is useful to make the connection to some normal-state quantities introduced in the previous section. The isotropic Eliashberg function is related to the state-dependent spectral function (28) via appropriate momentum averages at the Fermi energy

$$\alpha^2F(\omega) = \sum_k w_k \alpha^2F_k(\varepsilon = 0, \omega)\,, \tag{86}$$

while the isotropic coupling constant is given by

$$\lambda = \sum_k w_k \lambda_k\,. \tag{87}$$

Similarly, α^2F can be expressed in terms of the phonon linewidths derived in the limit $T \to 0$, Eq. (40), as

$$\alpha^2F(\omega) = \frac{1}{2\pi N(0)} \frac{1}{N_q} \sum_q \frac{\gamma_q}{\omega_q}\, \delta(\omega - \omega_q)\,, \tag{88}$$

which leads to the formula for the isotropic coupling constant

$$\lambda = \frac{1}{\pi N(0)} \frac{1}{N_q} \sum_q \frac{\gamma_q}{\omega_q^2}\,. \tag{89}$$

The dimensionless prefactor γ_q/ω_q in (88) can be interpreted as a measure of the coupling due to an individual phonon mode. The Eliashberg function is then given as a sum over all phonon branches and averaged over phonon momentum.

4.4.1 Coulomb effects

Our derivation up to now was based on the Fröhlich Hamiltonian, where the electronic subsystem is approximated by bands of noninteracting quasiparticles ignoring any Coulomb interaction. The largest consequences of the Coulomb interaction are supposed to be built into the quantities ε_k (and similar into ω_q). The residual Coulomb interaction among the quasiparticles can, however, not be completely neglected in the discussion of phonon-mediated superconductivity. It has a repulsive character and tends to reduce or completely suppress the pairing. The quantity analogous to the electron-phonon coupling constant λ is the Coulomb parameter

$$\mu = N(0) \langle\langle V_C(k, k') \rangle\rangle_{FS}\,, \tag{90}$$

which is a Fermi surface average of the effective screened Coulomb interaction $V_C(k, k')$. μ is of the order of 1 and thus not a small parameter. But because the electronic timescale is usually much smaller than the vibrational one, or equivalently electronic energies are much larger than phononic ones, only a significantly reduced Coulomb parameter enters the Eliashberg equations. It was shown by Morel and Anderson [22], that the Coulomb repulsion can be taken into account by replacing the kernel in the equation for the gap function by

$$\Lambda(i\omega_n - i\omega_{n'}) \to [\Lambda(i\omega_n - i\omega_{n'}) - \mu^*(\omega_c)] \Theta(\omega_c - |\omega_{n'}|)\,. \tag{91}$$

A cutoff ω_c is introduced which must be chosen to be much larger than phononic energies. The effective Coulomb parameter or *Morel-Anderson Coulomb pseudopotential* obeys a scaling relation

$$\mu^*(\omega_c) = \frac{\mu}{1 + \mu \ln(\varepsilon_0/\omega_c)} \,. \tag{92}$$

ε_0 denotes a characteristic energy scale of the electronic system, where the average matrix elements of the Coulomb interaction becomes small ($\varepsilon_0 \approx$ few ε_F). In practice, μ^* is commonly treated as a phenomenological parameter of the order of ≈ 0.1 for normal metals. A more satisfactory approach, which actually allows to incorporate the Coulomb effects from *first principles*, is the density-functional theory of superconductors [23]. As this is the topic of a separate lecture of this Autumn School, we will not discuss this method further.

4.4.2 Transition temperature T_c

The transition temperature T_c is solely determined by the material-dependent quantities $\alpha^2 F(\omega)$ and μ^*. A thorough numerical analysis of the isotropic gap equations was carried out by Allen and Dynes [24], who used a standard spectrum for $\alpha^2 F$ but varied λ and μ^* over a large parameter range. Their study revealed two important aspects. Firstly, they found that in a reduced parameter space ($\lambda < 2$ and $\mu^* < 0.15$) T_c can be well approximated by a T_c formula proposed originally by McMillan [25], but with a modified prefactor

$$T_c = \frac{\omega_{\log}}{1.2} \exp\left[-\frac{1.04(1+\lambda)}{\lambda - \mu^*(1+0.62\lambda)}\right] . \tag{93}$$

The prefactor contains a properly defined average frequency of the phonon spectrum weighted with the coupling strength

$$\omega_{\log} = \exp\left[\int d\omega \log(\omega) W(\omega)\right] , \tag{94}$$

with the normalized weight function

$$W(\omega) = \frac{2}{\lambda} \frac{\alpha^2 F(\omega)}{\omega} \,. \tag{95}$$

This T_c formula is a significant refinement of the BCS formula $T_c = 1.13\,\omega_D \exp(-1/\lambda)$ derived for the weak-coupling limit.
Secondly, while the T_c formula suggests that T_c approaches a finite value in the limit $\lambda \to \infty$, the isotropic gap equations do not possess a principle upper bound for T_c. Instead the asymptotic relationship

$$T_c \propto \sqrt{\lambda \langle \omega^2 \rangle} \tag{96}$$

holds, where $\langle \omega^2 \rangle$ is the second moment of $W(\omega)$.

5 Density functional theory approach

In the previous sections we have outlined the basic theory for the effects of EPC in the normal and superconducting state. Central quantities are the screened EPC matrix elements, which are not directly accessible from experiment. Thus it is desirable to have a computational scheme which allows materials-dependent predictions. The most common approach is based on density functional theory, which is briefly described in the following.

Density functional theory (DFT) goes back to the seminal works of Hohenberg, Kohn, and Sham [26, 27] and has been outlined in various reviews [28–30]. It provides a framework to map the complex many-body problem of interacting electrons moving in an external potential $v_{\rm ext}(\mathbf{r})$ onto a fictitious system of noninteracting electrons. Their wavefunctions obey a single-particle equation (Kohn-Sham equation) [27]

$$\Big(-\nabla^2 + v_{\rm eff}(\mathbf{r}) \Big) \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r}) \,. \tag{97}$$

The effective potential $v_{\rm eff}(\mathbf{r})$ is a functional of the density given as the sum of the external potential and a screening potential

$$v_{\rm eff}[n] = v_{\rm ext} + v_{\rm scr}[n] = v_{\rm ext} + v_H[n] + v_{XC}[n] \,. \tag{98}$$

The Hartree and exchange-correlation potentials v_H and v_{XC} are functionals of the density defined as the functional derivative of the Hartree and exchange-correlation energies, $E_H[n] = \int d^3r \int d^3r'\, n(\mathbf{r}) n(\mathbf{r}')/|\mathbf{r}-\mathbf{r}'|$ and E_{XC}. The density is determined by the wavefunctions via

$$n(\mathbf{r}) = \sum_i f_i \, |\psi_i(\mathbf{r})|^2 \,, \tag{99}$$

with f_i the occupation number of the single-particle state ψ_i. Eqs. (97) and (99) have to be solved self-consistently.

An expression for the EPC matrix elements is then derived within a linear-response scheme in the following way. For a solid, $v_{\rm ext}$ represents the sum of ionic potentials felt by the electrons. A small displacement of an atom evokes a perturbation in $v_{\rm ext}$, which translates via the self-consistent equations into a perturbation of $v_{\rm eff}(\mathbf{r})$

$$\begin{aligned} \delta v_{\rm eff}(\mathbf{r}) &= \delta v_{\rm ext}(\mathbf{r}) + \delta v_{\rm scr}(\mathbf{r}) = \delta v_{\rm ext}(\mathbf{r}) + \int d^3r'\, I(\mathbf{r},\mathbf{r}')\, \delta n(\mathbf{r}') \\ I(\mathbf{r},\mathbf{r}') &\equiv \frac{\delta v_{\rm scr}(\mathbf{r})}{\delta n(\mathbf{r}')} = \frac{\delta v_{\rm H}(\mathbf{r})}{\delta n(\mathbf{r}')} + \frac{\delta v_{\rm XC}(\mathbf{r})}{\delta n(\mathbf{r}')} = \frac{2}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta^2 E_{XC}}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \,. \end{aligned} \tag{100}$$

In first-order perturbation theory the variation of the single-particle wavefunctions is

$$\delta\psi_i(\mathbf{r}) = \sum_{j(\neq i)} \frac{\langle j|\delta v_{\rm eff}|i\rangle}{\varepsilon_i - \varepsilon_j} \, \psi_j(\mathbf{r}) \,. \tag{101}$$

Using a similar expression for $\delta\psi_i^*(\mathbf{r})$ gives

$$\delta n(\mathbf{r}) = \sum_{i \neq j} \frac{f_i - f_j}{\varepsilon_i - \varepsilon_j} \, \langle j|\delta v_{\rm eff}|i\rangle \, \psi_i^*(\mathbf{r}) \psi_j(\mathbf{r}) \,. \tag{102}$$

Eqs. (100) and (102) must be solved self-consistently. The results are the first-order variation of the density and the effective potential. The latter is then used for a calculation of the EPC matrix elements for a periodic crystal in the following way. One considers a periodic displacement of the ions from their equilibrium positions, $\mathbf{R}_{ls} = \mathbf{R}^0_{ls} + \mathbf{u}_{ls}$ of the form

$$u_{ls\alpha} = U^{\mathbf{q}}_{s\alpha} e^{i\mathbf{q}\mathbf{R}^0_{ls}} + (U^{\mathbf{q}}_{s\alpha})^* e^{-i\mathbf{q}\mathbf{R}^0_{ls}} \,, \tag{103}$$

where l denotes the unit cell, s specifies the ion inside a unit cell, and α indicates Cartesian coordinates. The wavevector $\mathbf{q}$ determines the periodicity. Applying the self-consistent procedure described above results in the linear response of the effective potential $\frac{\partial v_{\mathrm{eff}}(\mathbf{r})}{\partial U^{\mathbf{q}}_{s\alpha}}$ which is then used to calculate the electron-phonon matrix elements

$$g^{\mathbf{q}j}_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu} = \sum_{s\alpha} A^{\mathbf{q}j}_{s\alpha} \left\langle \mathbf{k}+\mathbf{q}\nu' \left| \frac{\partial v_{\mathrm{eff}}}{\partial U^{\mathbf{q}}_{s\alpha}} \right| \mathbf{k}\nu \right\rangle . \tag{104}$$

where again the transformation to the normal mode representation was performed (see Eq. (7)). The self-consistency procedure automatically takes into account the important screening effects. Eq. (104) thus enables to calculate the screened EPC matrix elements on a microscopic level, including their full momentum dependence and to resolve the contributions from different electronic bands and phononic modes.

The same perturbational approach can also be used to calculate the harmonic phonon spectrum without further approximations. This approach has been widely used to predict lattice dynamical and EPC properties from *first principles* for a large variety of materials, and has proven to be quite accurate in predicting the pairing strength in phonon-mediated superconductors.

We illustrate this *first principles* approach for two examples of superconductors with remarkable properties, the high-pressure superconductor H_3S and the multiband superconductor MgB_2.

The current record holder with the highest superconducting transition temperature is hydrogen sulfide with a T_c of 203 K [31]. This superconducting phase was reached by applying a huge pressure of more than 200 GPa to a nominally H_2S sample. It was soon recognized that the superconducting phase is actually a high-pressure modification of H_3S. Most remarkably, the experimental discovery was preceded by theoretical predictions of such high T_c based on the Eliashberg theory [32, 33]. DFT studies suggest that the high-pressure phase has a very simple cubic lattice structure which might be slightly distorted into a rhombohedral structure at lower pressures due to small shifts of the H atoms (see upper left panel in Fig. 8). Its electronic structure is characterized by strong covalent H–S bonds which support huge coupling constants of $\lambda \approx 2$, carried predominantly by the high-energy hydrogen vibrations. Therefore, also the effective phonon frequency $\omega_{\log}$ is large, which in conjunction with a large λ leads to the high T_c values. Fig. 8 shows the results of such a calculation for various pressures. Because a large part of the coupling is carried by hydrogen, a large isotope shift of T_c is expected when replacing hydrogen by deuterium. Because of the light mass of hydrogen, anharmonicity likely is important and may change some aspects of these predictions [34]. Nevertheless, these theoretical studies lend strong support to the view that H_3S is a phonon-mediated superconductor.

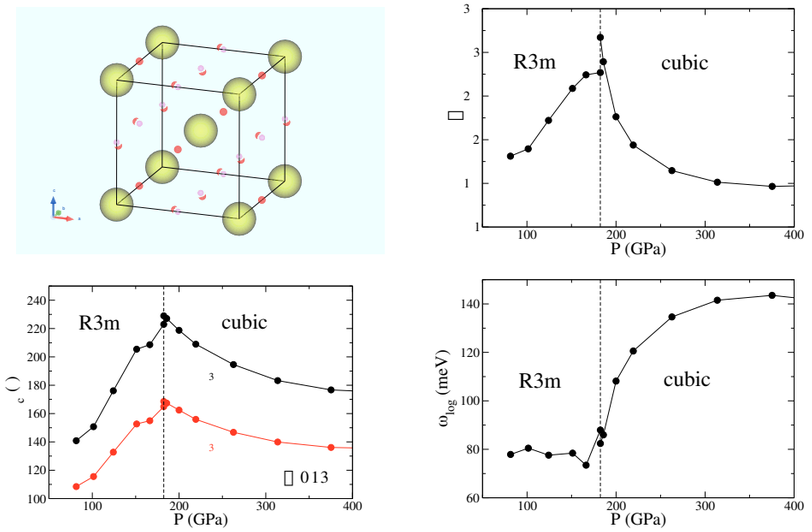


Fig. 8: *Upper left panel: predicted cubic structure of H_3S at high pressures. Large spheres denote S atoms, and small ones H atoms. The red circles indicate the small displacements of H atoms from their ideal cubic positions in the rhombohedral structure predicted for lower pressures. Right panels: variation of λ and $\omega_{\log}$ as function of pressure, respectively. Lower left panel: predicted T_c as calculated from the isotropic gap equations (with μ^* =0.13) as function of pressure. A large isotope shift is expected when replacing hydrogen with deuterium.*

In the previous section, we have restricted our discussion to the isotropic case. Anisotropic superconducting states can be handled using the full momentum dependence of the Eliashberg function (78). This has been done rarely in the past, as the fully anisotropic gap equations are difficult to solve. A special class of anisotropic superconductors are multiband superconductors, which possess several Fermi surface sheets. The superconducting gap can vary among the different sheets, but is approximately isotropic on a single sheet. In this case a partially averaged pairing function is appropriate

$$\alpha^2 F_{\nu\nu'}(\omega) = \frac{1}{N_{\nu'}(0)} \frac{1}{N_q} \sum_{\mathbf{q}j,\mathbf{k}} |g^{\mathbf{q}j}_{\mathbf{k}+\mathbf{q}\nu',\mathbf{k}\nu}|^2 \, \delta(\omega - \omega_{\mathbf{q}j}) \, \delta(\varepsilon_{\mathbf{k}\nu}) \, \delta(\varepsilon_{\mathbf{k}+\mathbf{q}\nu'}) \, . \tag{105}$$

The isotropic Eliashberg function is replaced by a matrix describing intraband and interband pairing contributions.

A textbook example of such a multiband superconductor is MgB_2. Here two types of electronic states are present at the Fermi level, σ and π states, which are derived mainly from the boron p states. Calculations of the band-resolved Eliashberg functions shown in Fig. 9 revealed that the pairing interaction is predominantly driven by the intraband σ–σ contribution. It originates from a strong coupling of σ states to in-plane B vibrations. This peculiar pairing interaction leads to a superconducting state with gaps of different magnitude for the σ and π Fermi surfaces,

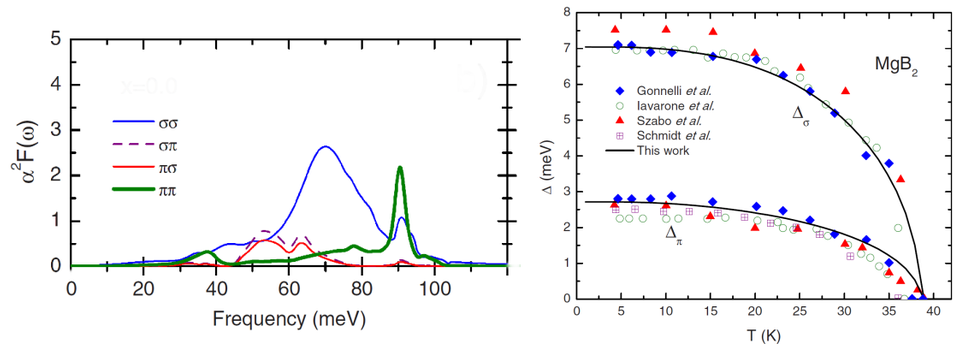


Fig. 9: *Left panel: Calculated band-resolved Eliashberg function for MgB_2. The dominant coupling originates from a large intraband σ–σ pairing interaction. Right panel: Calculated gaps of the σ and π Fermi surfaces as function of temperature. The large gap resides on the σ sheet. Symbols refer to experimental data [35–38]. Figures after [39].*

whose signature could be found, e.g., in specific heat measurements. Combination of *first principles* calculations and a two-band version of the Eliashberg equations were consistent with the experimental T_c of 39 K and could reproduce the measured temperature variation of the two gaps rather well (right panel of Fig. 9).

6 Summary

In this tutorial, an introduction to the theory of the electron-phonon interaction in metals was given. Focus was put on the renormalization properties of electronic and vibronic quasiparticles in the normal state, and on their role for the pairing interaction relevant for the superconducting state. This strong-coupling or Eliashberg theory has been tremendously successful in predicting material-dependent properties of various superconductors in great detail. Density functional theory provides a rather accurate *first principles* computational scheme to calculate the relevant electron-phonon vertex, which is one of the central quantities determining physical observables like electron renormalization, phonon linewidth, or phonon-mediated pairing interaction. Yet one has to keep in mind that the Eliashberg theory incorporates a variety of approximations. The current theoretical challenge is to extend its framework to include the usually neglected aspects of anharmonicity [40], and to quantify electron-phonon coupling effects in materials which are characterized by small electronic energy scales [41] and/or strong electron correlations [42].

Appendices

A Phonon quantization

Within the adiabatic approximation, statics and dynamics of the ions are governed by an effective potential

$$\Omega(\underline{\mathbf{R}}) = V_{ii}(\underline{\mathbf{R}}) + E_0(\underline{\mathbf{R}})\,, \tag{106}$$

where $E_0(\underline{\mathbf{R}})$ denotes the electronic ground-state energy for a given ion configuration $\underline{\mathbf{R}}$. The effective potential Ω builds the starting point of the microscopic theory of lattice dynamics, which has been outlined in a number of review articles [43–45].
Dynamical properties are derived by a systematic expansion of Ω for atom displacements $\mathbf{u}$ around a chosen reference configuration, $\mathbf{R}_i = \mathbf{R}_i^0 + \mathbf{u}_i$, leading to

$$\Omega(\underline{\mathbf{R}}) = \Omega(\underline{\mathbf{R}}^0) + \sum_{i\alpha} \Phi_a(i) u_{i\alpha} + \frac{1}{2} \sum_{i\alpha j\beta} \Phi_{\alpha\beta}(i,j) u_{i\alpha} u_{j\beta} + \dots\,. \tag{107}$$

Greek indices α and β denote Cartesian coordinates, while i and j are atom indices. The term of first order is the negative of the force acting on an atom in the reference configuration

$$F_{i\alpha} = -\left.\frac{\partial \Omega}{\partial R_{i\alpha}}\right|_0 = -\Phi_\alpha(i)\,. \tag{108}$$

It vanishes if one chooses as reference the equilibrium configuration, which minimizes Ω. The second-order coefficients are given by

$$\Phi_{\alpha\beta}(i,j) = \left.\frac{\partial^2 \Omega}{\partial R_{i\alpha} \partial R_{j\beta}}\right|_0 . \tag{109}$$

In periodic crystals, the atoms are characterized by two indices $i = (ls)$, which denote the unit cell (l) and the atoms inside a unit cell (s), respectively. For periodic boundary conditions, the Fourier transform of the force constant matrix is related to the dynamical matrix

$$D_{s\alpha s'\beta}(\mathbf{q}) = \frac{1}{\sqrt{M_s M_{s'}}} \sum_l \Phi_{\alpha\beta}(ls, 0s') e^{-i\mathbf{q}(\mathbf{R}^0_{ls} - \mathbf{R}^0_{0s'})}\,, \tag{110}$$

which determines the equation for the normal modes or phonons,

$$\sum_{s'\beta} D_{s\alpha s'\beta}(\mathbf{q}) \eta_{s'\beta}(\mathbf{q}j) = \omega^2_{\mathbf{q}j} \eta_{s\alpha}(\mathbf{q}j)\,. \tag{111}$$

$\omega_{\mathbf{q}j}$ and $\eta_{s\alpha}(\mathbf{q}j)$ denote the energy and polarization of the normal mode determined by the wavevector $\mathbf{q}$ and branch index j.
These quantities enter into the relationship between the atom displacements and the usual phonon annihilation and creation operators, $b_{\mathbf{q}j}$ and $b^\dagger_{\mathbf{q}j}$, describing quantized normal modes, as given in Eq. (7).

References

[1] G. Grimvall: *The Electron–Phonon Interaction in Metals*, Selected Topics in Solid State Physics, ed. by E. Wohlfarth (North-Holland, New York, 1981)

[2] M. Born and J.R. Oppenheimer, Ann. Physik **84**, 457 (1927)

[3] G.V. Chester and A. Houghton, Proc. Phys. Soc. **73**, 609 (1959)

[4] S.K. Sinha, Phys. Rev. **169**, 477 (1968)

[5] A.A. Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinski: *Methods of Quantum Field Theory in Statistical Physics* (Prentice-Hall, New Jersey, 1964)

[6] A.L. Fetter and J.D. Walecka: *Quantum Theory of Many-Particle Systems* (McGraw-Hill, San Francisco, 1971)

[7] G.D. Mahan: *Many-Particle Physics* (Plenum Press, New York, 1990)

[8] J. Jiang, S.S. Tsirkin, K. Shimada, H. Iwasawa, M. Arita, H. Anzai, H. Namatame, M. Taniguchi, I.Yu. Sklyadneva, R. Heid, K.-P. Bohnen, P.M. Echenique, and E.V. Chulkov, Phys. Rev. B **89**, 085404 (2014)

[9] A.B. Migdal, Sov. Phys. JETP **34**, 996 (1958)

[10] P.B. Allen and B. Mitrović, Solid State Physics **37**, 1 (1982)

[11] T. Holstein, Ann. Phys. **29**, 410 (1964)

[12] A. Eiguren, C. Ambrosch-Draxl, and P.M. Echenique, Phys. Rev. B **79**, 245103 (2009)

[13] P.B. Allen, Phys. Rev. B **6**, 2577 (1972)

[14] J. Bardeen, L.N. Cooper, and J.R. Schrieffer, Phys. Rev. **108**, 1175 (1957)

[15] D.J. Scalapino in R.D. Parks (ed.): *Superconductivity, Vol. 1* (Dekker, New York, 1969)

[16] J.P. Carbotte, Rev. Mod. Phys. **62**, 1027 (1990)

[17] J.P. Carbotte and F. Marsiglio: *Electron-Phonon Superconductivity* in K.H. Bennemann and J.B. Ketterson (eds.): *The Physics of Superconductors* (Springer, 2003)

[18] G.M. Eliashberg, Zh. Eksp. Fiz. **38**, 966 (1960); [Sov. Phys. JETP **11**, 696 (1960)]

[19] L.P. Gor'kov, Zh. Eksp. Teor. Fiz. **34** (1958) [Sov. Phys. JETP **7**, 505 (1958)]

[20] Y. Nambu, Phys. Rev. **117**, 648 (1960)

[21] P.W. Anderson, J. Phys. Chem. Solids **11**, 26 (1959)

[22] P. Morel and P.W. Anderson, Phys. Rev. **125**, 1263 (1962)

[23] L.N. Oliveira, E.K.U. Gross, and W. Kohn, Phys. Rev. Lett. **50**, 2430 (1988)

[24] P.B. Allen and R.C. Dynes, Phys. Rev. B**12**, 905 (1975)

[25] W.L. McMillan, Phys. Rev. **176**, 331 (1968)

[26] P. Hohenberg and W. Kohn, Phys. Rev. B**136**, 864 (1964)

[27] W. Kohn and L.J. Sham, Phys. Rev. A**140**, 1133 (1965)

[28] R.M. Dreizler and E.K.U. Gross:
Density Functional Theory: An Approach to the Quantum Many-Body Problem
(Springer-Verlag, Berlin, 1990)

[29] R.O. Jones and O. Gunnarsson, Rev. Mod. Phys. **61**, 689 (1989)

[30] R.G. Parr and W. Yang: *Density-Functional Theory of Atoms and Molecules*
(Oxford University Press, New York, 1989)

[31] A.P. Drozdov, M.I. Eremets, I.A. Troyan, V. Ksenofontov, and S.I. Shylin,
Nature **525**, 73 (2015)

[32] Y. Li, J, Hao, H. Liu, Y. Li, and Y. Ma, J. Chem. Phys. **140**, 174712 (2014)

[33] D. Duan, Y. Liu, F. Tian, Da Li, X. Huang, Z. Zhao, H. Yu, B. Liu, W. Tian, and T. Cui,
Sci. Rep. **4**, 6968 (2014)

[34] I. Errea, M. Calandra, C.J. Pickard, J.R. Nelson, R.J. Needs, Y. Li, H. Liu, Y. Zhang,
Y. Ma, and F. Mauri, Nature **532**, 81 (2016)

[35] R.S. Gonnelli, D. Daghero, G.A. Ummarino, V.A. Stepanov, J. Jun, S.M. Kazakov, and
J. Karpinski, Phys. Rev. Lett. **89**, 247004 (2002)

[36] M. Iavarone, G. Karapetrov, A.E. Koshelev, W.K. Kwok, G.W. Crabtree, D.G. Hinks,
W.N. Kang, E.M. Choi, H.J. Kim, H.J. Kim, and S.I. Lee,
Phys. Rev. Lett. **89**, 187002 (2002)

[37] P. Szabó, P. Samuely, J. Kačmarčík, T. Klein, J. Marcus, D. Fruchart, S. Miraglia,
C. Marcenat, and A.G.M. Jansen, Phys. Rev. Lett. **87**, 137005 (2001)

[38] H. Schmidt, J.F. Zasadzinski, K.E. Gray, and D.G. Hinks,
Phys. Rev. Lett. **88**, 127002 (2002)

[39] O. de la Peña Seaman, R. de Coss, R. Heid, K.-P. Bohnen, Phys. Rev. B **82**, 224508 (2010)

[40] I. Errea, M. Calandra, and F. Mauri, Phys. Rev. Lett. **111**, 177002 (2013)

[41] L.P. Gor'kov. Phys. Rev. B **93**, 054517 (2016)

[42] Z.P. Yin, A. Kutepov, and G. Kotliar, Phys. Rev. X**3**, 021011 (2013)

[43] M. Born and K. Huang: *Dynamical Theory of Crystal Lattices* (Clarendon Press, Oxford, 1954)

[44] A.A. Maradudin, E.W. Montroll, G.H. Weiss, and I.P. Ipatova: *Solid State Physics, Supplement 3* (Academic Press, New York, 1971), p. 1

[45] H. Boettger: *Principles of the Theory of Lattice Dynamics* (Physik Verlag, Weinheim, 1983)

16 Introduction to Superconducting Density Functional Theory

Antonio Sanna
Max-Planck Institut für Mikrostrukturphysik
Weinberg 2, 06120 Halle, Germany

Contents

1 Introduction **2**

2 Derivation of SCDFT **3**
2.1 Hamiltonian 3
2.2 Hohenberg Kohn theorem 4
2.3 The Kohn-Sham system 6
2.4 Decoupling approximations 8

3 Connection between SCDFT and Eliashberg theory **11**
3.1 A survey on Eliashberg theory of superconductivity 11
3.2 The Sham Schlüter connection 15
3.3 Analysis of a model 16
3.4 A simple SCDFT functional 18

4 Example: Superconductivity of a doped carbon-hydrogen nanotube **19**
4.1 Normal state properties 20
4.2 Solution of the gap equation and superconducting properties 21

5 Summary and conclusion **23**

E. Pavarini, E. Koch, R. Scalettar, and R. Martin (eds.)
The Physics of Correlated Insulators, Metals, and Superconductors
Modeling and Simulation Vol. 7
Forschungszentrum Jülich, 2017, ISBN 978-3-95806-224-5
http://www.cond-mat.de/events/correl17

1 Introduction

Density functional theories (DFTs) have proved to be a very convenient approach to deal with the many body problem in condensed matter. The original formulation of Hohenberg, Kohn, and Sham [1, 2] as well as several of its extensions [3–7] are now largely applied to study realistic systems and predict their properties with great success.

Density functional theories are all based on a Hohenberg-Kohn type of theorem that proves the existence of a one-to-one mapping between a set of chosen densities and the relative external potentials, implying that it is possible to achieve exact predictions of any observable without having to deal with anything more complex than the density itself.

All the complexity of the many body problem is transferred to the construction of a (universal) functional of the density. A DFT framework is practically useless until a good approximation for the universal functional is available. Clearly the process of functional construction is far from being straightforward. Also, for those used to diagrammatic methods, where approximations are usually expressed in terms of including/excluding some specific process, DFT functionals are sometimes rather obscure. Rarely one can deduce their domain of validity a priori from pure theoretical considerations.

On the other hand the advantage of DFTs is that, once a valid approximation for the functional is constructed, this functional is usually a simple object that can be easily evaluated implemented and applied. DFT algorithms are quite simple, incredibly fast, and very little is required by the user who can focus, instead, on the analysis of results.

Superconducting density functional theory (SCDFT) is an extension of DFT to account for the very peculiar symmetry breaking that occurs in a superconductor [8, 9]. Proposed in 1988 [10] by Oliveira Gross and Kohn it was later revisited [11, 12] to include the multi-component DFT of Kreibich and Gross [13], adding in this way the effect of nuclear motion.

In this lecture it is assumed that the reader has already a basic knowledge of superconductivity and the fundamentals of Green functions and field theory. There are excellent introductions to superconductivity, like DeGennes [9] who discusses in great details BCS theory and the physics of superconducting materials, while the book of Fetter and Walecka [14] is a great reference for the many body formalism.

For what concerns the superconducting state and the topic of this lecture the main references that have been used are the book of Vonsovsky, Izyumov, and Kurmaev [15] and for the SCDFT theory the original research papers [10–12] as well as the PhD theses of Kurth [16], Lueders [17], and Marques [18].

2 Derivation of SCDFT

2.1 Hamiltonian

The starting point of SCDFT is the non relativistic Hamiltonian for interacting electrons *and* nuclei.

$$H = H_e + H_{en} + H_n + H_{ext}, \tag{1}$$

where e stands for electrons, n for nuclei and ext for external fields.

$$H_e = \sum_\sigma \int d\boldsymbol{r}\, \psi^\dagger_\sigma(\boldsymbol{r}) \left[-\frac{1}{2}\nabla^2 - \mu\right] \psi_\sigma(\boldsymbol{r}) + \frac{1}{2}\sum_{\sigma\sigma'} \int d\boldsymbol{r} d\boldsymbol{r}' \psi^\dagger_\sigma(\boldsymbol{r})\, \psi^\dagger_{\sigma'}(\boldsymbol{r}') \frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|} \psi_{\sigma'}(\boldsymbol{r}')\, \psi_\sigma(\boldsymbol{r}) \tag{2}$$

where ψ are the electronic field operators, μ the chemical potential, and $d\boldsymbol{r}$ is an abbreviation for the 3D volume differential.

Nuclei need to be considered explicitly (not just as source of an external potential like in conventional DFT [1]) because in most known superconductors the ionic dynamics provides an essential part of the superconducting coupling

$$H_n = -\int d\boldsymbol{R}\, \Phi^\dagger(\boldsymbol{R}) \frac{\nabla^2}{2M} \Phi(\boldsymbol{R}) + \frac{1}{2}\int d\boldsymbol{R} d\boldsymbol{R}'\, \Phi^\dagger(\boldsymbol{R})\, \Phi^\dagger(\boldsymbol{R}') \frac{Z}{|\boldsymbol{R}-\boldsymbol{R}'|} \Phi(\boldsymbol{R}')\, \Phi(\boldsymbol{R}) \tag{3}$$

$$H_{en} = -\frac{1}{2}\sum_\sigma \int d\boldsymbol{R} d\boldsymbol{r}\, \psi^\dagger_\sigma(\boldsymbol{r})\, \Phi^\dagger(\boldsymbol{R}) \frac{Z}{|\boldsymbol{R}-\boldsymbol{r}|} \Phi(\boldsymbol{R})\, \psi_\sigma(\boldsymbol{r}) \tag{4}$$

where Φ are ionic field operators, M the mass, and Z the atomic number (assuming a single atom type to keep it simple).

The Hamiltonian still includes an external field that is necessary to break the *phase* symmetry and allow for a superconducting condensation [9]. In fact the normal state is a stationary point, and unless a continuous symmetry is broken in some way any perturbative approach based on this Hamiltonian will not lead to a superconducting state in perfect analogy with the theory of magnetism. From the knowledge of BCS [8, 19] and Eliashberg theory [15, 19, 20] it is expected that the most convenient way to introduce a symmetry breaking term that leads to superconductivity is to couple the system to an external superconductor[1] that can tunnel Cooper pairs in and out

$$H_{\Delta_{ext}} = \int d\boldsymbol{r} d\boldsymbol{r}'\, \Delta^*_{ext}(\boldsymbol{r},\boldsymbol{r})\, \psi_\uparrow(\boldsymbol{r})\, \psi_\downarrow(\boldsymbol{r}') + h.c. \tag{5}$$

This form of the symmetry breaking field has the great advantage of being extremely simple (involving only two field operators) and will prove also to be theoretically quite convenient. At the same time it has the disadvantage of introducing in the original Hamiltonian a Cooper pair source and sink such that the particle number is not fixed anymore.

In addition to the symmetry breaking external field in Eq. (5) one should also add an external field coupling to the electronic density

$$H_{v_{ext}} = \int d\boldsymbol{r}\, v_{ext}(\boldsymbol{r}) \sum_\sigma \psi^\dagger_\sigma(\boldsymbol{r})\, \psi_\sigma(\boldsymbol{r}) \tag{6}$$

[1] The symmetry breaking will be assumed of singlet type throughout this lecture.

and an external field that couples to the nuclei

$$H_{W_{ext}} = \int W_{ext}\left(\{\boldsymbol{R}_i\}\right) \prod_j d\boldsymbol{R}_j \, \boldsymbol{\Phi}^\dagger\left(\boldsymbol{R}_j\right) \boldsymbol{\Phi}\left(\boldsymbol{R}_j\right). \tag{7}$$

Unlike the electronic external potential, the ionic one couples all nuclei with each other (as it depends on the *set* of positions $\{\boldsymbol{R}_i\}$). This, not obvious, choice is made in order to be able, eventually, to construct a non interacting system of ions that actually behave like phonons under an external potential.

2.2 Hohenberg Kohn theorem

The SCDFT was introduced by Oliveira, Gross, and Kohn [10] as a generalization of finite temperature DFT to include the extra density (χ – called anomalous or superconducting) that couples with the proximity field Δ_{ext}. In its modern form [11, 12] SCDFT is based on the three densities

$$\rho\left(\boldsymbol{r}\right) = \mathrm{Tr}\left[\varrho_0 \sum_\sigma \psi_\sigma^\dagger\left(\boldsymbol{r}\right) \psi_\sigma\left(\boldsymbol{r}\right)\right] \tag{8}$$

$$\chi\left(\boldsymbol{r}, \boldsymbol{r}'\right) = \mathrm{Tr}\left[\varrho_0 \, \psi_\uparrow\left(\boldsymbol{r}\right) \psi_\downarrow\left(\boldsymbol{r}'\right)\right] \tag{9}$$

$$\Gamma\left(\{\boldsymbol{R}_i\}\right) = \mathrm{Tr}\left[\varrho_0 \prod_j \boldsymbol{\Phi}^\dagger\left(\boldsymbol{R}_j\right) \boldsymbol{\Phi}\left(\boldsymbol{R}_j\right)\right] \tag{10}$$

where ϱ_0 is the grand canonical density matrix (not to be confused with the electron density $\rho\left(\boldsymbol{r}\right)$)

$$\varrho_0 = \frac{e^{-\beta(H-\mu N)}}{\mathrm{Tr}\left[e^{\beta(H-\mu N)}\right]}, \tag{11}$$

N being the particle number operator.
SCDFT is based on a generalized Hohenberg-Kohn theorem at finite temperature, that states

1. There is a one-to-one mapping between the set of densities $\rho\left(\boldsymbol{r}\right)$, $\chi\left(\boldsymbol{r}, \boldsymbol{r}'\right)$, and $\Gamma\left(\{\boldsymbol{R}_i\}\right)$ onto the set of external potentials $v_{ext}\left(\boldsymbol{r}\right)$, $\Delta_{ext}\left(\boldsymbol{r}, \boldsymbol{r}'\right)$, and $W_{ext}\left(\{\boldsymbol{R}_i\}\right)$

2. There is a variational principle so that it exists a functional Ω that

$$\begin{aligned} \Omega\left[\rho_0, \chi_0, \Gamma_0\right] &= \Omega_0 \\ \Omega\left[\rho\ , \chi\ , \Gamma\ \right] &> \Omega_0 \quad \text{for} \quad \rho, \chi, \Gamma \neq \rho_0, \chi_0, \Gamma_0 \end{aligned} \tag{12}$$

   where ρ_0, χ_0, and Γ_0 are the ground state densities and Ω_0 the grand canonical potential.

The proof is a generalization of the finite temperature DFT proof of Mermin [7]. It is still worth to sketch it here. One defines the grand canonical functional as

$$\Omega\left[\varrho\right] = \mathrm{Tr}\left[\varrho\left(H - \mu N + \frac{1}{\beta}\ln\varrho\right)\right] \tag{13}$$

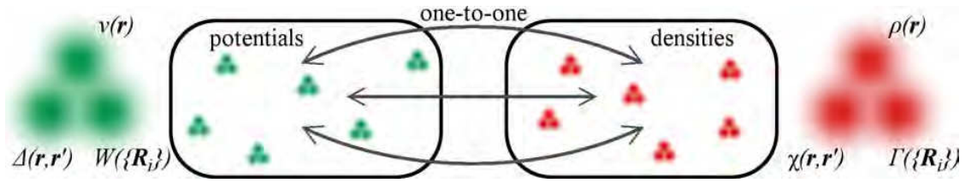


Fig. 1: *Correspondence between potentials (v, Δ, W) and densities (ρ, χ, Γ) in SCDFT.*

where ϱ are density matrices. If ϱ is the grand canonical density matrix ($\varrho \to \varrho_0$) then $\Omega[\varrho_0]$ is in fact the grand canonical potential of the system $-\frac{1}{\beta}\ln\left\{\mathrm{Tr}\left[e^{-\beta(H-\mu N)}\right]\right\} \equiv \Omega_0$. From this, it is straightforward to prove that, independently of the specific form of H, $\Omega[\varrho_0] < \Omega[\varrho]$ for any $\varrho \neq \varrho_0$ (refer to appendix A in Ref. [7]).

A second step is to prove by *reductio ad absurdum* that one set of densities can minimize two different sets of external potentials. So let us assume that both $v_{ext}, \Delta_{ext}, W_{ext}$ and $v'_{ext}, \Delta'_{ext}, W'_{ext}$ lead to the same densities ρ, χ, Γ. Clearly H and H' the two Hamiltonians corresponding to the two sets of external potentials are different[2] as are the corresponding density matrices ϱ and ϱ'. By using the minimum principle one can write the following set of relations

$$\begin{aligned}\Omega' \equiv &\mathrm{Tr}\left\{\varrho'_0\left[H' - \mu N\frac{1}{\beta}\ln(\varrho'_0)\right]\right\} < \mathrm{Tr}\left\{\varrho_0\left[H' - \mu N\frac{1}{\beta}\ln(\varrho_0)\right]\right\} \\ &= \Omega + \mathrm{Tr}\left\{\varrho_0\left[\left(H_{v'_{ext}} - H_{v_{ext}}\right) + \left(H_{\Gamma'_{ext}} - H_{\Gamma_{ext}}\right) + \left(H_{\Delta'_{ext}} - H_{\Delta_{ext}}\right)\right]\right\} \end{aligned} \tag{14}$$

where the last equation uses the assumption that the two systems have the same densities. The argument can be repeated starting from Ω, i.e., swapping primed and un-primed quantities, leading to another inequality

$$\Omega < \Omega' + \mathrm{Tr}\left\{\varrho'_0\left[\left(H_{v_{ext}} - H_{v'_{ext}}\right) + \left(H_{\Gamma_{ext}} - H_{\Gamma'_{ext}}\right) + \left(H_{\Delta_{ext}} - H_{\Delta'_{ext}}\right)\right]\right\}. \tag{15}$$

Summing Eq. (14) and (15) and inconsistency emerges

$$\Omega' + \Omega < \Omega + \Omega'. \tag{16}$$

This implies that the assumption was absurd and two different sets of potentials can not lead to the same set of densities. Establishing the one-to-one mapping between densities and potentials is the first part of the Hohenberg-Kohn theorem for SCDFT.

To prove the second part one has to notice that, since all observables are obviously functionals of the external potentials, they are also functionals of the densities (due to the one-to-one relation). Since this includes the thermodynamic potential Ω, one can rewrite the minimum principle proved above in terms of the densities (instead of using the density matrix) that is the second and last point of the Hohenberg-Kohn theorem for SCDFT.

[2] This is because the external potentials couple differently with the field operators. With this in mind one can easily extend the theorem to add extra densities and potentials [21, 22].

The fact that all observables are functional of the densities and that H is the sum of internal interactions (Eq. (1)) and couplings with external fields (Eqs. (5), (6), and (7)) allows to write $\Omega[\rho,\chi,\Gamma]$ as

$$\begin{aligned}\Omega[\rho,\chi,\Gamma] =& F[\rho,\chi,\Gamma] + \int d\boldsymbol{r}\, v_{ext}(\boldsymbol{r})\rho(\boldsymbol{r}) \\ &+ \int \Gamma(\{\boldsymbol{R}_i\})\, W_{ext}(\{\boldsymbol{R}_i\}) \prod_j d\boldsymbol{R}_j + \int d\boldsymbol{r} d\boldsymbol{r}'\, \Delta^*_{ext}(\boldsymbol{r},\boldsymbol{r}')\chi(\boldsymbol{r},\boldsymbol{r}') + c.c.\end{aligned} \tag{17}$$

which defines the universal functional $F[\rho,\chi,\Gamma]$. This functional is called universal because, not depending on the external potentials, it is system independent and uniquely fixed by the choice of the Hamiltonian and the chosen set of densities. Still, one retains an explicit dependence on the external potentials, because this coupling is necessary to bind the minimization to a specific physical problem.

2.3 The Kohn-Sham system

As for conventional DFT, in order to move from an exact one-to-one mapping derived in the previous section to a useful computational framework one needs to introduce the Kohn-Sham system [2], a non-interacting system with external potentials such that it is minimized by the same densities as the physical (interacting) one.
The thermodynamic potential of this support system is defined as

$$\begin{aligned}\Omega_s[\rho,\chi,\Gamma] =& T_{s,e}[\rho,\chi,\Gamma] + T_{s,n}[\rho,\chi,\Gamma] - \frac{1}{\beta}S[\rho,\chi,\Gamma] + \int d\boldsymbol{r}\, v_s(\boldsymbol{r})\rho(\boldsymbol{r}) \\ &+ \int \Gamma(\{\boldsymbol{R}_i\})\, W_s(\{\boldsymbol{R}_i\}) \prod_j d\boldsymbol{R}_j + \int d\boldsymbol{r} d\boldsymbol{r}'\, \Delta^*_s(\boldsymbol{r},\boldsymbol{r}')\chi(\boldsymbol{r},\boldsymbol{r}') + c.c.\end{aligned} \tag{18}$$

where T are kinetic energy functionals and S is the entropy functional; the Kohn-Sham "external" potentials must be chosen as

$$\begin{aligned} v_s(\boldsymbol{r}) &= v_{ext}(\boldsymbol{r}) + v_H(\boldsymbol{r}) + v_{xc}(\boldsymbol{r}) \\ \Delta_s(\boldsymbol{r},\boldsymbol{r}') &= \Delta_{ext}(\boldsymbol{r},\boldsymbol{r}') + \Delta_{xc}(\boldsymbol{r},\boldsymbol{r}') \\ W_s(\{\boldsymbol{R}_i\}) &= W_{ext}(\{\boldsymbol{R}_i\}) + W_H(\{\boldsymbol{R}_i\}) + W_{xc}(\{\boldsymbol{R}_i\}) \end{aligned} \tag{19}$$

where the subscript H stands for Hartree terms[3] and xc are the exchange-correlation potentials defined as

$$\begin{aligned} v_{xc}[\rho,\chi,\Gamma] &= \frac{\delta F_{xc}[\rho,\chi,\Gamma]}{\delta\rho} \\ \Delta_{xc}[\rho,\chi,\Gamma] &= \frac{\delta F_{xc}[\rho,\chi,\Gamma]}{\delta\chi} \\ W_{xc}[\rho,\chi,\Gamma] &= \frac{\delta F_{xc}[\rho,\chi,\Gamma]}{\delta\Gamma} \end{aligned} \tag{20}$$

[3] An Hartree contribution to Δ_s is not included because such a term does not exist in perturbation theory [15,19].

where

$$F_{xc}\left[\rho,\chi,\Gamma\right] = F\left[\rho,\chi,\Gamma\right] - T_{s,n}\left[\rho,\chi,\Gamma\right] - T_{s,n}\left[\rho,\chi,\Gamma\right] + \frac{1}{\beta}S\left[\rho,\chi,\Gamma\right]. \tag{21}$$

With these definitions, it is easy to see by functional differentiation of Eq. (18) and (17) that they are minimized by the same densities.

Having defined the Kohn-Sham potentials one can now write the Kohn-Sham equations for SCDFT. The ionic equation has the familiar form

$$\left[\sum_j \frac{\nabla_j^2}{2M_j} + W_s\left(\{\boldsymbol{R}_i\}\right)\right]\Phi_n\left(\{\boldsymbol{R}_i\}\right) = \mathcal{E}_n\Phi_n\left(\{\boldsymbol{R}_i\}\right), \tag{22}$$

where M_j is the mass of the atom with label j, $\mathcal{E}_n$ the eigenvalues and Φ_n the ionic eigenstates. The electronic equations are obtained from the electronic Hamiltonian

$$H_s = \sum_\sigma \int d\boldsymbol{r}\,\psi_\sigma^\dagger\left(\boldsymbol{r}\right)\left[\frac{-\nabla^2}{2} + v_s\left(\boldsymbol{r}\right) - \mu\right]\psi_\sigma\left(\boldsymbol{r}\right) + \int d\boldsymbol{r}d\boldsymbol{r}'\,\left[\Delta_s^*\left(\boldsymbol{r},\boldsymbol{r}'\right)\psi_\uparrow\left(\boldsymbol{r}\right)\psi_\downarrow\left(\boldsymbol{r}'\right) + h.c.\right]. \tag{23}$$

This Hamiltonian is non diagonal in the field operators because of the coupling induced by the anomalous potential Δ_s^*. It is diagonalized by a Bogoliubov-Valatin transformation [9]

$$\psi_\sigma\left(\boldsymbol{r}\right) = \sum_i \left[u_i\left(\boldsymbol{r}\right)\gamma_{i\sigma} - \mathrm{sgn}\left(\sigma\right)v_i\left(\boldsymbol{r}\right)\gamma_{i\sigma}^\dagger\right] \tag{24}$$

leading to the diagonalization conditions

$$\left[-\frac{\nabla^2}{2} + v_s\left(\boldsymbol{r}\right) - \mu\right]u_i\left(\boldsymbol{r}\right) + \int \Delta_s\left(\boldsymbol{r},\boldsymbol{r}'\right)v_i\left(\boldsymbol{r}'\right)\,d\boldsymbol{r}' = E_i u_i\left(\boldsymbol{r}\right) \tag{25}$$

$$-\left[-\frac{\nabla^2}{2} + v_s\left(\boldsymbol{r}\right) - \mu\right]v_i\left(\boldsymbol{r}\right) + \int \Delta_s^*\left(\boldsymbol{r},\boldsymbol{r}'\right)u_i\left(\boldsymbol{r}'\right)\,d\boldsymbol{r}' = E_i v_i\left(\boldsymbol{r}\right) \tag{26}$$

that are the electronic Kohn-Sham equation for SCDFT. Their mathematical form is well known in superconductivity literature as the Bogoliubov-deGennes (BdG) equations [9] mostly used, within the BCS model, to describe superconducting structures in real space. In SCDFT these equations become exact for the calculation of the total energy and the densities

$$\rho\left(\boldsymbol{r}\right) = 2\sum_i \left[\left|u_i\left(\boldsymbol{r}\right)\right|^2 f\left(E_i\right) + \left|v_i\left(\boldsymbol{r}\right)\right|^2 f\left(-E_i\right)\right] \tag{27}$$

$$\chi\left(\boldsymbol{r},\boldsymbol{r}'\right) = \sum_i u_i\left(\boldsymbol{r}\right)v_i^*\left(\boldsymbol{r}'\right)f\left(-E_i\right) - v_i^*\left(\boldsymbol{r}\right)u_i\left(\boldsymbol{r}'\right)f\left(E_i\right). \tag{28}$$

In absence of superconductivity both χ and Δ are zero and Eq. (25) becomes the conventional Kohn-Sham equation of DFT[4]

$$\left[-\frac{\nabla^2}{2} + v_s\left(\boldsymbol{r}\right) - \mu\right]\varphi_{n\boldsymbol{k}}\left(\boldsymbol{r}\right) = \xi_{n\boldsymbol{k}}\varphi_{n\boldsymbol{k}}\left(\boldsymbol{r}\right), \tag{29}$$

In the same limit Eq. (26) would be a DFT Kohn-Sham like equation with inverted eigenvalues, that can be seen as an equation for holes instead of electrons. The non-particle-conserving type of superconducting coupling, as a matter of fact, connects these two equations.

[4] Actually slightly more general because it would still include the full effect of temperature and ionic motion since it is still coupled with Eq. (22) via the potentials in (19) and (20).

2.3.1 Transformation to momentum space

Eq. (29) can be solved in the superconducting state (i.e., keeping the non-zero χ in the functional $v_s[\rho,\chi,\Gamma]$), introducing the corresponding eigenfunctions $\varphi_{n\boldsymbol{k}}(\boldsymbol{r})$. These would be different from the conventional KS orbitals, but can be used as a basis set to express the BdG equations in $\boldsymbol{k}$ space. Introducing the expansion

$$u_i(\boldsymbol{r}) = \sum_{n\boldsymbol{k}} u_{i,n\boldsymbol{k}}\,\varphi_{n\boldsymbol{k}}(\boldsymbol{r}) \tag{30}$$

$$v_i(\boldsymbol{r}) = \sum_{n\boldsymbol{k}} v_{i,n\boldsymbol{k}}\,\varphi_{n\boldsymbol{k}}(\boldsymbol{r}) \tag{31}$$

$$\Delta_s(\boldsymbol{r},\boldsymbol{r}') = \sum_{nn'\boldsymbol{k}\boldsymbol{k}'} \Delta_{s,nn'\boldsymbol{k}\boldsymbol{k}'}\,\varphi_{n\boldsymbol{k}}(\boldsymbol{r})\,\varphi_{n'\boldsymbol{k}'}(\boldsymbol{r}') \tag{32}$$

inserting into Eqs. (25), (26) and using the orthogonality of the basis set gives

$$\begin{aligned} \xi_{n\boldsymbol{k}}\,u_{i,n\boldsymbol{k}} + \sum_{n'\boldsymbol{k}'} \Delta_{s,nn'\boldsymbol{k}\boldsymbol{k}'}\,v_{i,n'\boldsymbol{k}'} &= E_i\,u_{i,n\boldsymbol{k}} \\ -\xi_{n\boldsymbol{k}}\,v_{i,n\boldsymbol{k}} + \sum_{n'\boldsymbol{k}'} \Delta^*_{s,nn'\boldsymbol{k}\boldsymbol{k}'}\,u_{i,n'\boldsymbol{k}'} &= E_i\,v_{i,n\boldsymbol{k}} \end{aligned} \tag{33}$$

a form of the BdG equations particularly useful for introducing approximations.

2.4 Decoupling approximations

The SCDFT formalism developed so far is exact but useless unless one is able to derive a valid approximation for the exchange correlation functional F_{xc} entering Eq. (20). However the problem introduced in Sec. 2.1 is a very complicated one and cannot be tackled without introducing key approximations to strip it to the bone. The goal of an ab-initio approach is to introduce controlled (or at least controllable) approximations. Luckily in the problem of interacting electrons and nuclei there is a lot of experience and many established approximations that can be introduced give results of broad validity.
The most important approximations, that will be discussed in this section, are meant to decouple as much as possible the many degrees of freedom (and densities) of the problem

1. Decouple electrons from ions separating the static and dynamic parts of the interaction, including the latter in a perturbative fashion.

2. Decouple the high energy chemical scale (responsible for bonding) from low energy pairing interactions (responsible for superconductivity).

2.4.1 Phonons and electron-phonon interaction

The formalism so far describes a set of interacting electrons and ions. This correlated electron-nuclear dynamics is enormously complex and far from being satisfactorily solved. However,

if one considers only systems close to their equilibrium it is reasonable to assume that atoms are locked to a lattice position and only small oscillations can occur. The assumption allows to treat ionic oscillations as a perturbation on a static field, greatly simplifying the problem. There is some excellent literature (like [23–25]) discussing these issue in great detail. Here the focus will be on the main approximations that are presently used in implementations of SCDFT.
A key approximation is to ignore the effect of superconductivity on the lattice dynamics and on the electron-phonon interaction. This, as the superconducting transition is usually of second order, is exact close to the critical temperature, where the superconducting density becomes infinitesimally small. This allows to study the lattice dynamics in the normal state.
To compute phonons and the electron-phonon interaction one usually relies on conventional Kohn-Sham density functional theory and the electron-phonon scattering matrix elements are defined as

$$g^{\nu}_{m\boldsymbol{k}+\boldsymbol{q},n\boldsymbol{k}} = \sqrt{\frac{\hbar}{2\omega_{\boldsymbol{q}\nu}}} \left\langle \varphi_{m\boldsymbol{k}+\boldsymbol{q}} \left| \Delta V^{\boldsymbol{q}\nu}_{scf} \right| \varphi_{n\boldsymbol{k}} \right\rangle \tag{34}$$

where $\boldsymbol{k}$ and $\boldsymbol{q}$ are the electron and phonon momenta, m and n Kohn-Sham band indices, $\varphi_{n\boldsymbol{k}}$ the Kohn-Sham states, ν is the phonon branch, $\omega_{\boldsymbol{q}\nu}$ the phonon frequency and $\Delta V^{\boldsymbol{q}\nu}_{scf}$ the variation in the Kohn-Sham potential due to the ionic displacement corresponding to the phonon mode. By means of density functional perturbation theory [25] these matrix elements can be computed accurately and at a reasonable computational cost for any bulk superconductor. The electron-phonon interaction of the Kohn-Sham system reads

$$\tilde{H}_{e-ph} = \sum_{mn\sigma} \sum_{\nu\boldsymbol{k}\boldsymbol{q}} g^{\nu}_{m\boldsymbol{k}+\boldsymbol{q},n\boldsymbol{k}} \sum_{\sigma} \psi^{\dagger}_{\sigma m\boldsymbol{k}+\boldsymbol{q}} \psi_{\sigma n\boldsymbol{k}} b_{\nu\boldsymbol{q}} = \sum_{\nu\boldsymbol{q}} \sqrt{\frac{\hbar}{2\omega_{\boldsymbol{q}\nu}}} \int d\boldsymbol{r}\, \Delta V^{\boldsymbol{q}\nu}_{scf}(\boldsymbol{r})\, \psi^{\dagger}_{\sigma}(\boldsymbol{r})\, \psi_{\sigma}(\boldsymbol{r})\, b_{\nu\boldsymbol{q}}\,, \tag{35}$$

where $\psi^{\dagger}_{\sigma n\boldsymbol{k}}$ and $\psi_{\sigma n\boldsymbol{k}}$ are creation and destruction operators for Kohn-Sham states and $b_{\nu\boldsymbol{q}}$ is a phonon operator.
The step of approximating the dynamic part of H_{en} with $\tilde{H}_{en}$ can certainly be justified empirically by its success in applications [26, 25] but is theoretically not very rigorous. The main supporting argument is essentially that if the Kohn-Sham band structure is close to the interacting one so will likely be their response to a lattice motion. Clearly if Kohn-Sham bands are far off from the interacting ones (like in strongly correlated systems) the Kohn-Sham electron-phonon coupling is also expected to be a poor approximation to the real one.

2.4.2 Band decoupling approximation

The electronic BdG Kohn-Sham equations (33) can be enormously simplified by assuming that the superconducting condensation will be a small perturbation to the non-superconducting system. As already pointed out in the previous section, since the superconducting transition is usually of second order the assumption becomes exact close to T_C so that it will not affect the estimation of T_C itself.
This assumption, first of all, implies that the superconducting transition will not induce a structural one, therefore $\Delta_s(\boldsymbol{r},\boldsymbol{r})$ should maintain the original lattice periodicity and the $\boldsymbol{k}$ quantum

number in Eq. (29) must be conserved [16, 17]. In other words the summations in equation Eqs. (30) and (31) should only run over the band index n and not over $\boldsymbol{k}$.
The summation over n means that the superconducting transition can still induce an *hybridization* between different bands corresponding to the same $\boldsymbol{k}$-point. However, unless bands are degenerate (or close to degeneracy with respect to the energy scale set by Δ_s that is of the order 10 meV) this hybridization must be extremely small. Therefore, apart for anomalous cases, one can introduce a second and stronger approximation by ignoring this superconductivity induced band hybridization effect. Then Eqs. (30) and (31) reduce to

$$\begin{aligned} u_i(\boldsymbol{r}) \equiv u_{n\boldsymbol{k}}(\boldsymbol{r}) &= u_{n\boldsymbol{k}}\varphi_{n\boldsymbol{k}}(\boldsymbol{r}) \\ u_i(\boldsymbol{r}) \equiv v_{n\boldsymbol{k}}(\boldsymbol{r}) &= v_{n\boldsymbol{k}}\varphi_{n\boldsymbol{k}}(\boldsymbol{r}), \end{aligned} \tag{36}$$

which implies $\Delta_{s,nn'\boldsymbol{k}\boldsymbol{k}'} \to \delta_{n\boldsymbol{k},n'\boldsymbol{k}'}\Delta_{s,n\boldsymbol{k}}$.
Inserting Eq. (36) into (33) one can formally solve these equations obtaining

$$u_{n\boldsymbol{k}} = \frac{1}{\sqrt{2}}\operatorname{sgn}(E_{n\boldsymbol{k}})\, e^{\phi_{n\boldsymbol{k}}}\sqrt{1+\frac{\xi_{n\boldsymbol{k}}}{|E_{n\boldsymbol{k}}|}} \tag{37}$$

$$v_{n\boldsymbol{k}} = \frac{1}{\sqrt{2}}\sqrt{1-\frac{\xi_{n\boldsymbol{k}}}{|E_{n\boldsymbol{k}}|}} \tag{38}$$

with $e^{\phi_{n\boldsymbol{k}}} = \Delta_s(n\boldsymbol{k}) / |\Delta_s(n\boldsymbol{k})|$ and $E_{n\boldsymbol{k}} = \pm\sqrt{\xi^2_{n\boldsymbol{k}} + |\Delta_s(n\boldsymbol{k})|^2}$. While the densities in Eq. (27) and (28) take on the simple form

$$\rho(\boldsymbol{r}) = \sum_{n\boldsymbol{k}}\left[1-\frac{\xi^2_{n\boldsymbol{k}}}{|E_{n\boldsymbol{k}}|}\tanh\left(\frac{\beta|E_{n\boldsymbol{k}}|}{2}\right)\right]|\varphi_{n\boldsymbol{k}}(\boldsymbol{r})|^2 \tag{39}$$

$$\chi(\boldsymbol{r},\boldsymbol{r}') = \frac{1}{2}\sum_{n\boldsymbol{k}}\frac{\Delta_s(n\boldsymbol{k})}{|E_{n\boldsymbol{k}}|}\tanh\left(\frac{\beta|E_{n\boldsymbol{k}}|}{2}\right)\varphi_{n\boldsymbol{k}}(\boldsymbol{r})\,\varphi^*_{n\boldsymbol{k}}(\boldsymbol{r}'). \tag{40}$$

The whole superconducting problem is now reduced to the construction of the matrix elements of the Kohn-Sham potential $\Delta_s(n\boldsymbol{k})$ that are defined by the solution of Eq. (20). The explicit dependence on χ in that equation can be substituted with a dependence on Δ_s (using Eq. (40))

$$\Delta_{xc} = \frac{\delta F_{xc}[\rho,\chi[\Delta_s,\rho,\Gamma],\Gamma]}{\delta\chi} \tag{41}$$

The above equation is a closed (self-consistent) equation for the Kohn-Sham potential and is usually called the SCDFT gap equation.
At this stage one needs some approximation for the $F_{xc}[\rho,\chi,\Gamma]$ functional. A simple functional will be derived in the following sections by first creating a link between SCDFT and many body perturbation theory.

3 Connection between SCDFT and Eliashberg theory

A DFT exchange-correlation functional contains information on the interacting behavior of the many body system. This type of information can be extracted from different approaches like models, solvable exact limits of the theory, empirical data, perturbative methods and more. Perturbation methods have the advantage of allowing for a systematic improvement, it is formally possible to reach arbitrary high accuracy, although in reality one is usually strongly limited by computational costs and convergence issues.
Nevertheless it is nice to construct an exchange correlation functional starting from many body perturbation theory and to keep exact control on what type of physics the functional will describe and what would be its limits.
It will be shown in Sec. 3.2 how to set up a link between many body perturbation theory and SCDFT. The focus will be on a specific many body approximation for superconductivity that goes under the name of Eliashberg theory and will be briefly reviewed in Sec. 3.1. Eventually, in Sec. 3.3, the exact link will be studied for a model superconductor.

3.1 A survey on Eliashberg theory of superconductivity

This section gives a formal introduction to Eliashberg theory of superconductivity that is the most popular theory used for superconductivity predictions, especially in the simplified form of the McMillan equation [27].
One can start from the Hamiltonian, Eq. (1). After decoupling the lattice dynamics as discussed in Sec. 2.4.1 only the electronic part and the electron-phonon interaction part $\tilde{H}_{en}$ introduced in Sec. 2.4.1 are retained

$$H \to H_e + H_{ee} + \tilde{H}_{en} + H_{ext} \tag{42}$$

where now H_{ext} contains the anomalous coupling in Eq. (5) as well as the static electron-ion coupling (as in normal DFT).[5]
H is then split in a zero-approximation H_0 plus an interaction part H_I. A convenient choice for the zero approximation is H_{ext} plus the Kohn-Sham Hamiltonian entering in Eq. (29)

$$H_s = \sum_\sigma \int d\boldsymbol{r}\, \psi^\dagger_\sigma(\boldsymbol{r}) \left[-\frac{\nabla^2}{2} + v_s(\boldsymbol{r}) - \mu \right] \psi_\sigma(\boldsymbol{r}), \tag{43}$$

while everything else goes into H_I. So

$$H_0 = H_s + H_{ext} \tag{44}$$

$$H_I = H_{ee} + \tilde{H}_{en} - H_{DC} \tag{45}$$

where the last term removes extra *xc* effects already included in H_s, therefore avoiding any *double counting*

$$H_{DC} = \sum_\sigma \int d\boldsymbol{r}\, \psi^\dagger_\sigma(\boldsymbol{r})\, v_s(\boldsymbol{r})\, \psi_\sigma(\boldsymbol{r}). \tag{46}$$

[5] Often in the literature on Eliashberg theory the external potential $\Delta_{ext}(\boldsymbol{r}, \boldsymbol{r}')$ is taken to be local ($\Delta_{ext}(\boldsymbol{r})$) here is introduced as non-local to be consistent with the SCDFT formalism.

The reference Kohn-Sham system to which we refer here is that of normal DFT and not of SCDFT (alternatively one could consider v_s as the SCDFT limit for $\chi = 0$, i.e., formally retaining the effects of the multi-component formalism and its dependence on Γ).
Unfortunately conventional many body perturbation theory [14] can not be directly applied to $H_0 + H_I$ because the particle source in Eq. (5) introduces new processes forbidden in a particle conserving theory.
There is a trick that allows to transform H back into a standard form: the Nambu-Gor'kov formalism. One defines two new electronic field operators

$$\bar{\psi}(\boldsymbol{r}) = \begin{pmatrix} \psi_\uparrow(\boldsymbol{r}) \\ \psi^\dagger_\downarrow(\boldsymbol{r}) \end{pmatrix} \tag{47}$$

$$\bar{\psi}^\dagger(\boldsymbol{r}) = \begin{pmatrix} \psi^\dagger_\uparrow(\boldsymbol{r}) & \psi_\downarrow(\boldsymbol{r}) \end{pmatrix}. \tag{48}$$

that still obey Fermionic commutation rules. With these two-component fields $\bar{\psi}$ one can rewrite H_0 and H_I as

$$H_0 = \int d\boldsymbol{r}\, \bar{\psi}^\dagger(\boldsymbol{r})\, \bar{H}_0(\boldsymbol{r}, \boldsymbol{r}')\, \bar{\psi}(\boldsymbol{r}') \tag{49}$$

$$\begin{aligned} H_I = &\int d\boldsymbol{r}\, \bar{\psi}^\dagger(\boldsymbol{r}) \left[\sum_{\nu \boldsymbol{q}} \sqrt{\frac{\hbar}{2\omega_{\boldsymbol{q}\nu}}} \int d\boldsymbol{r} \Delta V^{\boldsymbol{q}\nu}_{scf}(\boldsymbol{r})\, \bar{\sigma}_3 b_{\nu\boldsymbol{q}} - v_s(\boldsymbol{r}) \right] \bar{\psi}(\boldsymbol{r}) \\ &+ \frac{1}{2} \int d\boldsymbol{r} d\boldsymbol{r}' \left[\bar{\psi}^\dagger(\boldsymbol{r})\, \bar{\sigma}_3 \bar{\psi}(\boldsymbol{r}) \right] \frac{1}{|\boldsymbol{r} - \boldsymbol{r}'|} \left[\bar{\psi}^\dagger(\boldsymbol{r}')\, \bar{\sigma}_3 \bar{\psi}(\boldsymbol{r}') \right]. \end{aligned} \tag{50}$$

where $\bar{\sigma}_3$ is the Pauli matrix $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ and $\bar{H}_0$ is defined as

$$\bar{H}_0(\boldsymbol{r}, \boldsymbol{r}') = \begin{pmatrix} \left[-\frac{\nabla^2}{2} + v_s(\boldsymbol{r}) - \mu \right] \delta(\boldsymbol{r} - \boldsymbol{r}') & \Delta_{ext}(\boldsymbol{r}, \boldsymbol{r}') \\ \Delta^*_{ext}(\boldsymbol{r}, \boldsymbol{r}') & -\left[-\frac{\nabla^2}{2} + v_s(\boldsymbol{r}) - \mu \right] \delta(\boldsymbol{r} - \boldsymbol{r}') \end{pmatrix}. \tag{51}$$

The Hamiltonian in this new form does not feature anymore source terms for the new field $\bar{\psi}$, therefore the perturbative expansion for H_I will have exactly the same contributions (diagrams) as in conventional perturbation theory. The difference is that Green functions and self energy will have a 2×2 matrix structure and vertices will carry the extra $\bar{\sigma}_3$ terms. The Green function is defined as

$$\bar{G}(\tau\boldsymbol{r}, \tau'\boldsymbol{r}') := -\begin{pmatrix} \left\langle T\psi_{H,\uparrow}(\tau\boldsymbol{r})\, \psi^\dagger_{H,\uparrow}(\tau'\boldsymbol{r}') \right\rangle & \left\langle T\psi_{H,\uparrow}(\tau\boldsymbol{r})\, \psi_{H,\downarrow}(\tau'\boldsymbol{r}') \right\rangle \\ \left\langle T\psi^\dagger_{H,\downarrow}(\tau\boldsymbol{r})\, \psi^\dagger_{H,\uparrow}(\tau'\boldsymbol{r}') \right\rangle & \left\langle T\psi^\dagger_{H,\downarrow}(\tau\boldsymbol{r})\, \psi_{H,\downarrow}(\tau'\boldsymbol{r}') \right\rangle \end{pmatrix}, \tag{52}$$

where the H subscript refers to the Heisenberg picture, T is the time ordering operator for the imaginary time coordinate τ and the $\langle ... \rangle$ is used as short notation for the thermodynamic average as in Eq. (8). Comparing with Eq. (8) one can show that normal and anomalous densities can be easily extracted from $\bar{G}$

$$\rho(\boldsymbol{r}) = \lim_{\tau' \to \tau^+} \lim_{\boldsymbol{r}' \to \boldsymbol{r}} \bar{G}^{(11)}(\tau\boldsymbol{r}, \tau'\boldsymbol{r}') = \lim_{\boldsymbol{r}' \to \boldsymbol{r}} \frac{1}{\beta} \sum_i \bar{G}^{(11)}(\boldsymbol{r}, \tau'\boldsymbol{r}', \omega_i) \tag{53}$$

$$\chi(\boldsymbol{r}, \boldsymbol{r}') = \lim_{\tau' \to \tau^+} \qquad \bar{G}^{(12)}(\tau\boldsymbol{r}, \tau'\boldsymbol{r}') = \qquad \frac{1}{\beta} \sum_i \bar{G}^{(12)}(\boldsymbol{r}, \tau'\boldsymbol{r}', \omega_i) \tag{54}$$

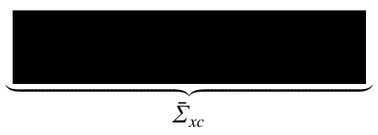
where the (ij) superscript indicates the matrix component of $\bar{G}$, and the second equality comes from the transform from imaginary time to Matsubara frequencies (ω_i) according to the standard relation[6] $\bar{G}(\boldsymbol{r}, \boldsymbol{r}', \omega_i) = \int d\tau\, e^{i\omega_n \tau} \bar{G}(0, \boldsymbol{r}, \tau \boldsymbol{r}')$.
$\bar{G}$ is obtained by solving the Dyson equation

$$\bar{G}(\boldsymbol{r}, \boldsymbol{r}', \omega_i) = \bar{G}_0(\boldsymbol{r}, \boldsymbol{r}', \omega_i) + \bar{G}_0(\boldsymbol{r}, \boldsymbol{r}', \omega_i)\, \bar{\Sigma}(\boldsymbol{r}, \boldsymbol{r}', \omega_i)\, \bar{G}(\boldsymbol{r}, \boldsymbol{r}', \omega_i), \tag{55}$$

where $\bar{G}_0(\boldsymbol{r}, \boldsymbol{r}', \omega_i)$ is the Green function corresponding to the non interacting Hamiltonian H_0. What is now left is to introduce an approximation for the self energy $\bar{\Sigma}$. Leaving aside (for the moment) contributions stemming from the double counting term H_{DC}, the first order contribution is

$\bar{\Sigma}^{(1)} =$  , (56)

where wavy lines are phonon propagators, dashed lines the (bare) Coulomb interaction and arrowed lines are the Green functions. The first term is an Hartree-like phonon driven electron-electron interaction. A careful analysis of this diagram shows that its contribution is rather small, essentially introducing a shift (of the order of the phononic energy scale) on the eigenvalues of H_0. Conventionally this contribution is not included in Eliashberg theory because it is irrelevant in comparison to the intrinsic error one commits computing the eigenvalues of H_0.[7]
The second term is the conventional Hartree diagram. It is exactly cancelled by a corresponding term in v_s since one assumes that the Kohn-Sham Hamiltonian leads to the same density of the interacting system.
The third and forth terms are exchange-like diagrams that contain those electron-electron scattering processes that are essential to superconductivity.
Quite clearly the first order set of diagrams in $\bar{\Sigma}^{(1)}$ would not be sufficient to obtain any reasonable result. Consider that electrons are now interacting only with the bare Coulomb interaction while in real materials (especially in metals) the electronic interaction is very effectively screened. Like in GW theory [28, 29] it is possible to boost the order of the approximation by dressing propagators and Green functions (the phonon propagator is already dressed as it is computed externally). Therefore defining the following approximation

$\bar{\Sigma} =$  $- \bar{\Sigma}_{DC}$ (57)

$\bar{\Sigma}_{xc}$

where $\bar{\Sigma}_{DC}$ is simply $\bar{\tau}_3 v_{xc}$.

[6] To keep the formalism as simple as possible convergence factors $e^{i\omega_i \eta}$ will be omitted [14].

[7] In the multi-component formalism discussed in Sec. 2.3 this term would be already included in v_s (since in Eq. (20) there is a functional dependence on the ionic density). However, Eliashberg theory usually starts from the conventional Kohn-Sham Hamiltonian where v_s is only a functional of the electronic density and ions only appear as a static external potential.

The computational cost to apply the approximation above would still be too expensive. Just looking at the Coulomb diagram essentially corresponds to a self consistent GW approach in the 2×2 Nambu-Gor'kov space! One should instead rely on the same approximations discussed in Sec. 2.4.2, assuming that the electronic states are already well described by the Kohn-Sham Hamiltonian and neglect inter-band hybridization. Essentially the self energy still is that of Eq. (57) but the diagonal part of the second diagram is removed together with $\bar{\Sigma}_{DC}$ (that was inserted in the first place to avoid the double counting of xc terms).
In the basis of Kohn-Sham states this self energy then takes the form

$$\bar{\Sigma}\left(n\boldsymbol{k},\omega_i\right)=-\frac{1}{\beta}\sum_j\sum_{m\boldsymbol{q}'}\bar{\sigma}_3\bar{G}\left(n\boldsymbol{k},\omega_i\right)\bar{\sigma}_3\left[\sum_\nu g^\nu_{m\boldsymbol{k}+\boldsymbol{q},n\boldsymbol{k}}D_\nu\left(\boldsymbol{q},\omega_i-\omega_j\right)+\bar{\sigma}_1W\left(n\boldsymbol{k},m\boldsymbol{k}+\boldsymbol{q}\right)\right], \tag{58}$$

where ω_i are the Matsubara frequencies, $\bar{\sigma}_1$ is the Pauli matrix $\begin{pmatrix}0&1\\1&0\end{pmatrix}$, $D_\nu\left(\boldsymbol{q},\omega_i-\omega_j\right)=-2\omega_{\nu\boldsymbol{q}}/\left[\left(\omega_i-\omega_j\right)^2+\omega^2_{\nu\boldsymbol{q}}\right]$ is the phonon propagator and W the screened Coulomb interaction. $\bar{G}\left(n\boldsymbol{k},\omega_i\right)$ is the Nambu-Gor'kov Green function that in momentum space is

$$\bar{G}\left(n\boldsymbol{k},\omega_i\right)=\int_0^\beta d\tau\, e^{-i\omega_i(\tau-\tau')}\int d\boldsymbol{r}\boldsymbol{r}'\,\varphi^*_{n\boldsymbol{k}}\left(\boldsymbol{r}\right)\bar{G}\left(\tau\boldsymbol{r},\tau'\boldsymbol{r}'\right)\varphi_{n\boldsymbol{k}}\left(\boldsymbol{r}'\right) \tag{59}$$

and is the solution of the Dyson equation

$$\bar{G}\left(n\boldsymbol{k},\omega_i\right)=\bar{G}_0\left(n\boldsymbol{k},\omega_i\right)+\bar{G}_0\left(n\boldsymbol{k},\omega_i\right)\bar{\Sigma}\left(n\boldsymbol{k},\omega_i\right)\bar{G}\left(n\boldsymbol{k},\omega_i\right), \tag{60}$$

where $\bar{G}\left(n\boldsymbol{k},\omega_i\right)$ is the Green function corresponding to the non interacting Hamiltonian H_0. The above equation is Eliashberg theory of superconductivity. Its solution is usually achieved by first expanding this matrix equation into Pauli matrices and separating it into components. The decomposition has the advantage to lead to a more explicit form of $\bar{G}$ and help the physical interpretation. $\bar{G}$ is expressed as

$$\bar{G}\left(n\boldsymbol{k},\omega_i\right)=\frac{\begin{pmatrix}i\omega_iZ\left(n\boldsymbol{k},\omega_i\right)+\left[\xi_{n\boldsymbol{k}}+\not{\chi}\left(n\boldsymbol{k},\omega_i\right)\right]&\phi\left(n\boldsymbol{k},\omega_i\right)\\\phi\left(n\boldsymbol{k},\omega_i\right)&i\omega_iZ\left(n\boldsymbol{k},\omega_i\right)-\left[\xi_{n\boldsymbol{k}}+\not{\chi}\left(n\boldsymbol{k},\omega_i\right)\right]\end{pmatrix}}{\left[i\omega_iZ\left(n\boldsymbol{k},\omega_i\right)\right]^2-\left[\xi_{n\boldsymbol{k}}+\not{\chi}\left(n\boldsymbol{k},\omega_i\right)\right]^2-\phi^2\left(n\boldsymbol{k},\omega_i\right)}, \tag{61}$$

where $\not{\chi}$ (not to be confused with the superconducting order parameter) shifts the non interacting energies, Z behaves as a mass term, and ϕ/Z is the function giving the superconducting gap (this interpretation is evident by analytically continuing $\bar{G}$ to the real frequency axis $i\omega_i\to\omega$). These are now scalar functions and here they are assumed to be real valued.[8]
The interested reader can refer to more specialized literature with extra focus on application and numerical implementation, like the classic Review of Carbotte [31] or the excellent review of Ummarino in this same lecture series [32].

[8] One can show that for the Hamiltonian (1) the ϕ, Z and $\not{\chi}$ functions satisfy a set of equations with real coefficients. In spite of this the solution could still be non trivially complex, as in the famous three crystal experiment [30], but this unusual situation will be neglected.

3.2 The Sham Schlüter connection

In Sec. 2 the SCDFT framework was derived and a set of quite general approximations was introduced in order to decouple the superconducting and the normal state problem. In the previous section the Eliashberg theory of superconductivity was reviewed and derived in the same approximations assumed for SCDFT.
Following the work of Marques [18], this section is devoted to set up a formal connection between the two theories, by extending the Sham Schlüter connection to SCDFT.
The connection is based on a Dyson equation similar to Eq. (60), however instead of using the normal state Kohn-Sham Hamiltonian H_0 (Eq. (49)) as reference Hamiltonian for the perturbation expansion, one should use the SCDFT Kohn-Sham system

$$\bar{H}_s(\boldsymbol{r},\boldsymbol{r}') = \begin{pmatrix} \left[-\frac{\nabla^2}{2} + v_s(\boldsymbol{r}) - \mu\right]\delta(\boldsymbol{r}-\boldsymbol{r}') & \Delta_s(\boldsymbol{r},\boldsymbol{r}') \\ \Delta_s^*(\boldsymbol{r},\boldsymbol{r}') & -\left[-\frac{\nabla^2}{2} + v_s(\boldsymbol{r}) - \mu\right]\delta(\boldsymbol{r}-\boldsymbol{r}') \end{pmatrix} \tag{62}$$

leading to the following form of the Dyson equation

$$\bar{G}(n\boldsymbol{k},\omega_i) = \bar{G}_s(n\boldsymbol{k},\omega_i) + \bar{G}_s(n\boldsymbol{k},\omega_i)\,\bar{\Sigma}_s(n\boldsymbol{k},\omega_i)\,\bar{G}(n\boldsymbol{k},\omega_i)\,, \tag{63}$$

that differs from Eq. (60) as $\bar{G}_s$ is the Green function of the SCDFT Kohn-Sham system and $\bar{\Sigma}_s(n\boldsymbol{k},\omega_i)$ is the self energy in which the double counting correction refers to SCDFT (not to DFT as in Eq. (57))

$$\bar{\Sigma}_s = \bar{\Sigma}_{xc} - \bar{\Sigma}_{DC}^{SC}\,, \tag{64}$$

where $\bar{\Sigma}_{DC}^{SC}$ reads

$$\bar{\Sigma}_{DC}^{SC}(n\boldsymbol{k}) = \begin{pmatrix} v_{xc}(n\boldsymbol{k}) & \Delta_{xc}(n\boldsymbol{k}) \\ \Delta_{xc}^*(n\boldsymbol{k}) & -v_{xc}(n\boldsymbol{k}) \end{pmatrix}. \tag{65}$$

Then one uses the fact that *both* $\bar{G}$ and $\bar{G}_s$ when inserted into Eq. (53) provide the exact density of the system. It is easy to see that in the Kohn-Sham basis this implies

$$\sum_i \sum_{n\boldsymbol{k}} \bar{G}^{(11)}(n\boldsymbol{k},\omega_i) = \sum_i \sum_{n\boldsymbol{k}} \bar{G}_s^{(11)}(n\boldsymbol{k},\omega_i) \equiv \sum_i \sum_{n\boldsymbol{k}} \frac{-(i\omega_i + \xi_{n\boldsymbol{k}})}{\omega_i^2 + \xi_{n\boldsymbol{k}}^2 + \Delta_s^2(n\boldsymbol{k})} \tag{66}$$

$$\sum_i \quad \bar{G}^{(12)}(n\boldsymbol{k},\omega_i) = \sum_i \quad \bar{G}_s^{(12)}(n\boldsymbol{k},\omega_i) \equiv \sum_i \quad \frac{-\Delta_s(n\boldsymbol{k})}{\omega_i^2 + \xi_{n\boldsymbol{k}}^2 + \Delta_s^2(n\boldsymbol{k})}\,, \tag{67}$$

where the second equality uses the explicit form of G_s that, belonging to a non interacting system, is simply $\frac{1}{i\omega_i \bar{I} - \bar{H}_s}$, with $\bar{I}$ being the $2x2$ identity matrix. These constraints used in Eq. (63) lead obviously to a set of conditions for v_{xc} and Δ_{xc}.
However, if Σ_{xc} is assumed in the Eliashberg approximation, the diagonal part of Eq. (65) should be dropped, as discussed in Sec. 3.1, and one single scalar condition is sufficient (all other matrix components would be redundant)

$$\begin{aligned} &\sum_i \Big[\bar{\Sigma}_{xc}^{(11)}(n\boldsymbol{k},\omega_i)\,\bar{G}_s^{(11)}(n\boldsymbol{k},\omega_i)\,\bar{G}^{(12)}(n\boldsymbol{k},\omega_i) + \bar{\Sigma}_{xc}^{(11)}(n\boldsymbol{k},\omega_i)\,\bar{G}_s^{(12)}(n\boldsymbol{k},-\omega_i)\,\bar{G}^{(11)}(n\boldsymbol{k},\omega_i) \\ &\quad -\bar{\Sigma}_{xc}^{(21)}(n\boldsymbol{k},\omega_i)\,\bar{G}_s^{(12)}(n\boldsymbol{k},\omega_i)\,\bar{G}^{(12)}(n\boldsymbol{k},\omega_i) + \Sigma_{xc}^{(12)}(n\boldsymbol{k},\omega_i)\,\bar{G}_s^{(11)}(n\boldsymbol{k},-\omega_i)\,\bar{G}^{(11)}(n\boldsymbol{k},\omega_i)\Big] \\ &= \sum_i \Big[\Delta_{xc}(n\boldsymbol{k})\,\bar{G}_s^{(11)}(n\boldsymbol{k},\omega_i)\,\bar{G}^{(11)}(n\boldsymbol{k},\omega_i) - \Delta_{xc}^*(n\boldsymbol{k})\,\bar{G}_s^{(12)}(n\boldsymbol{k},-\omega_i)\,\bar{G}^{(12)}(n\boldsymbol{k},\omega_i)\Big]. \end{aligned} \tag{68}$$

This can be further simplified by the assumption that the order parameter and the gap functions are real, which is usually the case even for unconventional superconductors, leading to the following form of the connection

$$\begin{aligned}\Delta^*_{xc}(n\boldsymbol{k}) = \tfrac{1}{\Xi(n\boldsymbol{k})}\tfrac{1}{\beta}\Bigg[& \sum_i \bar{\Sigma}^{(11)}_{xc}(n\boldsymbol{k},\omega_i)\,\bar{G}^{(11)}_s(n\boldsymbol{k},+\omega_i)\,\bar{G}^{(12)}(n\boldsymbol{k},\omega_i) \\ & + \sum_i \bar{\Sigma}^{(11)}_{xc}(n\boldsymbol{k},\omega_i)\,\bar{G}^{(12)}_s(n\boldsymbol{k},-\omega_i)\,\bar{G}^{(11)}(n\boldsymbol{k},\omega_i) \\ & - \sum_i \bar{\Sigma}^{(21)}_{xc}(n\boldsymbol{k},\omega_i)\,\bar{G}^{(12)}_s(n\boldsymbol{k},+\omega_i)\,\bar{G}^{(12)}(n\boldsymbol{k},\omega_i) \\ & + \sum_i \bar{\Sigma}^{(12)}_{xc}(n\boldsymbol{k},\omega_i)\,\bar{G}^{(11)}_s(n\boldsymbol{k},-\omega_i)\,\bar{G}^{(11)}(n\boldsymbol{k},\omega_i)\Bigg],\end{aligned} \tag{69}$$

where

$$\Xi(n\boldsymbol{k}) = \frac{1}{\beta}\sum_i\left[\bar{G}^{(11)}_s(n\boldsymbol{k},\omega_i)\,\bar{G}^{(11)}(n\boldsymbol{k},\omega_i) - \bar{G}^{(12)}_s(n\boldsymbol{k},-\omega_i)\,\bar{G}^{(12)}(n\boldsymbol{k},\omega_i)\right]. \tag{70}$$

Equation (69) above is the Sham-Schlüter connection for SCDFT as derived by Marques [18]. For *any* given many body self energy it returns the anomalous Kohn-Sham potential of SCDFT (Δ_{xc}) that, by construction, leads to the same anomalous density of the interacting system described by that self energy (in this work always assumed in the Eliashberg approximation).
Note that the equation has to be seen as a self-consistent equation for Δ_{xc} because it also enters the right hand side in the definition of $\bar{G}_s$ (see Eq. (66)–(67)), this equation is equivalent to Eq. (41), the gap equation of SCDFT, and its right hand side is therefore the functional derivative of F_{xc} with respect to the anomalous density. At the present stage it is not, however, an explicit function but expressed numerically in terms of Matsubara frequencies as it depends on $\bar{\Sigma}$ and is not useful for material studies. On the other hand it is a perfect tool to get more insights on the properties of the SCDFT Kohn-Sham system. This will be the topic of the next section where the Sham-Schlüter connection will be analyzed numerically for a model system.

3.3 Analysis of a model

In this section Eliashberg equations and Sham-Schlüter connection will be solved numerically for a simple system of electrons interacting only via an isotropic (momentum independent) electron-phonon coupling provided by a single Einstein phonon mode.

$$g^{\nu}_{m\boldsymbol{k}+\boldsymbol{q},n\boldsymbol{k}} = \sqrt{\frac{\lambda\omega_{ph}}{N_F}}, \tag{71}$$

where N_F is the density of states at the Fermi level, ω_{ph} is the energy of the Einstein mode and λ is the BCS-like electron-phonon coupling. For this example it will be assumed that $\lambda = 1$ and $\omega_{ph} = 60\ meV$. It is also assumed that the density of non-interacting states is constant

$$N(\xi) = \sum_{n,\boldsymbol{k}}\delta(\xi - \xi_{n\boldsymbol{k}}) = N_F\,. \tag{72}$$

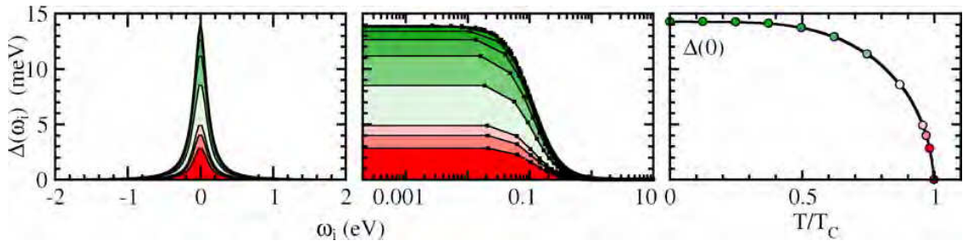


Fig. 2: *Exact anomalous potential $\Delta(\xi)$ corresponding to the Eliashberg approximation to the self energy computed for a model pairing interaction. Colors encode the temperature as indicated in the right-most panel.*

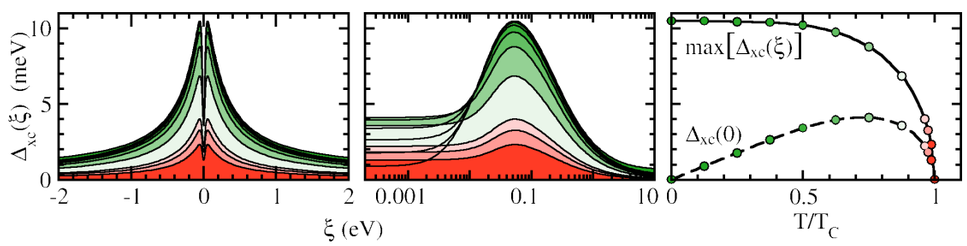


Fig. 3: *Exact Kohn-Sham anomalous potential $\Delta_{xc}(\xi)$ corresponding to the Eliashberg approximation to the self energy computed for a model pairing interaction.*

This model is completely isotropic meaning that Green functions and potentials will loose their dependence on $n\boldsymbol{k}$ and will only depend on $\xi_{n\boldsymbol{k}}$ the non interacting Kohn-Sham eigenenergy. Via Eq. (58) the coupling defined above enters the Sham-Schlüter connection (69) that can be solved numerically.[9]

The critical temperature T_C of the model is 80 K. The interacting Green function (see Eq. (61) is uniquely defined by the two functions $\Delta(n\boldsymbol{k}, \omega_i)$ and $Z(n\boldsymbol{k}, \omega_i)$, while in this particle hole symmetric model χ is zero. Furthermore, since the pairing in Eq. (71) is just a constant, Z and Δ do not depend on $\xi_{n\boldsymbol{k}}$ either but only on ω_i.

The superconducting gap $\Delta(n\boldsymbol{k}, \omega_i)$ is plotted in Fig. 2 as a function of the Matsubara frequency ω_i and temperature. This function decreases monotonically with $|\omega_i|$ a fact that is directly linked with the experimental evidence that the superconducting gap has its maximum close to the Fermi level. Together with T_C the gap at the Fermi level ($\Delta(\xi = 0, \omega = 0)$) is the most important property of a superconductor because it is connected with the superconductor excitation spectrum and therefore with its unique response properties. The temperature dependence of Δ is show in the right panel of Fig. 2 and has a standard BCS-like behavior.

In conventional DFT we are used to assume that the single particle spectrum of the Kohn-Sham system is quite similar to the interacting one and only in some pathological cases, like strongly correlated materials, the differences are really large. However, in SCDFT the properties of Δ_{xc}

[9] One can either proceed in two steps solving first the Dyson-Eliashberg equation (60) and then insert the resulting components of $\bar{\Sigma}$ and $\bar{G}$ into Eq. (69) that is eventually solved for Δ_{xc}. Alternatively, one can also directly solve Eq. (69) self consistently until convergence is achieved both for $\bar{\Sigma}$, $\bar{G}$, and $\bar{G}_s$.

are quite different from those of Δ although formally Δ_{xc} also plays the role of a superconducting gap for the SCDFT Kohn-Sham system. Like Δ, Δ_{xc} shows a sharp feature at low energy. The width of this peak is broader but one should consider that the energy width on the imaginary axis is not quite the same as on the physical axis. Moreover, and this is the most important aspect, the Kohn-Sham gap has a dip close to the Fermi energy, showing a non-monotonic behavior in energy. The dip is lower at low temperature so that in the zero temperature limit the superconducting gap at the Fermi level is actually zero. A completely unphysical property.
Not a conceptual problem as such, but just a reminder of the limitations of the Kohn-Sham system: the excitation spectrum of the superconducting Kohn-Sham system is not like that of the real interacting superconductor. The correct way to obtain the superconducting gap (or any other observable apart from density and total energy) in SCDFT is by constructing the functional of the densities for that observable. In this case the gap functional $\Delta[\rho,\chi,\Gamma]$. This was done for SCDFT [33, 34] but its construction would go beyond the goals of this lecture.

3.4 A simple SCDFT functional

As proved by Marques [18], the Sham-Schlüter connection can be used to derive a valid SCDFT functional (LM2005 [11, 12]). The main conceptual step is that G and G_s, while in general quite different, are connected by the sum rules (66) and (67) and it is likely not completely inaccurate to perform the substitution $G \to G_s$ into the Sham-Schlüter connection, both on explicit occurrences and on those coming from the G dependence of the self energy $\bar{\Sigma}_{xc}$ in Eq. (57). In this way the Sham-Schlüter connection becomes a closed equation in Δ_{xc}, i.e., an approximation for the SCDFT gap equation (41). In other word the substitution leads directly to an explicit expression for F_{xc} (or rather its functional derivative Δ_{xc}). The predictive power of such a simple functional was then improved by a set of procedures the details of which can be found in the original works [11, 12]. From a perturbative point of view this procedure is equivalent to choosing a many body self energy in first order (like in Eq. (56)) not in the non-interacting Green function $\bar{G}_0$ but in the SCDFT Kohn-Sham ones $\bar{G}_s$.
It is extremely important to observe that upon this transformation of Eq. (69) all the Matsubara summations can be carried out analytically. From this comes one of the main computational advantages of the SCDFT scheme: it does not involve numerical Matsubara integrations. The explicit form of Δ_{xc} reads

$$\Delta_{xc}(n\boldsymbol{k}) = \mathcal{Z}(n\boldsymbol{k})\Delta_{xc}(n\boldsymbol{k}) + \frac{1}{2}\sum_{n'\boldsymbol{k}'}\mathcal{K}(n\boldsymbol{k},n'\boldsymbol{k}')\frac{\tanh\left(\frac{\beta}{2}E_{n'\boldsymbol{k}'}\right)}{E_{n'\boldsymbol{k}'}}\Delta_{xc}(n'\boldsymbol{k}'), \tag{73}$$

where the two kernels $\mathcal{K}$ and $\mathcal{Z}$ are

$$\mathcal{Z}(n\boldsymbol{k}) = \sum_{n'\boldsymbol{k}'}\sum_{\nu}\left|g^{\nu}_{m\boldsymbol{k}+\boldsymbol{q},n\boldsymbol{k}}\right|^2\left[I'(\xi_{n\boldsymbol{k}},\xi_{n'\boldsymbol{k}'},\omega_{\boldsymbol{q}\nu}) + I'(\xi_{n\boldsymbol{k}},-\xi_{n'\boldsymbol{k}'},\omega_{\boldsymbol{q}\nu})\right] \tag{74}$$

$$\mathcal{K}(n\boldsymbol{k},n'\boldsymbol{k}') = W(n\boldsymbol{k},n'\boldsymbol{k}') - \frac{\sum_{\nu}\left|g^{\nu}_{n'\boldsymbol{k}+\boldsymbol{q},n\boldsymbol{k}}\right|^2\left[I(\xi_{n\boldsymbol{k}},\xi_{n'\boldsymbol{k}'},\omega) - I(\xi_{n\boldsymbol{k}},-\xi_{n'\boldsymbol{k}'},\omega)\right]}{\tanh\left(\frac{\beta}{2}\xi_{n\boldsymbol{k}}\right)\tanh\left(\frac{\beta}{2}\xi_{n'\boldsymbol{k}'}\right)}, \tag{75}$$

with $\boldsymbol{k}' \equiv \boldsymbol{k} + \boldsymbol{q}$, and where

$$
\begin{aligned}
I'(\xi,\xi',\omega) &= \frac{d}{d\xi} I((\xi,\xi',\omega) \\
I(\xi,\xi',\omega) &= J(\xi,\xi',\omega) - J(\xi,\xi',-\omega) \\
J(\xi,\xi',\omega) &= \left[f_\beta(\xi) + b_\beta(\omega)\right] \frac{f_\beta(\xi') - f_\beta(\xi-\omega)}{\xi - \xi' - \omega}
\end{aligned}
\tag{76}
$$

f_β and b_β being Fermi and Bose functions. In Eq. (75) the screened Coulomb interaction W is assumed to be static. More recent extension to this functional have included dynamical effects as well [35]. Also this LM2005 functional is not particularly accurate in describing the effect of the electron-phonon coupling, overestimating it at strong coupling and underestimating it in the weak coupling limit, presently more accurate functionals exist [33]. However this functional has the advantage of providing a gap function that does not feature the dip at the Fermi level discussed in the previous section; instead Δ_{xc} at the Fermi level is by experience quite close to the physical superconducting gap.

It should be noted that Eq. (73) has the form of a BCS gap equation, however unlike the BCS equation that relies on a model, this equation is derived from first principles and the kernels are not adjustable parameters but functionals of the densities. This functional dependence is implicit via a dependence on the Kohn-Sham orbitals and their matrix elements.

4 Example: Superconductivity of a doped carbon-hydrogen nanotube

There is a relatively large literature presenting applications of SCDFT using different approximations and functionals. Ranging from the study of classic superconductors [12, 36], strongly anisotropic systems [37–40], at high pressure [41, 34, 42], or investigations of different pairing mechanisms [43, 44, 35, 45], to cite only a few.

As discussed in the first part of this lecture, an SCDFT investigation starts from the characterization of the normal state. The normal state information is then used to construct the exchange correlation potential, or rather the kernels of the SCDFT gap equations (73). From the solution of the gap equation it is then possible to extract a broad variety of observables and properties of the superconductor.

To briefly exemplify this procedure it will be applied here to a realistic system, a crystal of hole doped, hydrogenated, carbon nanotubes. The structure is shown in the center of Fig. 4. Similar to graphane [46], it will be shown that this system also has a strong electron-phonon coupling and that it features nicely inhomogeneous superconducting properties. The possibility of its experimental synthesis however are not discussed here as neither the thermodynamic stability nor the effect of zero point motion nor the physical doping mechanism will be addressed.

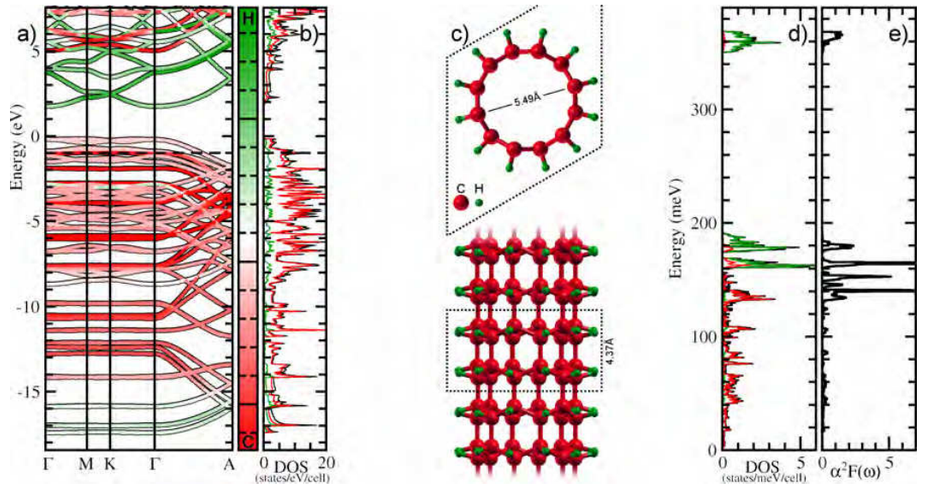


Fig. 4: *Left: Electronic bands (a) and density states (DOS - b) of an hydrogenated carbon nanotube. The colorscale in the band plot gives the atomic projection of the Kohn-Sham states on atomic orbitals, from red (100% C) to green (65% H). The Fermi level of the undoped crystal is at zero. A dashed line shows the position of the Fermi level used for the simulation, corresponding to hole doping. Center (c): views of the nanotube structure, the dashed line indicates the periodic unit cell of the crystal). Right: phonon density of states (d - also decomposed in H and C components) and $\alpha^2F(\omega)$ function (e). The latter is an average of the electron-phonon matrix elements (34) on the Fermi surface of the system (Eq. (77)).*

4.1 Normal state properties

The first step to apply SCDFT is to compute the normal state properties of the system: its Kohn-Sham eigenvalues and eigenstates, phonons, electron-phonon, and electron-electron coupling. From the band structure plot in Fig. 4 one can see that the undoped system is an insulator. A metallic state, a necessary precondition to superconductivity, could be realized, for example, by introducing boron substitutional impurities at the carbon sites, however here the doping will be simply induced by a rigid shift of the Fermi level (see Fig. 4a). While this over-simplified approach will neglect important effects as Kohn anomalies [47], it is still expected to provide a reasonable estimate of the superconducting coupling [46].

The phononic density of states is characterized by an high energy peak (350 meV) due to C-H bond stretching modes, a mid-energy region of hydrogen rocking modes, and a low energy part (below 150 meV) of C modes. The high-end of these carbon modes (C-C bond stretching) is the most relevant for the electron-phonon coupling, as shown by the $\alpha^2F(\omega)$ function

$$\alpha^2F(\omega) = \frac{1}{N_F}\sum_{n\mathbf{k}n'\mathbf{k}'}\sum_{\nu}\left|g^{\nu}_{n\mathbf{k},n'\mathbf{k}'}\right|^2\delta(\xi_{n\mathbf{k}})\,\delta(\xi_{n'\mathbf{k}'})\,\delta(\omega-\omega_{\mathbf{q}\nu}) \tag{77}$$

with $\boldsymbol{k}' \equiv \boldsymbol{k}+\boldsymbol{q}$.

Electronic states and phonon matrix elements are computed on a regular grid in momentum space, the size of which is set by convergence criteria and limited by the computational cost.

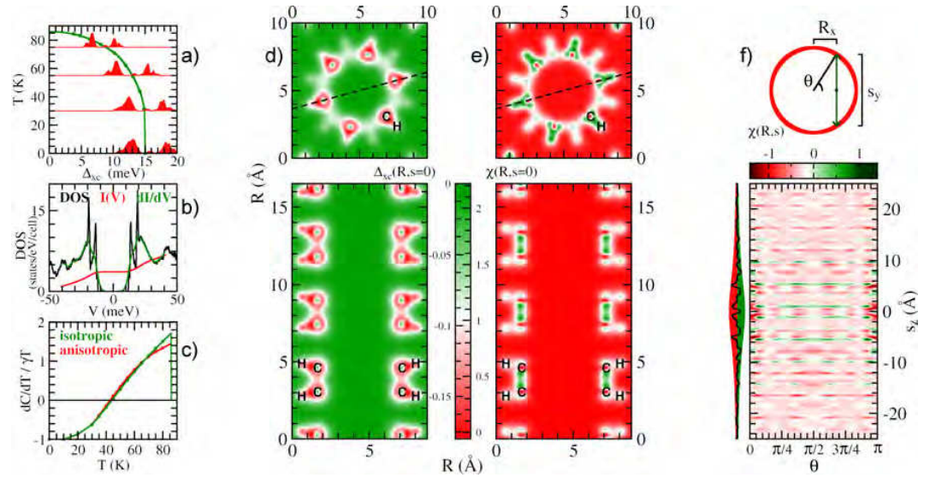


Fig. 5: *Left a): Temperature dependence of the average Kohn-Sham superconducting gap (green line) and its distribution function $P(\Delta)$ (red - arbitrary vertical scale). b) Density of states (black), tunneling current (red - shifted upward, arbitrary units) and differential conductance (green) as computed from Eq. (79). c) Difference of specific heat in the normal and superconducting state as a function of temperature in isotropic and anisotropic approximation. Center: real space anomalous potential $\Delta(\boldsymbol{R}, \boldsymbol{s})$ and order parameter $\chi(\boldsymbol{R}, \boldsymbol{s})$ as a function of the Cooper pair center of mass $\boldsymbol{R}$. Top: xy cut of the tube, bottom: vertical cut of the tube. Right f): the order parameter as a function of the electronic distance across the tube surface (on the C layer). $\boldsymbol{s}_z$ refers to the vertical distance on the tube and θ gives the position of the two electrons on the xy plane as shown by the cartoon above the plot. The curve at the left of this plot is the $(\boldsymbol{R} = 0, \boldsymbol{s}_z)$ cut highlighting the oscillatory behavior and its envelope function.*

However Eq. (73) requires a very accurate sampling especially of the states close to the Fermi level, because the kernels (74) are sharply peaked around $\xi = 0$. Therefore, an accurate interpolation scheme is necessary. As described in Ref [12] a convenient approach is to use a large set of random $\boldsymbol{k}$-points accumulated around the Fermi level and appropriately weighted. The properties of the corresponding states are then obtained by interpolation from calculations on regular grids: here a $10 \times 10 \times 20$ grid is used for the electronic states, a $2 \times 2 \times 10$ for the convergence of the potential- and $\boldsymbol{k}$-grids and a $1 \times 1 \times 2$ for the $\boldsymbol{q}$ grid. Calculations were done within the norm-conserving pseudopotential approximation [48, 25]. Coulomb matrix elements were computed with the Sham-Kohn approach [49], as described in Ref. [12].

4.2 Solution of the gap equation and superconducting properties

Solving the SCDFT gap equation (73) using the LM2005 functional gives a critical temperature of 86 K. Fig. 5a) gives the temperature dependence of the average gap on the Fermi surface (average of $\Delta_{xc}(\xi = 0)$) as well as its distribution function

$$P(\Delta) = \frac{1}{N_F} \sum_{n\boldsymbol{k}} \delta(\Delta - \Delta_{xc}(n\boldsymbol{k})) \, \delta(\xi_{n\boldsymbol{k}}) \tag{78}$$

showing that the Kohn-Sham gap is distributed on two peaks that at $T = 0$ are centered at about 12.8 and 18.5 meV. The larger gap corresponding to one band with higher C projection (easily recognizable in Fig. 4) as a consequence of the fact that C modes have a stronger coupling than H modes. This predicted multigap behavior is usually observable experimentally in different ways [39, 50]. Tunneling spectroscopy, for example, probes rather directly the excitation spectrum: a tunneling current can be modeled as the convolution of the DOS of the superconductor with that of the tip. In the simplest approximation [51]

$$I(V) \sim \sum_{n\boldsymbol{k}} \left[f_\beta(\xi_{n\boldsymbol{k}}) - f_\beta(\xi_{n\boldsymbol{k}} + V) \right] N_s(\xi_{n\boldsymbol{k}} + V) \tag{79}$$

where N_s is the DOS in the superconducting state. It shown in Fig. 5b), where the current is computed for T=10 K.

Although indirectly, thermodynamic measurements also give information about the gap distribution function. This can be observed in the temperature dependence of the entropy or, more conveniently, of the difference in specific heat

$$\Delta C = C - C_N = \frac{1}{T} \frac{d}{dT} (S - S_N), \tag{80}$$

where the electronic entropy is

$$S = -2k_B \sum_{n\boldsymbol{k}} \left\{ \left[1 - f_\beta(E_{n\boldsymbol{k}}) \right] \ln\left(1 - f_\beta(E_{n\boldsymbol{k}}) \right) + f_\beta(E_{n\boldsymbol{k}}) \ln\left(f_\beta(E_{n\boldsymbol{k}}) \right) \right\} \tag{81}$$

k_B being the Boltzmann constant. S_N is the corresponding quantity for the normal state ($\Delta_{xc} = 0$). ΔC is shown in Fig. 5c), both for a completely anisotropic calculation and for an isotropic one. In the anisotropic case ΔC tends to saturate towards T_C.

A more fundamental property that can be extracted from the solution of Eq. (73) is the Kohn-Sham potential Δ_{xc} and especially χ, the superconducting order parameter (40), that is the central object of theoretical superconductivity. These functions are discussed extensively in Ref. [38] and are the most rigorous representation of the concept of a Cooper pair. Instead of the $(\boldsymbol{r}, \boldsymbol{r}')$ dependence they are best plotted as a function of $\boldsymbol{R} = \frac{\boldsymbol{r}+\boldsymbol{r}'}{2}$ and relative distance $\boldsymbol{s} = \boldsymbol{r} - \boldsymbol{r}'$ as these can be interpreted as center of mass coordinate and electron-electron distance of the pair. The behavior of these functions is shown in the right side of Fig. 5. As discussed in Ref. [38] the $\boldsymbol{R}$ dependence of Δ_{xc} tends to highlight those regions in space most involved in the Coulomb renormalization, in this case clearly the C–H bonding region. On the other hand χ is larger where the phonon coupling is stronger, in the present case mainly involving the covalent C–C bond of the tube surface.

As a function of $\boldsymbol{s}$ these function show a typical oscillatory behavior. The envelope of the function has a maximum at $\boldsymbol{s} = 0$ and slowly decreases on the scale of the superconductor's coherence length. Fig. 5f) shows χ for two electrons located on the surface of the carbon tube, the oscillatory behavior is clearly visible as a function of $\boldsymbol{s}_z$, for vertical separation between the two electrons.

5 Summary and conclusion

SuperConducting Density Functional Theory was reviewed to focus on its relation to conventional Eliashberg many body theory. Starting from the formal construction of the two theoretical frameworks, a connection was set-up using the Sham-Schlüter method. This connection is used to extract information on exact features of the Kohn-Sham system that reproduces the superconducting density in the Eliashberg approximation and to develop a simple functional for SCDFT. The methodology was then applied to the study of a realistic system: a hole doped hydrogenated carbon nanotube, that proves to be a strong superconductor with a critical temperature above 80 K. While this is the single most important number that characterizes the superconductivity, there are many other properties that can be extracted from SCDFT at an affordable computational cost. In fact the theory gives access to a rich variety of observables as momentum and position dependent properties like the gap distribution function or the local order parameter.

Acknowledgements

I would like to thank Kay Dewhurst and Arkady Davydov for many helpful discussions during the preparation of this lecture.

References

[1] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964)

[2] W. Kohn and L.J. Sham, Phys. Rev. **140**, A1133 (1965)

[3] U. von Barth and L. Hedin, J. Phys. C: Solid State Phys. **5**, 1629 (1972)

[4] E.K.U. Gross and W. Kohn, Phys. Rev. Lett. **55**, 2850 (1985)

[5] T.L. Gilbert, Phys. Rev. B **12**, 2111 (1975)

[6] G. Vignale and M. Rasolt, Phys. Rev. Lett. **59**, 2360 (1987)

[7] N.D. Mermin, Phys. Rev. **137**, A1441 (1965)

[8] J. Bardeen, L.N. Cooper, and J.R. Schrieffer, Phys. Rev. **108**, 1175 (1957)

[9] P.G. de Gennes: *Superconductivity of Metals and Alloys* (Westview Press, 1966)

[10] L.N. Oliveira, E.K.U. Gross, and W. Kohn, Phys. Rev. Lett. **60**, 2430 (1988)

[11] M. Lüders, M.A.L. Marques, N.N. Lathiotakis, A. Floris, G. Profeta, L. Fast, A. Continenza, S. Massidda, and E.K.U. Gross, Phys. Rev. B **72**, 024545 (2005)

[12] M.A.L. Marques, M. Lüders, N.N. Lathiotakis, G. Profeta, A. Floris, L. Fast, A. Continenza, E.K.U. Gross, and S. Massidda, Phys. Rev. B **72**, 024546 (2005)

[13] T. Kreibich and E.K.U. Gross, Phys. Rev. Lett. **86**, 2984 (2001)

[14] A. Fetter and J.D. Walecka: *Quantum Theory of Many-Particle Systems* (Dover, New York, 1971 & 2003)

[15] S. Vonsovsky, Y. Izyumov, E. Kurmaev, E. Brandt, and A. Zavarnitsyn: *Superconductivity of Transition Metals: Their Alloys and Compounds*, Springer Series in Solid-State Sciences Series (Springer, London, 1982)

[16] S. Kurth Ph.D. thesis: *Exchange-Correlation Functionals for Inhomogeneous Superconductors* (Bayerische Julius-Maximilians Universität Würzburg, 1995)

[17] M. Lüders Ph.D. thesis: *Density Functional Theory for Superconductors, A first Principles Approach to the SC Phase* (Bayerische Julius-Maximilians Universität Würzburg, 1998)

[18] M. Marques Ph.D. thesis: *Density Functional Theory for Superconductors, Exchange and Correlation Potentials for Inhomogeneous Systems* (Bayerische Julius-Maximilians Universität Würzburg, 2000)

[19] J.R. Schrieffer: *Theory of Superconductivity* (W.A. Benjamin, New York, 1964)

[20] P.B. Allen and B. Mitrović: *Theory of Superconducting Tc*, *Solid State Physics*, Vol. 37 (Academic Press, 1983)

[21] A. Linscheid, A. Sanna, F. Essenberger, and E.K.U. Gross, Phys. Rev. B **92**, 024505 (2015)

[22] A. Linscheid, A. Sanna, and E.K.U. Gross, Phys. Rev. B **92**, 024506 (2015)

[23] A. Marini, S. Poncé, and X. Gonze, Phys. Rev. B **91**, 224310 (2015)

[24] F. Giustino, Rev. Mod. Phys. **89**, 015003 (2017)

[25] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi, Rev. Mod. Phys. **73**, 515 (2001)

[26] S.Y. Savrasov, D.Y. Savrasov, and O.K. Andersen, Phys. Rev. Lett. **72**, 372 (1994)

[27] W.L. McMillan, Phys. Rev. **167**, 331 (1968)

[28] L. Hedin, Phys. Rev. **139**, A796 (1965)

[29] F. Aryasetiawan and O. Gunnarsson, Reports on Progress in Physics **61**, 237 (1998)

[30] C.C. Tsuei, J.R. Kirtley, C.C. Chi, L.S. Yu-Jahnes, A. Gupta, T. Shaw, J.Z. Sun, and M.B. Ketchen, Phys. Rev. Lett. **73**, 593 (1994)

[31] J.P. Carbotte, Rev. Mod. Phys. **62**, 1027 (1990)

[32] G.A.C. Ummarino: Chap. 13 in E. Pavarini, E. Koch, and U. Schollwöck (eds.): *Emergent Phenomena in Correlated Matter*, Modeling and Simulation Vol. 3 (Forschungszentrum Jülich, 2013) `http://www.cond-mat.de/events/correl13`

[33] J.A. Flores-Livas and A. Sanna, Phys. Rev. B **91**, 054508 (2015)

[34] A.J. Flores-Livas, A. Sanna, and E. Gross, Eur. Phys. J. B **89**, 1 (2016)

[35] R. Akashi and R. Arita, Phys. Rev. Lett. **111**, 057006 (2013)

[36] A. Floris, A. Sanna, S. Massidda, and E.K.U. Gross, Phys. Rev. B **75**, 054508 (2007)

[37] A. Floris, G. Profeta, N.N. Lathiotakis, M. Lüders, M.A.L. Marques, C. Franchini, E.K.U. Gross, A. Continenza, and S. Massidda, Phys. Rev. Lett. **94**, 037004 (2005)

[38] A. Linscheid, A. Sanna, A. Floris, and E.K.U. Gross, Phys. Rev. Lett. **115**, 097002 (2015)

[39] A. Sanna, G. Profeta, A. Floris, A. Marini, E.K.U. Gross, and S. Massidda, Phys. Rev. B **75**, 020511 (2007)

[40] A. Sanna, A.V. Fedorov, N.I. Verbitskiy, J. Fink, C. Krellner, L. Petaccia, A. Chikina, D.Y. Usachov, A. Grüneis, and G. Profeta, 2D Materials **3**, 025031 (2016)

[41] G. Profeta, C. Franchini, N. Lathiotakis, A. Floris, A. Sanna, M.A.L. Marques, M. Lüders, S. Massidda, E.K.U. Gross, and A. Continenza, Phys. Rev. Lett. **96**, 047003 (2006)

[42] J.A. Flores-Livas, M. Amsler, C. Heil, A. Sanna, L. Boeri, G. Profeta, C. Wolverton, S. Goedecker, and E.K.U. Gross, Phys. Rev. B **93**, 020508 (2016)

[43] F. Essenberger, A. Sanna, A. Linscheid, F. Tandetzky, G. Profeta, P. Cudazzo, and E.K.U. Gross, Phys. Rev. B **90**, 214504 (2014)

[44] F. Essenberger, A. Sanna, P. Buczek, A. Ernst, L. Sandratskii, and E.K.U. Gross, Phys. Rev. B **94**, 014503 (2016)

[45] R. Akashi, K. Nakamura, R. Arita, and M. Imada, Physical Review B **86**, 054513 (2012)

[46] G. Savini, A.C. Ferrari, and F. Giustino, Phys. Rev. Lett. **105**, 037002 (2010)

[47] W. Kohn, Phys. Rev. Lett. **2**, 393 (1959)

[48] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G.L. Chiarotti, M. Cococcioni, I. Dabo, A.D. Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A.P. Seitsonen, A. Smogunov, P. Umari, and R.M. Wentzcovitch, J. Phys.: Condens. Matter **21**, 395502 (2009)

[49] L.J. Sham and W. Kohn, Phys. Rev. **145**, 561 (1966)

[50] R.S. Gonnelli, D. Daghero, D. Delaude, M. Tortello, G.A. Ummarino, V.A. Stepanov, J.S. Kim, R.K. Kremer, A. Sanna, G. Profeta, and S. Massidda, Phys. Rev. Lett. **100**, 207004 (2008)

[51] M. Tinkham: *Introduction to Superconductivity* (Dover, Mineola New York, 1996)

Index

Symbols
Φ-derivable approximations, 11.13

A
Abrikosov-Suhl resonance, 13.7
Anderson disordered model, 11.6
Anderson insulator, 2.7
Anderson model, 1.12–1.13, 9.20
    single-impurity, 13.6
    two-impurity, 13.9
Anderson molecule, 9.24
Anderson theorem, 13.14
Anderson transition, 3.14
anomalous Hall conductivity, 3.16
antiferromagnetism, 2.6
atomic limit, 4.23
atomic units, 4.28

B
band insulator, 2.5, 9.2
BCS gap equation, 13.14
BCS theory, 15.13
β function, 12.12, 12.18
Bethe ansatz solution, 12.17
Bethe-lattice, 1.17
Bohr magneton, 4.2
Brinkman-Rice transition, 6.27
bulk modulus
    iron, 8.22
    nickel, 8.18

C
canonical transformation, 9.8, 9.20
causality, 3.3
$CeIrIn_5$, 1.20
cerium, 1.18–1.20
charge density wave, 2.7
charge diffusion, 11.20
charge gap, 9.2
charge self-consistency, 9.36
Chern number, 3.10, 3.16
coherent potential, 11.4
compass model, 5.10
    generalized, 5.9
conductivity, 2.14, 3.3
configurational averaging, 11.4
conserving vertex, 11.25
constrained search, 8.3
    DFT, 8.4
    Hubbard density functional, 8.9
continued-fraction form, 7.7
continuity equation, 11.8, 11.20
conventional metal, 9.2
correlation narrowing, 6.26
Coulomb exchange, 4.4
Coulomb integral, 4.4
Coulomb pseudopotential, 15.21
cutoff, 12.11
    soft, 12.18

D
d^5 Mott insulator in Fe-based superconductors, 14.6
degenerate Hubbard model, 5.6
density functional
    DFT, 8.4
    Gutzwiller, 8.10
    Hubbard, 8.9
    single-particle, 8.5
density functional theory, 1.6–1.22, 10.2, 15.22
DFT+DMFT, 9.26, 9.35
DFT+U, 9.35
diffusion constant, 11.25
direct exchange, 4.8
dispersion kink, 15.8
dispersion relation, 2.4
DMFT, 9.26, 9.34
Doniach phase diagram, 12.3
double exchange, 4.18, 5.29
double-counting correction, 8.16, 9.29, 9.36
downfolding, 4.9, 4.29
Drude weight, 2.14, 3.6
dual boson, 10.17
dual fermion, 10.3
dynamical matrix, 15.26
dynamical mean-field theory, 1.7–1.9, 1.15–1.22, 7.4, 9.26, 9.34, 13.5

role of self-consistency, 13.7
two-orbital $e \times E$ Jahn-Teller problem, 13.15
connection with two-impurity model, 13.18
phase diagram, 13.18
strongly-correlated superconductivity, 13.16
dynamical response, 7.11
Dyson equation, 16.14

E
effective medium, 11.6
e_g-orbital Hubbard model, 5.4
e_g-orbital superexchange, 5.8
alternating order, 5.8
orbitons, 5.9
Einstein relation, 11.21
electrical conductivity, 11.19
electron self-energy, 15.6
electron-hole symmetry, 11.21
electron-phonon coupling, 15.2, 16.8
electron-phonon vertex, 15.2
electronic compressibility, 14.2, 14.17
electronic linewidth, 15.7
Eliashberg theory, 15.17, 16.11
energy band, 2.4
enhanced compressibility and superconductivity, 14.17
entanglement, 4.9
ergodic hypothesis, 11.6
exact diagonalization, 6.5
exchange
Coulomb, 4.4
direct, 4.8
double, 4.18
ferromagnetic superexchange, 4.16
kinetic, 4.7
orbital, 4.21
superexchange, 4.13
exchange integral, 4.4
exchange-correlation energy
DFT, 8.4
Hubbard, 8.9
non-interacting, 8.6
exchange-correlation hole, 3.11
expansion around mean field, 11.23

F
Falicov-Kimball model, 11.15
Fe-based superconductors, 14.2
Fermi liquid, 1.10, 1.11, 1.20, 13.2, 14.19
emergence of Kondo physics, 13.5
instability to superconductivity, 1.4
local Fermi liquids, 13.19
quasiparticle density of states, 13.19
quasiparticle scattering amplitude, 13.20
Fermi surface nesting, 6.13
Fröhlich Hamiltonian, 15.3
Friedel sum rule, 1.11–1.13
functional integral, 11.11, 11.14

G
gauge invariance, 11.20, 12.5
generalized gradient corrections, 8.22
Goodenough-Kanamori rules, 4.17, 4.19, 5.24
Green function, 6.4
ground-state projector, 3.13
Gutzwiller
approximation, 1.14, 8.14
wave function, 8.11
GW approximation, 1.7, 1.21–1.22, 16.14
gyromagnetic ratio, 4.2

H
Hartree-Fock, 1.3, 4.12, 9.16, 9.30
restricted, 1.3, 1.12
unrestricted, 1.3, 1.12, 1.13, 1.20
heavy fermions, 1.18, 1.20
Heisenberg model, 4.25, 7.3
high-pressure superconductor, 15.23
Hohenberg-Kohn theorem, 8.4
SCDFT, 16.4
Hohenberg-Kohn–Hubbard theorem, 8.9
Holstein model, 2.7
Hubbard alloy approximation, 1.15
Hubbard bands, 6.8, 13.7
Hubbard dimer, 9.10
Hubbard model, 2.6, 4.23, 7.4, 9.6
single-band, 13.6
two-band, 13.15
Hund's coupling, 5.6, 14.2
Hund's metal, 14.2
crossover

in Fe-based superconductors, 14.9
in models, 14.11
Hund's rules, 4.6
hybridization function, 10.12

I
imaginary-time Green function, 6.3, 15.5
impurity problem, 10.12
infinite dimensions, 1.17
interaction
Hubbard, 8.16
Hund's-rule, 8.16
isotropic Eliashberg function, 15.18
isotropic gap equations, 15.19
itinerant electron scenario for Fe-based superconductors, 14.3

K
kinetic exchange, 4.7
Kohn-Sham
equations, 8.26
Hamiltonian, 8.8
Hubbard equations, 8.10
potential, 8.8, 8.10
system, SCDFT, 16.6
theorem, 8.6
Kondo breakdown, 12.3, 12.16
Kondo effect, 1.12–1.13, 12.2, 12.4, 13.7
single impurity, 13.7
two impurity, 13.9
phase diagram, 13.10
pseudo-gap physics, 13.14
spectral properties, 13.10
Kondo lattice model, 13.5
Kondo temperature, 9.23, 13.7
Kramers-Kronig relation, 11.24
Kugel-Khomskii mechanism, 4.21, 5.15

L
Lanczos method
finite-temperature, 7.8
low-temperature, 7.14
microcanonical, 7.16
lattice constant
iron, 8.22
nickel, 8.18
lattice dynamics, 15.26
Lehmann representation, 6.4
linear response, 3.3
local moment, 12.2
normal phase of Fe-based superconductors, 14.6
long-range Coulomb interaction, 9.15
Luttinger theorem, 1.10, 1.11, 1.19–1.21
Luttinger-Ward functional, 11.12

M
magnetic moment
iron, 8.22
nickel, 8.18
magnetic moments, 6.8
many-body perturbation theory, 11.8
mean-field theory, 2.6, 6.8
metal-insulator transition, 4.24, 9.30
mobility edge, 2.11
model building, 9.28
Mott insulator, 1.5, 1.21, 1.23, 2.11, 5.2, 5.11, 6.14, 7.4, 9.4
favored by Hund's coupling, 14.15
Mott transition, 1.15, 1.16, 1.23, 3.7, 4.23, 13.4
multiband superconductor, 15.23
multiple-occupancy corrections, 11.8, 11.9
multiplet, 4.6

N
Nambu-Gor'kov formalism, 16.12
NiO, 1.21–1.22
nonlocal vertex, 11.19

O
on-site Coulomb interactions, 5.5
optical spectral weight, 5.13, 5.19, 5.23
orbital decoupling, 14.12, 14.16
orbital-ordering, 4.21, 9.37

P
partition function, 10.9
path integral formalism, 10.2
Pauli matrices, 4.30
perturbation theory, 12.8
singular, 12.9
phonon linewidth, 15.12
phonon self-energy, 15.11
poor man's scaling, 9.22
pseudo-fermion representation of spin, 12.5

Q
quantum critical point, 5.17, 5.18, 13.9
quantum Hall effect, 3.10
quantum Monte Carlo, 1.7, 1.9, 10.2, 10.11
quantum phase transition, 12.3
quasi-particle, 13.2
  dispersion
    iron, 8.23
    nickel, 8.20
  Kohn-Sham, 8.8
  Landau-Gutzwiller, 8.13, 8.15
  renormalization, 15.6

R
Racah parameter, 8.9, 8.16, 8.28
real-time dynamics, 7.20
renormalization group, 12.10
  perturbative, 12.11
response functions, 11.16
Ritz variational principle, 8.3
RKKY interaction, 12.2, 12.12
Ruthenates, 14.2

S
SCDFT
  functionals, 16.18
  Kohn-Sham system, 16.6
  Oliveira, Gross, Kohn theorem, 16.4
second quantization, 4.10
self-averaging property, 11.6
self-energy, 6.4, 16.13
selfconsistency, 12.18
  and dynamical mean-field theory, 12.18, 13.7
Sham-Schlüter connection, 16.15
Slater insulator, 9.4
Slater-Condon parameter, 8.17, 8.28
slave-spin mean-field, 14.4
Sommerfeld coefficient, 14.4
Soven equation, 11.5
spectral function, 7.7
spin liquid, 1.5, 1.23
spin screening, 12.2
spin-density wave, 2.6
spin-orbital entanglement, 5.26
spin-orbital superexchange, 5.11
  at charge dilution, 5.31
  at orbital dilution, 5.31
  for $KCuF_3$, 5.15
  for $LaMnO_3$, 5.18
  for $LaTiO_3$, 5.21
  for $LaVO_3$, 5.22
  for $NaTiO_2$, 5.27
spinon-orbiton separation, 5.28
staggered potential, 2.4
STM spectroscopy, 12.20
sum rule, 3.2, 3.9, 3.12
superconductivity, 15.12
superexchange, 4.13
superfluid density, 2.14
symmetry breaking, 13.12

T
T matrix, 11.5
t-J model, 5.2, 5.28, 5.29, 7.14
thermodynamic limit, 11.6
tight-binding Hamiltonian, 2.3
tilted clusters, 7.5
topological marker, 3.17
topological order, 1.23
transfer of spectral weight, 6.19
two-impurity Kondo problem, 12.15

V
V-representability
  Hubbard, 8.9
  non-interacting, 8.5
V_2O_3, 1.22
Velický-Ward identity, 11.8
Vollhardt-Wölfle-Ward identity, 11.8

W
Wannier functions, 3.14
width of the Hubbard bands, 14.15

X
X-ray emission spectronscopy and local moments, 14.6

Schriften des Forschungszentrums Jülich
Modeling and Simulation

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), 420 pages
   ISBN 978-3-89336-734-4

2. **Correlated Electrons: From Models to Materials**
   Lecture Notes of the Autumn School on Correlated Electrons 2012
   edited by E. Pavarini, E. Koch, F. Anders, and M. Jarrell (2012), 450 pages
   ISBN 978-3-89336-796-2

3. **Emergent Phenomena in Correlated Matter**
   Lecture Notes of the Autumn School on Correlated Electrons 2013
   edited by E. Pavarini, E. Koch, and U. Schollwöck (2013), 520 pages
   ISBN 978-3-89336-884-6

4. **DMFT at 25: Infinite Dimensions**
   Lecture Notes of the Autumn School on Correlated Electrons 2014
   edited by E. Pavarini, E. Koch, D. Vollhardt, A. Lichtenstein (2014), 450 pages
   ISBN 978-3-89336-953-9

5. **Many-Body Physics: From Kondo to Hubbard**
   Lecture Notes of the Autumn School on Correlated Electrons 2015
   edited by E. Pavarini, E. Koch, and P. Coleman (2015), 500 pages
   ISBN 978-3-95806-074-6

6. **Quantum Materials: Experiments and Theory**
   Lecture Notes of the Autumn School on Correlated Electrons 2016
   edited by E. Pavarini, E. Koch, J. van den Brink, G. Sawatzky (2016), 420 pages
   ISBN 978-3-95806-159-0

7. **The Physics of Correlated Insulators, Metals, and Superconductors**
   Lecture Notes of the Autumn School on Correlated Electrons 2017
   edited by E. Pavarini, E. Koch, R. Scalettar, and R. Martin (2017), 450 pages
   ISBN 978-3-95806-224-5